

---

---

# INEQUALITY OF OPPORTUNITY ACROSS ORIGINS IN FRANCE

## AND THE IMPLICATIONS OF COLLECTING ETHNIC DATA

---

---

Despite a considerable involvement of the state in the education and the tax system, levels of intergenerational mobility in France are among the lowest by OECD standards (Lefranc, 2018). Understanding the sources of inequality of opportunity between citizens and how different policies could affect its extent is thus of great interest. Besides, differentiated expected outcomes on the sole basis of ethnic origins conditional on parental socio-economic background would perpetuate earnings inequality between individuals of French and foreign origin over generations, in a way that could hardly be legitimized. It is necessary to quantify such an inequality of opportunity between the offspring of natives and immigrants, and to understand its potential determinants, as inceptive evidence to relevant policy recommendations.

## 1 Scientific context

***Intergenerational mobility.*** Using federal income tax records, Chetty et al. (2014) documented large variations in the correlation between the rank of children and their parents' in the income distribution across the United States. They investigated how spatial variations in intergenerational mobility are driven by the causal effect of living in a given commuting zone, and identified the features of areas producing good outcomes (Chetty and Hendren, 2018). Their latest contribution to this literature investigates racial disparities in economic opportunity (Chetty et al., 2020). Results notably reveal that Black Americans experience much lower rates of upward mobility than Whites, sustaining the income gap from one generation to the other. In France however, as no data linking individuals' incomes to their parents' have been collected, little is known about intergenerational mobility in earnings. The most advanced contributions on the issue relied on parental income predictions based on education variables, using several waves of the *Formation, Qualification, Profession* database (Lefranc, 2018). But this only allowed to draw the long-term evolution of the intergenerational elasticity. In addition, this estimator tends to be supplanted by correlations in terms of rank in the distribution, considered as more stable (Chetty et al., 2014). In France, the big picture remains to be drawn, and the links between intergenerational mobility and key factors such as ethnicity and spatial segregation still need to be evaluated.

***Segregation.*** Spatial segregation was shown by Chetty and Hendren (2018) to be associated with lower rates of upward mobility, and further results suggest that neighborhood quality would play a role in the perpetuation of the Black-White gap in the United-States (Chetty et al., 2020). But just like for intergenerational mobility, the French literature on spatial segregation according to ethnic origins is bounded by data constraints. Due to the restrictions on the collection of information on ethnicity in France, researchers had to rely either on nationality or place of birth to compute segregation indices so far. The main data sources that were used for that purpose are the Labor Force Survey and the Population Censuses (Préteceille, 2011; Safi, 2009). But their scope only allows to measure segregation for the biggest metropolitan areas,

and most importantly the reliance on country of birth and nationality restricts the analysis to first-generation immigrants, which impedes the relevance of these estimations.

***Ethnic data.*** The fact that the French literature lags behind the international standards when it comes to including ethnic considerations in socio-economic analyses questions the French position regarding the collection of ethnicity-related data. The variety of arguments raised by sociologists (Simon, 2008) and anthropologists (Zoïa, 2010) has not led to a consensus yet. On the one hand it could help academic research to design appropriate policies to reduce ethnic inequality, but on the other it could reinforce social cleavages. At this stage it is important to draw a comprehensive and comparative review of the hypotheses put forward by the different schools of thought, and to quantify the actual consequences that the implementation of ethnic data collection would have on the French society.

## 2 Research objectives

Intergenerational mobility and spatial segregation according to ethnic origins share the specificity that their precise measurement, and consequently their thorough analysis, is constrained by the limitations of French data. My Master's theses introduced new approaches that enlarge the scope of research possibilities related to these two topics in France, namely the use of the INSEE's Permanent Demographic Sample (EDP) to comprehensively investigate the extent and sources of intergenerational persistence, and the inference of ethnicity from an individual's last name rather than from her nationality or place of birth to compute estimations of segregation at an extremely granular level.

For the purpose of this project, I intend to exploit the potential of these contributions so as to carry more ambitious analyses than what the literature has proposed so far. Based on the precise measurement of intergenerational persistence and spatial segregation, I would quantify the extent to which the latter can explain differences in expected outcomes between children of natives and immigrants conditional on parental earnings. Indeed, in my second Master's thesis, the use of the EDP allowed me to identify that second-generation immigrants from Maghreb experience lower absolute upward mobility in earnings than children of French natives. This ethnic gap is reminiscent of that between Blacks and Whites in the United-States put forward by Chetty et al. (2020), but further investigation suggests that this discrepancy mostly stems from differences in terms of access to employment. These findings echo the well-documented hiring discrimination towards applicants of North-African origin observed on the French labor market, but Chetty and Hendren (2018) identified segregation as another potential channel to this phenomenon. Based on the estimated magnitude of the potentially heterogeneous effect of segregation on intergenerational mobility, I would then evaluate how different policies of segregation reduction would impact the ethnic gap conditional on parental earnings.

This analysis would be carried along with a clarification on the consequences of collecting ethnic data by the framing of a comprehensive review of the academic literature on these topics that was particularly fostered by demographers, sociologists, and anthropologists. To substantiate this exercise, I would proceed to a thorough comparison between the French and the British case in terms of labor market outcomes of immigrants before and after the inclusion of the ethnic question in the 1991 United Kingdom census.

### 3 Methodology

***Intergenerational mobility.*** In the direct continuation of my Master’s theses, I intend to contribute to the sparse French literature on intergenerational mobility by taking advantage of the EDP, whose continuous improvements in terms of scope and quality make it increasingly suitable for the study of mobility-related issues. It tracks all individuals born on specific dates, and notably gathers detailed information about their earnings from 2010 to 2016. The earnings of their parents can be inferred based on the socio-demographic variables collected in the population census waves that took place from 1968 to 1999. I would use machine learning methods to estimate a prediction model on the *Panel d’Actifs* (matched DADS data) of the EDP based on gender, age, year, socio-professional category, and place of work, and apply it to the parents of EDP individuals using their census information. To fill the gaps of the literature on the issue, I would start by providing the first estimations of intergenerational mobility in France in terms of rank in the income distribution. I would also investigate heterogeneity in terms of gender, ethnic origins, etc., as well as potential non-linearities in the relationship between parents’ and children’s income ranks. These preliminary descriptive evidence are a necessary baseline to the more in-depth analyses that the exploitation of the EDP would then allow me to conduct.

***Segregation.*** To overcome the measurement constraints of spatial segregation faced by the French literature, I developed an algorithm that can infer an individual’s origin based on her last name. Simply put, it compares in a flexible way the sub-sequences of letters forming the name whose origin has to be determined to the sub-sequences of letters appearing in well-defined corpora of names of French and Arabic origin. The parametrization of the algorithm is based on an axiomatic approach. Several robustness checks were performed and support that the algorithm correctly targets the population of interest. So far I only relied on the *Fichier Patronymique* to estimate segregation at the *département* level using variations across municipalities. But for the purpose of this project, I would apply this approach to more ambitious databases to obtain precise estimates of segregation at very granular spatial levels. There are two suitable data sources I consider to perform these estimations: the soon available *Répertoire Électoral Unique* that gathers the name and address of residence of 44.5 million French voters, and online phone books such as PagesBlanches, which would allow for estimations of comparable scale. By using the algorithm to infer one’s origin from her last name, and matching her address to a precise longitude and latitude (by webscraping OpenStreetMap for instance), I would be able to estimate segregation indices at virtually any geographical scale and to match it with the place of residence of EDP individuals so that their variation could be thoroughly analyzed.

***Segregation as a source of inequality of opportunity across origins.*** I then propose to combine such precise estimations of spatial segregation with the use of the EDP to draw policy-relevant conclusions by estimating the extent to which reducing the segregation level would narrow the ethnic gap conditional on parental earnings. To quantify this effect, I would take advantage of the panel structure of the fiscal data in the EDP to specify an event-study design. For simplicity of exposition, consider only individuals who faced a job loss and did not move until the end of the period, or until having found a new job. The event-study consists in regressing in a logistic model an unemployment dummy on a variable for the year relative to job-loss, on segregation at the municipality level, and on their interaction, by controlling for relevant covariates (gender, age, occupation, share of immigrants in the municipality, etc.) and

including time, individual, and municipality fixed effects. For each possible spell duration of the relative time variable, this would quantify the effect of segregation on the probability that the spell lasts at least this duration. In line with what is observed in the United-States, I expect the effect to be significantly negative. I would then attribute counterfactual unemployment spells to EDP individuals for alternative segregation levels, and re-evaluate the width of the conditional ethnic gap under these counterfactual conditions. This would allow to quantify the impact of different segregation-related policies on the extent of the conditional ethnic gap.

***Ethnic data.*** To compare the French and the British case, I would rely on the DADS variables imported in the EDP and on the British Labour Force Survey. Based on the precise information on social class and country of origin, I intend to compare in detail the yearly trends in terms of under-representation and segregation of certain origins into specific occupations before and after the implementation of the ethnic question in the British census. I would then take advantage of the inclusion of income variables in the Labour Force Survey from 1992 to deepen the analysis with income inequality considerations.

## 4 Main findings expected

***Intergenerational mobility.*** Based on a linear model to predict parental earnings, the results of my Master's theses suggest high levels of intergenerational persistence in France by OECD standards (in line with Lefranc (2018)), a higher persistence between fathers and sons and between mothers and daughters than across genders, and lower levels of intergenerational mobility for children of immigrants from Maghreb. I expect the estimations based on machine learning methods to predict parental income to confirm these results. I also expect the intergenerational persistence to be non-linear in parental earnings as it must be harder to break away from the socio-economic conditions inherited from one's parents when the latter either endure extreme poverty or earn a very high income.

***Segregation.*** The results of my second Master's thesis indicate that segregation indices are higher in the *départements* where large metropolitan areas are located. No segregation estimations at the municipality level that cover the whole French territory were yet provided by the literature, but I expect such estimations based on last names to corroborate this first insight with higher indices of segregation for municipalities located in large metropolitan areas.

***Segregation as a source of inequality of opportunity across origins.*** I expect spatial segregation to have a significantly negative impact on intergenerational mobility, especially for ethnic minorities. Indeed, the results of my second Master's thesis indicate that the higher intergenerational persistence for second-generation immigrants from Maghreb stems from differences in access to employment. Thus, segregation is likely to play a role in this relationship either *via* difficulties to build a diversified network, or *via* an impact on motivation and ambition through a feeling of exclusion. Chetty et al. (2020) showed that in the United-States, differences in neighborhood quality could explain up to 30% of the Black-White intergenerational gap. This estimation is obviously not directly transferable to the French case, but I expect that the ethnic gap conditional on parental earnings could be significantly narrowed by reducing segregation. I also expect the simulations to reveal that targeting the most segregated municipalities would be more effective than reducing segregation uniformly over the whole territory.

**Ethnic data.** While the positive consequences of the collection of ethnic data in terms of research advances are arguably likely, the repercussions such a reform can have on the inclusion of ethnic minorities are less clear. According to the publicly available OECD data starting in 2000, in both France and the United Kingdom the unemployment rate of foreign born follows the same trend as that of the whole population, but simply shifted upward in comparable proportions. It is likely that the consideration of more distant periods would reveal this broad trend to be robust to the inclusion of the policy, but I expect the more in-depth analysis to uncover heterogeneity according to both occupation and origins, and to disentangle diverging patterns in some cases.

## 5 Projected schedule

As I already have access to the EDP, I would start with the baseline study on intergenerational mobility. Meanwhile, I would conduct the administrative procedures to get access to the other databases. I expect to have access to the *Répertoire Électoral Unique* and/or to phone book data, and to the British Labour Force survey by early 2021. By then, I also intend to have generalized my algorithm to  $N$  corpora rather than only 2, and to have built corpora of names of the most represented origins in France. During 2021, I would then be able to study segregation patterns in detail, to evaluate their impact on the ethnic gap conditional on parental earnings, and to perform the policy simulations. I expect to start the literature review on the question of collecting ethnic data by the end of 2021, so that the quantitative analysis of the implications of the inclusion of the ethnic question in the 1991 United Kingdom census could be conducted in early 2022. I intend to have finalized most of these aspects of the project at the beginning of the third academic year so that part of the remaining time could be devoted to investigate other plausible determinants of intergenerational mobility differences across ethnic groups, like the characteristics of the place individuals grew up in, and spatial mobility in adulthood.

## References

- Chetty, Raj and Nathaniel Hendren**, “The impacts of neighborhoods on intergenerational mobility II: County-level estimates,” *The Quarterly Journal of Economics*, 2018, *133* (3), 1163–1228.
- , —, **Maggie R Jones, and Sonya R Porter**, “Race and economic opportunity in the United States: An intergenerational perspective,” *The Quarterly Journal of Economics*, 2020, *135* (2), 711–783.
- , —, **Patrick Kline, and Emmanuel Saez**, “Where is the land of opportunity? The geography of intergenerational mobility in the United States,” *The Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.
- Lefranc, Arnaud**, “Intergenerational earnings persistence and economic inequality in the long run: Evidence from French cohorts, 1931–75,” *Economica*, 2018, *85* (340), 808–845.
- Préteceille, Edmond**, “Has ethno-racial segregation increased in the Greater Paris metropolitan area?,” *Revue française de sociologie*, 2011, *52* (5), 31–62.
- Safi, Mirna**, “The spatial dimension of integration: Developments in segregation of immigrant populations in France from 1968 to 1999,” *Revue française de sociologie*, 2009, *50* (3), 521–552.
- Simon, Patrick**, “Les statistiques, les sciences sociales françaises et les rapports sociaux ethniques et de “race”,” *Revue française de sociologie*, 2008, *49* (1), 153–162.
- Zoïa, Geneviève**, “Faut-il avoir peur de l’ethnicité? Le cas français,” *Anthropologie et sociétés*, 2010, *34* (2), 199–223.