Anatomy of Inequality in France

Philippe Aghion, Vlad Ciornohuz, Maxime Gravoueille, Stefanie Stantcheva

July 1, 2019

1   Introduction

In this paper, we provide descriptive results about inequality and mobility in France over the period 2011-2015. Our contributions are based on a new panel data that allows us to track the tax returns of all French tax residents from year to year. This is the first time such panel data exists for France. Together with the companion paper, the current paper is a first step of a broader research project on the role of policy for entrepreneurship and innovation in France. The new available data will allow us to study the impact of tax reforms on the innovation behavior, as well as to look into models of optimal taxation that include innovation with its risk and externality dimensions. In this context, the scope of starting to look into the evolution of income for a broader population during this period rich in reforms before focusing on innovators is straightforward. Moreover, this preliminary research studies allow us to certify the validity of this newly constructed panel data.

We show that mobility over five years between 2011 and 2015 is quite low. An individual who in 2011 is in the bottom decile of the national income distribution has very little chance of making it above the third decile at all 5 years later and has a 60% chance of remaining in the bottom decile. And symmetrically, downward mobility is also very low. An individual who in 2011 is in the top decile will remain there with a 77% probability in 2015. The rank-rank slope at the national level between income in 2011 and income in 2015 is 0.83.

Mobility in the North of France over this 5 year period is lower; in addition, just a superficial look at the electoral results by region seem to indicate that departments which have lower mobility are also those with the highest share of far-right (Le Pen) votes. There are strong geographic variations in inequality too. Close to 30% of all top 0.1% earners live in the Paris region and 17% of all top 1% earners do.

If we look at mobility by age, we can see that the correlation between income in 2011 and income in 2015 is lower earlier in life, but by age 30, it is very stable and very high. This suggests that it is difficult to get another shot after age 30 in France.

Women are drastically less represented than men at all earnings levels above the 40th percentile. In the top 1%, less than 20% of earners are women. However, the representation of women at the top has increased between 2011 and 2015.

We can also track immigrants in France and see how well they do over time. 12% of all top 5%

earners in France are born abroad; and a bit more than 15% of all top 0.1% earners are. Immigrants.

2 Data

We use data from national income tax records covering the period 2011-2015, which refers to the year in which the income is earned[1]. We provide throughout the text a detailed description on how we construct our analysis sample from the raw data.

Sample definition Our panel dataset is constructed using data ont he full French population filling personal income tax records from the General Directorate of Public Finance (DGFIP). The coverage is high because filling an income tax record is mandatory in France, even when households are not taxed. These data contains detailed information on all incomes reported by households. It also contains socioeconomic variables on individuals such as gender, age, marital status, actual location and birth place. Two major contributions of this paper are (1) to tranform these household-level income tax returns to individual-level data, and (2) to transform these cross-section data to an individual-level panel.

Our benchmark sample consist of all individuals who are French fiscal resident and do not experience a change in their marital status. We restrict our analysis to these individuals as we can compute a coherent definition of income for them[2]. We only consider taxpayers and not dependants such as children. There are approximately 36 millions unique households for 48 millions unique taxpayers each year (see Appendix B for a detailed decomposition by year).

Variable definitions From the several income aggregates that are already available in the dataset, only two main ones are directly used for tax purposes. The first is the taxable income (*Revenu imposable - revimp*), which includes all income that is taxed through the income splitting system[3] and the progressive income tax schedule once different deductions are taken into account. These deduction include: *(i)* tax base deductions as the 10% deduction for personal expenses for wages and pensions, and the deductions in simplified regime and *(ii)* deductible expenses. The second one is the reference taxable income (*Revenu fiscal de référence - revfisc*). This aggregate is defined in order to have a better image of the entirety of the income of the households. This income aggregate is mainly used in income criteria for some exemptions or social benefits. The reference taxable income is a good proxy for the total declared income (we will use this two names interchangeably) inregardless of the type of tax treatment of the different types of income. We will mainly rely on this definition to define our income groups. Both the taxable income and the total declared income are defined at the household level. To define a coherent measure at the individual

---

[1]During our time period, France had an income tax system where incomes earned in year T where taxed in year $T + 1$, at the household level.

[2]Non fiscal residents have extraterritorial incomes we have not access to. Individuals changing their marital status have to make two declarations for each part of the year corresponding to a distinct marital status. The personal income tax schedule is then applied on each declaration separately and the corresponding amount of taxes are aggregate for the entire year, making it difficult to define coherent income measures.

[3]Système du Quotient Familial.

level, we divide the total declared income by the number of adults in the household[4]. Appendix C.1 provides a detailed description of income variables. We also look at labor incomes. We will mainly focus on wages which is computed at the individual level from labor activities (i.e excluding self-employed activities).

Finally, the dataset includes socioeconomic for each individuals: marital status [5], the number of fiscal shares, counties[6], year of birth, gender.

3 Social Mobility

We provide three different measure of income mobility: transition matrices, log-log regressions and rank-rank regression. We first apply these methods at the national level. Then, we use rank-rank regressions to analyse heterogeneity in income mobility among various group (by counties, sexe, age, etc.). In our analysis, the ranking is computed at the national level using the benchmark sample. We rank individuals using quantiles defined at the national level each year[7]. Note that estimations are done on a subsample of the benchmark sample, which contains individuals who are observed in the two years of interest. Sub-population estimations (by county, gender) also use the national level ranking.

3.1 Transition matrices at the national level

Transition matrices We first start by analyzing transition matrices at the individual level. Here, states x and y are respectively income groups in 2011 and 2015, defined as deciles of income. In Figure 1, it is graphically represented on the horizontal and vertical axes. Each cell gives the probability for someone in income group x in 2011 to be part of income group y in 2015. As an example, the cell in the top-left corner of Figure 1 tells us that 0.75% of individuals from the first decile of income (P1-P10) in 2011 reached the tenth decile (P90-P100) in 2015, that is the highest decile of the 2015 income distribution. One important thing to notice here that the diagonal of this matrix shows the probabilities of staying in each decile of income.

For every income group, the most probable outcome is to remain in the same income group between 2011 and 2015. Indeed, diagonal elements are the highest probabilities of all columns, which means that remaining in the same decile is the most probable outcome for each 2011 income group.

Focusing on this diagonal, one can notice that the probabilities of staying in the same income group (i.e. the probabilities of no mobility at all) get higher as one moves towards extreme groups (the lowest and highest deciles of income). On Figure 1 we see that someone belonging to the fifth decile of income (P40-P50) in 2011, just below the median, will be in this same decile in 2015 with a

---

[4]We assign two income shares for couples (*married* individuals) and one income share for single individuals.

[5]There are five possible status in France. *Pacses* and *Married* are defined as couples. *Divorced, Widower/widow* and *unmarried* are defined as singles

[6]*Departements*. Note that we define the county of residence as the county where the tax return has been filled and treated by the administration
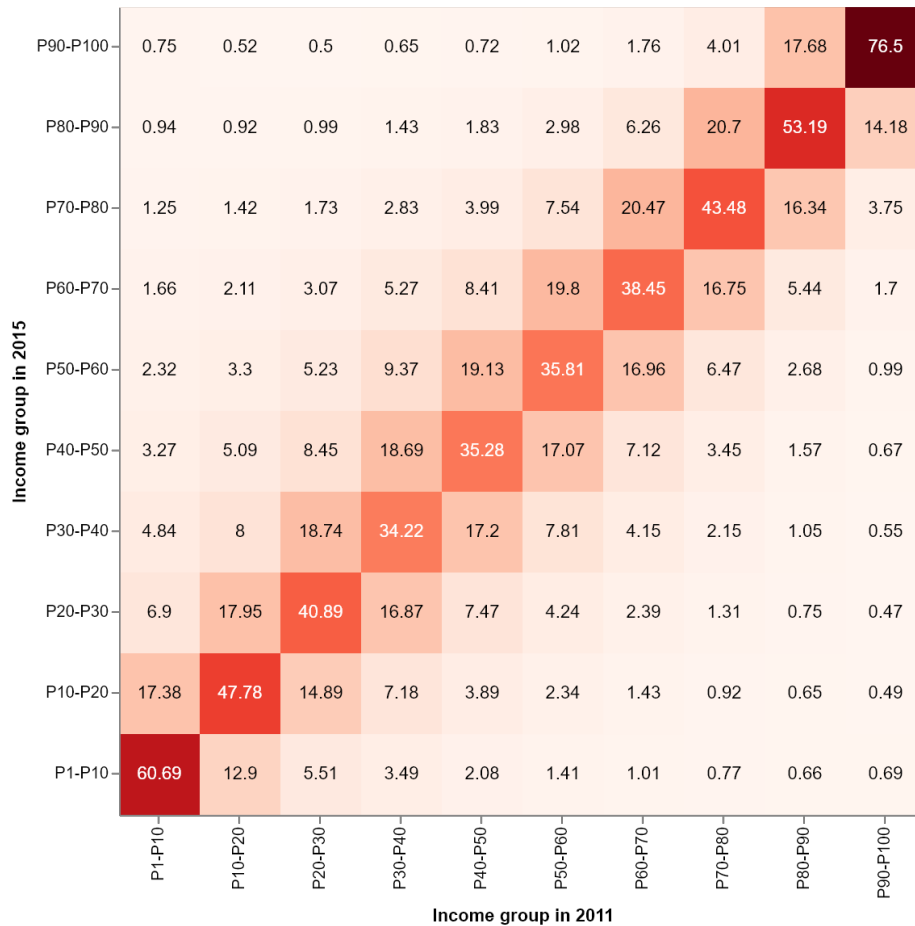
[7]In our baseline analysis, we use percentiles to rank individuals. If one unique group cannot be constructed, for example if the bottom 5% of the population has zero income, we compute the expected rank for this population and set their rank to the mean rank in $t-1$, i.e 2.5.

probability of 35.28%. In contrast, individuals belonging to the first (P1-P10) and last (P90-P100) deciles of income in 2011 will respectively be in these same deciles in 2015 with probabilities of 60.69% and 75.5%.

This probability of staying in the same income group is a direct mesure of persistence of incomes. These results show that mobility is much higher (twice higher) at the middle of the distribution than at the bottom and the top of the distribution of income. Middle class individuals are more likely to move upward and downard than the bottom and the top groups of income. 2 shows a finer decomposition of top income groups' transition matrix. This is basically the same transition matrix than Figure 1, but focusing on the very top quantiles of income.
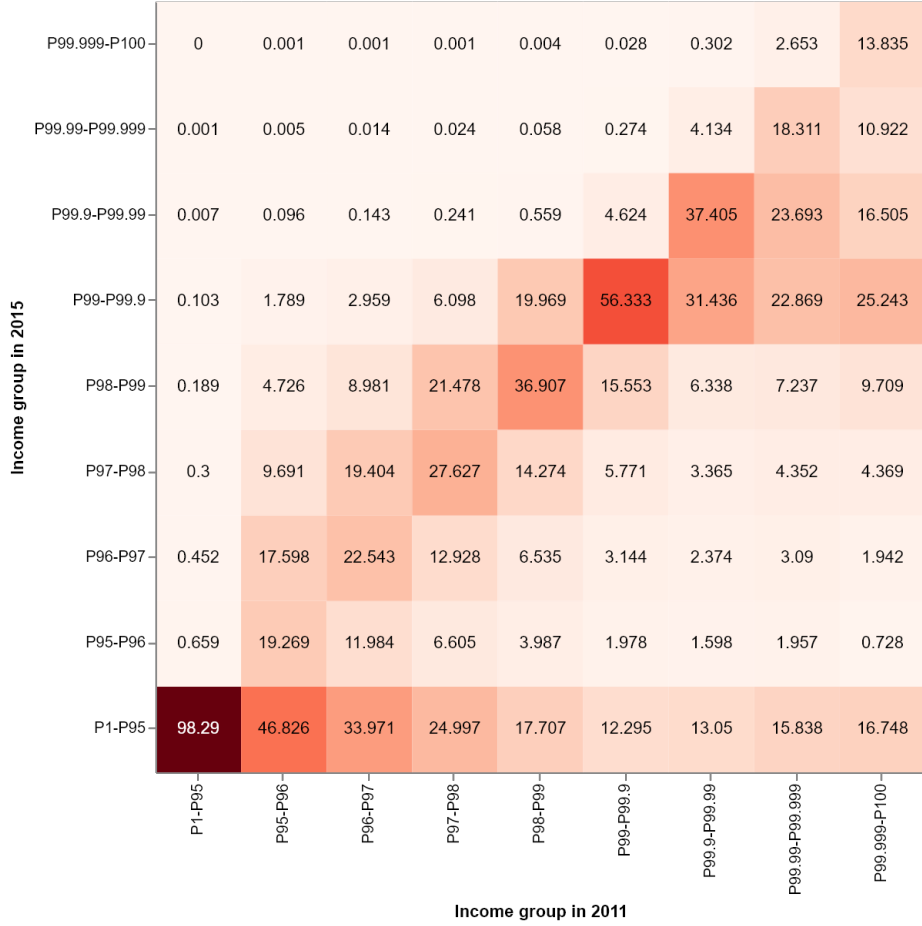
Again, probabilities are higher on and around the diagonal, which means that top income groups are more likely to stay in the same quantile of income than to move. Even at the top, immobility is more likely than mobility.

FIGURE 1: GENERAL DECOMPOSITION

Income group in 2015	P1-P10	P10-P20	P20-P30	P30-P40	P40-P50	P50-P60	P60-P70	P70-P80	P80-P90	P90-P100
P90-P100	0.75	0.52	0.5	0.65	0.72	1.02	1.76	4.01	17.68	76.5
P80-P90	0.94	0.92	0.99	1.43	1.83	2.98	6.26	20.7	53.19	14.18
P70-P80	1.25	1.42	1.73	2.83	3.99	7.54	20.47	43.48	16.34	3.75
P60-P70	1.66	2.11	3.07	5.27	8.41	19.8	38.45	16.75	5.44	1.7
P50-P60	2.32	3.3	5.23	9.37	19.13	35.81	16.96	6.47	2.68	0.99
P40-P50	3.27	5.09	8.45	18.69	35.28	17.07	7.12	3.45	1.57	0.67
P30-P40	4.84	8	18.74	34.22	17.2	7.81	4.15	2.15	1.05	0.55
P20-P30	6.9	17.95	40.89	16.87	7.47	4.24	2.39	1.31	0.75	0.47
P10-P20	17.38	47.78	14.89	7.18	3.89	2.34	1.43	0.92	0.65	0.49
P1-P10	60.69	12.9	5.51	3.49	2.08	1.41	1.01	0.77	0.66	0.69

Income group in 2011

Notes: The figure shows transition probabilities by deciles between 2011 and 2015. The ranking is computed at the national level using the benchmark sample. The probabilities are computed using individuals present both in 2011 and 2015. We rank individuals each year using deciles for the total declared income by adult shares, defined at the national level.

FIGURE 2: FINE DECOMPOSITION AT THE TOP

Income group in 2015	P1-P95	P95-P96	P96-P97	P97-P98	P98-P99	P99-P99.9	P99.9-P99.99	P99.99-P99.999	P99.999-P100
P99.999-P100	0	0.001	0.001	0.001	0.004	0.028	0.302	2.653	13.835
P99.99-P99.999	0.001	0.005	0.014	0.024	0.058	0.274	4.134	18.311	10.922
P99.9-P99.99	0.007	0.096	0.143	0.241	0.559	4.624	37.405	23.693	16.505
P99-P99.9	0.103	1.789	2.959	6.098	19.969	56.333	31.436	22.869	25.243
P98-P99	0.189	4.726	8.981	21.478	36.907	15.553	6.338	7.237	9.709
P97-P98	0.3	9.691	19.404	27.627	14.274	5.771	3.365	4.352	4.369
P96-P97	0.452	17.598	22.543	12.928	6.535	3.144	2.374	3.09	1.942
P95-P96	0.659	19.269	11.984	6.605	3.987	1.978	1.598	1.957	0.728
P1-P95	98.29	46.826	33.971	24.997	17.707	12.295	13.05	15.838	16.748

Income group in 2011

Notes: The figure shows transition probabilities by quantiles between 2011 and 2015. The ranking is computed at the national level using the benchmark sample. The probabilities are computed using individuals present both in 2011 and 2015. We rank individuals each year using quantiles for the total declared income by adult shares, defined at the national level.

3.2 Log-log analysis at the national level

Log-log regression We then regress the log income Y in period t on the log income Y in period $t-1$.

$$\log(Y_{i,t}) = \alpha + \log(Y_{i,t-1}) + \epsilon_i$$

We restrict our sample to individuals with a positive income. Figure 3 shows that the relationship between the log income in 2011 and the log income in 2015 is non-linear. We divide the sample into three sub-population to estimate the income mobility. We see that the bottom 10 percent as a larger persistence of incomes (0.2221) than the rest of the population: 0.8855 for P10-P90 and 0.7416 for the top 10 percent. It suggests that the bottom of the distribution experience less social mobility.

FIGURE 3: LOG-LOG ANALYSIS AT THE NATIONAL LEVEL
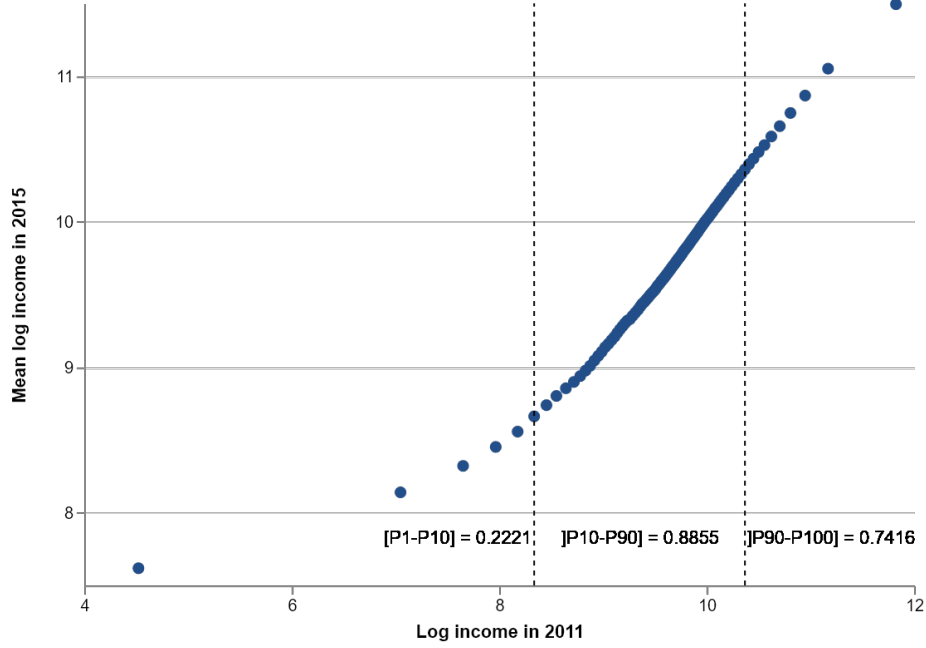
Notes: The figure shows the results from the log-log regression. The binscatter is based on the ranking computed at the national level using the benchmark sample. We rank individuals each year using percentiles for the total declared income by adult shares, defined at the national level. The estimation is done on individuals who are observed in the two years of interest. Slopes are estimated using an OLS regression on the microdata.

3.3   Rank-rank analysis at the national level

We begin by examining the income mobility at the national level. Let $R_{i,t}$ denote the income[8] for individual i at period t. Our baseline analysis consists in a rank-rank regression where the slope and the intercept are estimate by regressing $R_{i,t}$ on $R_{i,t-1}$.

$$R_{i,t} = \alpha + \beta R_{i,t-1} + \epsilon_i$$

We first define β measures as the degree of *relative income mobility*. It denotes the difference between the expected ranks of an individual starting at the top and an individual starting at the bottom of the income distribution. Since the relation is linear ($\bar{r}_{100} - \bar{r}_0 = 100 \times \beta$), a large β is associated to a low relative income mobility. Then, we define $\bar{r}_q = E[R_t | R_{t-1} = q]$ as the expected rank at time t of a person with rank q at time $t - 1$. It is a measure of *absolute income mobility*.

Relative income mobility   Figure 4 shows the relative income mobility between year 2011 and year 2015. The difference between the expected ranks in 2015 of individuals starting at the top and at the bottom of the income distribution in 2011 is equal to 83.21, suggesting little income mobility.

---

[8]Define as the total declared income by adult share.A married person will have two shares while a single person will have one share.

Absolute income mobility We can define the expected income mobility using the slope and intercept from the rank-rank regression. For example, the expected income rank in 2015 for an individual in the 25th percentile is $\bar{r}_{25} = \alpha + \beta \times 25 = 9.52 + 0.83 \times 25 = 30.27$.

FIGURE 4: RANK-RANK REGRESSION
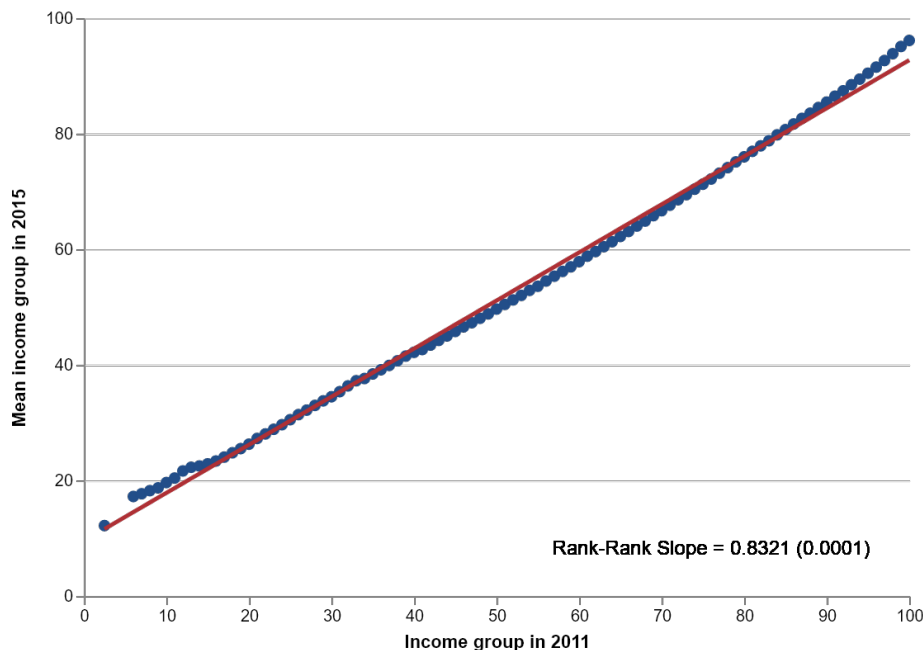
Notes: The figure shows non-parametric binned scatter plot of the relationship between total declared income of an individual between 2011 and 2015. The ranking is computed at the national level using the benchmark sample. We rank individuals each year using percentiles for the total declared income by adult shares, defined at the national level. The estimation is done on individuals who are observed in the two years of interest. The slope and best-fit line are estimated using an OLS regression on the microdata. Standard errors are reported in parentheses.

3.4 The Geography of Social Mobility and Inequality

3.5 Inequality

We present here the geographic dimension of income inequalities in France. Figure 5 shows the average individual income of each French county. This simple map already underlines the existence of sharp geographic disparities of income. The vast majority of French counties show average incomes that lie between 15000 and 21000 euros a year, while only three counties have an average income higher than 27000 euros. These three counties represent Paris and its western suburbs.

More generally, counties where major cities are located (e.g. Rhne with Lyon, Bouches-du-Rhne with Marseille, etc.) seem to be better off than rural counties. However, the finer administrative division of counties within the Parisian agglomeration shows that sharp disparities can exist inside large urban areas. Indeed, we see that the county with the highest average income (Paris) lies next to one of the counties with the lowest average income (Seine-Saint-Denis).

FIGURE 5: AVERAGE INCOME BY COUNTY

(15000, 17000]
(17000, 19000]
(19000, 21000]
(21000, 23000]
(25000, 27000]
(27000, 29000]
(31000, 33000]
(33000, 35000]
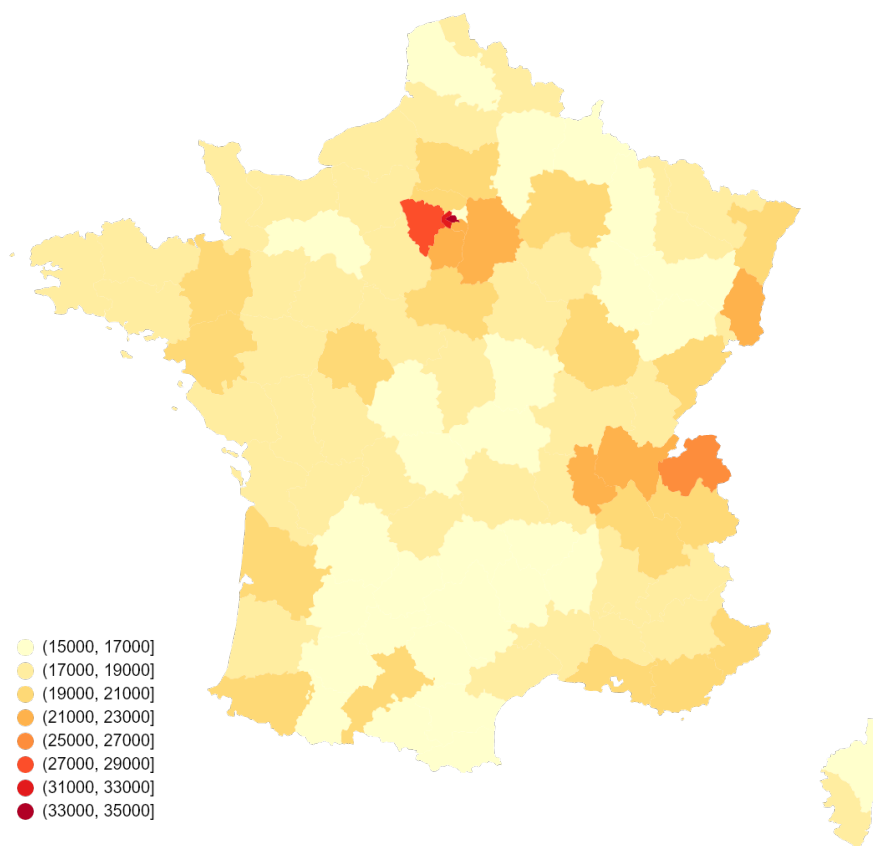
Notes: The figure shows the average total declared income adult shares at the individual level in 2015. We use the benchmark sample to compute the average total declared income adult shares.

Figure 6 shows the top five place of residence of top income earners in 2015. There is a large concentration of top income earners in Paris (county ZIP code equal to 75), where almost one third of the top 0.1% income earners live. We also seen that the top three place of residence are in the same region (le-de-France).

FIGURE 6: Where do Top Income Earners Live?
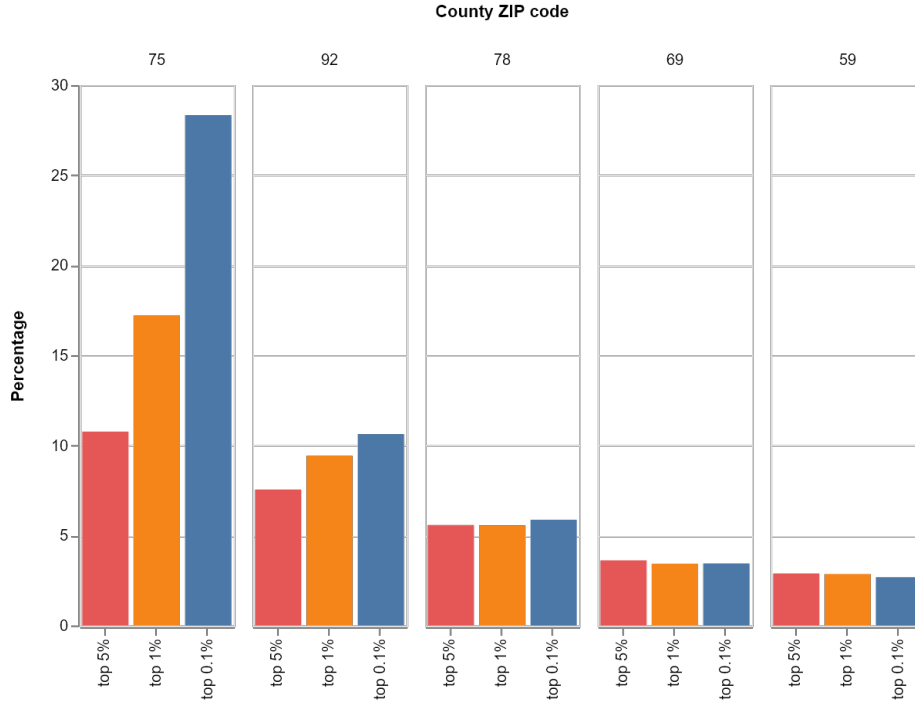
Notes: The population includes individuals that are part of a top income household. The income groups are defined using the RFR and are computed at the household level.
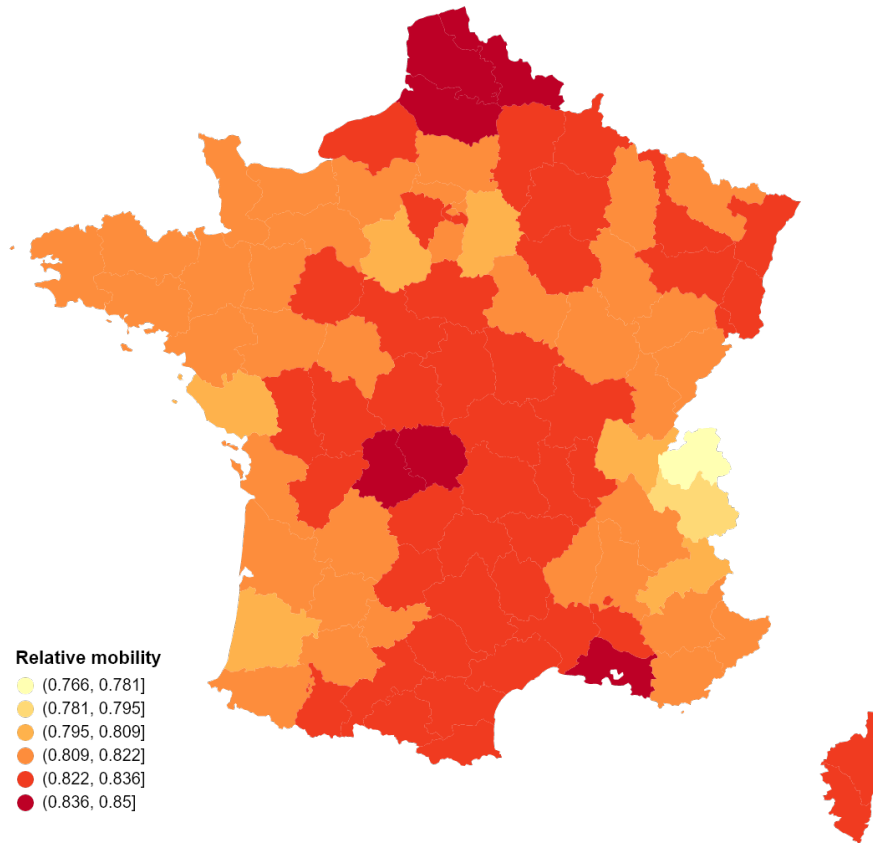
3.6 Mobility

We replicate the same analysis of mobility as above, but at the county level (*departement*). Income groups are computed at the national level, but we provide estimates at the county level. We fix location using the county in the first period of observation (denoted by c).

$$R_{i,t} = \alpha + \beta R_{i,c,t-1} + \epsilon_i$$

Relative income mobility   Figure 7 shows the relative income mobility by county. It exhibits geographic variation, with an outlier county in the east of France. It might be due to the proximity of Switzerland and labour mobility between the two countries. Nonetheless, counties which have difficult economic situation seem to have less relative income mobility (high β).
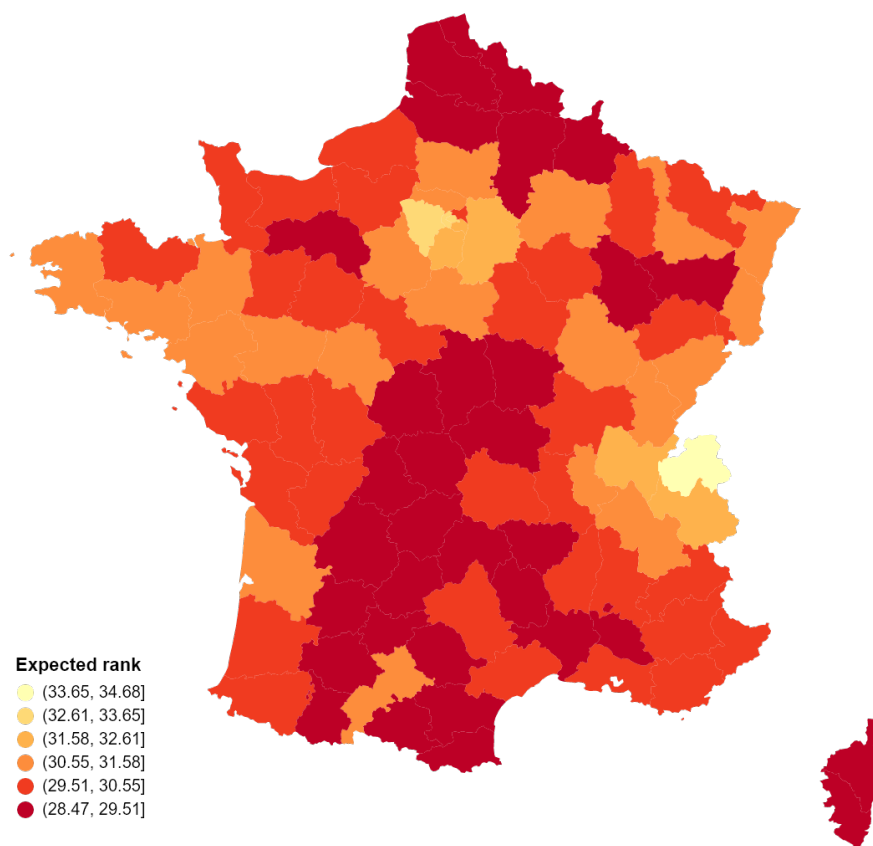
Absolute income mobility   Figure 8 shows the absolute income mobility by county for someone starting at the 25 percentile in 2011. There is a large geographic variation, where county which have good economic situation have higher absolute income mobility than county with bad economic situation.

FIGURE 7: RELATIVE MOBILITY BY COUNTY

Relative mobility
- (0.766, 0.781]
- (0.781, 0.795]
- (0.795, 0.809]
- (0.809, 0.822]
- (0.822, 0.836]
- (0.836, 0.85]

Notes: The figure shows the slope from the rank-rank regression, at the county level. The ranking is computed at the national level using the benchmark sample. We rank individuals each year using percentilesfor the total declared income by adult shares, defined at the national level. The estimation is done on individuals who are observed in the two years of interest. Slopes are estimated using an OLS regression on the microdata.


Figure 8: Absolute Mobility by County

Expected rank
- (33.65, 34.68]
- (32.61, 33.65]
- (31.58, 32.61]
- (30.55, 31.58]
- (29.51, 30.55]
- (28.47, 29.51]

Notes: Notes: The figure shows the relative income mobility of an individual in the 25th percentile in 2011, at the county level. The ranking is computed at the national level using the benchmark sample. We rank individuals each year using percentiles for the total declared income by adult shares, defined at the national level. The estimation is done on individuals who are observed in the two years of interest. Intercepts and slopes are estimated using an OLS regression on the microdata.

Figure 9 shows the relative mobility of income for individuals who change of county of residence between 2011 and 2015 - (a) Movers - and individuals who stay in the same county - (b) Non-movers. On both maps, the darker the color the lower is the relative mobility. Relative mobility is higher for people who move from their initial county. This suggests that geographic mobility (here identified with "movers") is associated with opportunities of income mobility.

FIGURE 9: RELATIVE MOBILITY BY COUNTY FOR MOVERS AND NON-MOVERS
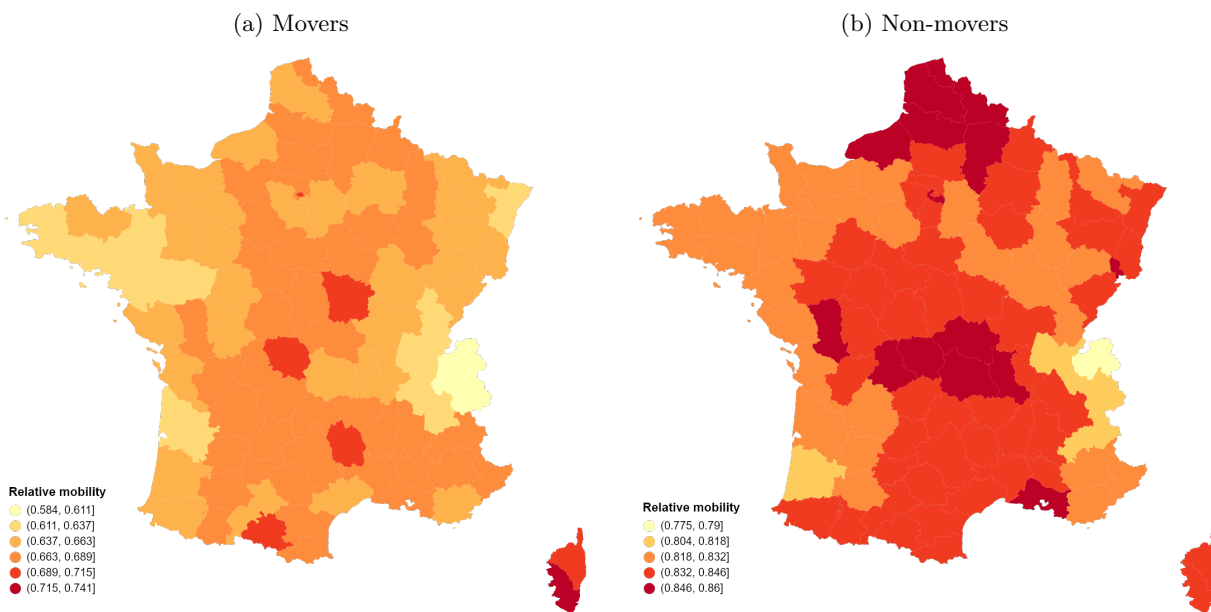
(a) Movers

(b) Non-movers

Notes: Figures shows the slope from the rank-rank regression, at the county level. Panel (a) shows the slope for individuals who are not in the same county in 2011 and 2015, while panel (b) shows the slope for individuals who are in the same county in 2011 and 2015. The ranking is computed at the national level using the benchmark sample. We rank individuals each year using percentiles for the total declared income by adult shares, defined at the national level. The estimation is done on individuals who are observed in the two years of interest and who are in the same county in 2011 and 2015. Slopes are estimated using an OLS regression on the microdata.

4   Social Mobility By Age

Figure 10 shows the relative mobility between 2011 and 2015 by group of age (set in 2011). The higher the score of relative mobility, the lower the social mobility of individuals within a given group of age.

This curve is globally concave with a steep surge for individuals younger than 30 years old. This tends to show that income mobility is much higher for younger individuals, and that the opportunities of social mobility decrease sharply after 30 years old. It reflects the transition from unstable careers with potentially greater opportunities of mobility at a young age to stable career paths after 30.
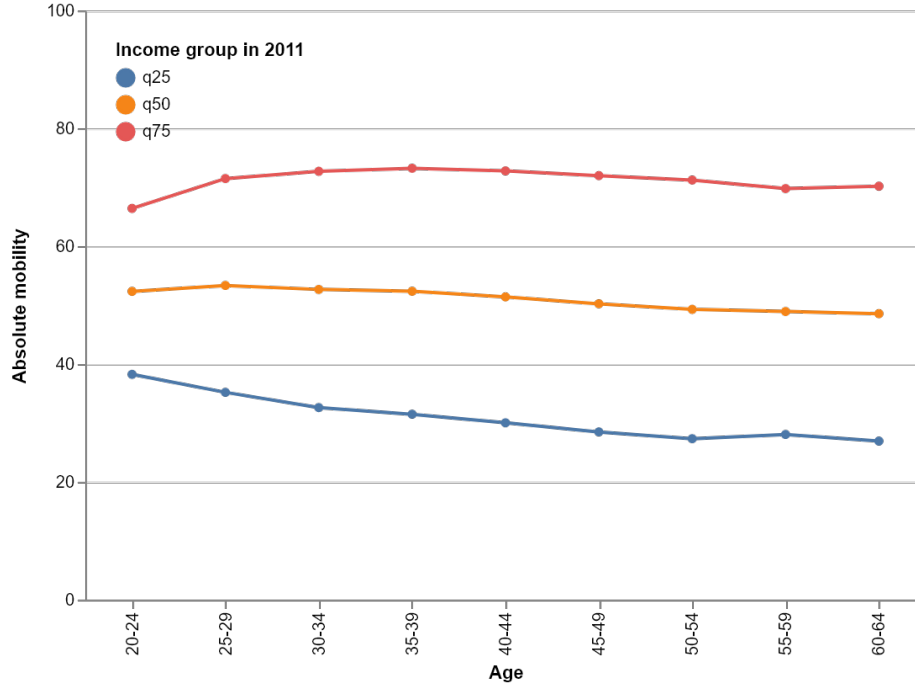
Figure 11 plots the absolute mobility by group of age (set in 2011) for individuals who belong to the 25th, 50th and 75th percentiles of income in 2011. It highlights once again the fact that social mobility is higher for younger agents. The expected percentile of income in 2015 is further from the starting percentile (i.e. higher mobility) for individuals who belong to the youngest groups of age in 2011.

FIGURE 10: RELATIVE MOBILITY BY AGE
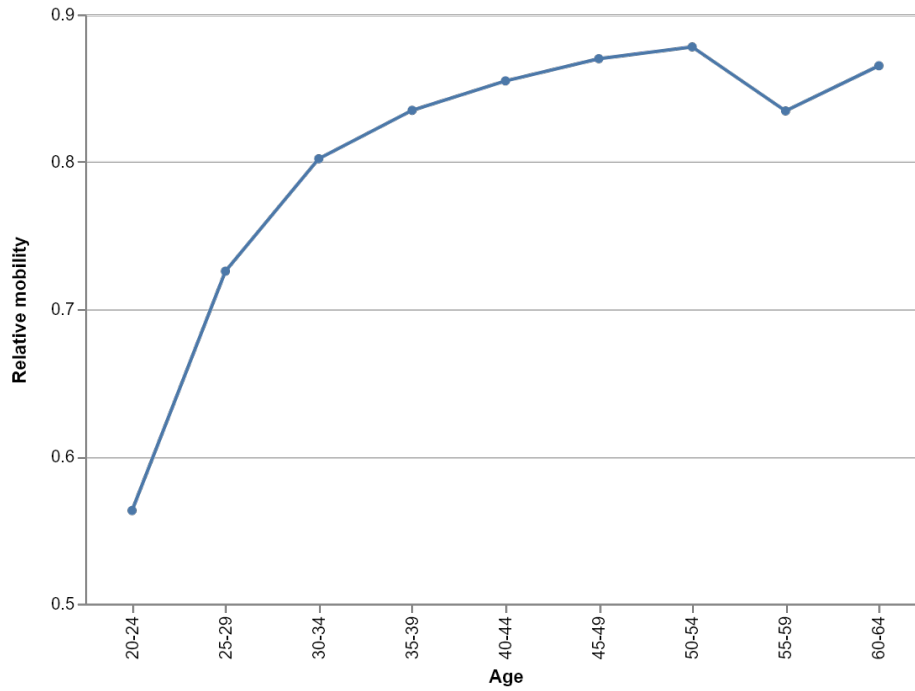
Notes: The figure shows the slope from the rank-rank regression, by group of age in 2011. The ranking is computed at the national level using the benchmark sample. We rank individuals each year using percentiles for the total declared income by adult shares, defined at the national level. The estimation is done on individuals who are observed in the two years of interest. Slopes are estimated using an OLS regression on the microdata.

FIGURE 11: ABSOLUTE MOBILITY BY AGE

Notes: The figure shows the absolute mobility computed using slopes and intercepts from the rank-rank regression, by group of age in 2011. The ranking is computed at the national level using the benchmark sample. We rank individuals each year using percentiles for the total declared income by adult shares,defined at the national level. The estimation is done on individuals who are observed in the two years of interest. Slopes are estimated using an OLS regression on the microdata.

5   Social Mobility and Inequality by Gender

5.1   Inequality

We begin our analysis by exploring the inequality between men and women to access top wage earning groups. Figure 12 panel (a) shows in which proportion men and women are represented in a given income group in 2015. We see that women are overrepresented in the bottom of the distribution and largely underrepresented in the top of the distribution. Panel (b) shows the difference in percentage point between 2011 and 2015 by income groups. We see a small improvement in the proportion of women in the top of the distribution.

Next, in 13 we investigate the geographic inequality in wage earnings between men and women. Panel (a) shows the ratio between the mean wage for a woman and the mean wage for a man by county. Panel (b) shows the ratio between the sum of wages earnings for women and the mean wage for men by county. We see that women earn on average far less than men, with some geographic disparities.

FIGURE 12: REPRESENTATION OF MEN AND WOMEN AT DIFFERENT LEVELS OF EARN-INGS

(a) In 2015

(b) Evolution between 2011 and 2015
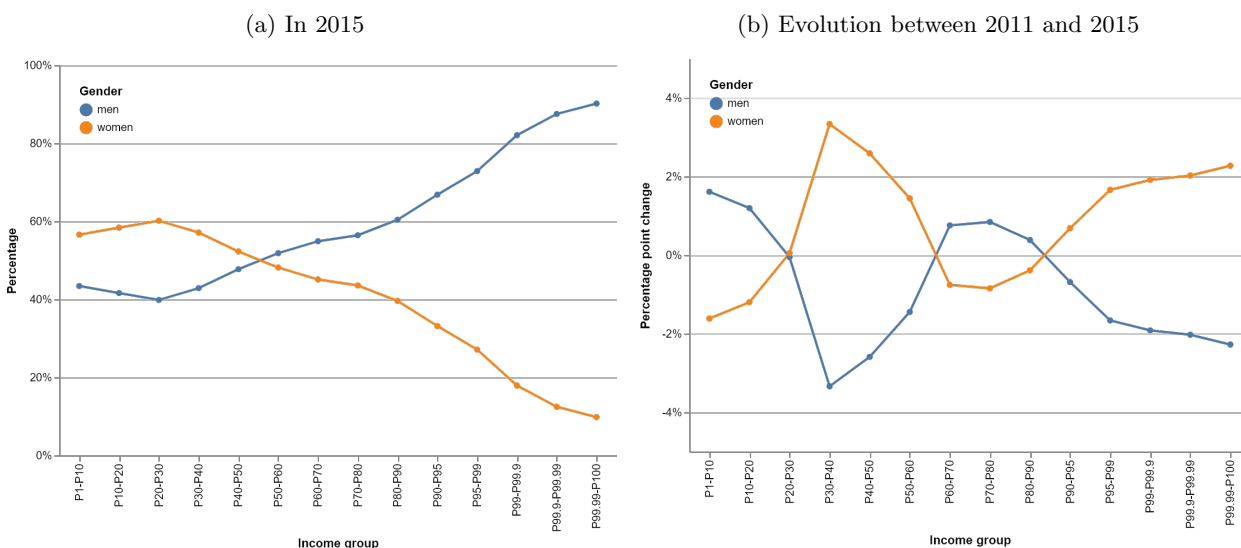
Notes: Figures shows the proportion of men and women by income groups. The ranking is computed at the national level using the benchmark sample. We rank individuals each year using percentiles for wages, defined at the national level. Panel (a) shows the cross-section distribution in 2015, while panel (b) shows the percentage point change between 2011 and 2015.

FIGURE 13: EARNINGS OF MEN AND WOMEN BY COUNTY

(a) Ratio of Mean Earnings of Women and Men

(b) Ratio of the Total Earnings of Women and Men
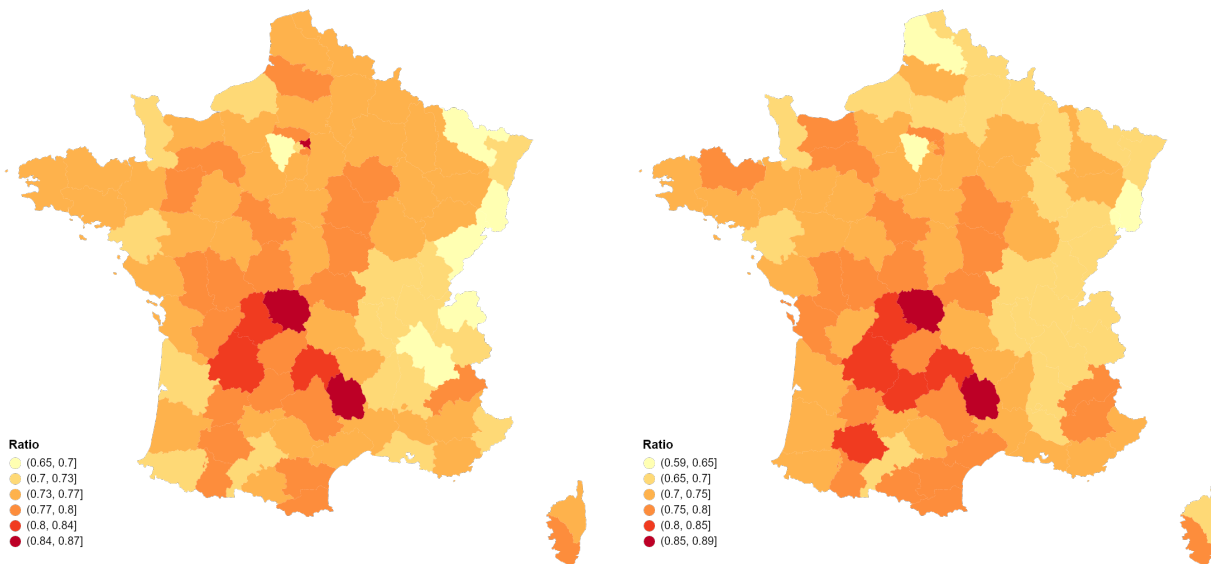
Notes: Figures shows the ratio of wage earnings between women and men, at the county level. Panel (a) shows the ratio between the mean wage for a woman and the mean wage for a man, while panel (b) shows the ratio between the sum of wage earnings for women and the sum of wage earnings for men. Statistics are computed using the benchmark sample.

5.1.1 Social mobility

We estimate the expected rank in 2015 with respect to the rank in 2011 for wage income by sexe. We see no large difference between men and women expected rank, except at the bottom of the distribution. It suggests that women do not have less mobility than men, but are less likely to be in top earning groups. In other words, once women access high top earning earning group, they have the same social mobility as men.
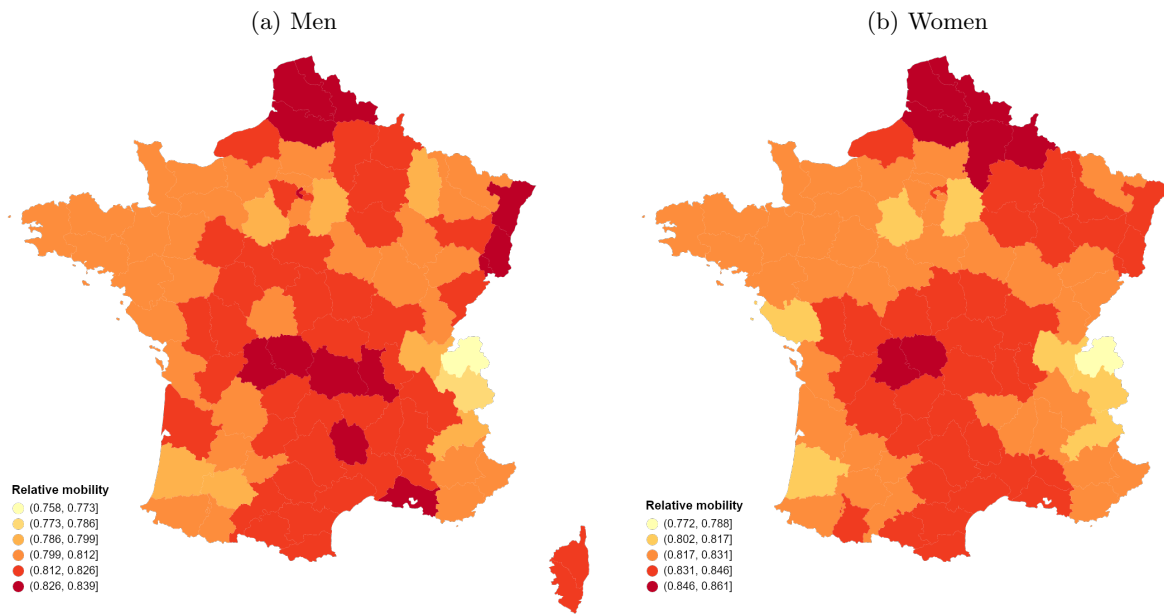
FIGURE 14: RANK-RANK CORRELATION FOR MEN AND WOMEN
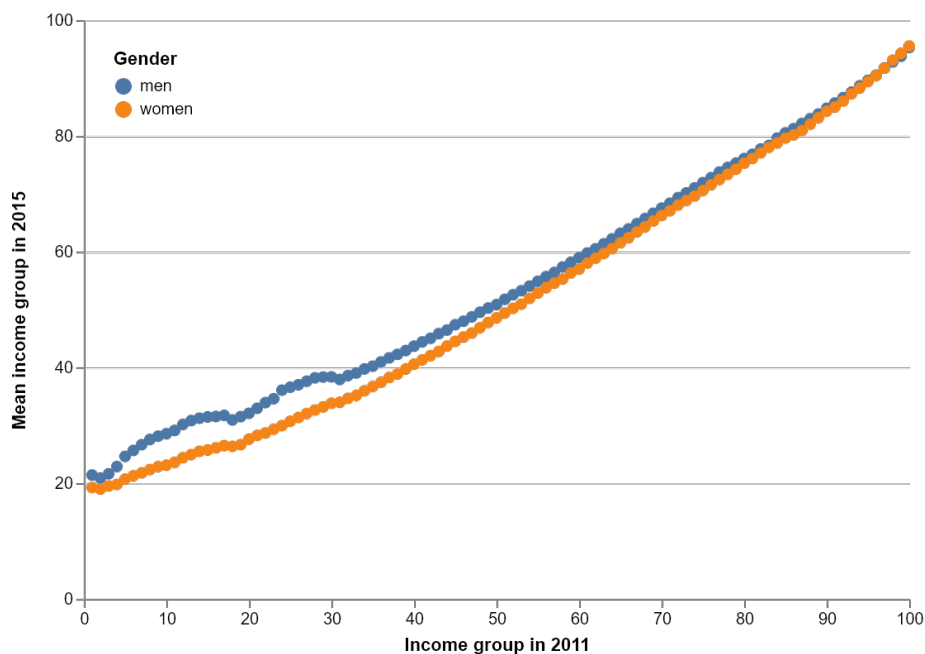
Notes: The figure shows non-parametric binned scatter plot of the relationship between wage earnings of an individual between 2011 and 2015, by gender. The ranking is computed at the national level using the benchmark sample. We rank individuals each year using percentiles for the total declared income by adult shares, defined at the national level. The estimation is done on individuals who are observed in the two years of interest. The slope and best-fit line are estimated using an OLS regression on the microdata. Standard errors are reported in parentheses.

Figure 15 presents the relative mobility of men and women by county. Again the darker the color, the lower the relative mobility. We see that the geographic disparities of men' and women' mobility are quite similar, with slightly darker counties for men. This suggests that men and women experience social mobility in the same areas.

FIGURE 15: RELATIVE MOBILITY BY GENDER AND COUNTY

(a) Men

(b) Women

Relative mobility
(0.758, 0.773]
(0.773, 0.786]
(0.786, 0.799]
(0.799, 0.812]
(0.812, 0.826]
(0.826, 0.839]

Relative mobility
(0.772, 0.788]
(0.802, 0.817]
(0.817, 0.831]
(0.831, 0.846]
(0.846, 0.861]

Notes: Figures shows the slope from the rank-rank regression, at the county level. Panel (a) shows the slope for men, while panel (b) shows the slope for women. The ranking is computed at the national level using the benchmark sample. We rank individuals each year using percentiles for the total declared income by adult shares, defined at the national level. The estimation is done on individuals who are observed in the two years of interest and who are in the same county in 2011 and 2015. Slopes are estimated using an OLS regression on the microdata.

References

Aghion, Philippe, Vlad Ciornohuz, Maxime Gravoueille, and Stefanie Stantcheva, "Reforms and Dynamics of Income: Evidence Using New Panel Data," Working Paper.

APPENDIX

A   Figures

TBC

B   Summary statistics

TABLE 1: BASE SAMPLE COUNTS BY YEAR

Year	Number of Unique Households	Number of Unique Taxpayers
2011	35,510,590	47,941,354
2012	35,791,624	48,235,007
2013	36,165,891	48,657,231
2014	36,417,607	48,938,838
2015	36,621,298	49,141,249

B.1 Variables

C Variable definitions

C.1 Income

FIGURE 16: DEFINITIONS OF THE FISCAL INCOME AND THE TAXABLE INCOME

> **Total declared income (progressive tax schedule)**
>
> \-
>
> **Different tax base deductions**
>
> =
>
> **Brut global income**
>
> \-
>
> **Deductible expenses**
>
> =
>
> **Net global income**
>
> \-
>
> **Special tax bases deductions**
>
> =
>
> **Taxable Income**
>
> +
>
> **Income taxed at a flat rate**   **Exempt income (fully or partially)**   **Deduction: retirement savings contribution**
>
> =
>
> **Fiscal taxable income**