Intergenerational Mobility, Gender Differences and the role of Out-Migration: new evidence from Spain

P. Javier Soria Espín[a]

Supervisor: Thomas Piketty[b]

Referee: Clara Martínez-Toledano[c]

Abstract

In this paper, I provide the first estimates of (income) intergenerational mobility in Spain using administrative data linking parents and children through tax returns and the rank-rank approach. Exploiting the richness of the data, I estimate relative and absolute mobility at various geographical levels. The results show that income mobility in Spain is higher than in the United States and Italy but lower than in Switzerland and Scandinavian countries. Geographical variation in mobility rates is remarkable in Spain but smaller than in other compared countries. In addition, daughters have systematically worse outcomes in both relative and absolute mobility measures than sons. Exploiting the high comparability degree of siblings that have the exact same values of observable characteristics, I document a positive quasi-causal effect of out-migration on various upward mobility outcomes.

Keywords: intergenerational mobility, gender, out-migration, inequality, public economics

JEL codes: E24, J16, J31, J61, J62, R1

---

[a] Paris School of Economics - Thesis submitted to obtain the degree of Master in Public Policy and Development (PPD).

[b] Paris School of Economics & World Inequality Lab (WIL).

[c] Imperial College Business School & World Inequality Lab (WIL).

To my grandparents, who keep supporting me from heaven.

Acknowledgments

I am very grateful to my supervisor Thomas Piketty for giving me the opportunity to work under his supervision, supporting this project from the very first moment and for his precious advises to keep researching on these topics. I am also indebted to my referee Clara Martínez-Toledano, without her guidance an useful comments this master thesis would not be the same. Equally important, I am really thankful to Octavio Medina, Kiko Llaneras and the *Fundación Felipe González* for allowing me to collaborate in the fantastic *Atlas de Oportunidades* project.

On a more personal note, I have no words to express my enormous gratitude to Martín Fernández Sánchez, who has guided and supported me from the very first day I arrived to the Paris School of Economics. Martín has always been a mirror to look at for inspiration and his friendship has been key in my journey towards the PhD. Additionally, I am really grateful to have had the opportunity of meeting my amazing colleagues of the PPD 21' cohort. In particular, I thank Marina Cauchy, Yann David, Emma Faure, Florentine Friedrich, Éloi Flamant, Chiara Loriaux, Paola Mallia, Max Molaro, Paolo Monteiro de Macedo, Gillian O'Connell, Ana Rivera, Pablo Sánchez and Maria Souza for their fantastic friendships and constant mutual support during these two years. I want to thank as well the very good friends I have made in APE, especially Charles de Fornel, Ruggero Doino, María José González, Mario Herrera, Donia Kamel and Rodrigo Pazos. Also, I am obliged to Ander Iraizoz, Eric Tesckhe, Juliet-Nil Uraz and all the other PhD students I have met; they have been a very healthy source of inspiration during this last year.
I do not want to miss the opportunity to thank all my friends from the Collège d'Espagne for all the good moments lived together in such a difficult year marked by the Covid-19 pandemic. In this sense, I am immensely thankful to Javier Carbonell, Júlia Codina, Michele Gentile, Paz Ferrer, Berna León and Ainhoa Ugartetxe for having shared with me the best moments of the year and for having wholeheartedly supported me in every aspect of my life. In addition, I want to thank Belén Yu Irureta-Goyena for her genuine interest in this master thesis and for her useful comments.

Finalmente, estoy infinitamente agradecido a mi madre, a mi padre, a mi hermano Álvaro y a mis amigos de Murcia y Valencia por haber siempre creído en mí, incluso cuando ni siquiera yo era capaz, y por haber apoyado todas mis decisiones personales y académicas para conseguir mis objetivos. Todos mis éxitos os los debo a vosotros.

Contents

List of Figures

List of Tables

1 Introduction

Inequality is one of the most pressing issues both in social sciences and policy debate. The huge increase in income inequality over the last decades has raised public concern about the deterioration of equality of opportunities (Alvaredo et al., 2018; Piketty, 2020). An excellent indicator of equality of opportunities is the intergenerational mobility of income, since it measures the extent to which the income of parents influences the income of their children as adults. Hence, a society with high levels of intergenerational mobility is one where an individual's economic success is less dependent on the socioeconomic status of their parents and which, consequently, provides more opportunities to its members.

Intergenerational mobility is important for several reasons. Firstly, it is a matter of fairness, derived from the fact that inequality tends to persist over generations. Families from the highest deciles of the income distribution can transmit a wide range of benefits such as better education and childhood environments (Chetty and Hendren, 2018a,b; Chetty et al., 2020), appropriate health practices (Abel, 2008; Chetty et al., 2016b; Matthew and Brodersen, 2018), large economic inheritances (Korom, 2016; Fessler and Schürz, 2018) and high levels of social and cultural capital (Bourdieu, 1987, 2011) that families at the lowest deciles cannot. These disparities tend to linger over time and, as a consequence, opportunities remain substantially diverse for the children growing up in top-income families in comparison to those growing up in low-income families. Secondly, a high level of intergenerational mobility is not only desirable in terms of fairness, but also in terms of economic efficiency: the loss of talent due to fewer opportunities of disadvantaged backgrounds children is detrimental to innovation (Aghion et al., 2017; Bell et al., 2019) and growth (Van der Weide and Milanovic, 2018).

Despite its relevance, there is little reliable empirical evidence on intergenerational mobility. This is mainly due to the shortage of data that allow researchers to (a) exhaustively link parents and their children and (b) to build robust measures of permanent income for both generations. In fact, classical estimates of intergenerational mobility in Western countries are often based on survey data, small samples of parental-children cohorts or a broad use of imputed income (Bratberg et al., 2017; Blanden, 2009; Black and Devereux, 2010; Solon, 1999). These limitations, together with differences in measurement, have prevented us from reaching decisive conclusions from traditional cross-country comparisons (Solon, 2002). In their seminal paper, Chetty et al. (2014) overcome these problems by exploiting extensive administrative data linking millions of parents and children through tax records and using a rank-rank specification to estimate intergenerational mobility. This rank-rank approach consists in regressing the child's percentile rank in the income distribution of the children of a specific cohort on their parents' percentile rank in the income distribution of parents with children from that cohort. The estimate of this regression is a good representation of the influence of the family percentile on their descendants' one since the rank-rank relationship appears to be almost perfectly linear.

This new approach, together with an increased access to similar administrative datasets, has triggered a new wave of comparable studies on intergenerational mobility in some countries such as Australia (Deutscher and Mazumder, 2020), Canada (Connolly et al., 2019), Denmark (Eriksen and Munk, 2020), Italy (Acciari et al., 2019), Sweden (Heidrich, 2017) or Switzerland (Chuard and Grassi, 2020), which has produced more precise estimations of intergenerational mobility, especially for children cohorts born in the late 1970s and the 1980s.

Nevertheless, the range of country studies using this new framework remains limited. To the best of my knowledge, no prior study has investigated income mobility across generations in Spain using this new approach and comprehensive administrative data. This country is an interesting scenario to study intergenerational mobility that can help to improve our understanding of the factors shaping it: Spain is a very decentralized state formed by autonomous communities (the equivalent of regions) that have a high degree of power in very relevant areas of public policy such as health, education, transports or housing policies. This is a very particular institutional structure that could help to clarify the effects of local policies on intergenerational mobility in quasi-experimental settings. Furthermore, with the particular exception of Italy (Acciari et al., 2019), no prior work has analyzed other European contexts beyond Nordic countries.

The existing literature on intergenerational mobility in Spain is limited in time span and geographical coverage and survey data. It mainly uses other variables different from income and relies on small samples. Several studies carry out historical analyses (focused on the 18th and 19th centuries) that examine social mobility in some Spanish regions (namely Madrid, Valencia and Guadalajara) and use archive data on literacy, education or family occupation (Santiago-Caballero, 2011; Santiago Caballero et al., 2018; Beltran Tapia and de Miguel Salanova, 2021). Next, other studies develop innovative methodologies to analyze income mobility exploiting the socio-economic information conveyed in surnames. Collado et al. (2012) analyzes two regions, Cantabria & Murcia, focusing on the long-run intergenerational mobility of occupation and education over the 20th century. Güell et al. (2015) also take advantage of surnames information exploiting the 2001 census data of Cataluña to analyze educational mobility over generations. Regarding the analysis of intergenerational mobility of education, De Pablos Escobar and Gil Izquierdo (2016) explore the 2005 Spanish Intergenerational Transmission of Poverty survey and show a huge improvement in the access to education and completion rates for women over the past century but it does not translate to significant gains in the labor market. There are very few studies that examine income mobility (and not other outcome variables) across generations. In this sense, Cervini-Plá (2015) analyzes the intergenerational elasticity of income for children cohorts born between the 1950s and the 1970s exploiting survey data from the Spanish Income and Living Conditions survey. She shows that Spain is located between high and low mobility countries. In companion papers, the author provides potential explanations for this fact such as social referral to fill jobs, intergenerational persistence of occupation or a more intense process of assortative mating (Cervini-Plá, 2012; Cervini Plá and Ramos, 2013). Yet, these works are based on a small sample for children cohorts way before the 1980s. In sum, the limitations of these studies render the intergenerational mobility estimates for Spain rare, imprecise and, importantly, not comparable to the country studies circumscribed in the so-called new wave of literature on intergenerational mobility.

In this paper, I provide the first estimates of intergenerational mobility of income in Spain using the rank-rank approach and rich administrative data that link parents and children through tax declarations, following Chetty et al. (2014). To this end, I exploit a new dataset from the Spanish Tax Agency[1] including 2.7 millions of children born between 1980 and 1990 that are matched to their parents through income tax declarations. For the parents' households (observed in 1998), there is information about their gross and net income both at the individual level and their new household. For children (observed as adults in 2016), there is also information about their gross and net income both at the individual level and their new household. Exploiting the richness of this dataset, I estimate relative and absolute mobility at various geographical levels (national, regional, provincial and municipal), providing a detailed picture of the geographic variation in intergenerational mobility. In addition, I investigate gender differences in these estimates at the same geographical levels, something under-explored in the recent literature of intergenerational mobility. Next, I explore the relationship between out-migration and upward mobility. In particular, I exploit the high comparability degree of siblings that have the exact same values of observable characteristics to estimate the quasi-causal effect of leaving the home province on various upward mobility outcomes.

[1] This dataset has been retrieved from the Spanish Tax Agency by the project *Atlas de Oportunidades*, jointly funded by the *Fundación Felipe González* and *Fundación COTEC*

I find that the relationship between mean child percentile and family percentile is noticeably linear along the family income distribution except for the Top 10%, where it rises slightly faster. The estimated slope of this relationship, the rank-rank slope (RRS), is my main measure of relative mobility since it summarizes to which extent the outcomes of the low-income families are different from the outcomes of high-income families. At the national level, the RRS is 0.195 for Spain. This means that a 10 percentile point increase in family percentile is associated, on average, with a 1.95 percentile increase in a child's income percentile. Regarding absolute mobility, I estimate that a child coming from families located in the 25th percentile of the family income distribution (this is, below median families[2]) is expected to attain, on average, the 45th percentile of their own income distribution as an adult. I call this measure absolute upward mobility (AUM). In addition, I show that the probability of reaching the top quintile as an adult coming from a bottom quantile family, P(Q5|Q1), is 12.27%. On the other hand, the probability of staying at the top quantile as an adult coming from a top quintile family is 33%. Furthermore, I calculate the relative probability of getting to the top of the distribution depending on family background. This is a measure of intergenerational mobility at the very top of the distribution. For instance, among the children that make it to the Top 1% as adults, 9% of them come from Top 1% households whereas this percentage is 20.65% for those coming from Bottom 10% households. Therefore, it is 24 times more likely to end up in the Top 1% as an adult coming from a Top 1% household than from a bottom-decile household. This finding shows the disproportionate advantage given to children by top-percentile families to finish in the very top of the distribution as adults. Moreover, I estimate the probability that a child earns more than their parents. In Spain, this probability is 40.39%. Where does Spain stand among other comparable countries? These intergenerational mobility estimates place Spain somewhere in the middle between high mobility countries as Australia or Switzerland and low mobility ones as the United States or Italy.

Family income has a clear influence on children's income as adults but the region in which they grow up can substantially enlarge or shrink this family income dependency. Three main facts arise from the analysis of geographical differences in intergenerational mobility. Firstly, there is a high variance in mobility estimates across Spanish provinces. However, this within-country variation is smaller than the one estimated by similar studies doing geographical analyses as the United States, Italy or Switzerland. The most mobile areas tend to be located in the North/North-East of the

[2]Given the linearity of the rank-rank relationship, a family located in the 25th percentile of the parental income distribution is a good representation of the average family below the median

country whereas the less mobile ones are mainly located in the South/South-West. The region with the highest level of both absolute and relative mobility is Cataluña, with mobility rates on the levels of Scandinavia. The regions with the lowest levels of absolute and relative mobility are Andalucia and Canarias, with absolute mobility estimates similar to Southern United States ones. Secondly, there is a positive association between relative and absolute mobility. Areas that show high levels of absolute mobility (i.e., that provide more opportunities to children coming from poor families) tend to display high rates of relative mobility (i.e., the income gap between children born in poor families and rich families is smaller) as well, with the important exception of Canarias. Thirdly, there is a negative association between income inequality (as measured by the Gini Index) and absolute mobility measures and a positive one with relative mobility measures. Hence, provinces that have higher levels of income inequality tend to be less mobile both in relative and absolute terms, confirming the existence of a *Great Gatsby Curve* within Spain. This result adds within-country evidence to the literature exploring the negative relationship between intergenerational mobility and income inequality (Corak, 2013).

Daughters have systematically worse intergenerational mobility outcomes than sons both in relative and absolute terms and across different geographical levels. For the country as a whole, I find a RRS of 0.211 for women and 0.179 for men, which indicates a higher influence of family income on the adulthood income for daughters than for sons. In fact, daughters who grew up in median-income households end up, on average, at the 46th percentile while the sons of those same families reach the 52th percentile. This corresponds to an average income gap of €2,796 (a 13% of the national mean income). Furthermore, I document a persistent and heterogeneous gender gap at every geographical level. Gender gaps tend to be greater in provinces that present low relative mobility (higher RRS) but there is mixed evidence regarding the association between gender gaps and absolute mobility. There is no conclusive evidence concerning the relationship between provincial income and gender gaps.

Finally, I identify a positive association between out-migration[3] and upward mobility, regardless of children's parental income. In addition, the vast majority of children that leave their home province (*movers*) migrate to richer provinces than the home one, which points towards an economically driven migration. Examining the family origins of these movers to richer provinces, I uncover an U-shaped pattern, which shows that the majority of this type of movers come from either relatively poor or relatively rich households. This finding reinforces the economically driven out-migration argument since

[3]The migration that takes place within a country, i.e. from one geographical area, whatever it is, to another are within the same country

the children that move out to richer provinces are *predominantly* those that have the higher incentives to do so, either to improve their well-being ("looking-for-opportunities" migration) or to maintain themselves at the top of the social ladder ("keep-the-status" migration). In an attempt of better capturing the effect of leaving the home province on upward mobility, I exploit the high comparability degree of siblings that have the exact same values of observable characteristics to estimate the quasi-causal effect of out-migration across various upward mobility outcomes. I estimate that moving out from the home province increases, on average, child income by almost €8000 and the child rank in their own income distribution by 16 percentiles.

The rest of the paper is organized as follows: Section 2 defines the measures of intergenerational mobility that I estimate. Section 3 describes the administrative dataset, explains the reasons behind the selection of the analysis sample and shows some descriptive statistics. Section 4 presents the main results at the national level for both relative and absolute mobility and put them in an international context. Section 5 delves into the geographical analysis. Section 6 is studies the relationship between out-migration and intergenerational mobility. Section 7 concludes.

2 Measures of Intergenerational Mobility

The measures of intergenerational mobility are usually divided in two broad categories: relative mobility and absolute mobility. The first one reflects how different the outcomes of children from rich and poor families are while the second measures the extent to which children from a specific background (for instance, low income families) are better off than their parents when they become adults. The interest of each type of measures depends on the policy goal. From a Rawlsian standpoint, absolute mobility is probably the most interesting measures since they typically focus on how well the children of the poorest families end up. However, from a pure equality of opportunity point of view, relative mobility measures might be more interesting: even if all children are better-off (perfect absolute mobility), people would be still concerned about the fairness of the system if the they are always locked in the same part of the income distribution (null relative mobility). Hence, we see that both categories of measures are complementary and answer to different political questions.

Therefore, in this section I discuss the measures of relative and absolute mobility that I estimate in this paper. I follow the measures of Chetty et al. (2014), widely used in the recent literature on intergenerational mobility, to ensure the comparability of my estimates with the rest of the available country studies.

2.1 Relative Mobility

Relative mobility has been at the center of the traditional literature in the field. In particular, the vast majority of studies have estimated the so-called intergenerational income elasticity (IGE), which measures the influence of parental income on the income of their children as adults and it is mathematically described as follows:

$$\begin{gathered} Y_i = \alpha_i + \beta_{IGE}(X_i) + \epsilon_i \\ \beta_{IGE} = IGE = \rho_{XY}\frac{SD(Y_i)}{SD(X_i)} \end{gathered} \tag{1}$$

where Y_i is the child log income and X_i is the parent log income. β_{IGE} measures the difference in log income between children of high versus low income parents. Yet, Chetty et al. (2014) show that this measure is very unstable because the income distribution is not well approximated by a bivariate log-normal distribution. more specifically, the relationship between log parental and children income appears to be highly non-linear, which implies that the β_{IGE} is quite sensitive to the point of measurement in the income distribution. In addition, this specification rules out observations with zero income and the authors demonstrate that different ways of dealing with zero incomes yield substantially different estimates [4]

To solve these problems, Chetty et al. (2014) propose the use of the rank-rank specification, which consists in regressing the children income percentile rank as an adult on their parents' income percentile rank, as explained in the following equation (2):

$$\begin{gathered} C_i = \alpha_i + \beta_{Rank}(P_i) + u_i \\ \beta_{Rank} = \rho_{PC} \end{gathered} \tag{2}$$

C_i is the child percentile rank and P_i is the parent percentile rank (both in their respective distributions). Hence, β_{Rank} measures the association between a child's position in the income distribution and their parents' position in the distribution. This is a good measure of relative mobility since the slope $\beta_{Rank} * 100$ represents the difference between the income percentile of the children from the poorest and the richest parents. I will call this measure the rank-rank slope (RRS) throughout the rest of the paper. A high value of the RRS means that the difference in terms of income percentiles

[4] See Chetty et al. (2014) Section IV.A Subsection 1

between these two children is large, which implies a low level of relative mobility. On the contrary, a small RRS means that this difference is reduced and therefore relative mobility is high. For the sake of illustration, Figure A.1 in Appendix shows a graphical summary of the possible values for this measure of relative mobility. In this figure, the RRS would correspond to the slope of the black line.

Chetty et al. (2014) show that measuring income using percentiles produces more robust estimates. It reduces the influence of anomalous data and mitigates life cycle bias, since the income percentile tends to stabilize earlier in life than the income level. Importantly, they show that this relationship is mostly linear, which makes the estimate of equation (2) a robust summary of intergenerational mobility. Also, the rank-rank approach leads to the same standard deviation for children and parents income because both have an uniform distribution, which abstracts the RRS from changes in income inequality across generations.

Inspired by the approach of Acciari et al. (2019), I calculate what they call the Top Mobility ratio (TMR). The goal is to evaluate whether mobility at the top is different from mobility in the rest of the distribution due to the potential extra advantage given by very rich families. To this end, first I compute the rank-rank slope by running the rank-rank equation (2) on the top decile of the parental distribution (β_{9099}). Then, I run it on the bottom 90% and get (β_{8089}). Finally, I define top mobility ratio as follows:

$$TMR = \frac{\beta_{90-99}}{\beta_{80-89}} \tag{3}$$

The higher the TMR, the greater the rank persistent over generations in the top decile compared to the rest of the distribution. In addition, even if most of the recent studies show a strong linearity of the rank-rank relationship, some of them have found that it becomes slightly non-linear around the top decile of the parental income distribution (Acciari et al., 2019; Deutscher and Mazumder, 2020). Therefore, this ratio accounts for this potential non-linearity that may appear at the very top.

In this line, I also calculate an additional relative mobility measure to further explore the potential extra privilege of growing up in top percentile families: the relative probability of ending up in the Top 1% as an adult coming from a Top 1% household rather than from a bottom 10% household. The question behind this relative likelihood is the following one: how easy is to get to the very top of the distribution for the children of top percentile families in comparison to those from bottom decile families?. I call this measure relative probability of getting to the top (RPT).

To calculate the RTP, I follow three steps. Firstly, among the children that make it to the Top 1% as adults, I calculate the percentage of them coming from the Top 1% of the parental income distribution and the percentage of them coming from the Bottom 10% of the same distribution. Secondly, I calculate deviations in representation using a perfectly equal society as a benchmark and divide them to obtain the final relative likelihood. If we divide the parental distribution of income in percentiles, the probability of ending up in a given percentile as adults should be 1% for the children of each percentile. However, imagine, for instance, that the percentage of children in the Top 1% coming from top percentile households as well results to be 9%. This means that the children from top percentile households are 9 times over-represented in their adulthood Top 1%. Now imagine that the percentage of children in the Top 1% coming from bottom decile households is 4%. In a perfectly equal society, this figure should be 10% since the bottom decile represents 10% of the parental income distribution and this is what I use as a benchmark. This implies that bottom decile children are under represented: only a 40% (0.4) of them (4/10) are in the Top 1%. Therefore, the value of the RTP is the result of dividing the extra representation of top children (9) over the under-represented share of bottom decile children (0.4), which is 22.5. This indicates that it is 22.5 times easier to get to the Top 1% as an adult coming from a Top 1% family than from a bottom decile one [5]. In mathematical terms, the RPT can be expressed as follows:

$$RPT = \frac{\frac{[P(T1|T1)]}{1}}{\frac{[P(T1|B10)]}{10}} \tag{4}$$

where $P(T1|T1)$ is the probability of ending up in the Top 1% as an adult coming from the Top 1% and $P(T1|B10)$ is the probability of ending up in the Top 1% as an adult coming from the Bottom 10%.

2.2 Absolute Mobility

As previously mentioned, the focus of absolute mobility is to examine the adulthood performance of children coming from a specific part of the parental income distribution, especially from the bottom deciles of it. Following Chetty et al. (2014), I use two main measures to evaluate absolute mobility in Spain: (a) absolute upward mobility (AUM) and (b) the probability of reaching the top quintile coming from the bottom quintile (PQ1Q5).

[5]In addition, I do the same for the relative probability of getting to the Top 10% coming from a Top 10 % family rather than from a Bottom 10% one.

Absolute upward mobility (AUM) is defined as the mean rank (in the national child income distribution when they become adults) of the children that grow up in families at the 25th percentile of the parents income distribution. In other words, this statistic shows how far the children from families below the median end up in their adulthood income distribution[6] . Mathematially the AUM be described as follows:

$$\bar{R_{25}} = \alpha_i + \hat{\beta}(P_{25})$$
$$\bar{R_{25}} = E[C_i|P_i = 25] \tag{5}$$

The AUM can be easily retrieved from equation (2) in the following way

$$\bar{R_{25}} = \alpha + 25 * \beta = 50 - 25 * \beta \tag{6}$$

Th probability of reaching the top quintile coming from the bottom quintile (PQ1Q5) is obtained by calculating the share of children that end up in the highest quintile of the income distribution as adults growing up in bottom quintile families and it can be expressed as:

$$P(Q5|Q1) = P\{C_i \geq 80|P_i < 20\} \tag{7}$$

Beyond this two main measures of absolute mobility, I also calculate the probability of a child of earning more than their parents (PEM) by parental income rank once the two distributions have been re-scaled by their mean to capture aggregate growth over the period. More precisely, at each percentile p, this measure is defined as follows:

$$P_{c>p} = P\{Y_i \geq X_i * (\frac{\bar{Y_i}}{\bar{X_i}})|P_i = p\} \tag{8}$$

where Y_i is child income and X_i is parental income. In international comparisons, this measure can be a good complement since comparisons based on positional indexes (as the ones described above) can be affected by different levels of inequality. If inequality is higher in one country than in another, it will be more difficult to climb the income distribution ladder for the individuals of that country compared to the ones in the country with less inequality. For instance, one percentile increase in the high-inequality country will require a higher increase in income than in the low-inequality country because the distance

[6]If the rank-rank relationship is linear, the mean rank of children with below-median parental income equals the average rank of children with parents at the 25th percentile of the national income distribution

between percentile ranks is bigger. Nevertheless, this measure can be also problematic since, typically, parental income is measured at older ages than the children's adulthood income[7]. Since parents are older than children when their incomes are observed, there is a mechanical tendency towards capturing a lower percentage of children earning more than their parents just because it is harder, in general terms, to have a higher income than parents when children are only on their early 30's. Furthermore, this measure appears to be sensitive to disparities in local income distributions (Chetty et al., 2014). To sum up, the estimation of this measure is still interesting but it has to be interpreted bearing in mind the aforementioned caveats.

2.3 Measures for Within-Country Geographic Comparisons

the second part of the paper is devoted to the analysis of geographical differences in intergenerational mobility across in Spain. To this end, I employ the same approach and measures that at the national level but taking into account the geographic origin of children. Therefore, I run the same rank-rank regression from equation (2) but the dependent variable (R_{ig}) represents the mean percentile rank in the national distribution for a child i that grow up in a geographical area g (region, province, municipality) and the independent variable reflects its parental rank in the national distribution of parental income. Mathematically:

$$
\begin{gathered}
R_{ig} = \alpha_g + \beta_g(P_i) + u_{ig} \\
\beta_g = \rho_{PR}
\end{gathered}
\tag{9}
$$

Importantly, I keep ranking both children and parents based on their positions in the national income distribution (rather than the distribution within their region,province or municipality). As long as linearity holds, relative and absolute mobility at different geographical levels can be approximated can be approximated using the RRS (β_g) and the AUM (R_{25g}) measures :

$$
\bar{R}_{25g} = \alpha_g + 25 * \beta_g \tag{10}
$$

[7]This is usually the case in many recent studies. However, in my current data, I cannot check for this since there is no information about the age of parents.

3 Data

3.1 Description and source of the dataset

The data I exploit in this paper comes from the *Atlas de Oportunidades* project, jointly funded by the *Felipe González* and *COTEC* foundations. Inspired by the work of Chetty et al. (2014), the database focuses on the cohort of individuals born during the 1980s. In particular, it combines economic information from parents whose children were born from 1980 to 1990 with the economic information of those children 18 years later. This is achieved by combining the income tax returns of parents in 1998 and the income tax returns of their offspring in 2016, when they are adults[8]. The database also contains income information from the main types of tax declaration in Spain: *Modelo 100* (personal income tax for labor activities) and *Modelo 190* (mainly for self-employment activities). For the parents (observed in 1998), there is information about their gross (before tax and transfers) and net income both at the individual and household level as well as their location. For the children (observed in 2016), there is also information about their net income both at an individual level and the level of the new household. In addition, for the children, the database has information about their geographical origin, location as adults, gender, marital status, type of tax declaration (joint or individual) and the source of income (labor or self-employed). In total, the dataset includes 2,712,065 children, which represents more than 70% of the children born during that decade in Spain (see Figure A.2 in Appendix).

To the best of my knowledge, there are no other existing databases for Spain combining administrative economic information for millions of parents and their children to analyze intergenerational mobility. Consequently, thanks to this rich administrative data I can estimate, for the first time in Spain, relative and absolute mobility measures at various geographical levels (national, regional, provincial, municipal and post code), providing a detailed picture of geographic differences in income mobility for this country[9]. However, this database has some limitations. Firstly, in Spain there are two fiscal regimes: the special one (for Basque Country and Navarra) and the general one (for the rest of Spain)[10]. Therefore, since the data comes from the general Spanish tax Agency, it only covers 1998

[8]These tax returns are facilitated by the Spanish Tax Agency to the *Atlas de Oportunidades* project and they are not public since they include sensitive personal information of millions of individuals.

[9]An important point that should be made is that the detailed estimates presented in this paper represent the necessary first step in the search for causal mechanisms explaining upwards mobility and its geographical variation. I aim to study causal effects during the PhD but it is beyond the scope of this master thesis.

[10]To obtain more information on the origins, legal implementation and economic consequences of the special Basque regime see this and this

households under the general regime. This is, households living in Spain outside of the Basque Country and Navarra regions and filling tax returns in 1998 claiming a children born between 1980 and 1990 as descendant. Secondly, the raw database does not include information on the children of parents who did not file income tax returns in 1998. These households were probably among the poorest in Spain. The minimum personal income threshold to be obligated to declare was 550,000 *pesetas*, about €3,300 per year. Thirdly, parental income is observed, together with location, in 1998, when the children are between 8 and 18 years old. This is convenient because between these ages the vast majority of the children in Spain live with their parents, ensuring that we observe parental income while children are growing up at home[11]. Yet, since the dataset does not include other tax returns different from those of 1998 and 2016, I have no way of knowing how long they have been living there or when they moved out, if they did so[12]. Finally, there were some families for which the post code was not identified in the matching process. In particular, there are about 6,200 families with zip code "00000" that were eliminated, but they represent only 0.3% of the total.

3.2 Sample Selection

Two main potential sources of bias have been discussed in the literature addressing measurement issues in intergenerational mobility.

The first one is attenuation bias. Solon (1992) show that the intergenerational mobility estimates are smaller when based on one year of income for both generations relative to estimates based on an average of several years of income records. This is because using an income measure based only on one year can be a noisy representation of lifetime income, and attenuates the relationship between parents and children income percentile (RRS) due to classic measurement error. Unfortunately, in my current dataset, I am not able to check whether this estimate changes when averaging several years of income records since I only have one tax declaration for parents and one for children. However, as empirically shown by (Chetty et al., 2014) in their Figure III, the change in RRS is negligible when one year of family income is used compared to when several years are used.

The second one is life-cycle bias, which indicates that measuring child income at early ages can underestimate the influence of family income, because children with high lifetime incomes have steeper

[11]Young Europeans leave the family home at an average age of 26 years, while in Spain it is 29.3 years, according to data published by (2018). Therefore, it is very unlikely that the children whose parental income is observed in 1998 were living out of the family household.

[12]I only know whether the child lives in the same geographical area in 2016 as the one in which they grew up or not.

earnings profiles when they are young (especially those pursuing higher education studies). Put differently, child income tends to stabilize at older ages, when the income growth perspectives are smaller than those of younger adults. As discussed in the previous section, this concern is partially mitigated by the use of percentiles rather than income levels because an individual's income percentile stabilizes earlier in their life than their income level (Chetty et al., 2014). However, the rank-rank approach is not fully exempt from this potential bias. Checking for this bias, the authors found that intergenerational mobility estimates show very little life-cycle bias when the child income is measured after the age of 30, which reinforces the conclusion of Haider and Solon (2006). Importantly, the few recent studies exploiting administrative datasets for different countries also show slight life-cycle bias when they use child income as adults on their 30's (Deutscher and Mazumder, 2020; Connolly et al., 2019; Eriksen and Munk, 2020; Acciari et al., 2019; Heidrich, 2017). Thus, this is not a huge concern in my current dataset since I have seven generations of children that are 30 or older when they are observed in 2016.

To minimize as much as possible this potential second bias, I restrict my analysis sample to the cohorts born between 1980 and 1986 because they are 30 to 36 years old when their income is observed in 2016. Importantly, this delimitation virtually rules out another potential issue: co-residence bias. In some other studies (Acciari et al., 2019; Chetty et al., 2014), parental income is measured when many of the children are in their early 20's but they still live with their parents. However, this selection of children is probably not a random selection of the population of children: co-residing with their parents at those ages appears to be highly dependent on being a college student, especially in the case of Italy (Acciari et al., 2019). Also, this is significantly predominant across families with low educated parents. Consequently, since college graduates coming from low educated families are one of the engines of upward mobility (because they are more likely to have high lifetime income profiles), the use of children between those ages still living with their parents may induce an endogenous overestimation of intergenerational mobility rates. In my sample selection, this is not a concern since parental income is observed when children are between 8 and 18 years old, substantially minimizing the possibility of over-sampling college students when parental income is measured. Then, the children income is observed as adults when they are 30 and 36 years old, an age range where the vast majority of individuals have finished higher-education studies, for those who decided to pursue them (see Table A.1 in Appendix to see the age structure of the data).

I call this selection the core sample (see Figure A.2 in Appendix). On top of this first restriction and to avoid noisy estimations, I leave out the individuals from the autonomous cities of Ceuta and

Melilla due to important sample-size limitations [13]. Furthermore, I discard those parents and children whose type of tax returns is labeled as "receives capital income" because there is no data about their actual income (they represent 2.4% of the total dataset). I also discard those individuals whose type of tax return is labeled as "other" for the same reason (they represent around 7% of the total dataset). After these restrictions, the core sample size has 1,492,107 observations (this is, matches of children and their parents)[14].

3.3 Descriptive Statistics

In Table A.2 (see Appendix), I present summary statistics for the core sample. On average, parents household gross[15] income in 1998 is €27,113 . For Parent 1, this figure is €20,805 and €6263 for Parent 2. The dataset does not include any information on the gender of each parent. Looking at the share of parents with positive income, all I know is that Parent 1 is usually the top earner (84.8% of times). Parent 1 income is almost always positive (99.5%) whereas Parent 2 income is positive only in the 26.9% of cases. Regarding the children statistics, the mean age of children in 2016 is almost 33 years. In addition, the vast majority of children are either married or single, being this last option more predominant. On average, child individual income in 2016 is €20,557 and €25,668 at the household level. In terms of gender, men earn on average €933 more than women.

4 National Results & International Comparisons

In this section, I firstly present the main results through estimations of the aforementioned measures of relative and absolute mobility, looking at differences by child gender. Next, I put these estimates in context comparing them with the recent literature on intergenerational mobility that uses similar children cohorts and administrative income data.

Regarding the income definitions used in the following analysis, I use parents household gross income as a measure for parental income and child individual gross income as a measure for child income[16]. There are two reasons for this selection. Firstly, I choose total household income for parents because

[13]Therefore, the core sample includes 46 provinces of the 15 remaining regions ("Comunidades Autonomas") since the basque Country and Navarra are not present in the dataset and I get rid of Ceuta and Melilla

[14]When looking at the total number of parents I observe that it is smaller than the total number of children. This is because two or more children might come from the same household (siblings)

[15]When I refer to income I intend gross income (before tax and transfers) unless otherwise specified

[16]In Appendix Figure A.3, density plots for both parents household income and child individual income are shown.

it is the best measure of the economic conditions under which the children grew up, rather than using only father or mother income. Secondly, I choose child individual income (and not household) to avoid capturing the effect of assortative mating. If I were to use child household income in the analysis, I could not separate the influence of the parental income from the influence of the income that the child's partner brings into the new household, which would generate uninformative estimates of income mobility. Furthermore, previous studies have estimated that, on average, about 50% of the covariance between parents' income and child family's income can be attributed to the person the child is married to, which indicates a high sensitivity of the household income level to the assortative mating process (Cervini-Plá, 2012; Cervini Plá and Ramos, 2013).

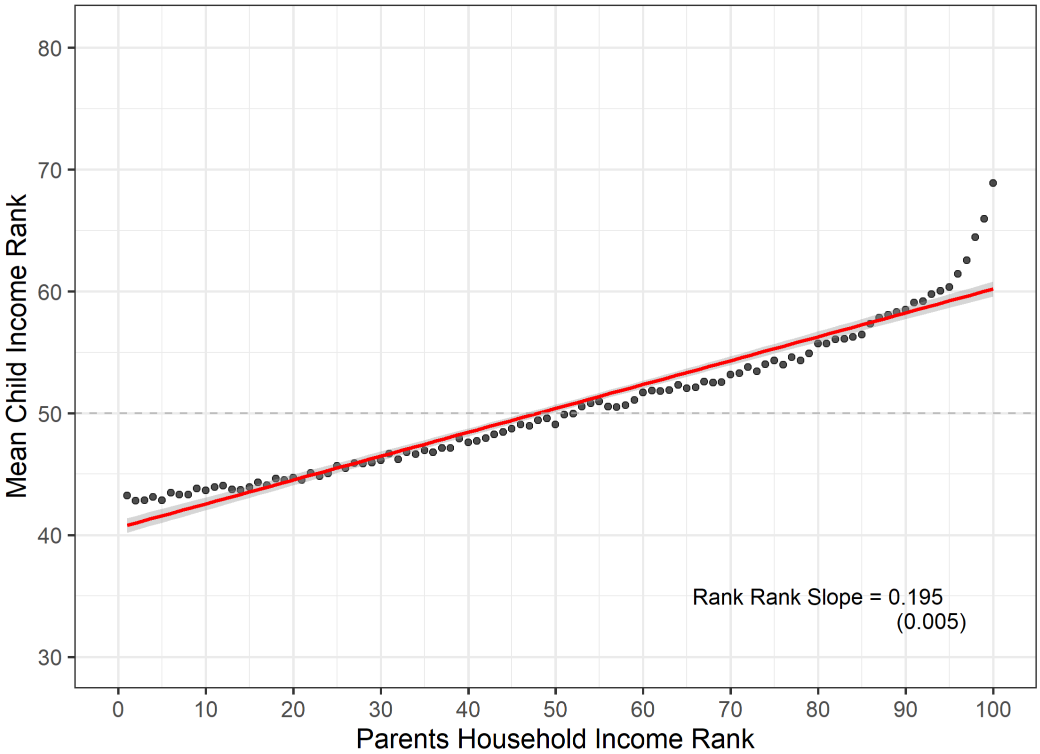
4.1 Relative Mobility Estimates

I start my analysis by measuring the income rank of parents as their percentile in the national distribution of the parents household income and the income rank of children as their percentile in the national distribution of the child individual income. Once I have ranked both parents and children by percentiles in their own income distribution, I calculate the mean (and median) percentile rank of children by parental income rank. Following the rank-rank approach described in Section 2, I regress the mean child percentile on the parents household income percentile to obtain the national rank-rank slope (RRS) estimate. The following Figure 1 presents a graphical summary of these first steps[17].

The first result that can be retrieved from this figure is that the relationship between parental and children percentiles is almost perfectly linear except for the very top percentiles of the distribution where it raises slightly faster, indicating higher rank persistence at the very top. This almost perfect linearity together with a small deviation at the very top is also present in many of the recent country studies mentioned before (Acciari et al., 2019; Heidrich, 2017; Deutscher and Mazumder, 2020; Eriksen and Munk, 2020). For the sake of completeness, I repeat the same process for median child rank and I compare it with the mean child rank to see whether the relationship with parental income rank changes (see Appendix Figure A.5). The relationship between median child rank and parental rank is still fairly linear, but it becomes much more non-linear at the top, starting from the 80th percentile. This disparity between median and the mean is due to the fact that the conditional distributions of child ranks are very skewed. In the bottom decile, the majority of children are located in the very bottom

[17] Figure A.4 places the rank-rank relationship for Spain within the relative mobility framework presented in Figure A.1

Figure 1: Association between Children's and Parents' Percentile Ranks in Spain



percentiles while at the top decile they are mostly concentrated in the very top percentiles (see Figure A.6 in Appendix). Consequently, the median is lower than the mean for low parental percentiles and bigger than the mean for high parental percentiles. These conditional distributions produce a steeper relationship between median child rank and parental rank compared to the one between the mean child rank and the parental rank. To exploit the advantages of the almost perfectly linear relationship in summarizing intergenerational mobility measures (and to make the estimates comparable with other country studies) I always use mean (and not median) child percentiles in this paper.

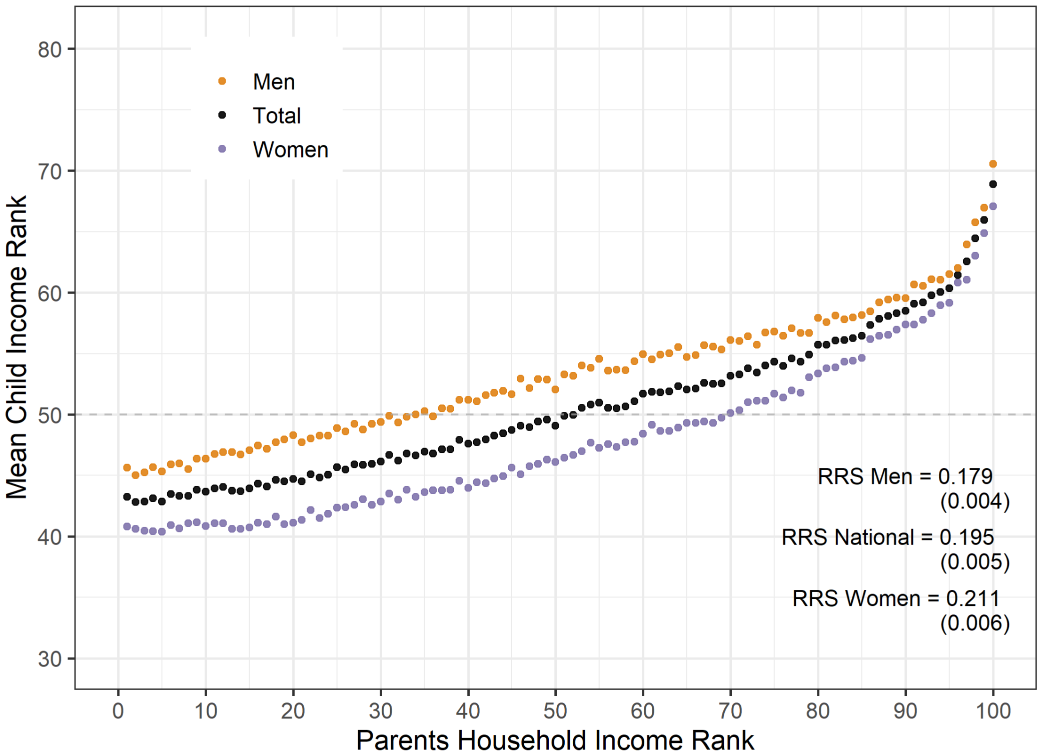
The national estimate of the RRS is 0.195, as shown in Table A.3, Column (1). Therefore, one percentage point (pp) increase in parents rank is associated with a 0.195 pp increase in the mean child rank[18]. This implies that the mean rank for children whose parents are in the top rank is 19.5 percentiles higher than the mean rank for children whose parents are in the bottom rank[19]. To give a sense of the magnitude of differences, Appendix Figure A.7 reports the mean individual child income by parental rank. A child from a family located in the median of the distribution is expected to earn roughly €19,345. This average income is €16,877 for children coming from families at the 10th percentile whereas for children coming from families in at Top 1% it is almost €23,000higher (€39,602).

[18]Alternatively, an increase of 10 pp in parental rank is associated with a 1.95 pp increase in the mean child rank

[19]This corresponds to the distance between the top and the bottom of the red fitting line

The data reveals a clear gender difference: relative mobility is higher for sons than for daughters. Figure 2 presents the rank-rank relationship at the national level by gender. As reported in Table A.3 Columns (2) & (3), the RRS for men (0.179) is lower than the one for women (0.211), indicating a higher influence of family household income on the adulthood income (less relative mobility) for women than for men. Daughters who grew up in median income households end up, on average, at the 46th centile while the sons of those same families finish at the 52th centile. This corresponds to an average income gap of €2796 for children growing up in families at the median[20]. Interestingly, a gender gap is observed for all parental ranks, although its extent dramatically shrinks for the Top 5% of the parental income distribution, which are families above €61,471.

Figure 2: Association between Children's and Parents' Percentile Ranks in Spain by Gender



To account for a potential extra persistence at the very top of the distribution (present in this data), I proposed in Section 2 a metric called top mobility ratio (TMR), following Acciari et al. (2019). The estimates by gender of this measure are reported in the bottom part of the Table A.3. The national TMR is 5.45 and, although the estimate for men (5.81) is higher than for women (5.16), the difference is rather small, which is in line with the tiny gender gap observed in this top part of the distribution. This high value of the TMR confirms that rank persistence is much stronger at the top of the income

[20]Figure A.8 in Appendix summarizes the average child income by gender ordered by parental rank. Also, this gender gap remains fundamentally unchanged over the different cohorts included in the core sample, as reported in Figure A.9

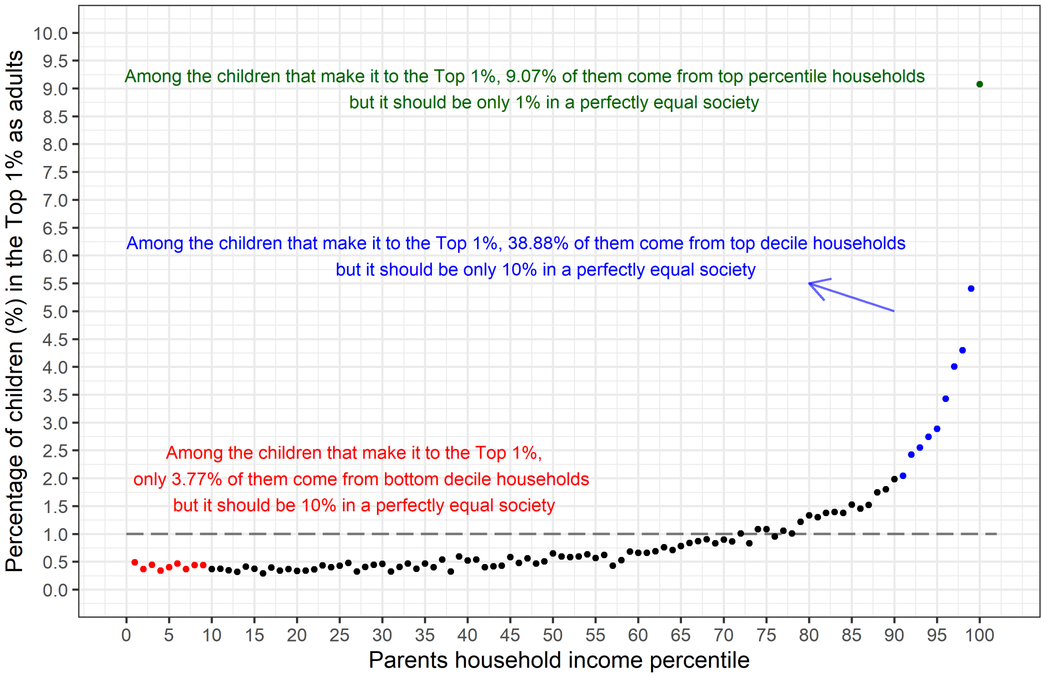
distribution[21]. To put the Spanish TMR in context, it is worth mentioning that the Italian TMR is 3.7 at the national level (Acciari et al., 2019). This implies that rank persistence in the top decile is higher in Spain than in Italy, the only country for which this metric is estimated. In Spain, the RRS for the bottom 90% of the income distribution (the denominator of the TMR) is 0.17 and the RRS for the Top 10% (the numerator) is 0.92. This value of the numerator means that there is a difference of almost 10 percentiles between the mean child rank of children from families in the the 99th percentile and one the of children from families in the 90th, showing a remarkable level of inequality of opportunities within the top decile.

To further investigate this extra advantage given by top-percentile families, in Section 2 I proposed a new metric: the relative probability of getting to the top (RPT). Among the children that are in the Top 1% as adults, in Figure 3, I report the percentage of them coming from Top 1% households (green), from Top 10% households (blue) and from Bottom 10% households (red). We see that the probability of a child to end up in the Top 1% coming from a top percentile household is 9.07% whereas this figure is only 3.77% for a child growing up in a bottom decile household. Therefore, the RTP in this case is 24: it is 24 times more likely to get to the top as adult coming from a Top 1% household than from a Bottom 10% household. Furthermore, comparing the percentages of children coming from Top 1% and Top 10% households I obtain that its 2.3 times easier to get to the top percentile as an adult coming from a Top 1% family than from a Top 10% in general, a reflection of the high rank-rank persistence at the very top reflected by the TMR estimates. A summary of estimates from different versions of the RTP disaggregated by gender can be found in Appendix Table A.4 [22]. Looking at this table, a clear pattern emerges: the estimates for women are always larger than those for men, indicating a higher influence of coming from a top family to end up in the top of the distribution for daughters compared to sons.

[21] As I have shown in Figure A.6, this is induced by the fact that the income distribution is skewed to the right and hence percentiles are further apart at the top compared to the middle. In addition, percentiles are closer to each other at the very bottom bottom, which explains the flattening of the rank-rank relationship for the first percentiles.

[22] Similar graphs showing the percentages used for the calculation of these RTPs can be found from Figure A.10 to Figure A.12

Figure 3: Relative Probability of Getting to the Top 1% (RPT-Top 1%)



4.2 Absolute Mobility Estimates

The first main measure of absolute mobility that I estimate is the absolute upward mobility (AUM), which is defined as the mean rank for children whose parents are at the 25th percentile of the national parental income distribution ($\bar{C_{25}} = E[C_i|P_i = 25]$). The national AUM is 45. This means that a child growing up in a family below the median is expected to end up in the 45th percentile of their own income distribution as an adult. Therefore, on average, kids coming from families below the median improve their position in the income distribution but remain below the median of the child income distribution. As reported in the second row of Table 1, there is a significant gender gap in this measure as well: sons end up, on average, 6 percentiles above daughters in their adulthood income distribution. In addition, Appendix Figure A.13 shows that the national estimate fluctuates very little around the 45th percentile for most of the generations included in the analysis and the gender gap remains essentially unchanged.

The second main measure of absolute mobility is the probability of reaching the top quintile as an adult coming from a bottom quintile household (P(Q5|Q1)). In a perfectly equal society (i.e., a society where the parental income had no influence on their children income), this probability should be 20% since every child would have the same likelihood to end up in the top quintile coming from any quintile. However, in Spain, this probability is 12% at the national level[23]. Figure 4 graphically summarizes the

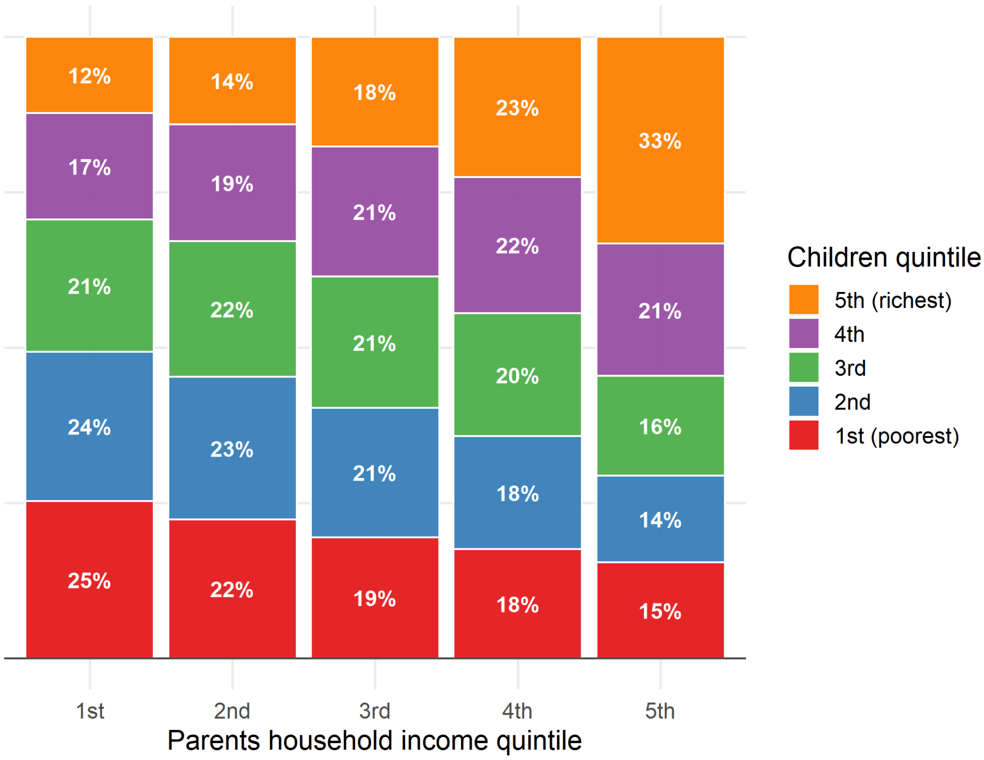
[23]This probability is called "rags to riches" or "American Dream measure" in other recent studies (Corak, 2020; Chuard

quintile transition matrix [24]. To put this jump in the intergenerational income ladder in context, it can be said that 12% of children coming from families whose household is below €11,190 (bottom parental quintile) finish with an average individual income above €29,223 as adults (top child quintile). Again, the analysis shows a remarkable gender gap: the probability of moving to the top from the bottom is almost 5pp higher for sons than for daughters (Appendix Figure A.14). On the other hand, the probability of growing up in a top quintile and staying in the same quintile as adults (P(Q5|Q5)) is quite high: 33%. In addition, 1 out of 4 children that are raised in a bottom quintile family stay in the bottom quintile as adults (P(Q1|Q1)). Importantly, the measure of interest (P(Q5|Q1)) as well as the other transition probabilities are stable across the different cohorts of children (Appendix Table A.15). Furthermore, the gender gap barely varies across these generations (Appendix Table A.16).

and Grassi, 2020)

[24] A classical quintile transition matrix table with 2 decimals can be found in Appendix TableA.5

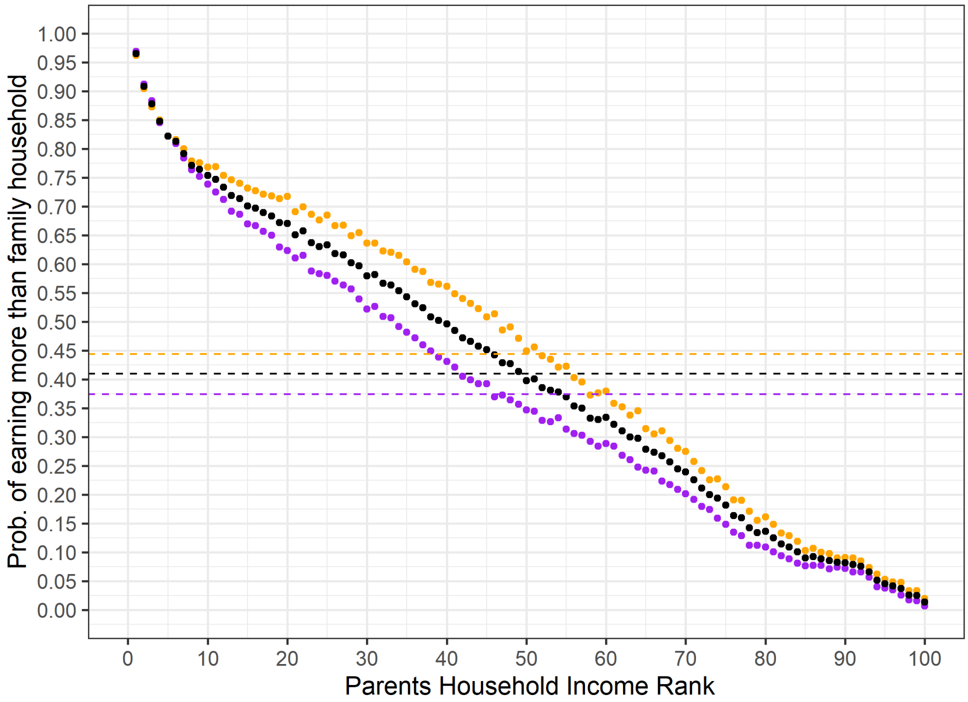
Figure 4: Probability of Ending Up in the Top Quintile Coming from a Household in the Bottom Quintile at the National Level



Beyond the main absolute mobility measures, I calculate as well the probability of a child to earn more than their parents as an adult (PEM). In particular, at each parental rank, I compute the percentage of children that individually earn more than both parents (household income). The results are reported in the following Figure 5. Up to the 40th parental rank, half of the children earn more than their parents. On average, this share is 40.39% for the total population (black dotted line), 44.38% for sons (orange dotted line) and only 37.42% for daughters (purple dotted line)[25]. This gender gap is appreciable for most of the parental income distribution (from the 10th parental rank to the 85th one). This is probably not surprising. For children coming from bottom decile households, it is much easier to earn more than their parents, especially in Spain where labor productivity and wages have steadily improved during the last 25 years. For children coming from Top 5% households the story is just the opposite: it is very difficult to overcome extremely high parental household incomes when an individual is only in their early 30's. Therefore, there is little room for gender disparities at the very bottom and very top of the income distribution.

[25] Appendix Figure A.17 shows the same probability but comparing children income with Parent 1 (father) income only

Figure 5: Probability of Earning More than Their Parents by Parental Rank (by Gender)



At last, Table 1 summarizes national estimates disaggregated by child gender of the different intergenerational mobility measures presented in this section. Three main findings can be highlighted. Firstly, the almost perfect linearity exhibited by the mean child income percentile and the parents income percentile allows to produce useful estimates of relative and absolute mobility in Spain through the use of the rank-rank linear regression. In particular, from this regression we can summarize absolute mobility (AUM) and relative mobility (RRS) with only two parameters. Secondly, the moderate non-linearity observed in the top percentiles is well accounted for by the top mobility ratio (TMR). This measure indicates a higher income (rank) persistence between parents and children in the top decile of the distribution compared to the rest of the distribution. In this line, the relative probability of getting to the top (RPT) provides a good sense of the relative advantage of growing up in top percentiles families in comparison to the bottom ones when it comes to finish in the top of the distribution as an adult. Thirdly, across all measures of both relative and absolute mobility, sons appear to have better economic outcomes than daughters. On average, sons experience a higher degree of equality of opportunities (lower RRS), end up higher in the adulthood income distribution (larger AUM & P(Q5|Q1)) and are more likely to earn more than their parents (higher PEM) compared to daughters. Furthermore, this gender gap is also present when focusing in the dynamics of the top of the distribution (TMR & RTP measures), but it is much more reduced .

Table 1: Summary of Intergenerational Mobility Measures in Spain (by Gender)

Measure	National	Men	Women
Main measures			
RRS	0.196	0.179	0.211
AUM	45th	48th	42th
P(Q5|Q1)	12.2	15.1	10.2
Additional measures			
TMR	5.45	5.81	5.16
RTP	24.04	23.33	25.60
PEM	40.39%	44.38%	37.42%

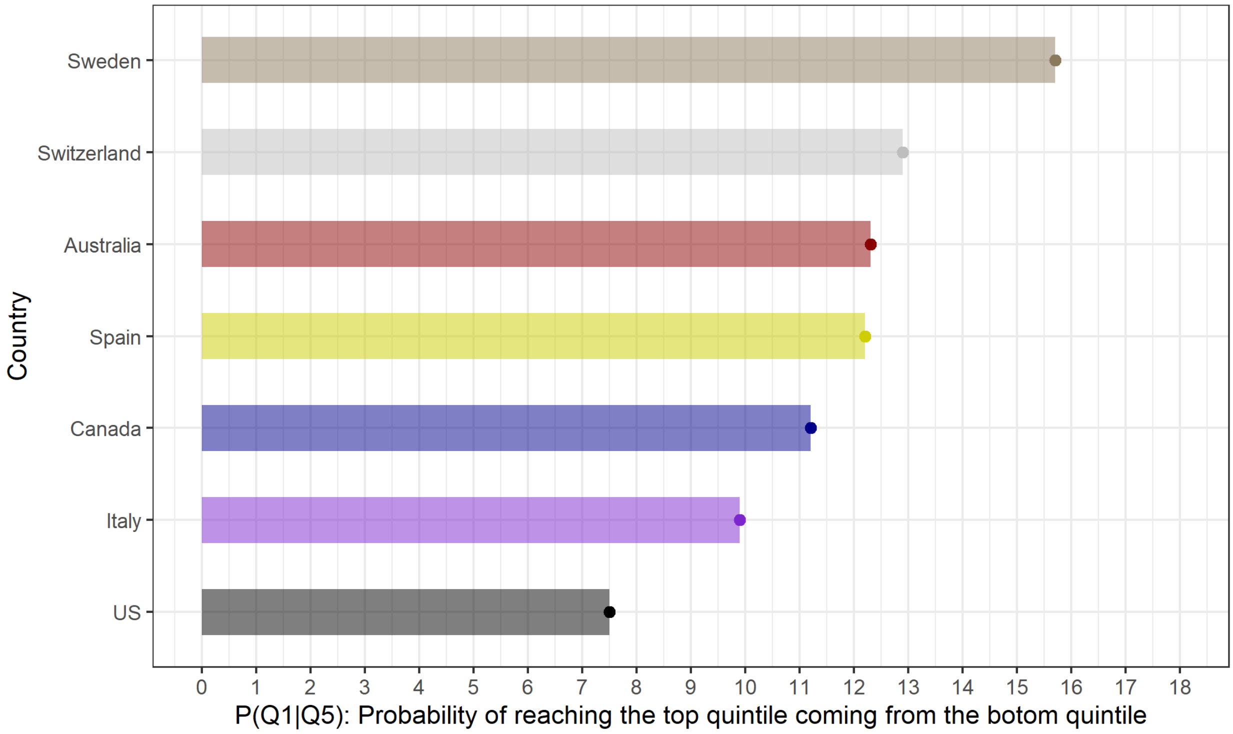
4.3 International Comparisons

Where does Spain stand in the global picture of intergenerational mobility? To better understand the extent of intergenerational mobility in Spain, Table 2 puts the above estimates in an international context. An important caveat is that I restrict my international comparison to those country studies using (i) similar children cohorts, (ii) the rank-rank approach and (iii) income data coming from administrative sources. Following these selection criteria, the resulting comparable countries are Australia (Deutscher and Mazumder, 2020), Canada (Connolly et al., 2019), Denmark (Eriksen and Munk, 2020), Italy (Acciari et al., 2019), Sweden (Heidrich, 2017), Switzerland (Chuard and Grassi, 2020) and the United States (Chetty et al., 2014). An additional caveat is that for Denmark and the United States, the rank-rank estimates use both parents and children household income instead of only individual income as the rest of the studies[26]

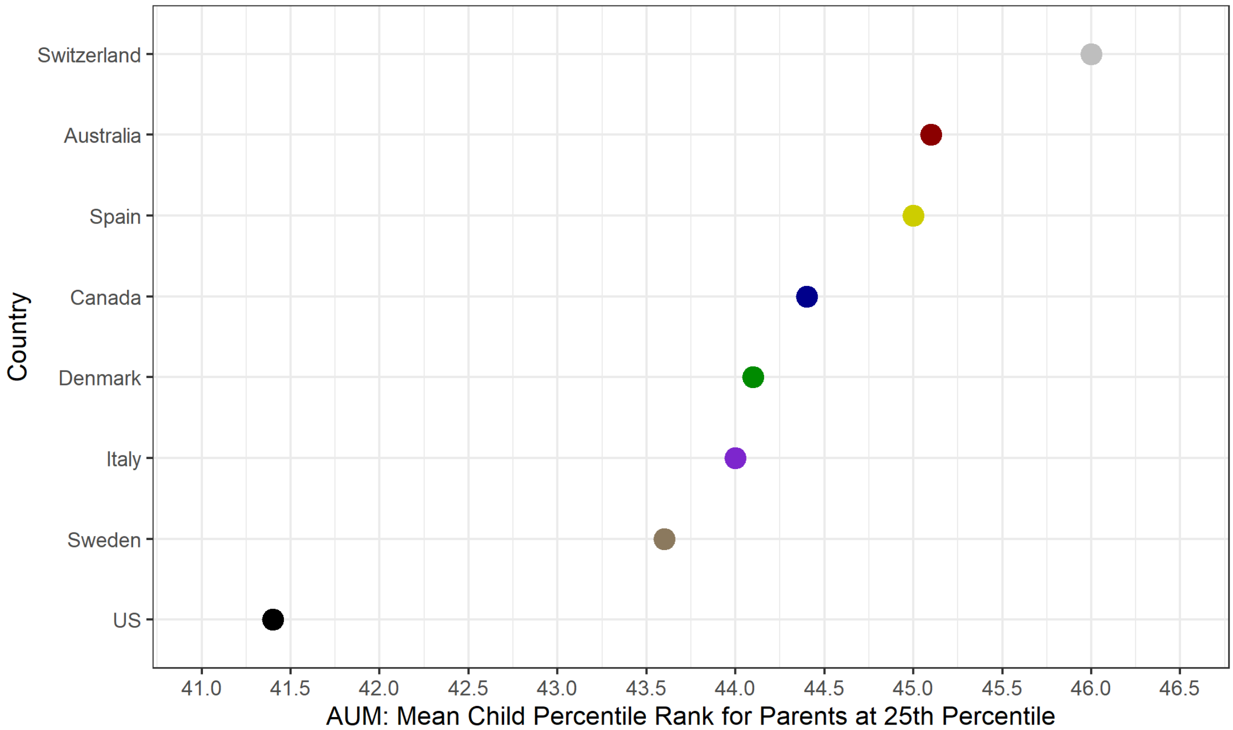
Figure 6 focuses only on absolute mobility and compares the estimates of Spain with the ones provided by these country studies. The top sub-figure reports cross-country estimates of the probability of reaching the top quintile coming from the bottom family quintile (P(Q5|Q1)). The bottom sub-figure does the same but for absolute upward mobility (AUM) estimates. The first result that arises from this comparison is the significant heterogeneity in absolute mobility estimates across countries, especially in the probability of climbing to the top. The second result is that Spain is somewhere in

[26]In the case of Switzerland, they use both definitions of children income (household and individual) and show that estimates barely change.

the middle between high upwards mobility countries as Australia or Switzerland and very low upward mobility ones as the United States or Italy.



(a) P(Q5|Q1)



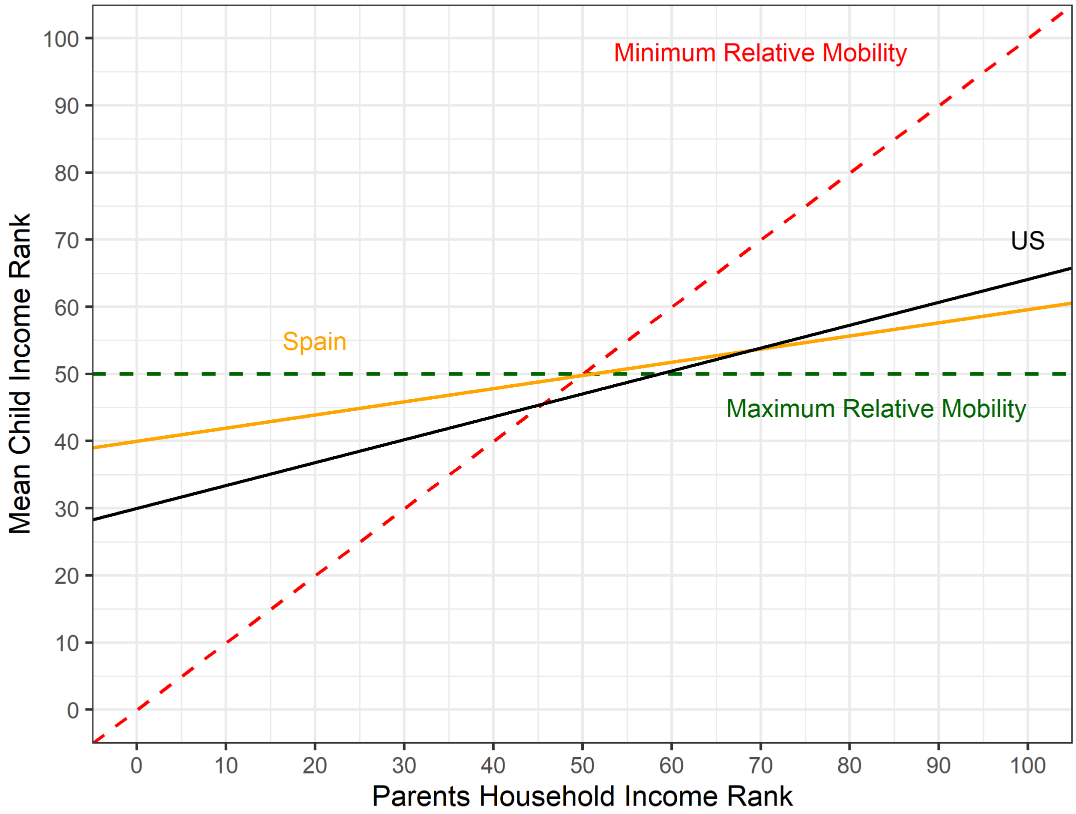
(b) AUM

Figure 6: Absolute Mobility Estimates in International Perspective

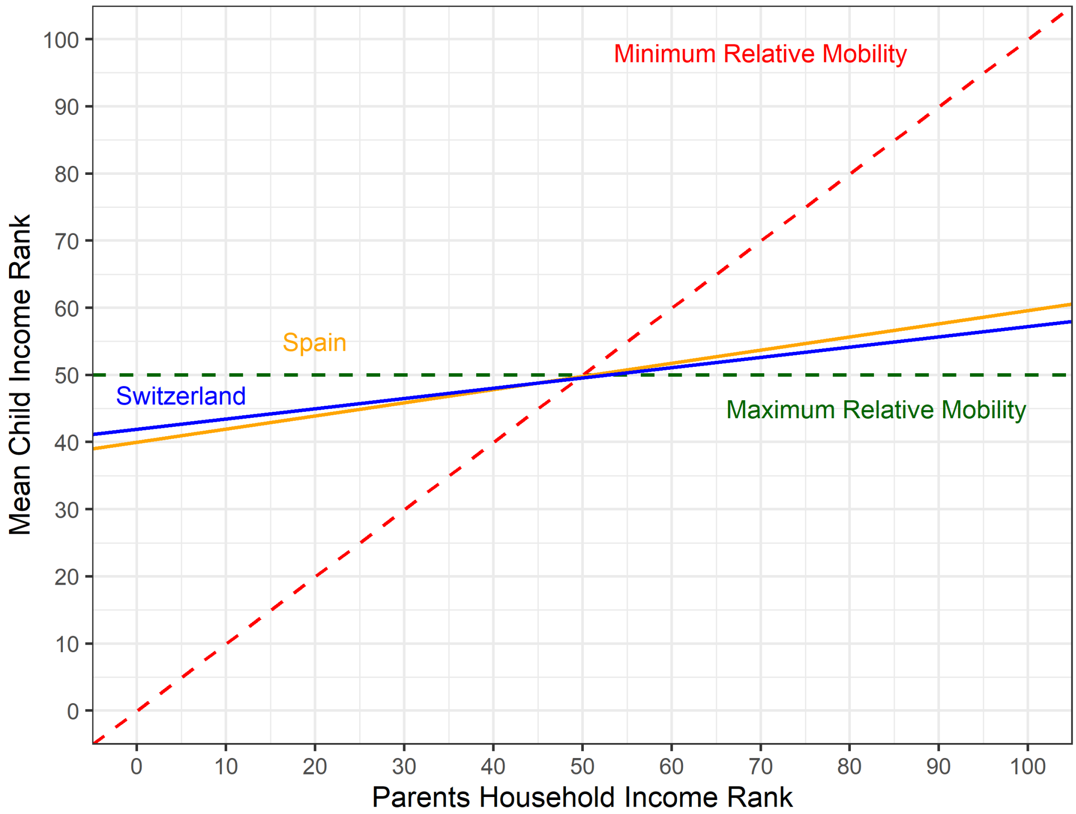
To better interpret the behavior of relative mobility in Spain, Figure 7 compares the rank-rank slope (RRS) across countries. In the top sub-figure, I plot the fitting lines of the rank-rank regression for Spain (RRS of 0.196) and for the country with the lowest level of relative mobility (highest RRS, 0.341), the United States, within the framework presented in Appendix Figure A.1. Spain appears to be much more egalitarian than the United States from a pure equality of opportunities standpoint[27] since the distance in percentiles between children coming from top-rank families and bottom rank families is, on average, lower in Spain than in the United States. In fact, we can graphically see that the American line is closer to the "minimum relative mobility" benchmark (dotted red line) than the Spanish line. Furthermore, children in Spain end up, on average, higher in their own income distribution than children in the United States until the 70th percentile. From this point to the top of the distribution of parental income, American top kids reach, on average, a higher percentile in their own distribution than the Spanish ones. This disparity before and after the 70th percentile is actually explained by the observed lower level of equality of opportunities in the United States: children from top parental deciles become much richer than the Spanish equivalent children but those coming from the bottom 70% are left with less upward mobility opportunities than bottom 70% Spanish children.

In the bottom sub-figure, I plot the fitting lines of the rank-rank regression for Spain and for the country with the highest level of relative mobility (lowest RRS, 0.156), Switzerland, within the framework presented in Appendix Figure A.1. In this case, the story is just the opposite. Although the difference between the Swiss line (in blue) and the Spanish one (in orange) are smaller than the difference between the American and Spanish lines, we can see that Spain is a less egalitarian country in terms of equality of opportunities than Switzerland. Approximately, from the median of the parental income distribution, Swiss children coming from bottom half families reach, on average, a higher percentile in their own income distribution than bottom half Spanish children. The opposite is true for children coming from the top half of the parental distribution. In addition, I present a similar figure but comparing Spain to a very similar country, Italy (see appendix Figure A.18). We see that Spain is slightly more egalitarian then Italy. The interpretation is the same that for the Spain vs. United States comparison but gaps are much smaller.

[27] following the definition of relative mobility provided in Section 2



(a) Spain vs. United States



(b) Spain vs. Switzerland

Figure 7: Relative Mobility Estimate (RRS) in International Perspective

Finally, the last column of Table 2 presents some estimates of the probability of earning more than both parents (PEM). As I explained in Section 2, this measure has some advantages and drawbacks when it comes to international comparisons. Among the comparable studies, only Connolly et al. (2019) provide a similar estimate for Canada. We see that in that country the share of children earning more than their parents is, on average, roughly 22pp higher than in Spain. This huge gap is somewhat surprising since both countries have fairly similar estimates for other absolute mobility measures. The discrepancy is probably driven by differences in the age at which parental income is measured. Measuring parental income when parents are very old or very young has clear consequences on this metric: it is easier for early-30s-children to earn more than their parents when these are young than when they are old because they are more likely to have a lower income. Hence, it cannot be disentangled whether the difference between the Canadian and the Spanish rate is due to a higher degree of upwards mobility or to measurement biases. This lack of similar estimates for this measures calls for its homogenization, with measures of children and parents income at the same ages to properly study whether the former are doing better than later.[28]

Table 2: International Comparison of Intergenerational Mobility Estimates

Country	Author	Correlation	Children Cohorts	RRS	P(Q5|Q1)	AUM	PEM
Sweden	Heidrich, 2017	Child-Parents	1968-1976	0.197	15.70	43.60	-
Denmark	Eriksen & Munk, 2020	Family-Parents	1973-1977	0.211	-	44.10	-
Australia	Deutscher & Mazumder, 2020	Child-Parents	1978-1982	0.215	12.30	45.10	*
Canada	Conolly et al. 2019	Child-Parents	1980-1982	0.242	11.20	44.40	62.5%
United States	Chetty et al. 2014	Family-Parents	1980-1982	0.341	7.50	41.40	-
Italy	Acciari et al. 2019	Child-Parents	1972-1983	0.251	9.90	44.00	*
Spain	This Study	Child-Parents	1980-1986	0.196	12.20	45.00	40.39%
Switzerland	Chuard & Grassi, 2020	Both	1982-1984	0.156	12.90	46.00	38.1% (father)

Notes: This table reports intergenerational mobility estimates for comparable country studies. "Correlation" describes the type of income measures used in the rank-rank estimations. *Child-Parents* means that individual child income is regressed on total parents household income. *Family-Parents* means that child household income is regressed on total parents household income. *Both* means that both options are estimated. "Children Cohorts" describes the children cohorts of birth. "RRS" stands for rank-rank slope. "P(Q5|Q1)" is the probability of ending up in the top quintile coming from the bottom quintile (no available estimate for Denmark). "AUM" is absolute upwards mobility (the mean child rank for children coming from families at the 25th percentile). "PEM" reflects the average probability of earning more than both parents (in the case of Switzerland, only the father). *Italy and Australia compute slightly different measures: the probability that children earn 50% more than their parents. They present this statistic by parental income rank but not average rates and therefore they are not reported here. " - " means not available.

[28] as in the case of Switzerland, but they calculate the probability of earning more than the father/mother separately

5 The Geography of Intergenerational Mobility

The goal of this section is to investigate the influence of growing up in a particular area on intergenerational mobility. To better understand this exercise, consider two kids coming from families in the same percentile of the national parental income distribution but living in different geographical areas of the country. Then, the question is: where do these children, with similar family backgrounds but growing up in different areas, end up in their adulthood income distribution?. Generalizing this idea, I estimate the same measures of relative and absolute mobility as before but classifying children by the area in which they grew up[29]. To do so, I focus on provinces as the main level of geographical disaggregation to ensure comparability with recent studies exploring geographic heterogeneities. However, I also provide estimates for a higher geographical level (regions) and for a lower one (municipality).

This analysis exploiting geographical variation also allows me to analyze the relationship between relative and absolute mobility (do high relative mobility regions report high levels of absolute mobility as well?), further examine the gender gap patterns (what correlates with gender gaps?) and explore whether there is a "Great Gatsby" curve in Spain (is there a correlation between income inequality and intergenerational mobility?) while keeping constant the same general institutional setting.

5.1 A Brief Description of Spain

Spain is a country divided in 17 regions, known as "comunidades autonomas" (autonomous communities) and the 2 autonomous cities of Ceuta and Melilla ("ciudades autonomas") that have a high degree of independence in important areas of public policy as health, education or transports. Then, these regions are divided in a total of 50 provinces plus the 2 aforementioned autonomous cities. A provincial map of Spain where regions are highlighted in colors can be found in the Appendix Figure A.19. As I described in Section 3, due to different fiscal regimes, basque Country and Navarra are not included in this dataset. Also, Ceuta and Melilla are eliminated due to sample size limitations. Therefore, the core sample includes 15 regions and 46 provinces, which corresponds to the rest of Spain. Appendix Table A.6 reports summary statistics at the regional level for these 15 regions. We see that the richest regions in terms of parental income are Madrid, Cataluña and Balears whereas the poorest are Extremadura,

[29]The underlying assumption in this exercise is that a child grow up in a specific area if they were observe in that area when parents claimed him as descendant in the 1998 tax returns (this is, when parental income is observed). Then, when the child is an adult and does their own tax return in 2016, I observe whether they live in the same are in which they grew up or not.

Castilla la Mancha and Andalucia. In the case of child income, Madrid and Cataluña keep the top 2 of richest regions and the third one is now Aragon. The poorest one by child income are Canarias, Extremadura and Andalucia. In terms of population, Andalucia, Madrid and Cataluña are the top three regions in terms of children origin[30]

Next, to illustrate provincial disparities across generations, Appendix Figure A.20 shows heatmaps of both parents and child mean incomes by provinces.[31]. In general terms, we see that the richest provinces in terms of parental income tend to be also the richest in terms of child income and same thing for the poorest provinces. A surprising case is the one of Canary Islands, which has fairly high-income provinces when parental income is ranked but they appear to become low-income provinces when child income is examined.

5.2 Geographic Variation in Mobility Rates

5.2.1 Relative Mobility Differences

I begin my analysis by examining regional and provincial differences in relative mobility with a special focus on gender disparities among children. Figure 8 summarizes relative mobility estimates across regions. In Figure 8a, I plot the rank-rank relationship at the regional level. Companion Figure 8b reports the regional estimates of these slopes, showing remarkable gender disparities. Two main facts can be highlighted from these estimations. Firstly, there is a substantial variation in relative mobility across regions. Some regions have a remarkably high level of relative mobility, on the level of Switzerland, as Cataluña or La Rioja whereas other regions experience much higher parental income persistence levels, as Extremadura o Andalucia, similar to the ones of Southern United States. A surprising case in this respect is Madrid. Alongside with Cataluña, this is the richest region of the country, but appears to have a much lower level of equality of opportunities (relatively high RRS). This contrast with Cataluña is related to different income inequality levels across generations, something that I will further explore in this section. Secondly, there are substantial gender gaps in most of the regions: daughters systematically experience lower levels of relative mobility (higher RRS, in purple) than men (lower RRS, in orange). However, their magnitudes are very diverse. In Cataluña or Baleares

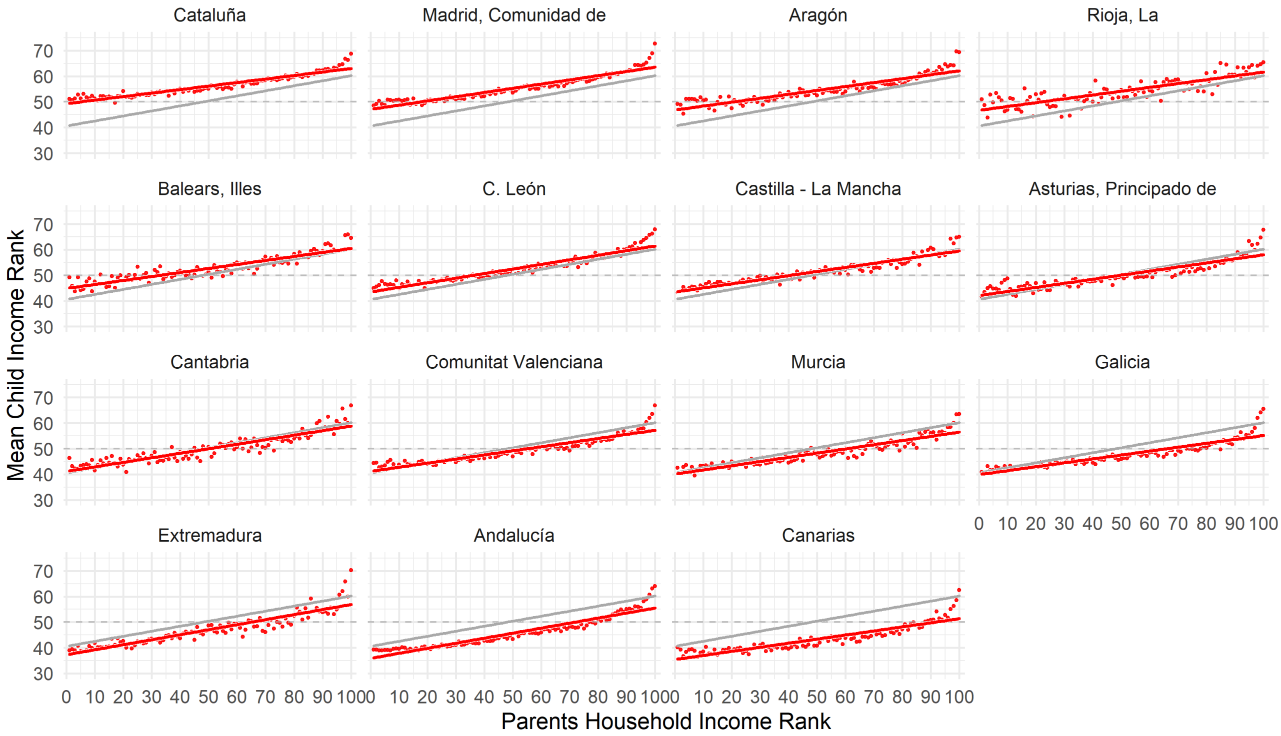
[30]Again, I assign a child to an specific area based on where they were initially claimed by their parents in their 1998 tax return. I assume that they grew up in the same area.

[31]The bluer the area is, the higher the income is in that area. The redder the area is, the lower the income is in that area

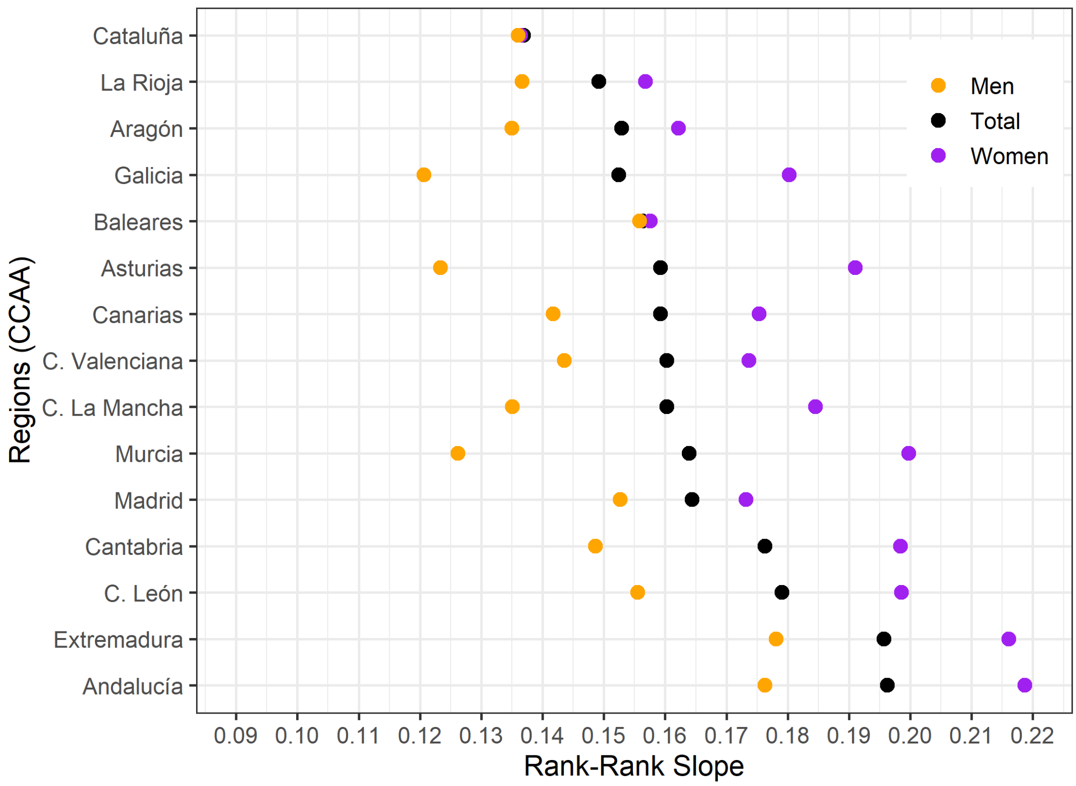
the gender gap is almost non-existent, while in Murcia, Asturias or Galicia is acute. In other words, the difference between daughters from top-percentile families and those from bottom-percentile families in terms of the mean percentile they reach is larger than the difference for sons.

Similarly, Appendix Figure A.21 summarizes relative mobility estimates across provinces. Figure A.21a shows the rank-rank relationship at the provincial level. Figure A.21b presents the provincial estimates of these slopes, which also show remarkable gender disparities. The same facts are also true at this level of disaggregation [32]

[32]For some provinces (Soria, Tarragona and Girona) daughter RRS are lower (larger mobility) than for sons. In the case of Girona and Tarragona, this is not surprising since they belong to Cataluña, a region with an almost non-existent gender gap.



(a) Rank-Rank Association at the Regional Level

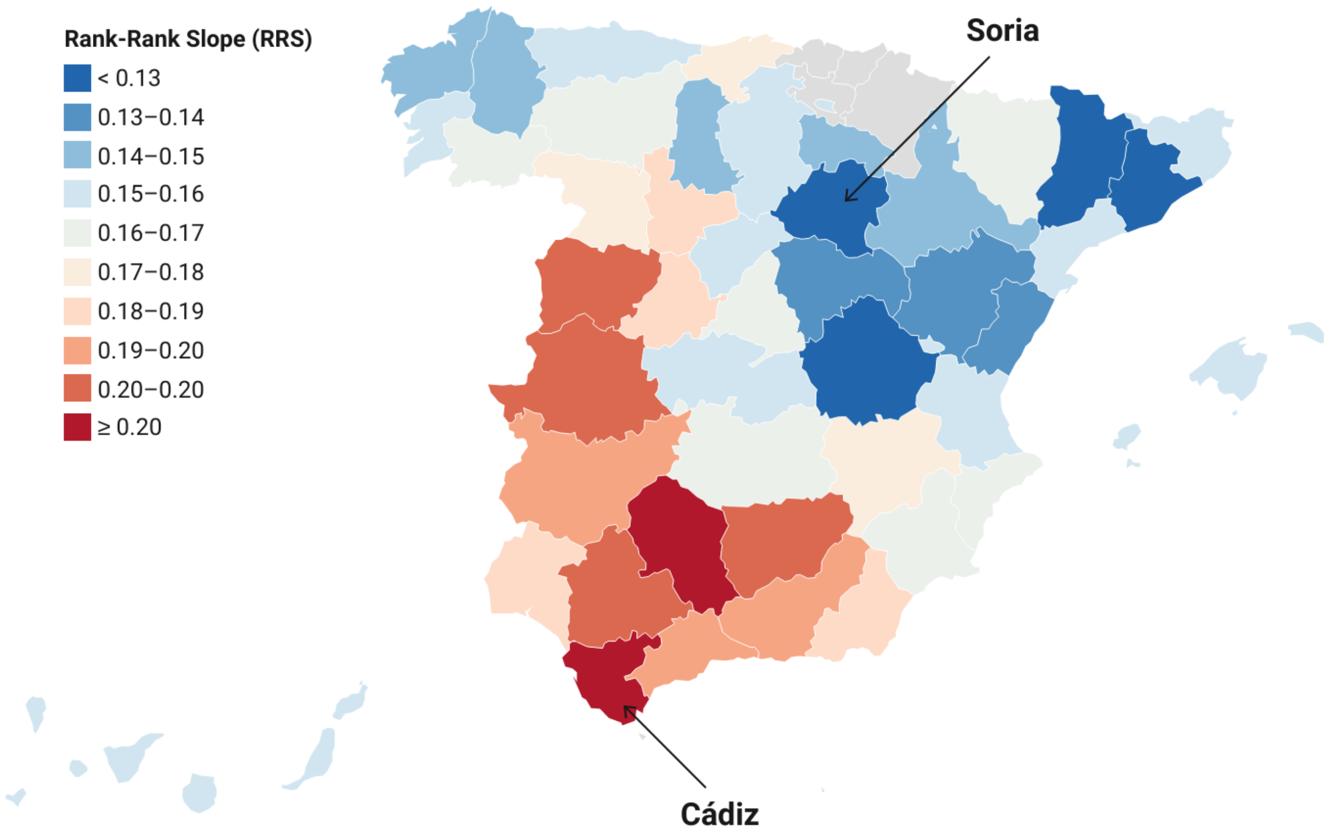


(b) RRS by Province - Gender Differences

Figure 8: Relative Mobility Across Regions

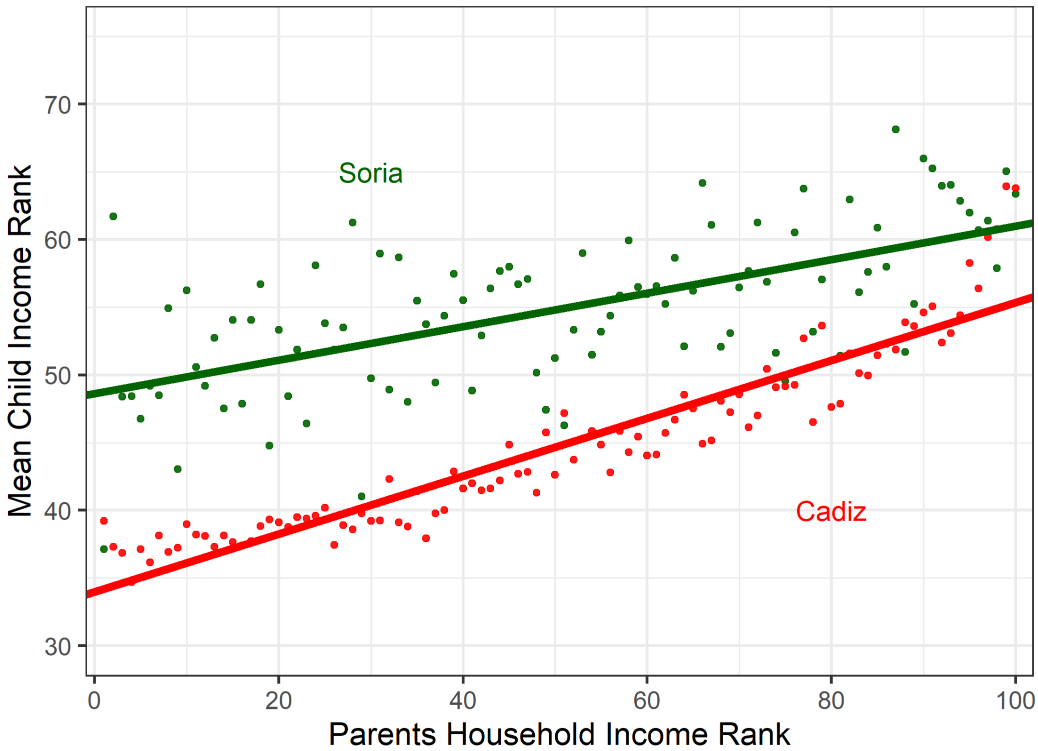
For provinces, I build a heatmap (Figure 9) where geographic patterns in relative mobility can be easily identified. The most mobile provinces in relative terms are located from the Center to the North-East of the country, whereas the less mobile ones are mainly located in Extremadura and Andalucia. The magnitude of these differences is significant. Consider the province with the highest level of relative mobility, Soria (Castilla y Leon; RRS of 0.123), and the one with the lowest, Cadiz (Andalucia; RRS of 0.213). For children growing up in Soria, the difference between the mean percentile achieved by children from bottom rank families and the one reached by children coming from top-rank families is 12 (percentiles). This difference doubles in the case of Cadiz (9 percentiles larger).

Figure 9: Heatmap of Rank-Rank Slope (RRS) Estimates at the Province Level



To have a better insight of what big differences in relative mobility can imply, I take a closer look to the comparison between Soria and Cadiz in the following Figure 10. Firstly, on top of having a more egalitarian distribution when adults, children from Soria end up, on average, in a higher percentile in virtually all the parental ranks. This shows that the expected economic outcomes for almost every child growing up in Soria are better than those doing so in Cadiz. Secondly, both provinces show a fairly linear rank-rank relationship. However, in Cadiz there is a clear non-linearity at the very top, as in the national distribution. This indicates that for the case of Cadiz not only there is a significant divergence between the top and the bottom, but also that the extra advantage given by top percentiles families is greater than in Soria.

Figure 10: Rank-Rank Relationship for Children from Soria (highest relative mobility) and from Cadiz (lowest relative mobility)



Finally, I calculate the relative probability of ending up in the Top 1% for each province (RPT) to further investigate how the extra advantage provided by top-income families varies across provinces. To better understand the construction of this measure, in Figure 13a I show a heatmap of the numerator of the RTP[33]: the probability of being in the Top 1% as an adult coming from a Top 1% family by province, $P(Top1\%|Top1\%)$[34]. In Figure 13b, I present a heatmap of the denominator of the RTP: the probability of being in the Top 1% as an adult coming from a Bottom 10% family by province, $P(Top1\%|Bottom10\%)$.

Before examining the differences in the value of the RPT, a striking result from these heatmaps is worth mentioning: the disproportionate advantage of growing up in a Top 1% household in Madrid Among the Madrilean children that end up in the top percentile as adults, 17% of them come from top

[33] A reminder of the RTP formula: $RPT = \frac{\frac{[P(T1|T1)]}{1}}{\frac{[P(T1|B10)]}{10}}$. For details, see Section 2

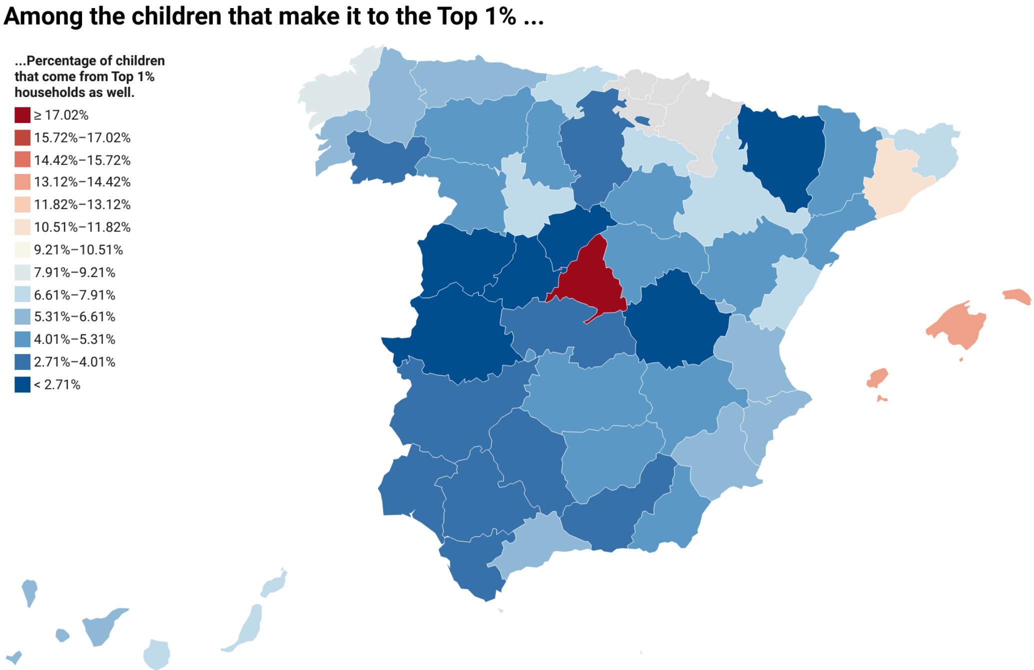
[34] This probability is equivalent to the percentage of children that get to Top 1% as an adult coming from a Top 1% households. Same for the following probability

percentile households, whereas among Ávila top-percentile children this figure is less than 2%[35]. This means that in Madrid it is more relevant to grow up in a Top 1% household to end up in the same percentile as an adult than in Avila, implying a higher influence of parental top incomes in Madrid compared to Avila[36]. In fact, this Madrid-rest-of-Spain remarkable difference can be seen in the final value of the RPT. In Appendix Figure A.22, I plot the value for this relative likelihood for each region. In Madrid, it is 71 times easier to get to the Top 1% coming from a Top 1% household rather than for a Bottom 10% one.

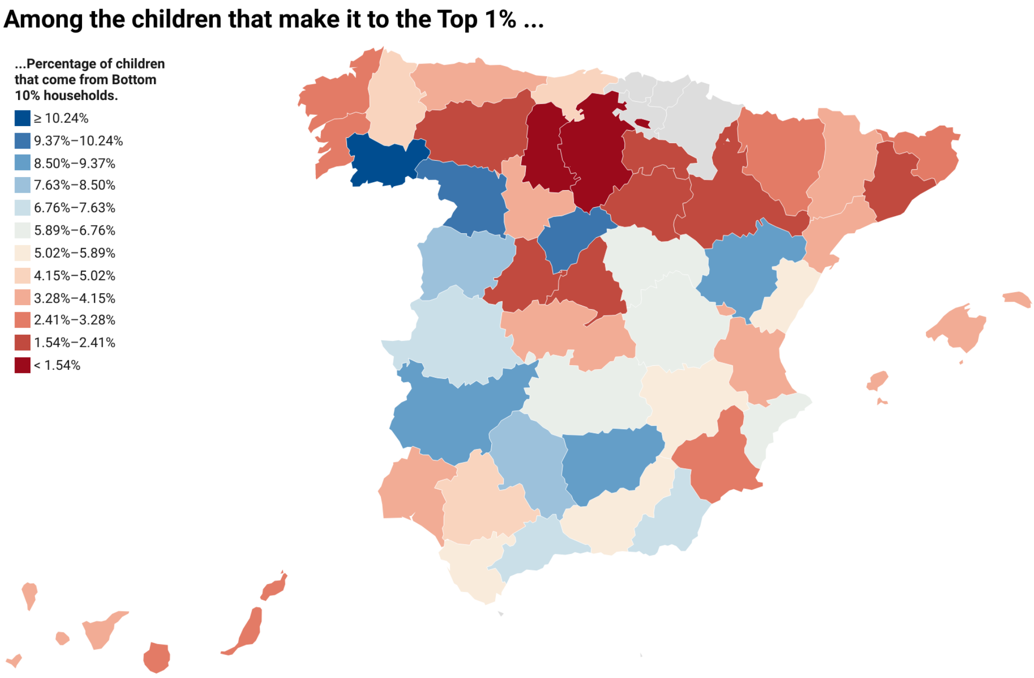
Interestingly, this contrasts with a medium value of the RRS (see Figure A.21b). Given this enormous advantage provided by top-income families children in Madrid, the RRS, a measure of *relative* mobility, should be larger (less mobility). This discrepancy is why the RPT measure is necessary. At the very top of the distribution, the rank-rank relationship becomes slightly non-linear. Therefore, the RRS does not capture as precisely as in the rest of the distribution the difference in outcomes between very rich and very poor children. Yet, with this measure focused on comparing the very top with the very bottom, this higher income persistence observed in the top parental percentiles is better accounted for, complementing the RRS estimates.

[35] Ávila is the province with the lowest percentage. As a benchmark, remember that these percentages should be just 1% in a perfectly egalitarian society.

[36] For children ending up in the Top 1% coming from the Bottom 10% the story between Madrid and Avila is the opposite: children from bottom-percentile families are much more under-represented among Madrilean top percentile adulthood kids than in Avila



(a) Heatmap of $P(Top1\%|Top1\%)$



(b) Heatmap of $P(Top1\%|Bottom10\%)$

Figure 11: Relative Probability of Getting to the Top 1% by Province (numerator, Fig. 13a, and denominator, Fig. 13b)

Notes: Top figure interpretation: in the population of children that make it to the Top 1% as adults from a specific province, this map shows the percentage of these children that come from Top 1% households as well. Bottom figure interpretation: in the population of children that make it to the Top 1% as adults from a specific province, this map shows the percentage of these children that come from Bottom 10%

5.2.2 Absolute Mobility Differences

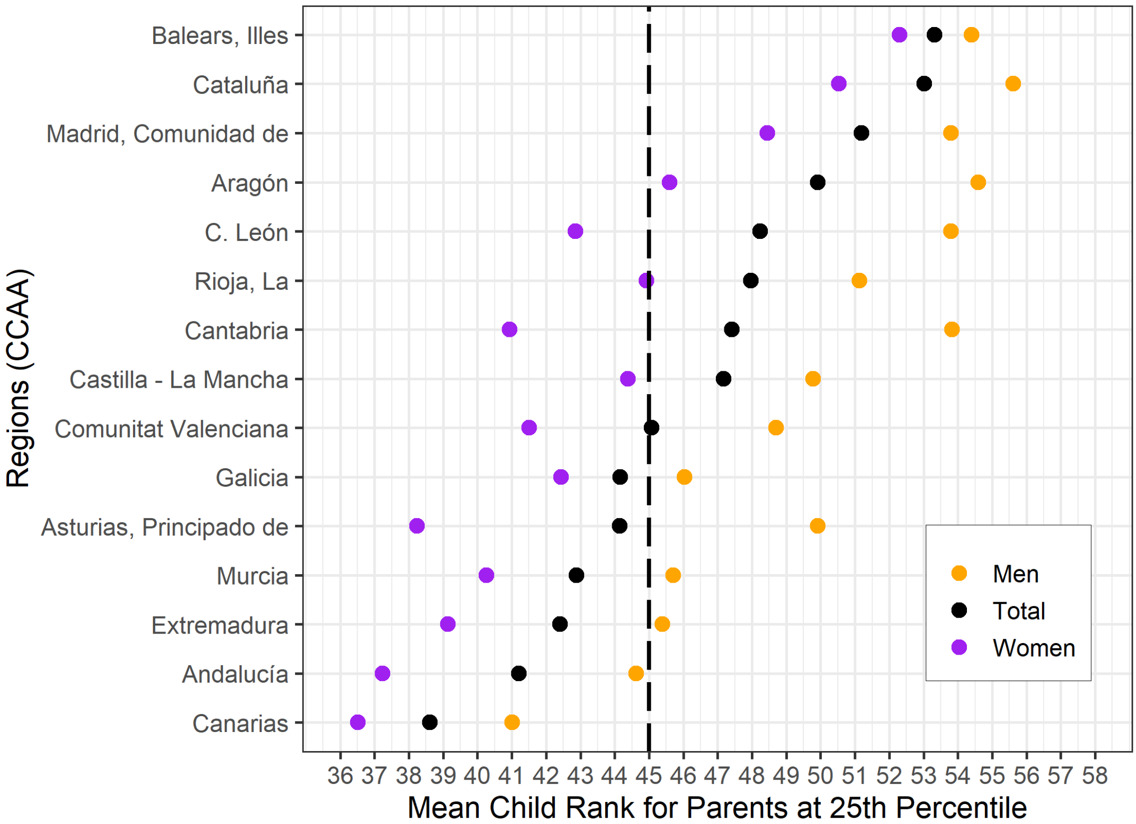
Next, I continue my geographic analysis by examining regional and provincial differences in absolute mobility estimates disaggregated by child gender. Figure 12 shows the regional estimates for the main absolute mobility measures used in the analysis. On the top, Figure 12a presents the absolute upward mobility (AUM) estimates, while bottom Figure 12b plots the probability of ending up in the top quintile coming from bottom quintile families (P(Q5|Q1)). In addition, Appendix Figure A.23 shows full quintile transition matrices for all regions. The same two main findings as in relative mobility differences are observed here as well. Firstly, there is a remarkable variation in upward mobility opportunities across the country: while the probability of reaching the top quintile from the bottom one is around 19% in Cataluña, it is only 8.6% in Canarias. Similarly, the expected rank for children coming from below-median households is the 53th percentile in Cataluña[37] whereas children in Canarias do not even achieve, on average, the 39th percentile. Secondly, there is a persistent and heterogeneous gender gap across regions and measures: daughters tend have less probabilities of climbing to the top quintile and end up in a lower position of their own income distribution when adults.

As for relative mobility, I present absolute mobility estimates in heatmaps to better explore geographical trends.[38]. The first finding that comes out from these maps is that estimates from both measures follow a geographical trend. The most mobile provinces in absolute terms tend to be located in the North/North-East of the country whereas the less mobile ones are mainly located in the South/South-West.[39]. The second interesting finding is that they seem to follow the same geographical patterns as relative mobility estimates (see Figure 9), something that I further explore in the next subsection. This also seem to be the case for the additional measure of absolute mobility that I use to complement the analysis, the percentage of children that earn more than their parents (PEM), whose heatmap can be found in Appendix Figure A.25.

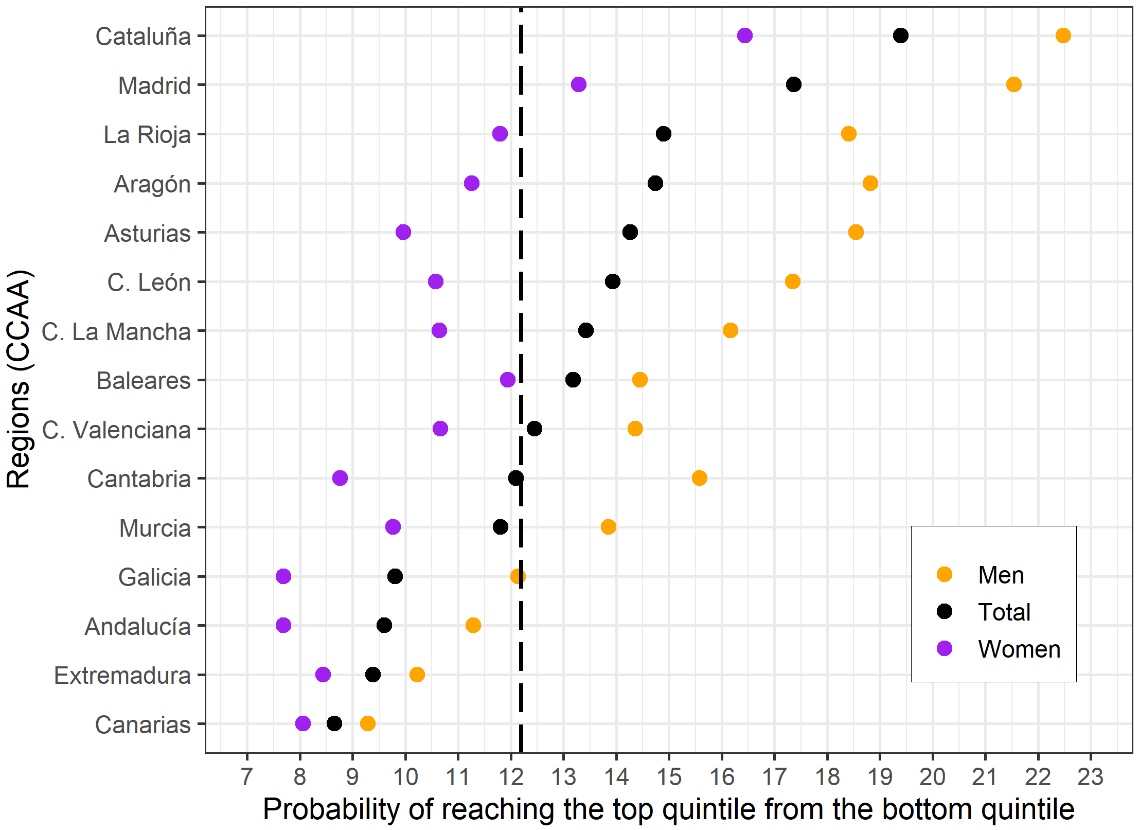
[37]They actually surpass the median of their own adulthood income distribution

[38]A detailed summary of the provincial estimates for these measures can be found at Appendix Figure A.23

[39]Perhaps the only exception is the west part of Galicia for the second measure (Figure 12b)

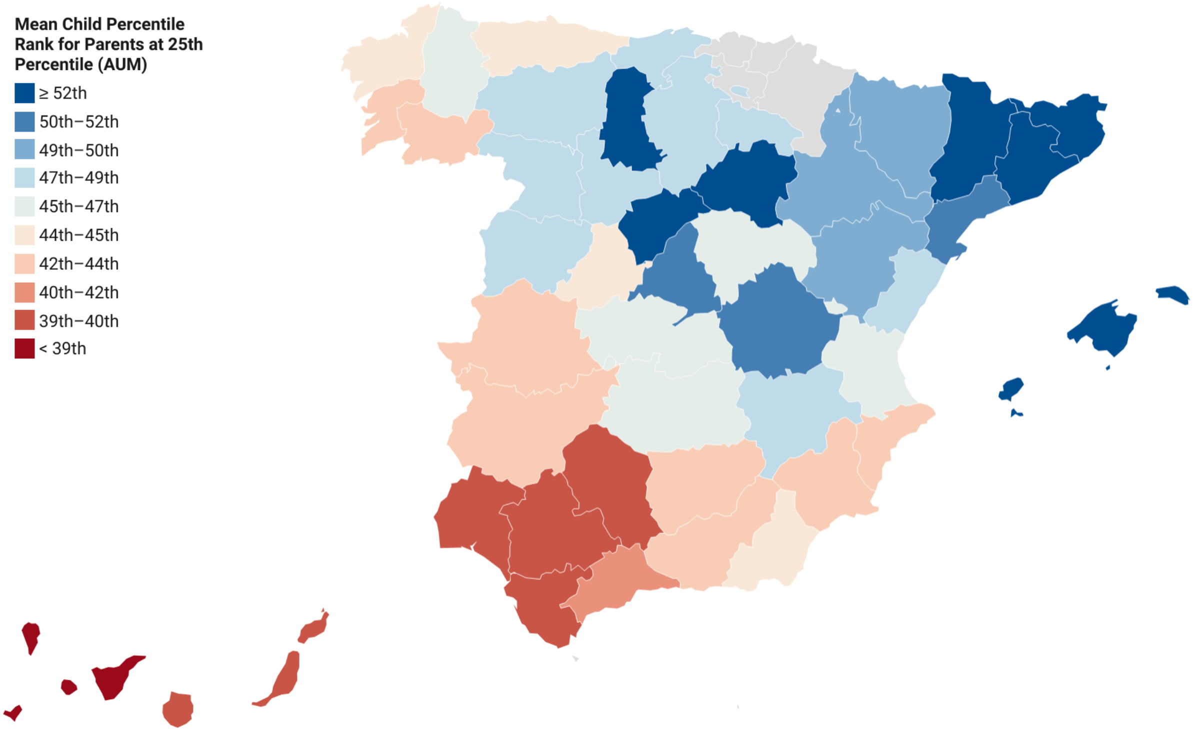


(a) Absolute Upwards Mobility (AUM) across Regions

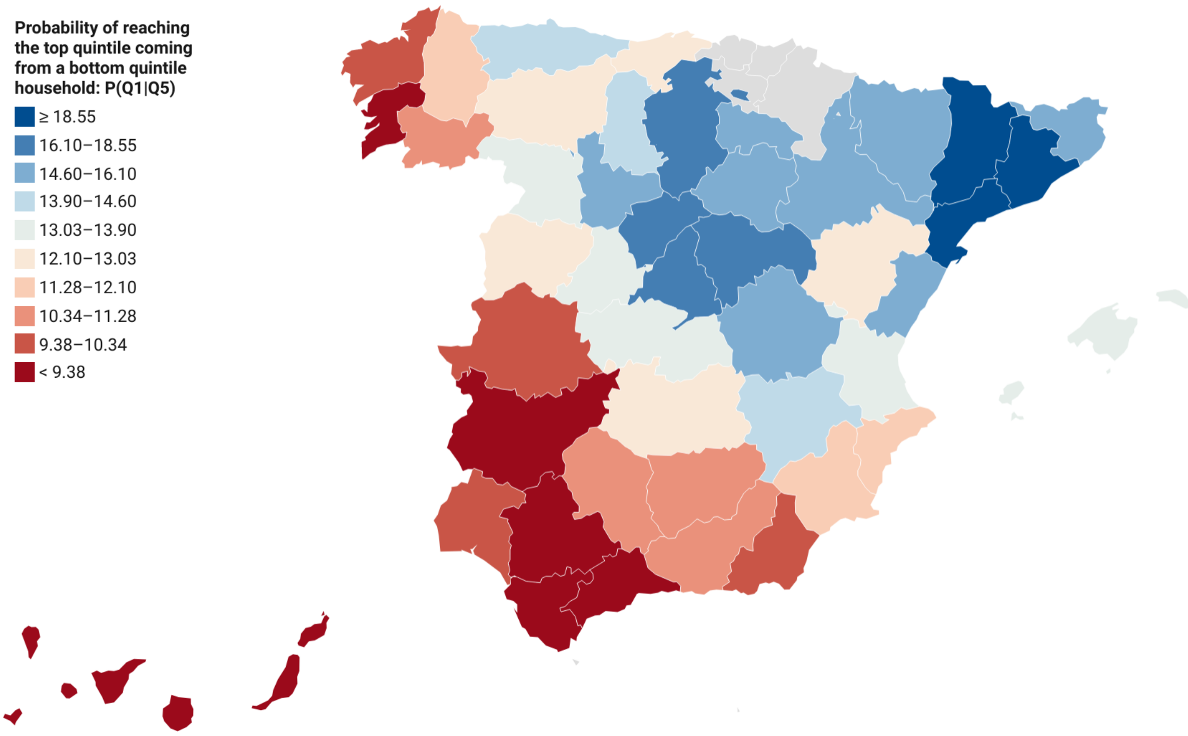


(b) P(Q5|Q1) across Regions

Figure 12: Absolute Mobility Estimates across Regions



(a) Heatmap of the AUM at the Province Level



(b) Heatmap of the P(Q5|Q1) at the Province Level

Figure 13: Heatmaps of Absolute Mobility Estimates at the Province Level

Notes: the bluer the more upwardly mobile the province is. The redder the less upwardly mobile the province is

How big is the geographic variation in Spain compared to the one documented by other country studies?. The following Table 3 provides a summary of the main relative and absolute mobility estimates at the province level for some interesting countries: Switzerland (the most egalitarian in relative terms), United States (the less egalitarian in relative terms) and Italy (the most similar country to Spain among the comparable ones). In absolute terms, we see that, although it is quite large, the gap between the most mobile province at the least mobile one (the range) is not as high as in Italy or the United States for both AUM and P(Q5|Q1). For this last metric, the range is even smaller than in Switzerland. In relative terms, Spain has the lowest gap between the most mobile province and the less mobile one. Therefore, taking into account international comparisons both at the national (Table 2) and provincial level (Table 3), we can say that Spain in aggregate terms is somewhere in the middle of the global scene of income mobility. Yet, within-country differences in terms of equality of opportunities (RRS) and upward mobility (AUM;P(Q5|Q1) are relatively small compared to other countries.

Table 3: International Comparisons of Geographic Variation in Intergenerational Mobility Estimates

	AUM				**P(Q5|Q1)**				**RRS**			
	Min	Mean	Max	Range	Min	Mean	Max	Range	Min	Mean	Max	Range
Switzerland	42	46	50	8	8	12.9	22	14	0.10	0.15	0.22	0.12
Spain	37	45	53.8	16.8	8.5	12.2	20.2	11.7	0.12	0.19	0.22	0.10
Italy	36.4	44	62.7	26.3	4.4	10	37.2	32.8	0.09	0.25	0.25	0.16
United States	26	41.4	65	39	2.2	41.4	47	44.8	0.06	0.34	0.50	0.44

Notes: This table reports a summary of intergenerational mobility estimates at the province/community zone level (US) for comparable country studies."RRS" stands for rank-rank slope. "P(Q5|Q1)" is the probability of ending up in the top quintile coming from the bottom quintile (no available estimate for Denmark). "AUM" is absolute upwards mobility (the mean child rank for children coming from families at the 25th percentile).

5.3 Are High Relative Mobility Areas Also Highly Mobile in Absolute Terms?

In this subsection, I study the interaction between absolute mobility and relative mobility exploiting the above regional and provincial variation. To this end, I regress different absolute mobility estimates on their corresponding rank-rank slopes (RRS) at both geographical levels. The results of these estimations are reported in the following Table 4. To make results easier to interpret, dependent variables are scaled so they reflect the change in absolute mobility measures when the RRS changes by 0.01 (and not by one unit)[40]. The reason is that the RRS estimates vary from a minimum of 0.12 to a maximum of 0.22. As a consequence, analyzing the changes in absolute mobility when the RRS changes by 1 unit is much less informative (and somehow unrealistic as the value of the RRS cannot go beyond 1) than doing so when it changes by 0.01, which is a common variation among provinces and regions.

The results from these regressions indicate that there is a negative and statistically significant association between relative and absolute mobility at both geographic levels. Focusing on provinces (Columns 4-6), we see that a 0.01 increase in the RRS (hence a *decline* in relative mobility) is associated with a decrease of 1.27 percentiles in the expected percentile for children growing up in below-median families (AUM), of 0.01pp in the share of children earning more than their parents (PEM) and a reduction of 0.82pp in the probability of climbing to the top quintile coming the bottom quintile (P(Q5|Q1)). Put differently, regions and provinces that tend to have high levels of relative mobility also experience high levels of absolute mobility and this is true across a variety of measures. The same relationship between relative and absolute mobility is found as well by the recent country studies of the literature (Chetty et al., 2014; Acciari et al., 2019; Chuard and Grassi, 2020; Eriksen and Munk, 2020; Deutscher and Mazumder, 2020; Heidrich, 2017; Connolly et al., 2019). Appendix Figure A.26 graphically shows the results of these regressions at the province level, where provinces are colored by parental income quintile.

5.4 Gender Gaps & Geographical Heterogeneity

Next, I further investigate the geographical patterns of gender gaps (i.e., the difference in intergenerational mobility estimates between sons and daughters), something that has been under-explored in the literature.

To start with, in the following Figure 14 I show heatmaps for the gender gaps across different absolute and relative measures at the province level. As a general rule, the redder the higher the gap between sons and daughter and the bluer the lower it is[41]. Some interesting findings emerge from this graphical analysis. Firstly, the gender gaps do not seem to follow the same geographical patterns as the estimates of relative and absolute mobility. In particular, the aforementioned North/North-East vs. South/South-West divide is not present as a general rule, as for some measures southern provinces

[40] Essentially, I divide dependent variables by 100

[41] Note that for all measures except the RRS, the gender gap is defined as the estimate for men minus the estimate for women. for the RRS, it is defined the other way around because a higher value means worse relative mobility levels.

Table 4: The Association between Relative and Absolute Mobility

	Dependent variable: Absolute Mobility Measures					
	(1)	(2)	(3)	(4)	(5)	P(6)
	AUM	PEM	P(Q1|Q5)	AUM	PEM	P(Q1|Q5)
	Regions	Regions	Regions	Provinces	Provinces	Provinces
RRS	−1.252*	−0.0115*	−1.035**	−1.279***	−0.0106***	−0.829***
	(0.643)	(0.0561)	(0.416)	(0.238)	(0.021)	(0.168)
Constant	0.670***	0.060***	0.300***	0.674***	0.058***	0.265***
	(0.106)	(0.009)	(0.068)	(0.039)	(0.003)	(0.027)
Observations	15	15	15	46	46	46
R^2	0.226	0.244	0.322	0.396	0.350	0.356
Adjusted R^2	0.166	0.186	0.270	0.383	0.335	0.342

Note: *p<0.1; **p<0.05; ***p<0.01

show less gender disparities than northern provinces[42]. Secondly, even if there are no clear geographical cleavages in gender gaps, there is an enormous heterogeneity across the country. Therefore, we see that daughters not only have systematically worse outcomes than men but also that this gap can hugely vary depending on the province in which children grow up, which is in line with the conclusions drawn by Chetty et al. (2016a) for the United States.

Since geographical patterns in gender gaps appear to be different from those depicted in subsections 5.2.1 and 5.2.2, I further study what correlates with gender gaps of different measures. To this goal, I regress these gender gaps with (i) their own provincial estimate of the same measure, (ii) the provincial estimate of other measures and (iii) with parental and child income at the provincial level. The results from these regressions can be found in Appendix Table A.7. Looking at the estimations, the relationship between AUM and PEM and their respective gender gaps is not statistically significant. Put differently, higher mobility in absolute terms, for these measures, do not appear to be correlated with the magnitude of the gender gap. On the contrary, this is the case for the P(Q5|Q1) and for the RRS. More precisely, there is a positive and statistically significant association between the estimates of these measures and their respective gender gaps. Therefore, for instance, provinces with a higher

[42] With some exceptions for two measures that I discuss later on

RRS (less relative mobility) appear to have larger gender gaps in relative mobility between son and daughters. Furthermore, for these two measures (RRS and the P(Q5|Q1)) there is also a positive and statistically significant association between parental and child income, but the coefficients are rather small. To sum up, there is mixed evidence regarding the association between gender gaps and absolute mobility since some measures appear to be correlated with them and other do not. On the other hand, gender gaps tend to be higher when relative mobility is lows (higher RRS). Also, there is no conclusive evidence concerning the relationship between provincial income (either parental or child) and gender gaps.

5.5 The *Great Gatsby* Curve in Spain

Finally, I explore whether there is a "Great Gatsby Curve"[43] in Spain and briefly discuss the interaction between income inequality and intergenerational mobility. To this goal, I regress the estimates from different relative and absolute measures on the provincial Gini Index, created using parental income. Table 5 reports the results of these regressions. I document a negative and statistically significant association between income inequality (as measured by the Gini Index) and absolute mobility measures and a positive one with relative mobility measures[44]. Provinces that have higher levels of inequality tend to be less mobile both in relative and absolute terms, confirming the existence of a *Great Gatsby* Curve within Spain[45]. This result is in line with evidence showing a positive relationship between income persistence and income inequality across different countries (Corak, 2013).

Why is there a negative relationship between income inequality and intergenerational mobility? A priori, one could think that inequality negatively impacts intergenerational mobility: families from the highest deciles of the income distribution are able to transmit a wide range of benefits (in terms of education, health, social networks, etc) that bottom deciles families are not able to. These initial disparities in many dimensions tend to persist over generations and, consequently, opportunities remain quite different for the children growing up in top-income families in comparison to those growing up in low income families. In few words, it is harder to climb a ladder when the rungs are farther apart.

[43]The "Great Gatsby Curve" is the name originally given by Alan Krueger to the negative relationship between income inequality and intergenerational mobility during their presentation at the Center for American Progress in 2012. It is inspired by the main character, Jay Gatsby, of the famous Francis Scott Fitzgerald's novel "The Great Gatsby".

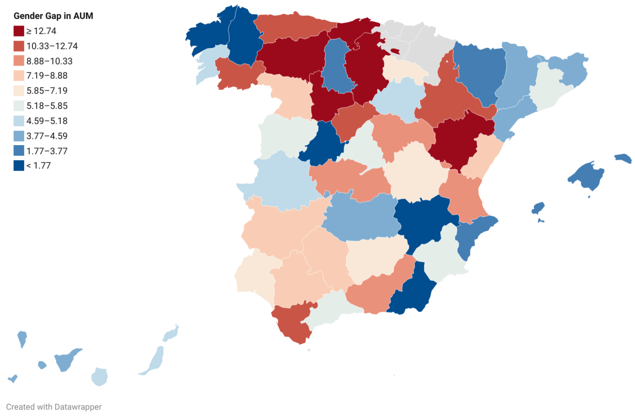
[44]Remember that a higher RRS means less relative mobility.

[45]A graphical representation of the Great Gatsby Curve across different measures is presented in Appendix Figure A.27.

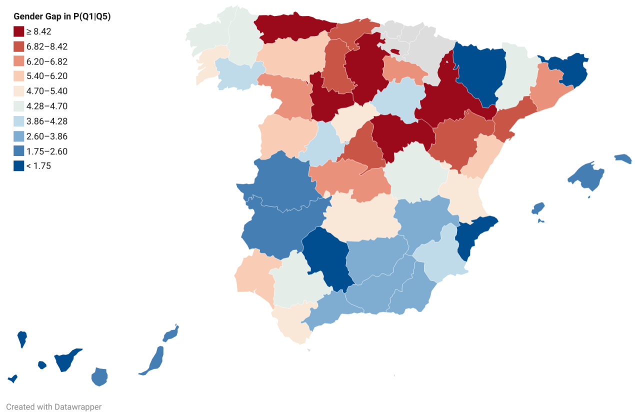
However, the causality could be the other way around: a low level of intergenerational mobility may enlarge existing initial inequalities. Hence, further research is needed to disentangle the direction of the causality between inequality and intergenerational mobility. Some important first steps have been done by studies jointly analyzing the interaction between inequality, education, neighborhood characteristics and intergenerational mobility (Chetty and Hendren, 2018a,b; Chetty et al., 2020; Bingley et al., 2021)

Table 5: The *Great Gatsby* Curve in Spain - Province Level

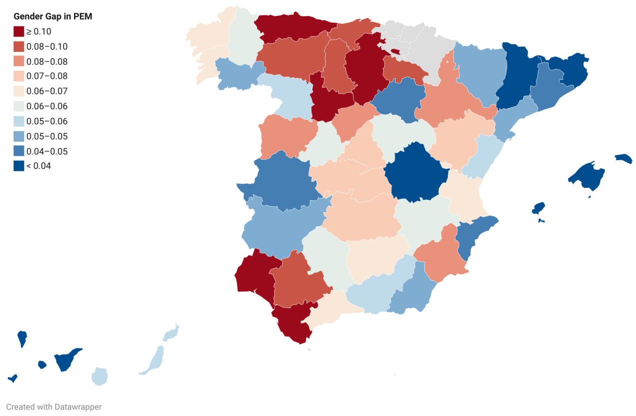
	Dependent variable:			
	AUM	PEM	P(Q5|Q1)	RRS
	(1)	(2)	(3)	(4)
Gini Index	-0.749^{**}	-0.008^{***}	-0.583^{***}	0.003^{**}
	(0.299)	(0.003)	(0.207)	(0.002)
Constant	75.765^{***}	0.731^{***}	35.697^{***}	0.029
	(11.652)	(0.100)	(8.061)	(0.061)
Observations	46	46	46	46
R^2	0.127	0.190	0.156	0.103
Adjusted R^2	0.107	0.171	0.136	0.082
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$



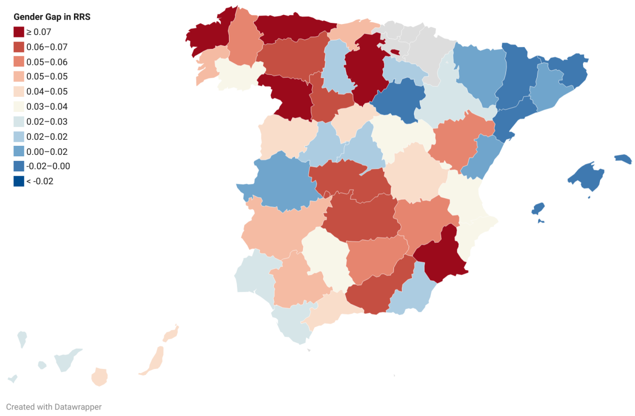
(a) Heatmap of the Gender Gap in AUM



(b) Heatmap of the Gender Gap in P(Q5|Q1)



(c) Heatmap of the Gender Gap in PEM



(d) Heatmap of the Gender Gap in RRS

Figure 14: Heatmaps of Gender Gaps at the Province Level

6 Intergenerational Mobility & Out-Migration

In previous sections I documented a high level of geographical heterogeneity in both absolute and relative mobility rates. In addition, gender gaps appear to hugely vary across areas as well. Even if the magnitude of geographical variation in mobility rates is not as extreme as in the United States, a similar conclusion can be drawn: in Spain there are some "lands of opportunities" (i.e. areas that provide more equal opportunities and promote upward mobility) and some areas of persistent inequality (where family background substantially influences children's outcomes). In addition, Appendix Figure A.20 showed remarkable income disparities across provinces in terms of both parental and child income.

Therefore, in this final section, I focus on the interaction between out-migration[46] and intergenerational mobility outcomes. In particular, I analyze whether leaving the home province leads to higher rates of absolute mobility and, if so, whether children migrate to more economically prosperous places in the hope of having access to better opportunities. To this goal, I firstly divide the children in two broad categories: *movers* and *stayers*. A child is defined as a mover if in their 2016 tax returns they declare to live in an area different to the one they were observed in their parents 1998 tax declaration. Similarly, a child is defined as a stayer if in their 2016 tax returns they declare to live in the same area as the one they were observed in their parents 1998 tax declaration[47]. Following these definitions, I find that, in the core sample (with a total of 1,492,107 children), 9.73% of children live outside of their home region (CCAA), 12.86% live outside their home province and 31.32% live outside their home municipality. Again, I concentrate on provincial out-migration to ensure comparability with similar analyses studying the relationship between within-country migration and upward mobility carried out in the recent literature, namely in Italy (Acciari et al., 2019) and the United States (Chetty et al., 2014; Chetty and Hendren, 2018a,b).

6.1 The Relationship between Out-Migration and Absolute Upwards Mobility

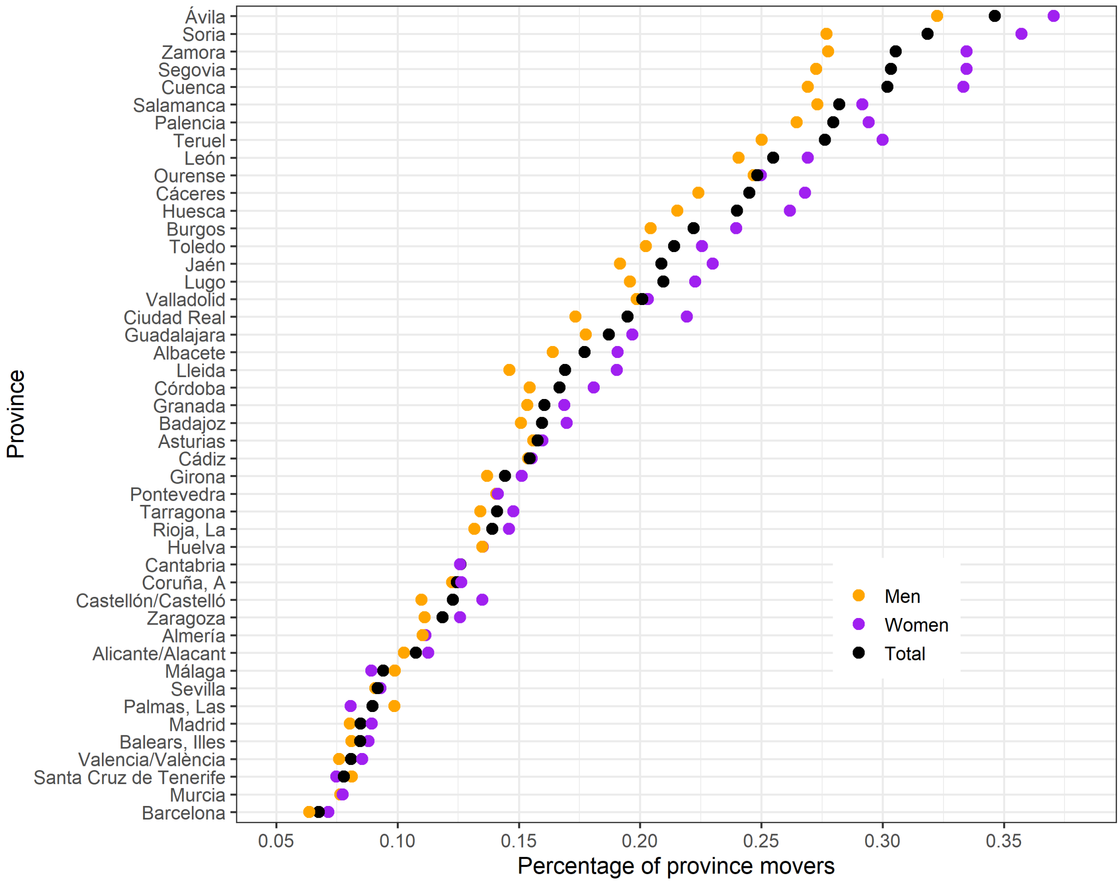
As discussed, to study the relationship between out-migration and (absolute) upward mobility, I use the province as a main level of geographical disaggregation. This is very useful since there is a high

[46]Out-migration refers to the action of leaving the home area to move to another one within the same country.

[47]A caveat that should be done regarding the following analysis is that I cannot know when exactly children migrate to other provinces nor if they moved out with or without their parents. I cannot know this because I only observe parents' location when they fill their tax returns in 1998 but not anytime later, including the year 2016, when their kids fill their own tax returns and declare their location (which is also the only time I observe their location after 1998)

out-migration variation across provinces in the core sample, as shown in Figure 15. In addition, this figure shows that daughters tend to migrate more than their male counterparts but this gap is very small in most provinces.

Figure 15: Percentage of Movers by Province and Gender



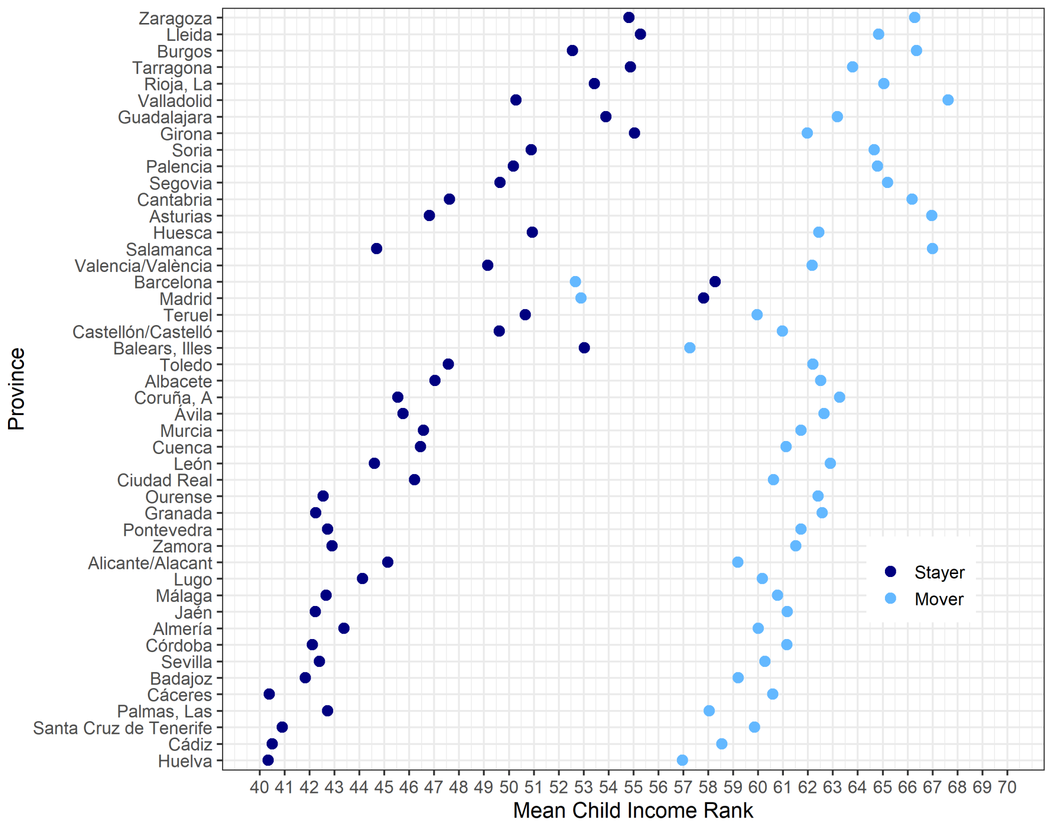
To start exploring differences in outcomes between stayers and movers, I estimate the mean child rank achieved by each group at the province level. The results are presented in the following Figure 16. A striking pattern comes out: for the vast majority of provinces, the children that move out end up, on average, in a higher position of their own income distribution compared to those who stay in their home province[48]. The two important exceptions are Madrid and Barcelona, where the contrary is true. However, this is not particularly surprising: since these two provinces are the richest ones in Spain, moving *out* from them almost always implies that a child is going to a poorer province[49], which probably provides less economic opportunities than the vibrant areas of Madrid and Barcelona. Furthermore, in general terms, there is a positive and statistically significant association between out-

[48]As shown in Appendix Figure A.29, this result are not affected by child gender: sons and daughters that move also end up higher in their income ladder as adults compared to those sons and daughters who stay

[49]The only exception to this would be if a child moves from Barcelona to Madrid because this last one is slightly richer

migration and upward mobility: leaving the home province is associated with ending up around 10 percentiles higher in the child income distribution (see Appendix Table A.8 and Figure A.28). Further evidence of this relationship dividing provinces by income quantile is provided in Appendix Figure A.30 & Table A.9 and at the municipality level, dividing by income decile, in Appendix Figure A.31 & Table A.10.

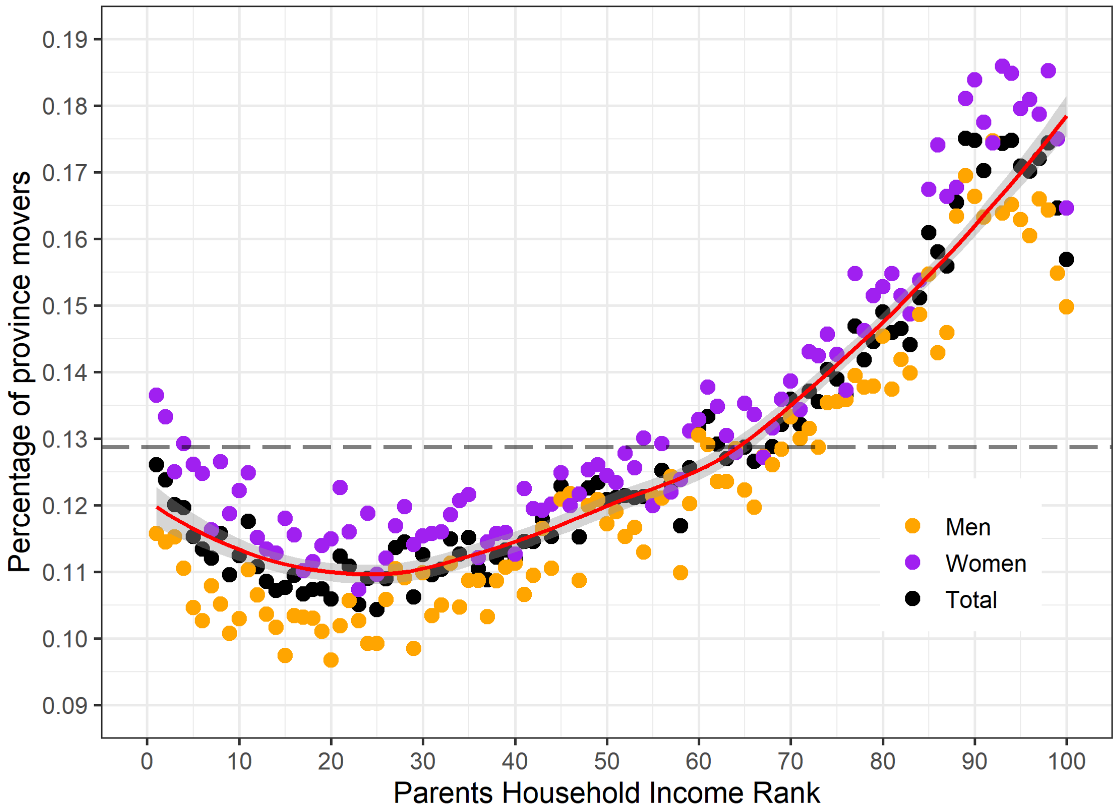
Figure 16: Mean Child Percentile of Stayers and Movers by Province



Nevertheless, these better mean outcomes of movers could be explained by more favorable initial conditions (the influence of coming from rich families), and not by the opportunities offered in the destination place (the effect of destination places). In fact, it appears that children coming from top deciles families tend to emigrate slightly more than those coming from parents in the middle and bottom of the income distribution, independently of the child gender (Figure 17). Hence, the question that I proceed to explore is: are these better outcomes explained by more favorable initial conditions (family background) or by the economic advantages of the destination place?

To answer this, I re-estimate the mean child rank achieved by movers and stayers but controlling for parental income rank. Figure 18 plots the results. Even if children from rich families tend to emigrate more, on average, movers reach higher income ranks than stayers all over the parental income distribution. In fact, the gap between movers and stayers is greater for children coming from families at the middle and bottom part of the parental income distribution. This is the first finding of the out-migration analysis: being a mover is associated with a higher degree of upward mobility for virtually all provinces and independently of the family income rank.

Figure 17: Percentage of Province Movers by Parental Income Rank



Notes: The dotted horizontal line shows the overall rate of out-migration at the province level (12.86%)

Moreover, to better understand the destination effects compared to family background effects in determining upward mobility outcomes, I analyze the economic patterns of out-migration. More precisely, I calculate the share of children moving to a richer province to see how predominant this type of migration is[50]. Hence, among those children that migrate, I classify them in two subgroups: those who move to a province with higher mean income than the home one and those that go to a poorer province than the home one. The results of this classification are presented in Figure 19. I color the home provinces of children by income quintile to give a clearer vision of the migration pattern. The

[50] I define a rich province as a province with high parental income because it is the income that children observe most of their life. In other words, when children grow up they know which provinces are richer and which are poorer, because it is well-known where the best universities and job opportunities are and this barely changes between the adolescent of children and their early 30s, when I observe their new location. Consequently, to keep things realistic and restricted to the information contained in this dataset, I use mean parental income in a province to proxy its income level

Figure 18: Mean Child Percentile of Stayers and Movers by Parental Income Rank

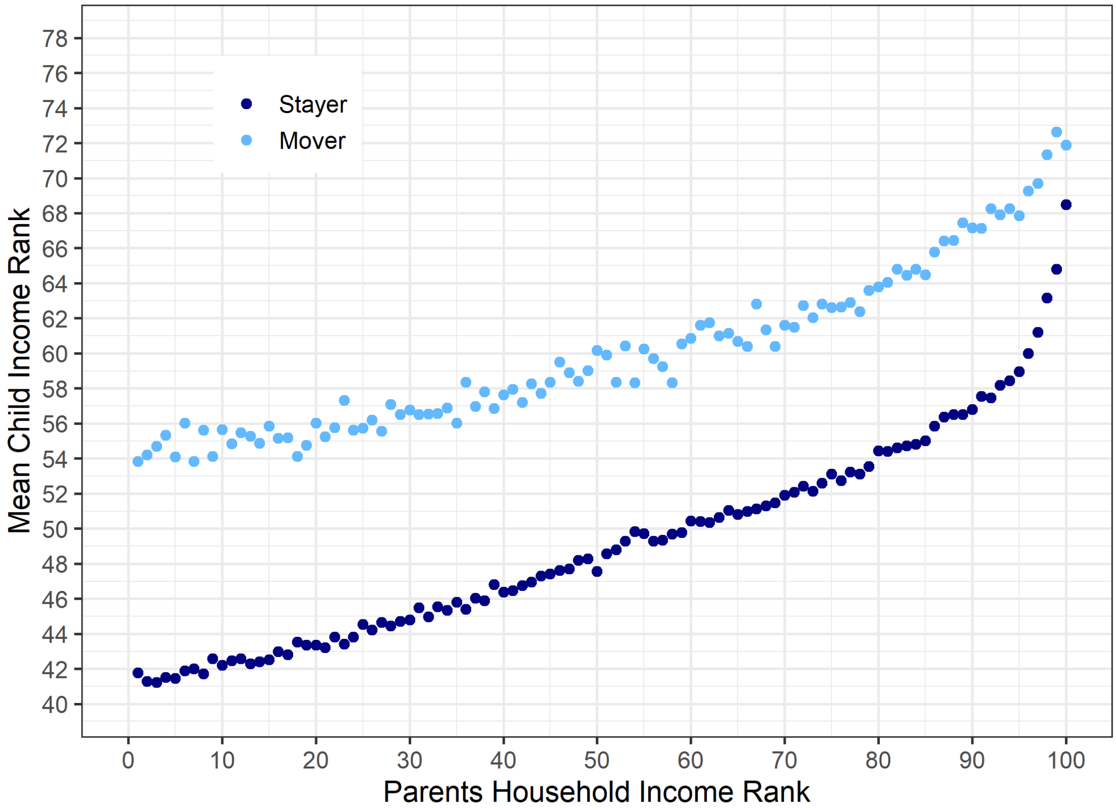
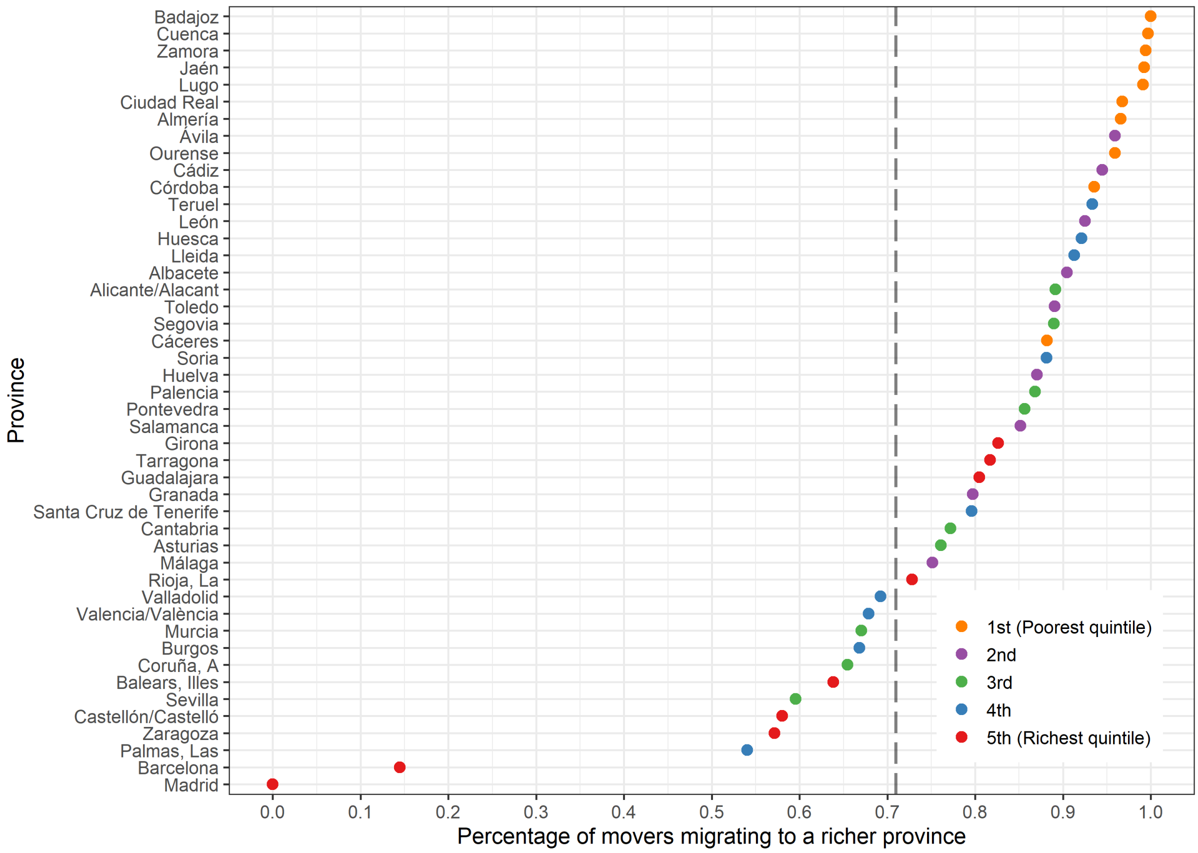


figure shows that the majority of movers migrate to richer provinces (in particular, a 70.96% of them, reflected by the dotted gray line). This is the second main finding of the analysis: among the children that decide to leave, most of them move to richer provinces (70.96%), especially those coming from relatively poor provinces, which points towards an economically driven migration[51]. Furthermore, Figure 20 shows that the mean child percentile achieved by movers that migrate to a richer province (movers-to-richer-province) is higher than the one achieved by those children that just migrate (movers in general, or "just-movers") and much larger than stayers' mean percentile. All this suggests that going to richer places is even better in terms of upward mobility and reinforces the economically driven migration argument[52].

[51] I claim this because the majority of children that migrate do so to richer provinces but this does not mean that economic reasons are the only ones determining the decision of migration nor that all out-migrants are solely driven by an economic rationale.

[52] As mentioned before, the only two exceptions are Madrid and Barcelona. In the case of Madrid, there is no green data point because it is the richest province of the country and therefore no one can migrate to a richer province. In the case of Barcelona we see that moving out only leads to better upward mobility on average if the child goes to a richer province, which is Madrid (the only province richer than Barcelona). For both Madrid and Barcelona, we observe that stay is better than just move out, which gives more credibility to the argument of rich places providing better

Figure 19: Percentage of Movers that Migrate to a Richer Province



However, a similar concern as before arises: the positive effect of moving to a richer province in terms of upward mobility may be driven by parental income and not by destination place effects. To explore this issue, I estimate again the mean child rank achieved by stayers, just-movers and movers-to-richer-province controlling for parental income rank. Figure 21 shows that both type of movers reach higher mean income ranks than stayers all over the parental income distribution, which indicates, again, that moving out (especially to higher income provinces) brings more opportunities of upward mobility regardless of the family income of these migrant children.

opportunities and higher upward mobility outcomes

Figure 20: Mean Child Percentile of Stayers, Just-Movers and Movers-to-richer-provinces by Province

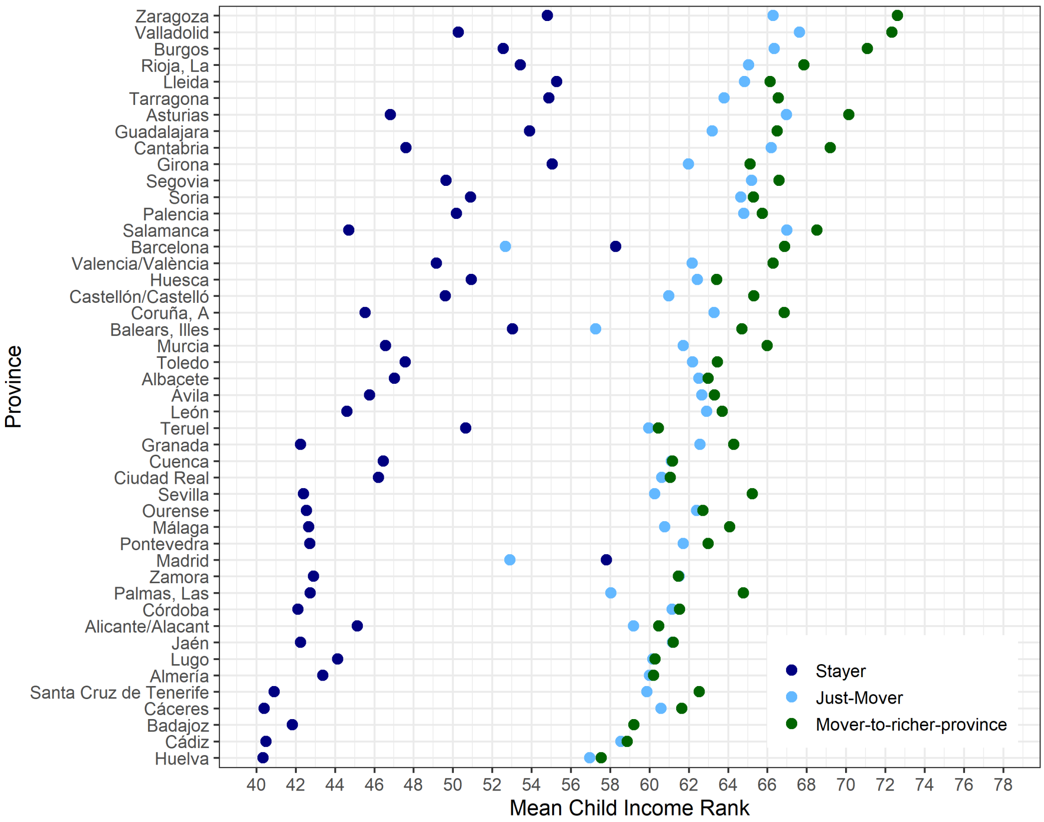


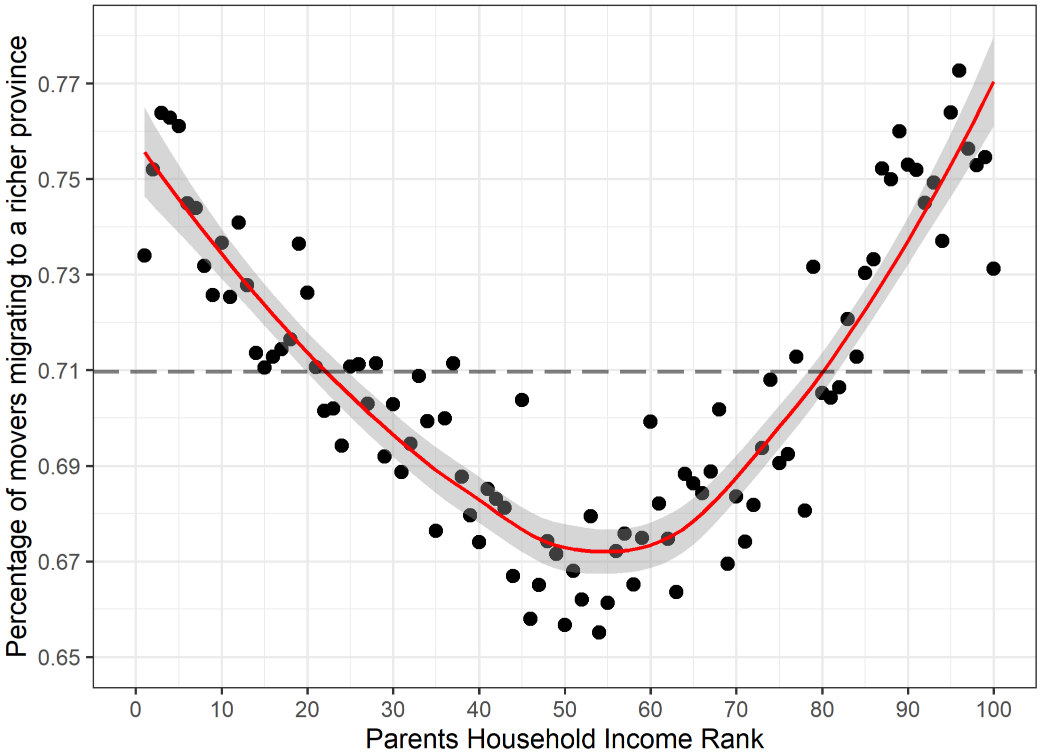
Figure 21: Mean Child Percentile of Stayers, Just-Movers and Movers-to-richer-province by Parental Income Rank

At this point, it seems to be clear that out-migration and absolute upward mobility are positively associated. Also, I show that the benefits of leaving the home province are the highest for those that move to richer provinces. However, a last question remains under-explored: who is benefiting from this economically driven out-migration that brings the best upward mobility outcomes?. Also, does family income (percentile) explain this majoritarian moving-to-richer-provinces migration trend?

There are three possible scenarios. Firstly, it could be that, among those who move to richer provinces, most of them come from bottom deciles families. This would indicate the existence of what I call a "looking-for-opportunities" migration pattern in which children coming from relatively poor families leave for more prosperous places to improve their well-being. Secondly, we could observe an absence of correlation between family background and the pattern of migration. In other words, migrating to richer provinces could have no relation with parents' position in the income distribution. Thirdly, we could think about an scenario in which, among those who move to richer provinces, the majority of them come from top deciles families. This could point to what I call a "keep-the-status" migration pattern: children from top-income families migrate to richer places not because they *need* better opportunities but because they are more likely go to elite private colleges that tend to be located in the richest provinces of the country, fill top jobs facilitated by family/social networks in richer places or they may have a stronger preference for more vibrant and culturally dynamic places than their home provinces.

Examining the family origins (parental income rank) of movers to richer provinces, I uncover an U-shaped pattern in which the majority of this type of movers (i.e.,the economically driven ones) come from either relatively poor or relatively rich households (Figure 22). This interesting pattern is the third main finding of the out-migration analysis: for those movers going to richer provinces, a "looking-for-opportunities" migration of bottom-percentiles children (first scenario) coexists with a "keep-the-status" migration of top-deciles children (third scenario). Following the above reasoning, this finding adds more credibility to the economically driven out-migration since the children that move out to richer provinces are *predominantly* those that have the higher incentives to do so, either to improve their well-being ("looking-for-opportunities" migration) or to maintain themselves in the top of the social ladder ("keep-the-status" migration)

Figure 22: Percentage of Province Movers That Migrate to Richer Provinces by Parental Income Rank



6.2 Isolating Out-Migration Effects: Analysis of Siblings

To complement the previous analysis and, in an attempt to better capture the effect of out-migration on upward mobility, I compare the outcomes not only of children coming from the same parental income percentile but from the very same households: the siblings. In particular, I analyze the differences across various absolute mobility measures between comparable couples of siblings in which one of the siblings leaves their home province (mover sibling) and the other stays in it (stayer sibling). The reason behind the use of siblings is that the closest counterfactual I can have to a child that moves is their sibling that stays. In other words, studying similar siblings couples is the closest I can be to know what would have been the upward mobility outcome of a mover child had they stayed in their home province. Therefore, bearing in mind some important caveats that I will further discuss, the following analysis provides the best approximation to the causal effect of out-migration on upward mobility given the current information contained in this dataset.

In total, the core sample contains 462,869 siblings (or approximately 231,434 pair of siblings). Among this overall pool, I only select those pairs in which one of them moves out the home province and the other one stays in it. This reduces the total number of siblings to 66,683 (or approximately

33,341 couples of siblings). To ensure the highest possible degree of comparability among these siblings, I restrict the sample again to those couples of siblings with exactly the same age, marital status, gender, number of kids, province of origin and parental income[53], which are all the observable characteristics of children included in the dataset. After this selection procedure, I end up with a subset of 548 siblings that have the exact same characteristics, grow up in the same household and only differ in the fact that one migrates to another province and the other stays.

Formally speaking, I follow a matching process in which I assume that selection into treatment (i.e., moving to anther province) mainly depends on a set of observable characteristics (age, marital status, gender, number of kids, province of origin and parental income). Hence, conditional on these observable characteristics, selection into treatment is random (conditional independence assumption). In addition to this assumption, matching designs rely on common support, this is, that there is a similar amount of individuals with the same characteristics in both populations (treated and untreated). In this case, my matching process ensures perfect common support, as there is an equal number of individuals in the treated and control groups (274 children in each one) with the exact same characteristics (see Appendix Tables A.11 & A.12). Since I am comparing siblings that grow up in the very same household, I further assume that this comparison controls as well for some potentially relevant unobserved characteristics as parenting behavior, income shocks while the children grow up, etc. However, there are other observable factors (as education) or unobservable factors (personality traits) that could influence at the same time the decision to leave the home province and upward mobility outcomes, violating the conditional independence assumption. Despite not being able to fully rule out this concern with my current data, the fact that siblings have the exact same characteristics may be indicative of a very similar sensitivity to, at least, unobservable factors (as parenting behavior). Finally, Appendix Table A.13 shows that both treated (mover sibling) and control (stayer sibling) are perfectly balanced in terms of mean characteristics.

Therefore, under these assumptions and matching design, I can calculate the (quasi-causal) average treatment effect (ATE) of leaving the home province on upward mobility by comparing the mean outcomes of the mover siblings and stayer siblings. Mathematically, the ignorability assumption that allows the computation of the ATE can be expressed as follows:

[53] These last two are guaranteed by the fat that both siblings live together with the same family according to the parents' 1998 tax returns.

$$Y_0, Y_1 \perp D|X \tag{11}$$

which implies that:

$$\begin{aligned} E(Y_0|D, X) &= E(Y_0|X) \\ E(Y_1|D, X) &= E(Y_1|X) \end{aligned} \tag{12}$$

where Y_0 is the outcome of the child if not treated (i.e, if they stay in the home province), Y_1 is the outcome of the child if treated (i.e, if they leave the home province), D is a dummy variable indicating the treatment (i.e., leaving the home province) and X represents the range of observable children's characteristics I have (age, marital status, gender, number of kids, province of origin and parental income). Then, using the assumption described in equation (11), a simple comparison of means provides the ATE:

$$\begin{aligned} ATE &= E(Y_1|D = 1, X) - E(Y_0|D = 0, X) \\ ATE &= E(Y_1|X) - E(Y_0|X) \\ ATE &= E(Y_1) - E(Y_0) \end{aligned} \tag{13}$$

Hence, I calculate the ATE for different outcome variables by estimating the following linear regression using OLS.

$$Y = \beta X + \alpha_{ATE} D + u \tag{14}$$

where α_{ATE} captures the ATE of leaving the home province. The results are presented in the following Table 6. The top of the table shows the estimated ATE while the bottom part reports the mean outcome values for mover siblings and stayer siblings[54]. The first finding that arises from this table is that leaving the home province brings better upward mobility outcomes, as a higher percentile rank or a higher probability of earning more than parents (Columns 1-4-5-6), and larger incomes, both at the individual and household level (Columns 2 and 3)[55]. Focusing on child income percentile, one of the main outcomes examined in this paper, we see that the effect of leaving the home provinces on the percentile achieved by a child is 16, which is a substantial jump in the child income distribution. Furthermore, moving out of the home province increases, on average, the child income by almost €8000,

[54]The observed characteristics estimates (β) are not shown to keep the table shorter and easier to interpret

[55]It should be noted that the positive effect on child household income is driven by married children since single children individual and household income is, by definition, the same.

an income jump corresponding to roughly 38% of the mean child income of the core sample (Appendix Table A.2).

Table 6: Average Treatment Effect (ATE) Estimates of Out-migration on Upward Mobility and Incomel

	Dependent variable:					
	Child Rank	Child Indiv. Income	Child HH. Income	PEM Family	PEM Parent 1	PEM Parent 2
	(1)	(2)	(3)	(4)	(5)	(6)
ATE	16.040***	7,985.174***	8,601.984***	0.172***	0.186***	0.066**
	(2.367)	(1,447.022)	(1,755.225)	(0.042)	(0.042)	(0.032)
Constant	51.201***	20,412.510***	22,763.030***	0.339***	0.442***	0.803***
	(1.674)	(1,023.199)	(1,241.132)	(0.029)	(0.030)	(0.022)
	Mean values of dependent variables for Mover Siblings and Stayer Siblings:					
Mover Sibling	67.24	28397.69	31365.02	0.51	0.63	0.87
Stayer Sibling	51.20	20412.51	22763.03	0.34	0.44	0.80
Observations	548	548	548	548	548	548
R^2	0.078	0.053	0.042	0.030	0.035	0.008
Adjusted R^2	0.076	0.051	0.040	0.028	0.033	0.006

Note: *p<0.1; **p<0.05; ***p<0.01

7 Conclusions

In this paper, I provide the first estimates of (income) intergenerational mobility in Spain using administrative data linking parents and children through tax returns and the rank-rank approach. Exploiting the richness of the data, I estimate relative and absolute mobility at various geographical levels (national, regional, provincial and municipal), giving a detailed picture of geographic variation in intergenerational mobility. The results show that Spain is located somewhere in the middle between high-mobility countries such as Australia or Switzerland and low-mobility ones such as the United States or Italy. The geographical analysis reveals that, on top of parental income, the province of origin substantially determines children's economic outcomes as adults. Spain presents a high level of variation in mobility estimates, but this within-country variation is lower than in the United States, Italy and, in some measures, Switzerland. The most mobile areas tend to be located in the North/North-East of the country whereas the less mobile ones are mainly located in the South/South-West. The region with the highest level of both absolute and relative mobility is Cataluña, with mobility rates on the levels of Scandinavia. The regions with the lowest levels of absolute and relative mobility are Andalucia and Canarias, with absolute mobility estimates similar to Southern United States ones. In addition, I document a positive association between relative and absolute mobility: areas that show high levels of absolute mobility tend to also display high rates of relative mobility. I find a negative association between income inequality and absolute mobility measures and a positive one with relative mobility measures, confirming the existence of a Great Gatsby Curve within Spain.

The gender analysis shows that daughters have systematically worse intergenerational mobility outcomes than sons both in relative and absolute terms and across different geographical levels. To give a sense of this gap, daughters who grew up in median-income households end up, on average, at the 46th percentile while the sons of those same families reach the 52th percentile. This corresponds to an average income gap of €2,796 (a 13% of the national mean income). Furthermore, gender gaps tend to be greater in provinces that present low relative mobility (higher RRS) but there is mixed evidence regarding the association between gender gaps and absolute mobility. There is no conclusive evidence concerning the relationship between provincial income and gender gaps.

Finally, I identify a positive association between out-migration and upward mobility, regardless of children's parental income. The vast majority of children that leave their home province migrate to richer provinces than the home one, which points towards an economically driven migration. Examining the family origins of these movers to richer provinces, I uncover an U-shaped pattern, which shows that the majority of this type of movers come from either relatively poor or relatively rich households. This finding reinforces the economically driven out-migration argument since the children that move out to richer provinces are predominantly those that have the higher incentives to do so, either to improve their well-being ("looking-for-opportunities" migration) or to maintain themselves at the top of the social ladder ("keep-the-status" migration). In an attempt of better capturing the effect of leaving the home province on upward mobility, I exploit the high comparability degree of siblings that have the exact same values of observable characteristics to estimate the quasi-causal effect of out-migration across various upward mobility outcomes. I estimate that moving out from the home province increases, on average, child income by almost €8000 and the child rank in their own income distribution by 16 percentiles.

One of the main problems of the cross-country comparisons in intergenerational mobility is the lack of datasets combining fiscal information from the same cohorts of parents and children with socio-economic information of the same individuals in a wide range of countries. The World Bank has taken an important step in this direction by creating the Global Database on Intergenerational Mobility (GDIM). This centralized database allows for a systematic comparison of educational mobility trends across different generations and countries. Without a similar project for income mobility, our knowledge about the evolution of intergenerational mobility and its (causal) determinants remains restricted to a limited number of country studies. On a more methodological note, there is a need of developing income relative measures that abstract from the different inequality levels across the compared countries. Changes over time in the rank-rank slope (the main measure used to estimate relative mobility) are not informative, since it can be disentangled whether they are driven by worse outcomes of children from top-income families or by better outcomes of children from low-income ones.

Appendix A Figures & Tables

Figure A.1: Graphical Illustration of the Rank-Rank Approach to Estimate Relative Robility

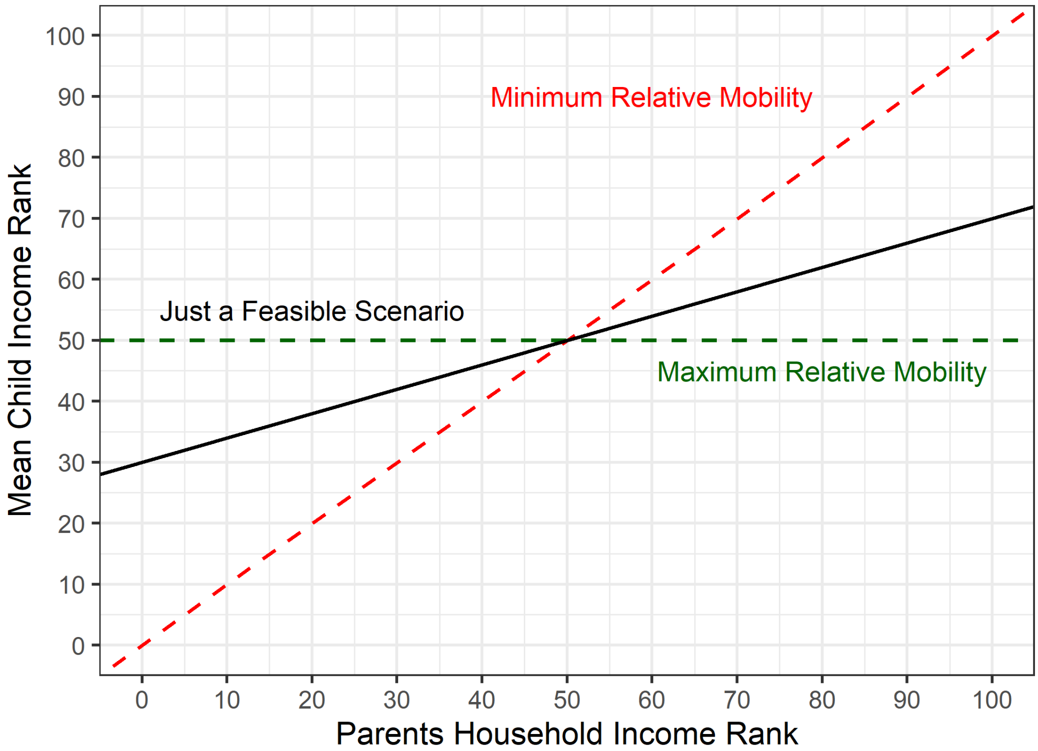


Figure A.2: *Atlas de Oportunidades* Project Dataset Timeline

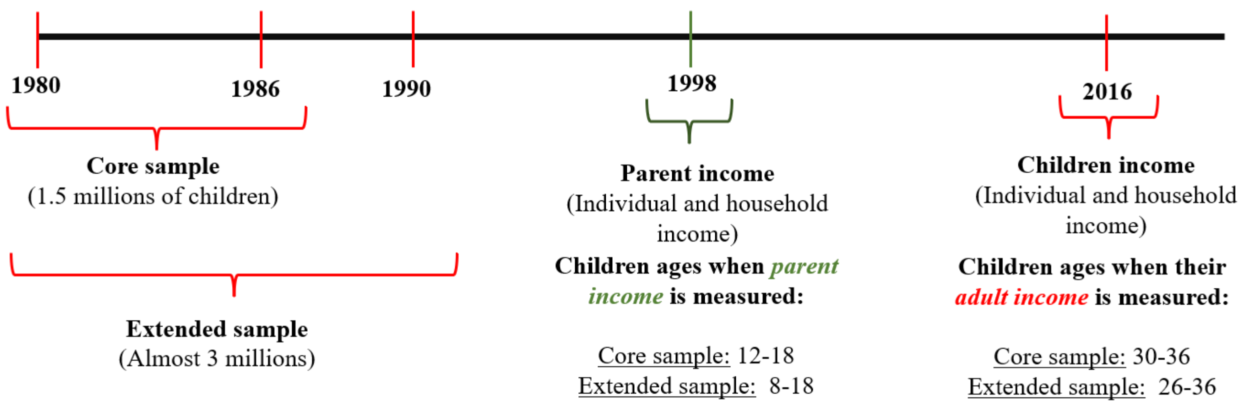


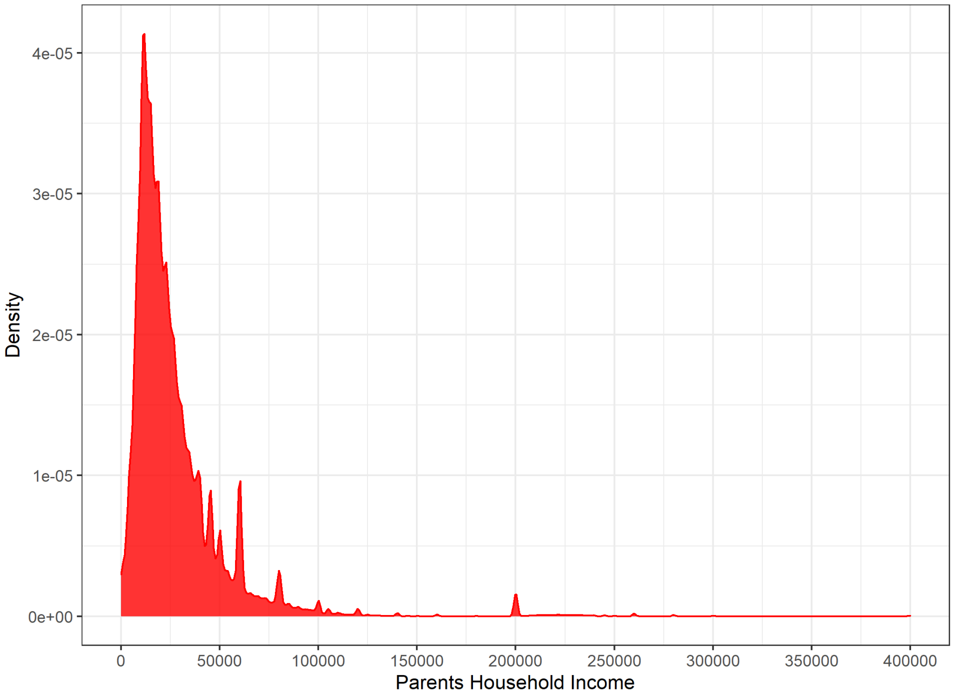
Table A.1: Description of Child Age Depending on the Year of Observation

	Child age										
2016 (adult child)	26	27	28	29	30	31	32	33	34	35	36
1998 (in parents' hh)	8	9	10	11	12	13	14	15	16	17	18
Cohort	1990	1989	1988	1987	1986	1985	1984	1983	1982	1981	1980

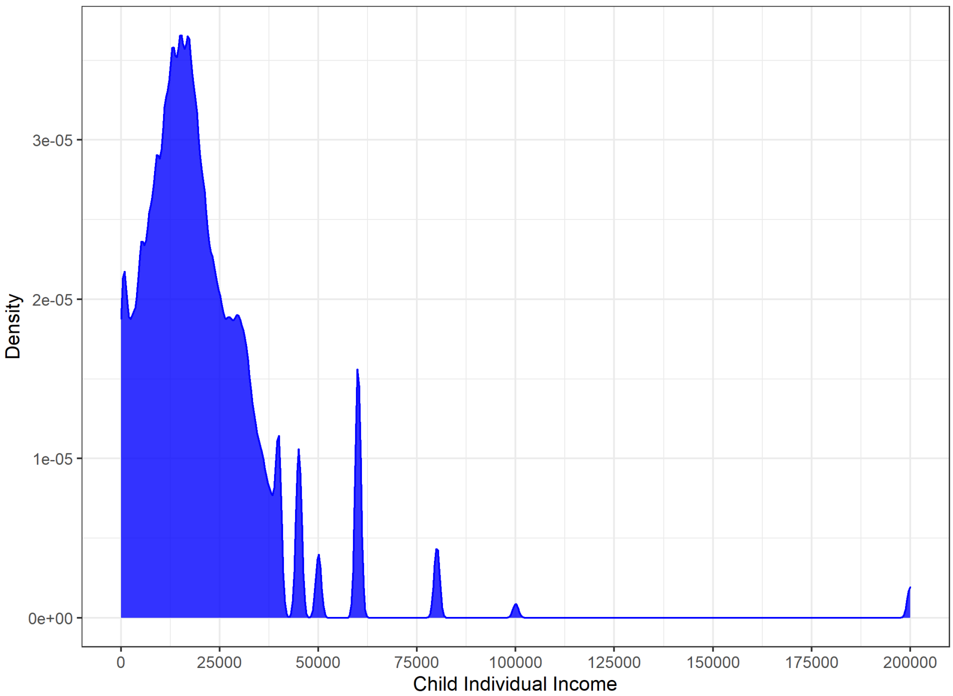
Notes: the sample selection used for the analysis includes children born between 1980 and 1986

Table A.2: Descriptive Statistics of the Core Sample

	Value
Parents in 1998	
Mean Parents Household Income	27113.92
Mean Parent 1 Income	20850.76
Mean Parent 2 Income	6263.16
Share Positive Income - Parent 1	99.5%
Share Positive Income - Parent 2	27%
Share Positive Income - Both Parents	26.9%
Parent 1 Income > Parent Income	84.8%
Children in 2016	
Cohorts	1980-1986
Mean Age (in 2016)	32.69
Men Share	50.7%
Women Share	49.3%
Married Share	36.9%
Single Share	61.1%
Divorced Share	1.9%
Widowed Share	0.4%
Mean Child Individual Income	20557.58
Mean Child Household Income	25668.43
Mean Child Individual Income Men	22004
Mean Child Individual Income Women	19071
Sample Size	1,492,107



(a) Child Individual Income



(b) Parents Household Income

Figure A.3: Parental and Child Income Densities

Table A.3: Relative Mobility Estimates by Gender at the National Level

	Dependent variable:		
	Mean Child Income Percentile		
	National	Men	Women
Parents Household Income Percentile	0.195***	0.179***	0.211***
	(0.005)	(0.004)	(0.007)
Constant	40.604***	44.113***	37.090***
	(0.308)	(0.257)	(0.390)
TMR	5.45	5.81	5.16
Observations[1]	1,492,107	756,293	735,813
R^2	0.933	0.944	0.910
Adjusted R^2	0.932	0.943	0.909

Note: *p<0.1; **p<0.05; ***p<0.01

[1] Observations here refer to the number of individuals used to build the 100x100 percentile matrix that is then employed to estimate the rank-rank slopes (RRS)

Figure A.4: Association between Children's and Parents' Percentile Ranks in Spain with Minimum and Maximum Indicators

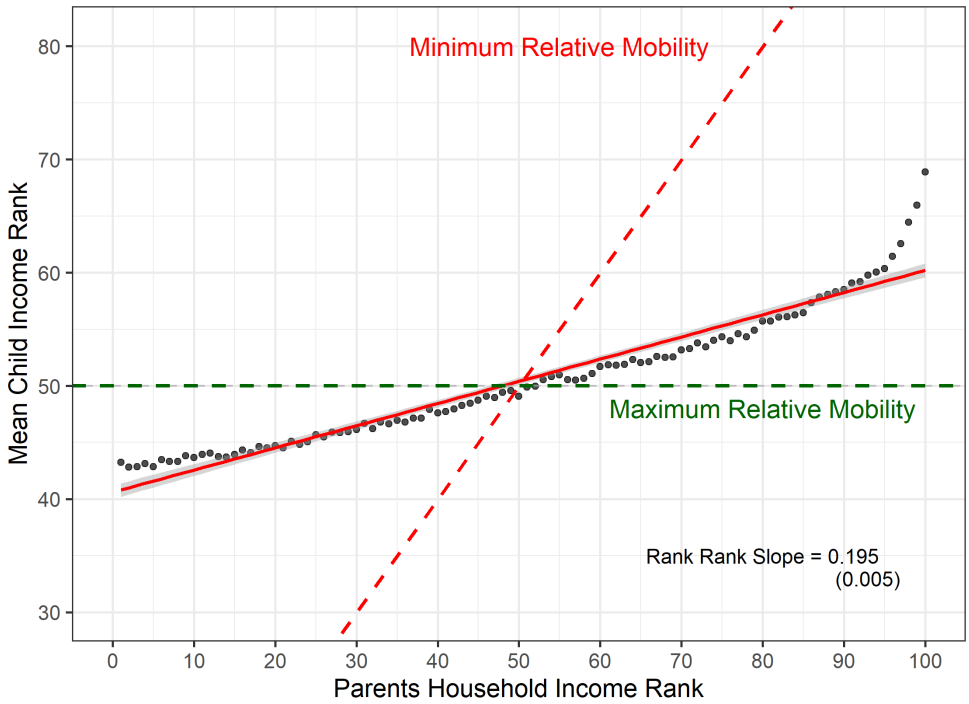


Figure A.5: Mean and Median Child Income by Parental Income Rank

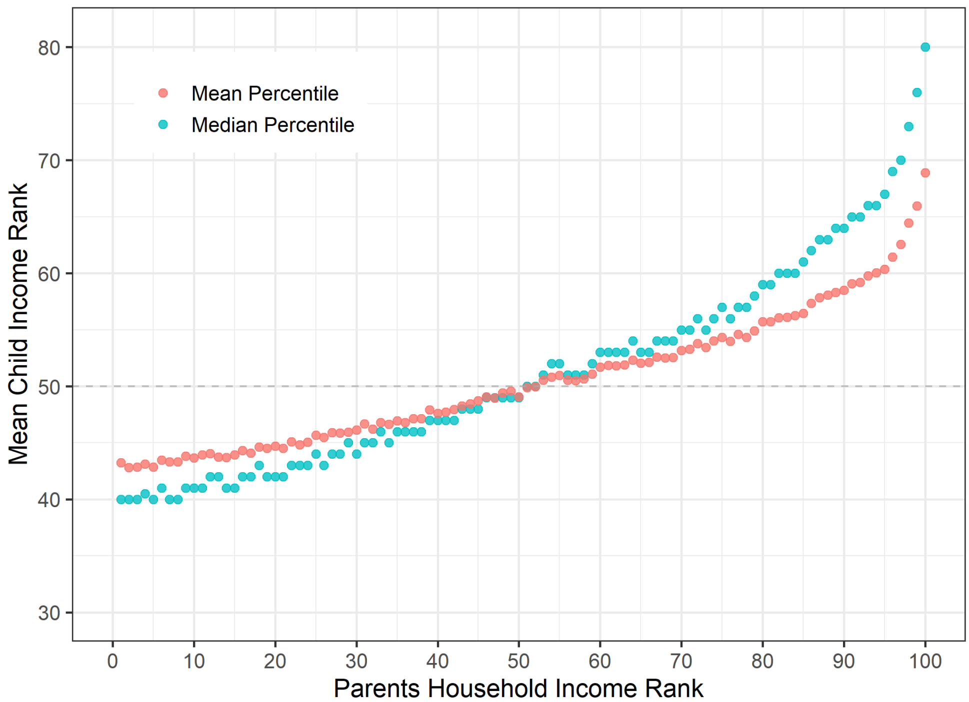


Figure A.6: Conditional Distributions of Child Ranks at Bottom and Top Parental Income Deciles

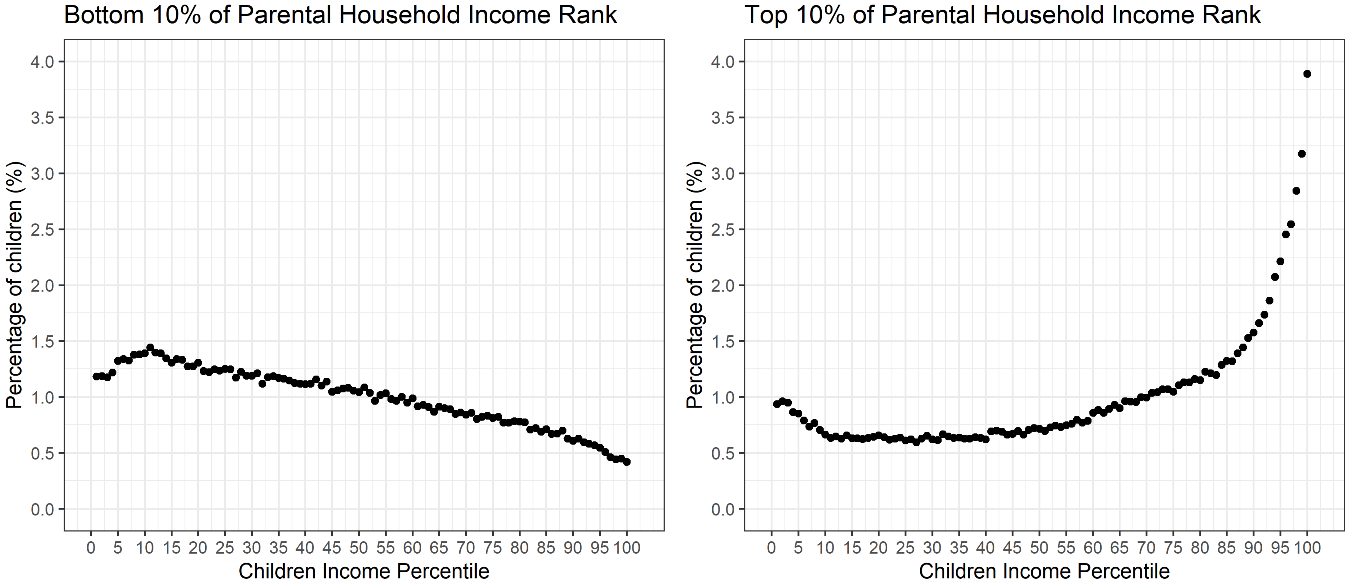


Figure A.7: Mean Child Income by Parental Income Rank

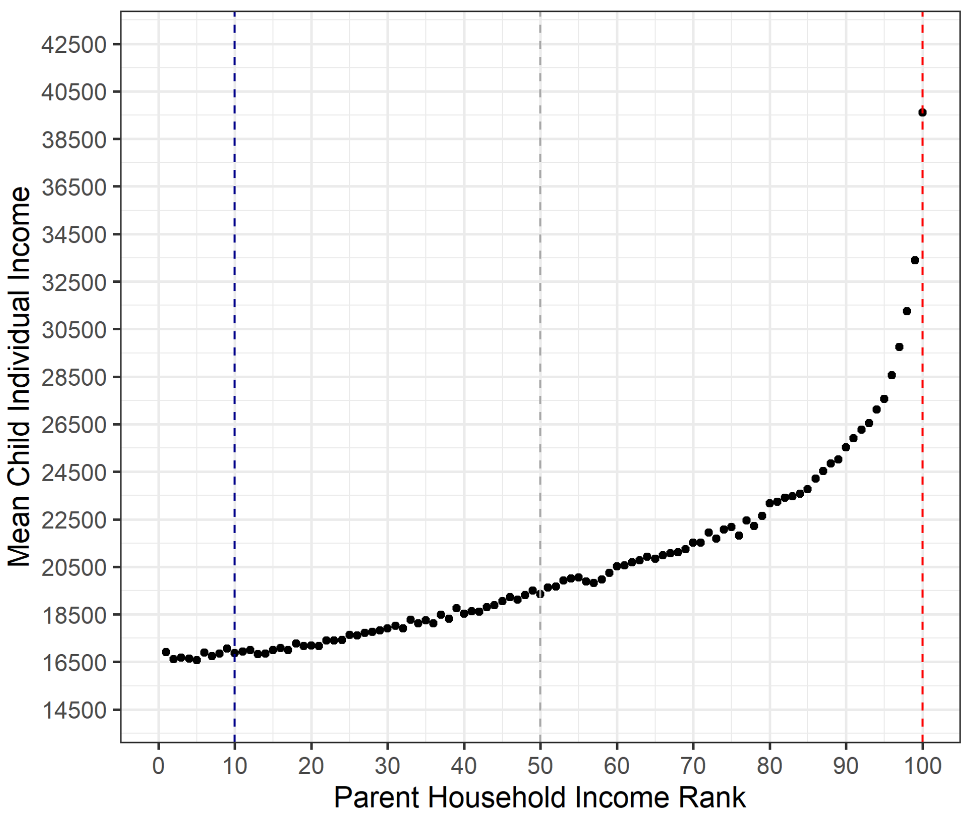


Figure A.8: Mean Child Income by Parental Income Rank and Child Gender

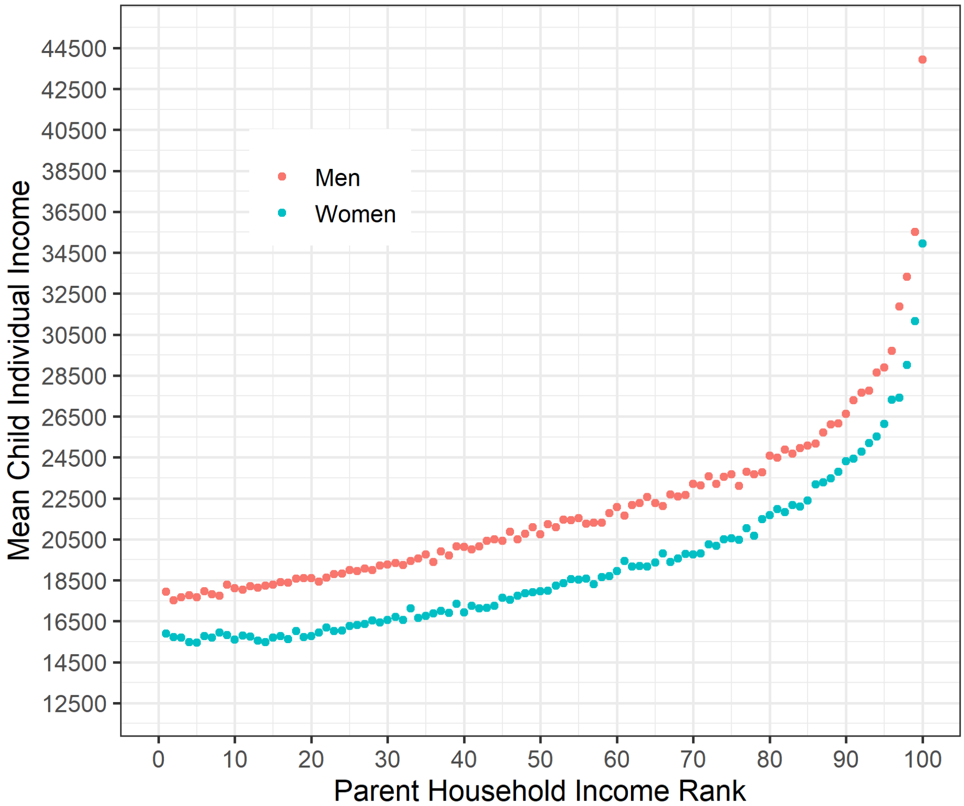


Figure A.9: Mean Child Income by Parental Income Rank, Child Gender and Cohort

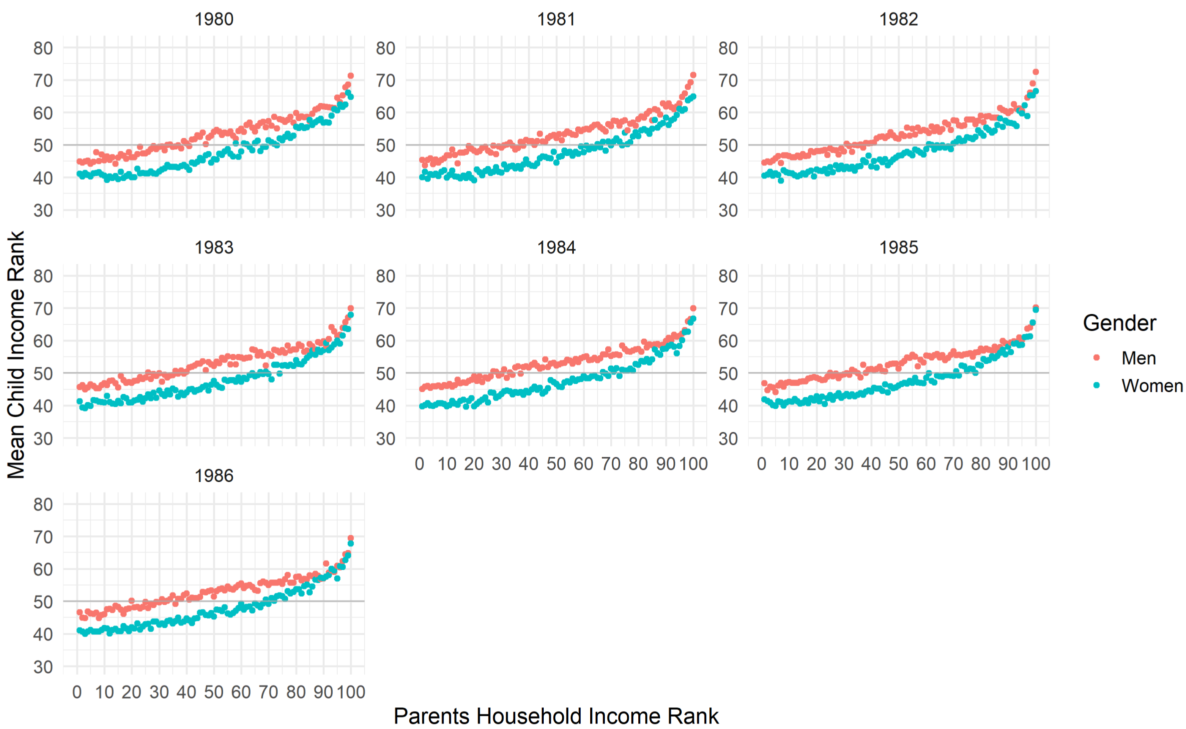


Table A.4: Relative Probability of Getting to the Top of the Distribution (RTP)

	(1)	(2)	(3)	(4)
Children from the:	Top 1%	Top 1%	Top 10%	Top 10%
Gender	T1/B10	T10/B50	T10/B10	T10/B50
Total	24.05	9.41	5.22	4.23
Men	23.33	8.98	4.77	3.85
Women	25.60	10.27	5.99	4.90

Figure A.10: Relative Probability of Getting to the Top 1% (RPT) - Column (2) from Table A.4

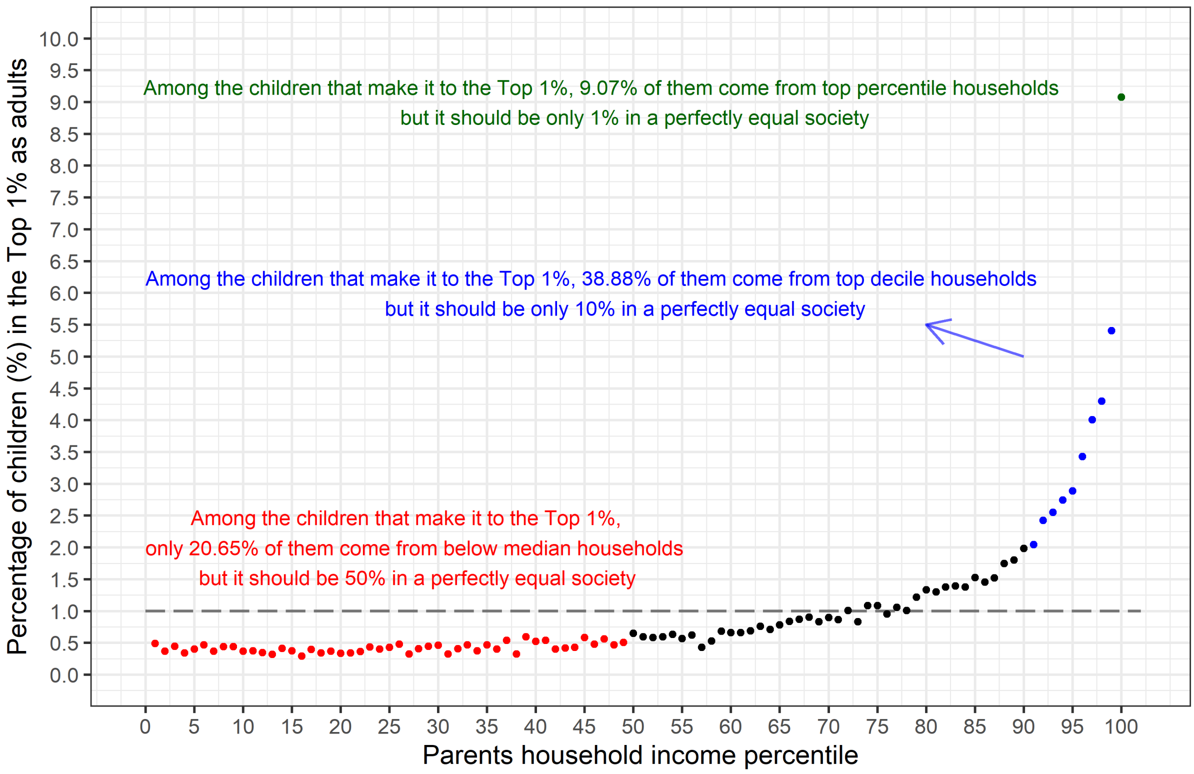


Figure A.11: Relative Probability of Getting to the Top 10% (RPT) - Column (3) from Table A.4

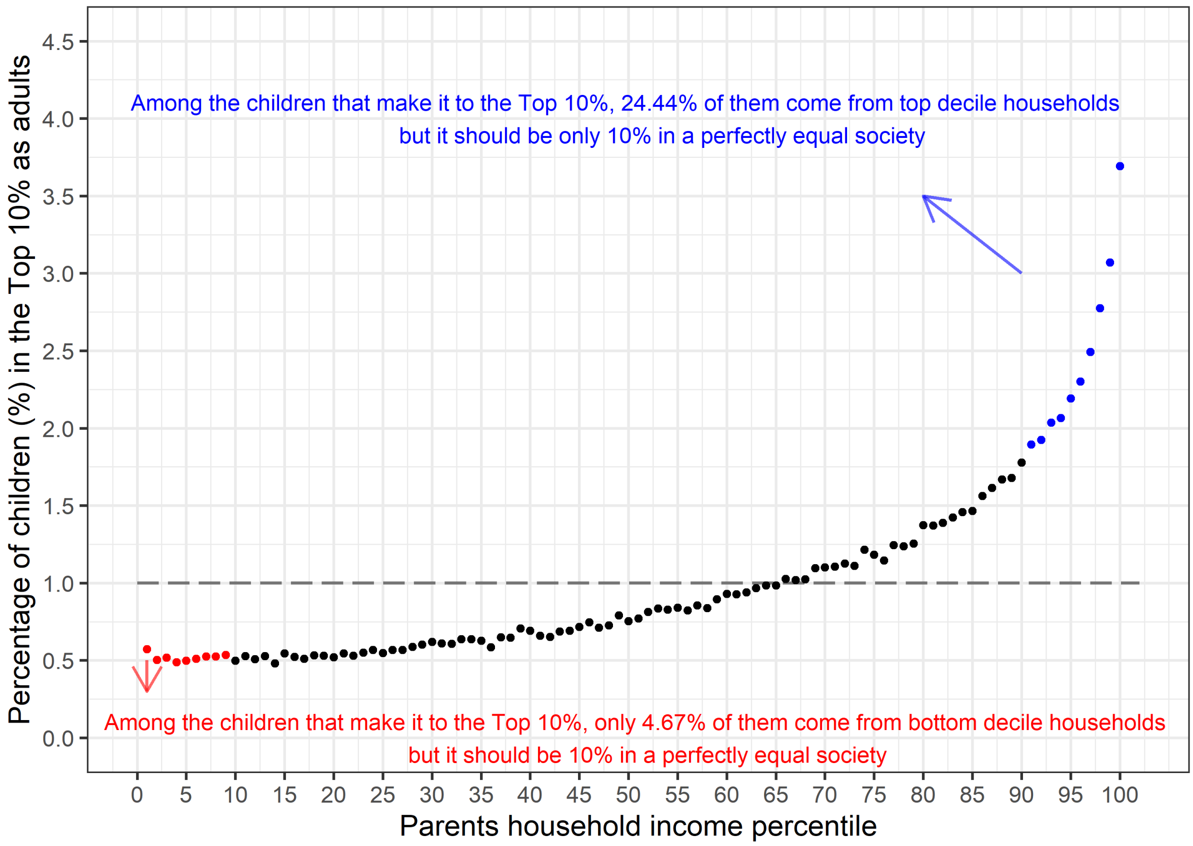


Figure A.12: Relative Probability of Getting to the Top 10% (RPT) - Column (4) from Table A.4

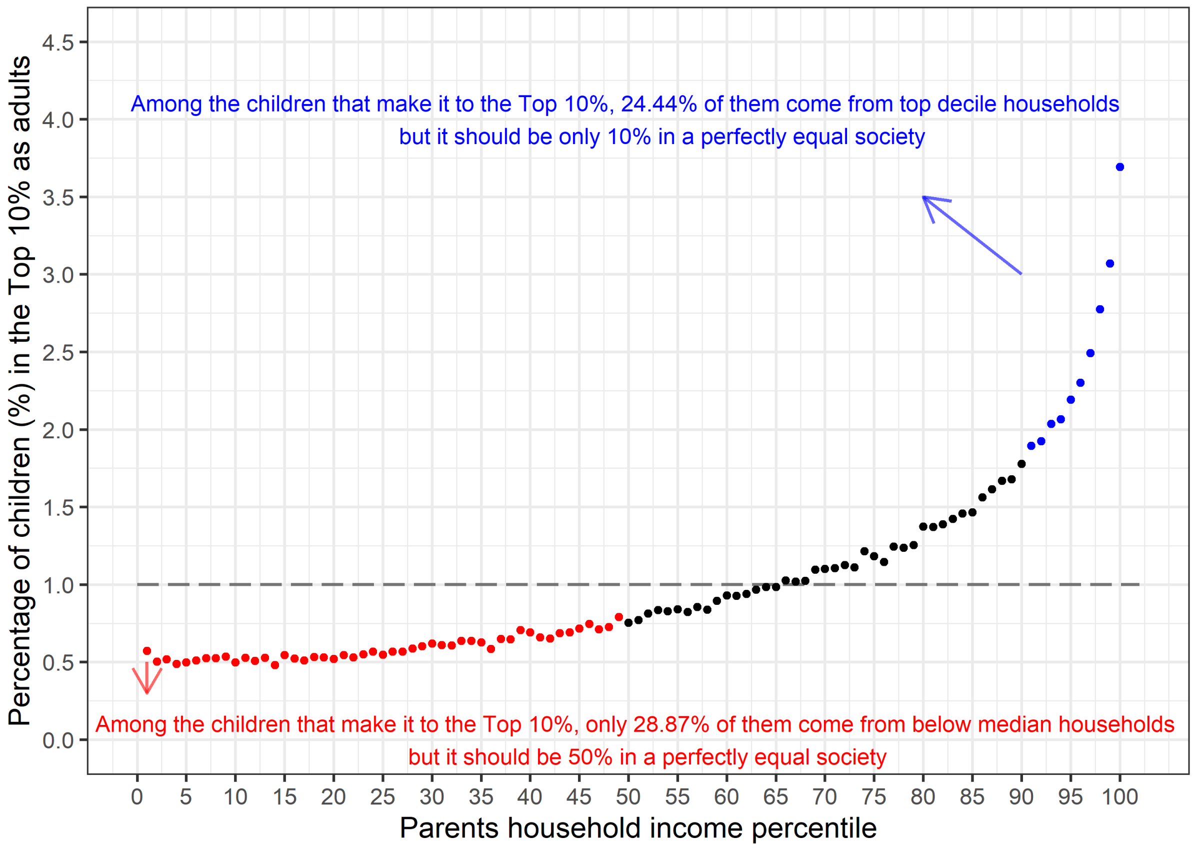


Figure A.13: Evolution of the AUM at the National Level

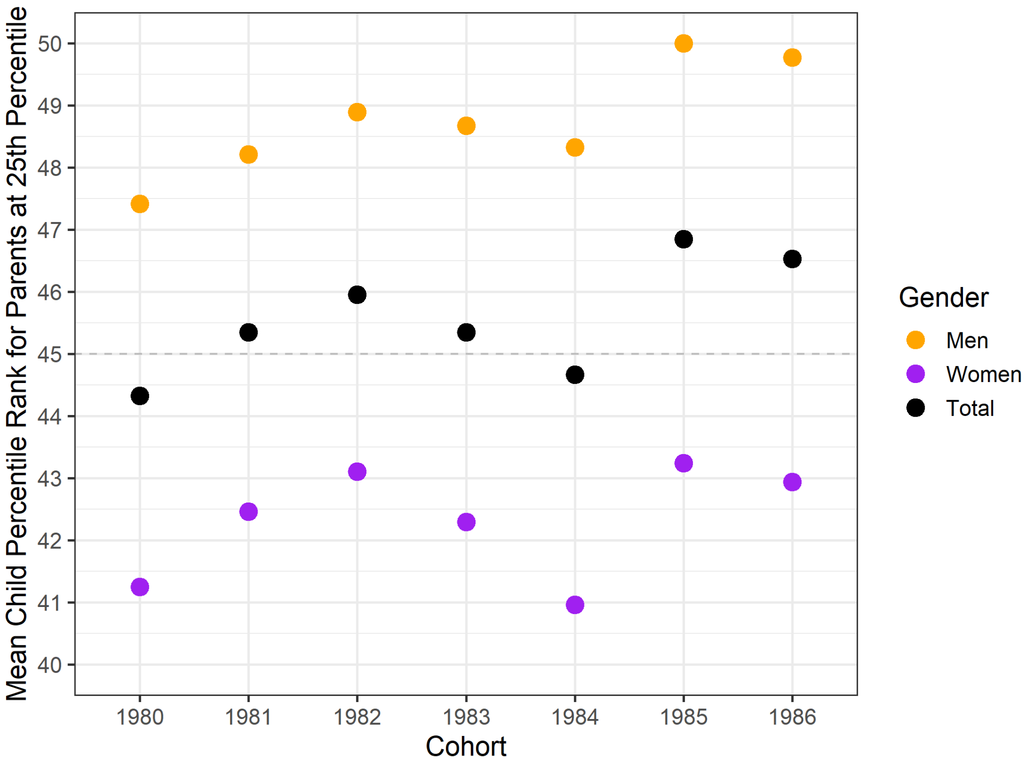


Table A.5: National Quintile Transition Matrix

	Parents household quintile				
Child quintile	1	2	3	4	5
1	25.28	22.31	19.43	17.54	15.41
2	24.02	22.96	20.85	18.17	13.96
3	21.25	21.83	21.09	19.78	16.02
4	17.15	18.74	20.92	21.88	21.29
5	12.27	12.13	17.68	22.60	33.29

Figure A.14: Probability of Ending Up in the Top Quintile Coming from a Household in the Bottom Quintile at the National Level (by Gender)

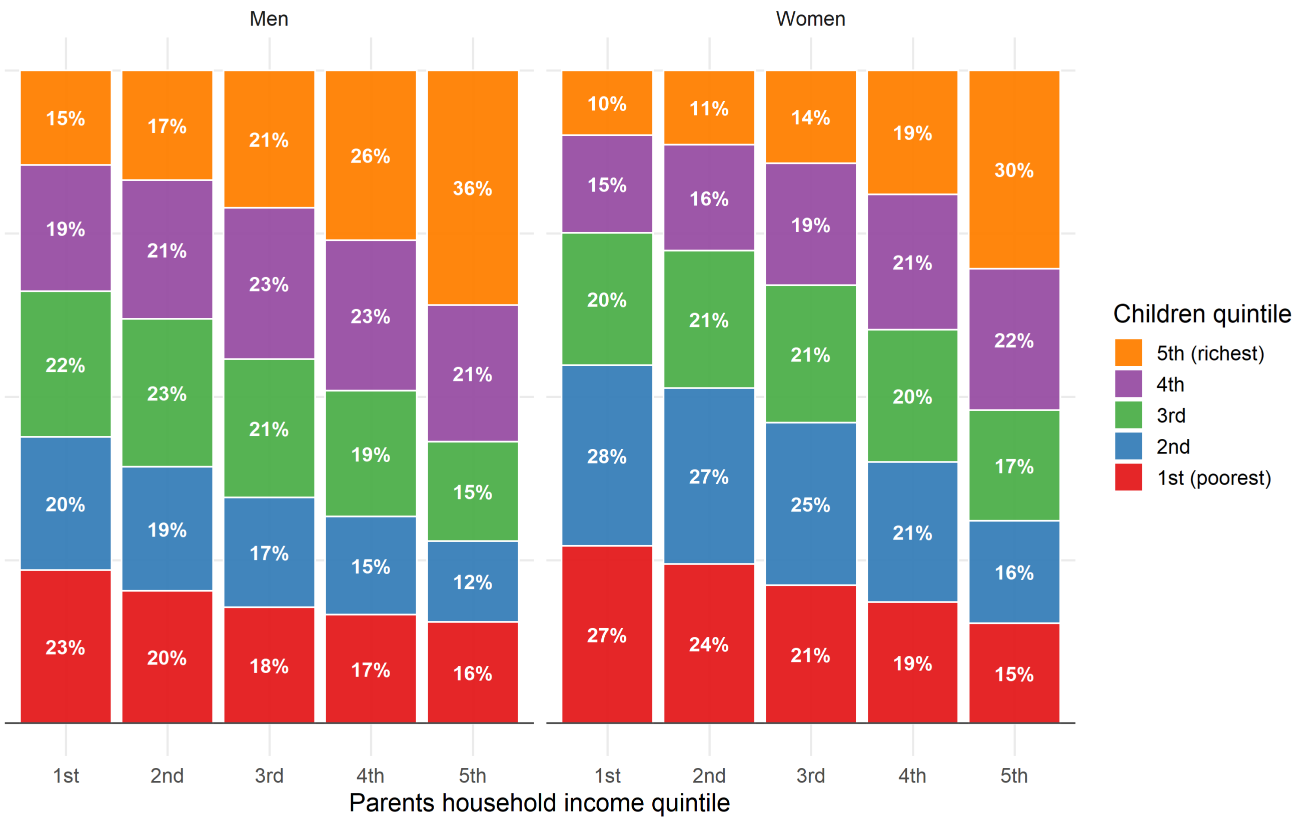


Figure A.15: Probability of Ending Up in the Top Quintile Coming from a Household in the Bottom Quintile at the National Level (by Cohort)

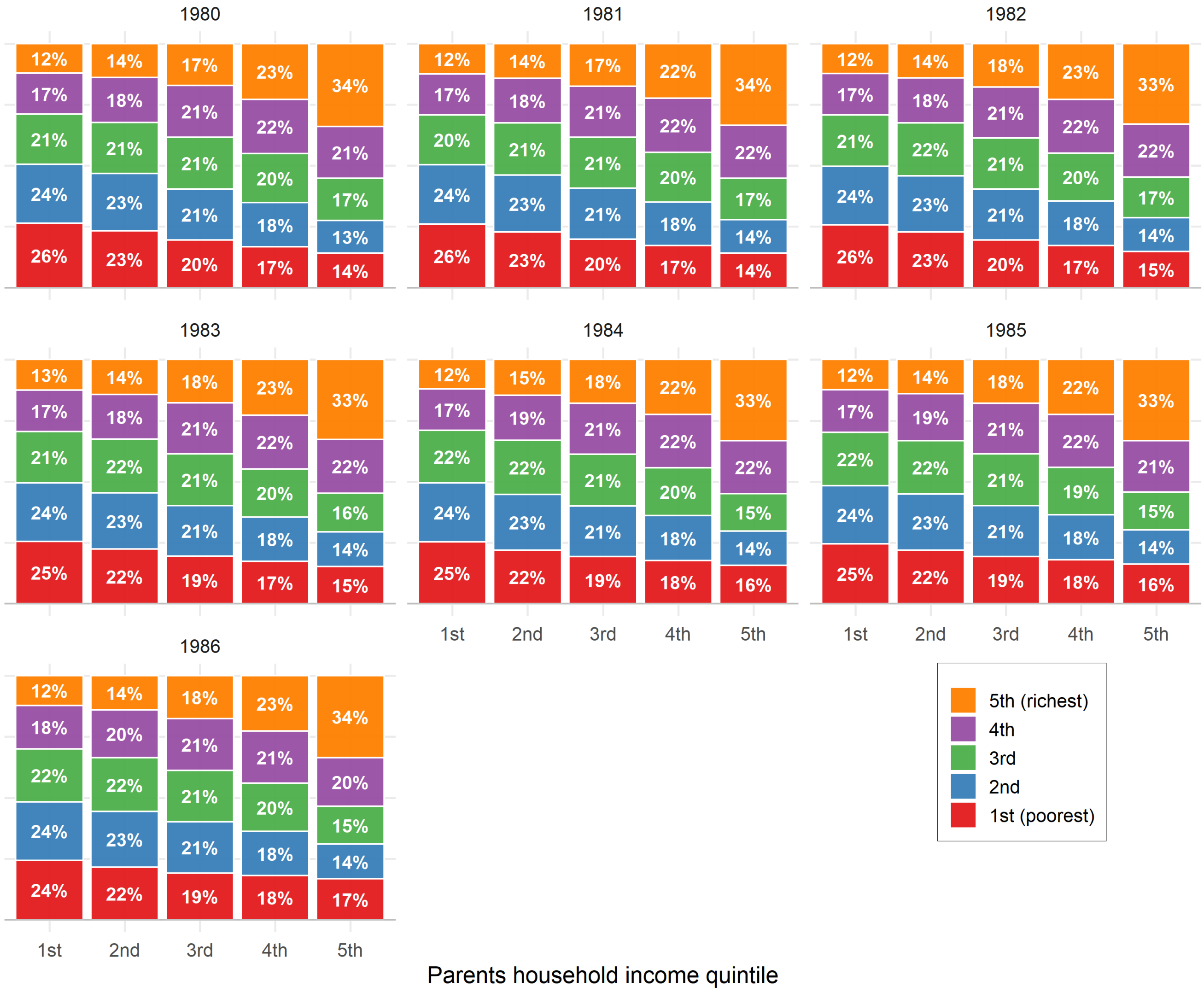


Figure A.16: Probability of Ending Up in the Top Quintile Coming from a Household in the Bottom Quintile at the National Level (by Gender and Cohort)

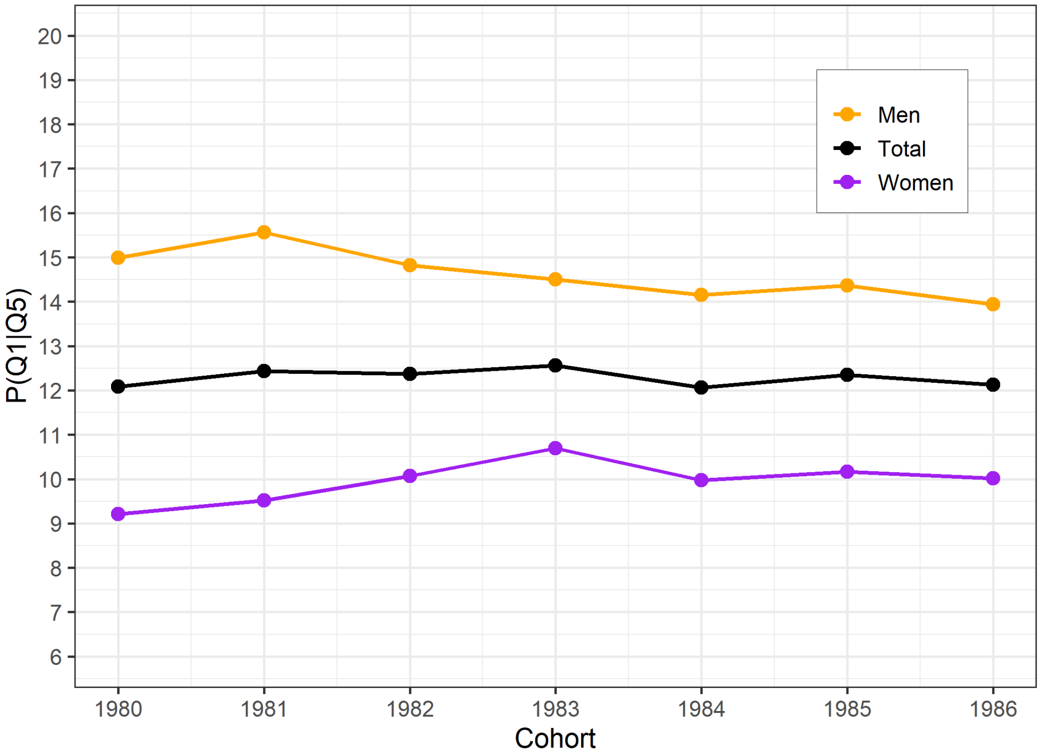


Figure A.17: Probability of Earning More than Parent 1 by Parental Rank (by Gender)

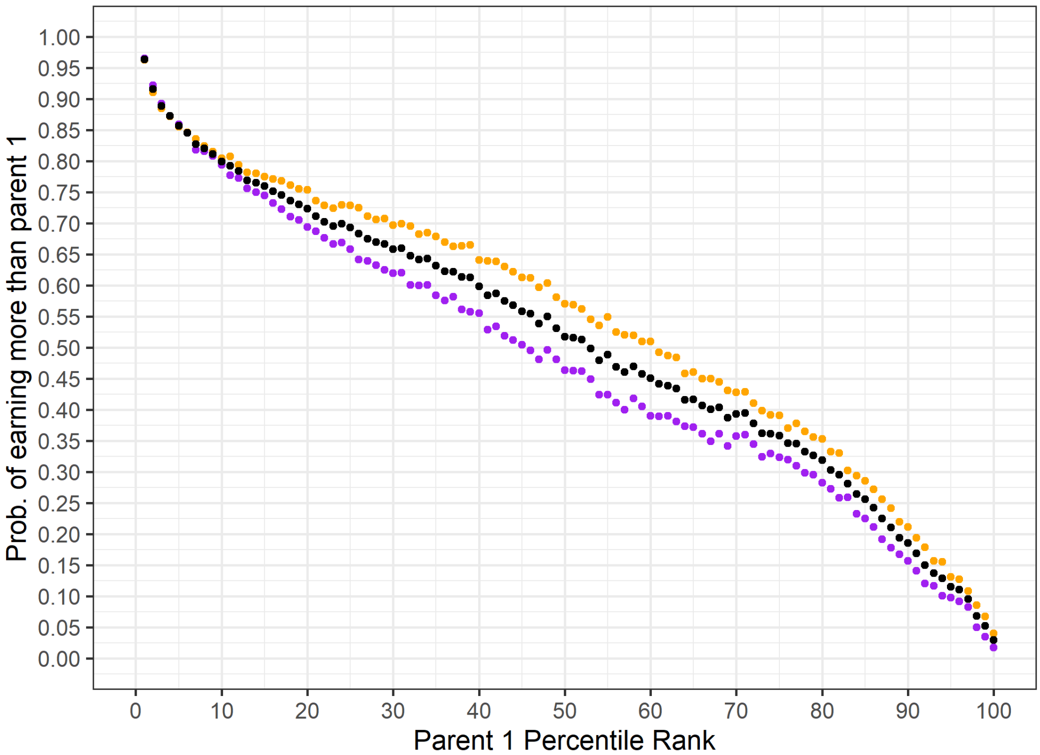


Figure A.18: Relative Mobility Estimates (RRS): Spain vs. Italy

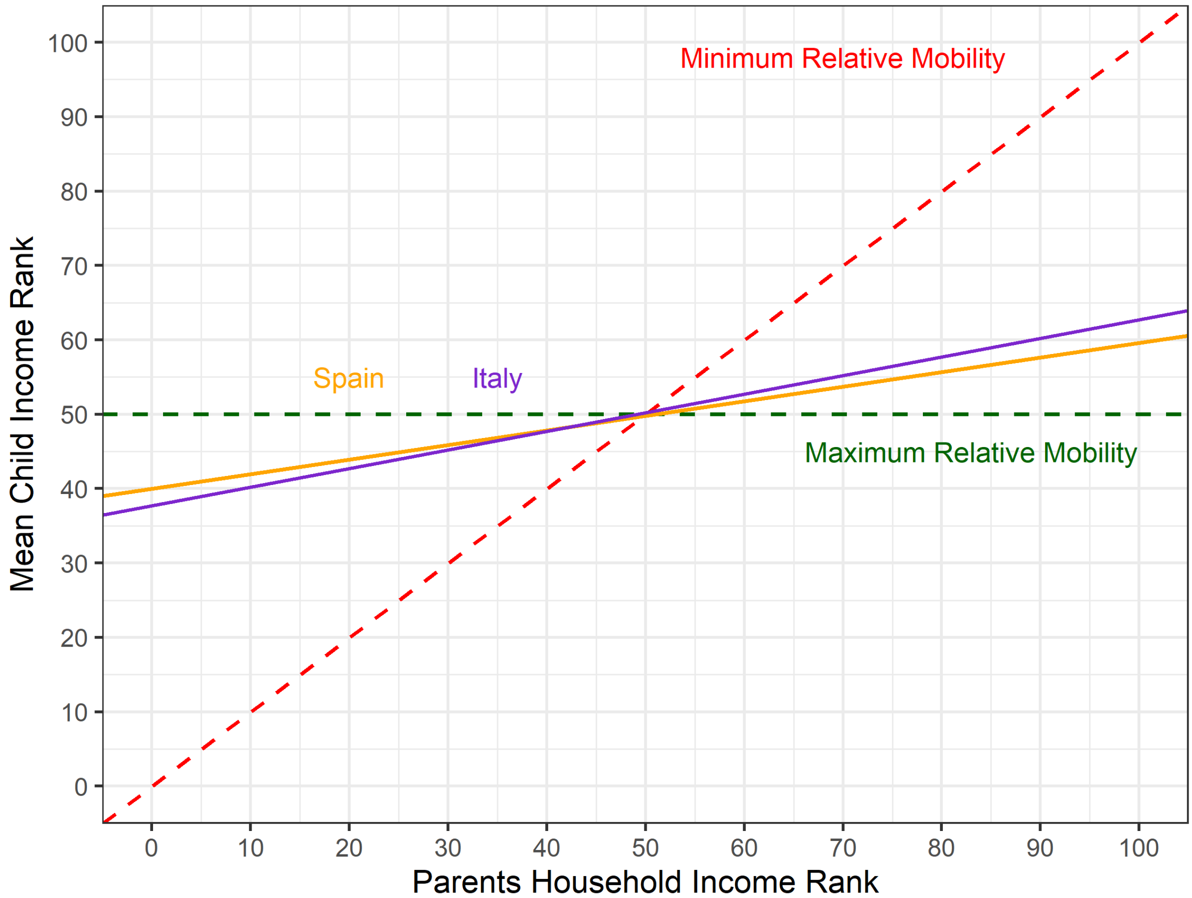
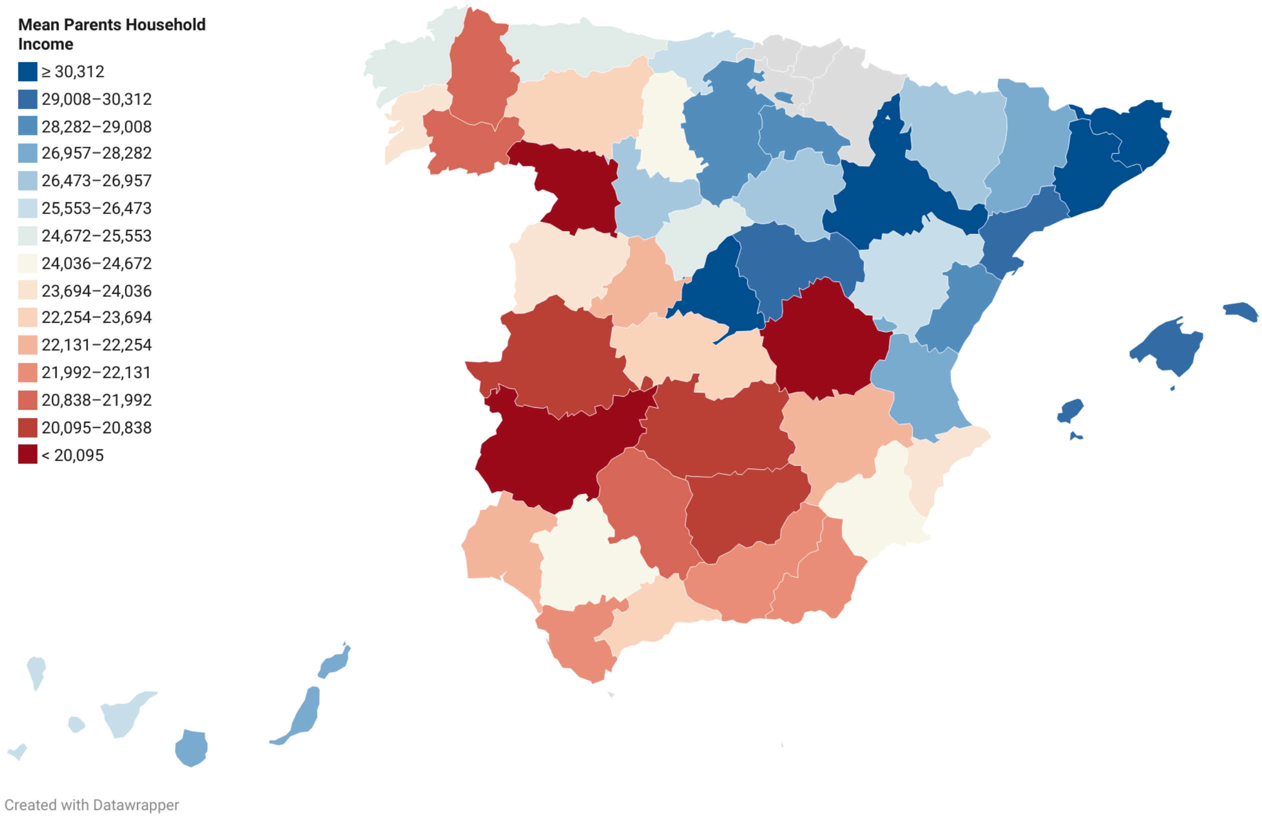
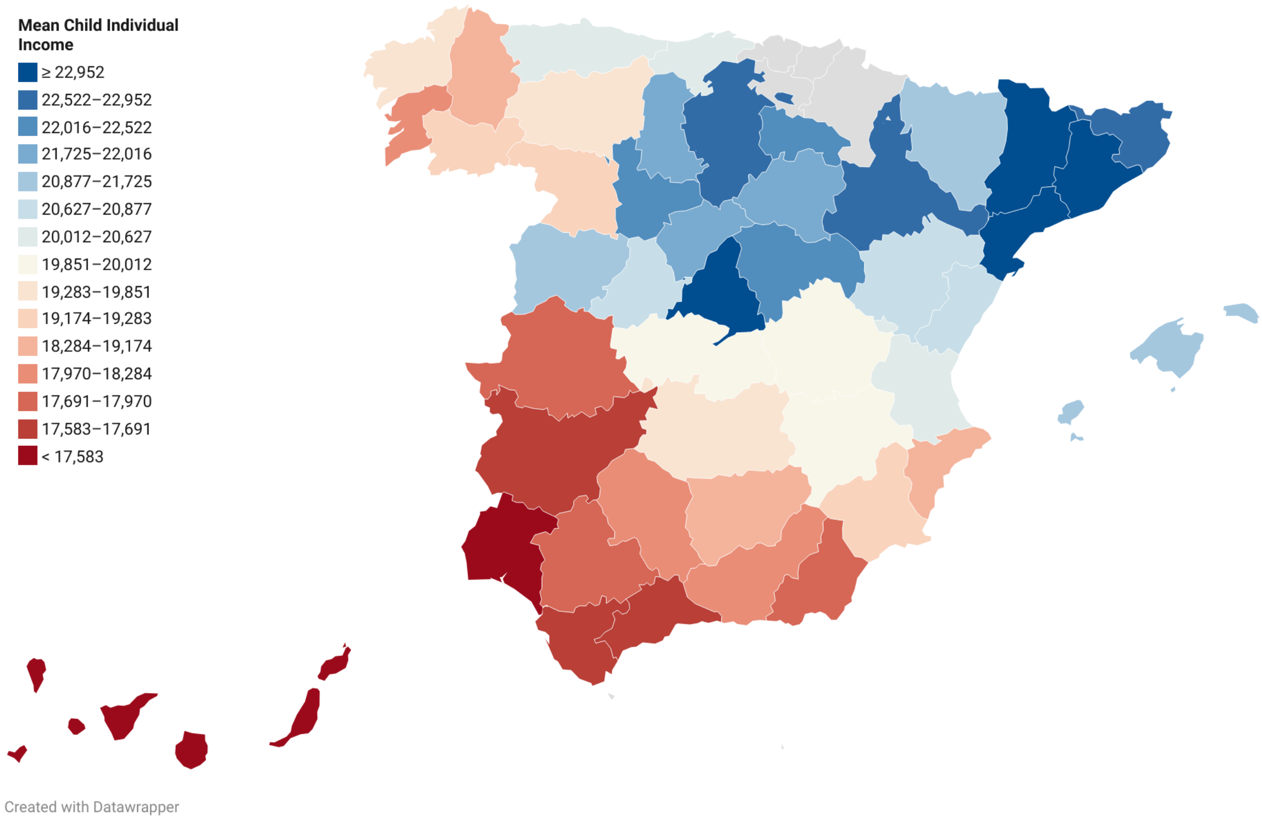


Figure A.19: Map of Spain



(a) Mean Parental Income

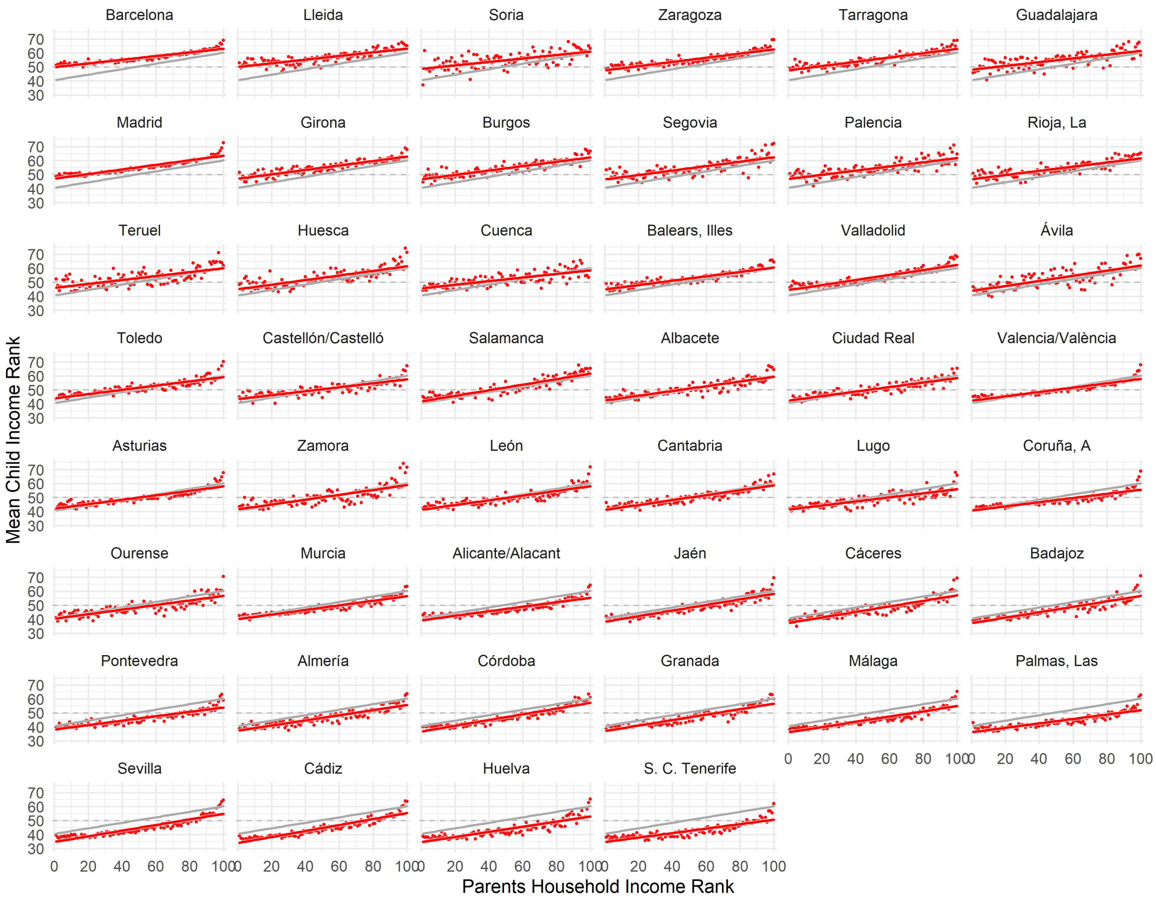


(b) Mean Child Individual Income

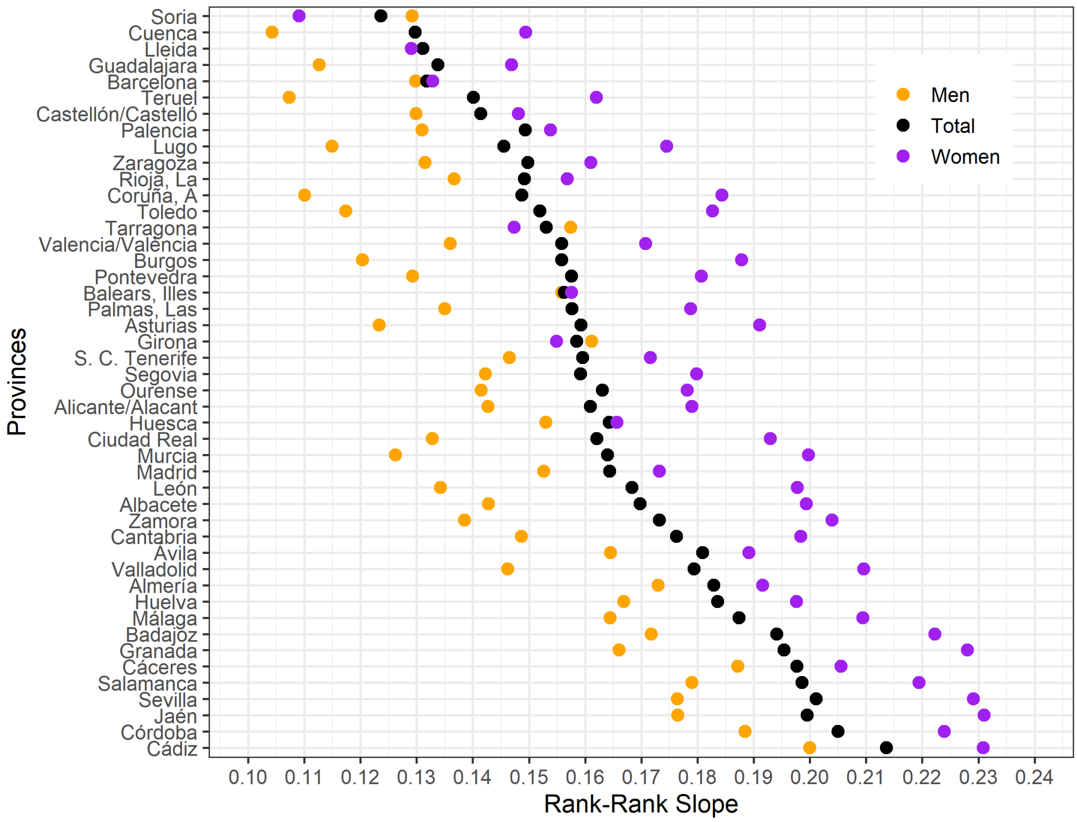
Figure A.20: Heatmaps of Parental and children Incomes by Province

Table A.6: Summary Statistics at the Regional Level (CCAA)

	Andalucía	Aragón	Asturias	Baleares	Canarias	Cantabria	C.La Mancha	C.y León	Cataluña	C.Valenciana	Extremadura	Galicia	Madrid	Murcia	La Rioja
Mean Parental Income	22337.10	29216.32	25393.28	29751.93	26347.45	25553.42	22193.19	24875.37	33038.69	26135.96	19549.02	23624.87	35928.18	24036.09	28317.52
Median Parental Income	16996.90	23004.86	20995.25	21997.41	19001.84	19998.39	16004.36	19997.03	24998.98	19004.05	13998.72	17996.07	26000.44	18001.81	21997.10
Mean Child Income	17781.07	22256.53	20332.76	21242.52	17337.45	20012.20	19988.33	21159.45	23562.15	19717.67	17742.53	18818.15	24517.08	19186.35	22032.99
Median Child Income	14998.11	19003.11	17003.07	17998.95	14001.73	17001.36	17003.98	17999.91	20996.14	16998.13	15000.29	16000.12	20003.12	16000.52	19001.46
% of total popul.	0.21	0.03	0.03	0.02	0.04	0.02	0.05	0.07	0.14	0.10	0.03	0.07	0.14	0.03	0.01
Mean Age	32.73	32.63	32.87	32.37	32.66	32.81	32.59	32.78	32.54	32.66	32.67	32.85	32.72	32.67	32.66
% Women	0.47	0.51	0.49	0.51	0.51	0.50	0.49	0.50	0.51	0.50	0.47	0.51	0.50	0.49	0.52
% Men	0.53	0.49	0.51	0.49	0.49	0.50	0.51	0.50	0.49	0.50	0.53	0.49	0.50	0.51	0.48



(a) Rank-Rank Association at the Province Level



(b) RRS by Province - Gender Differences

Figure A.21: Relative Mobility Across Provinces

Figure A.22: Relative Probability of Ending Up in the Top 1% (RTP-Top1%) by Region

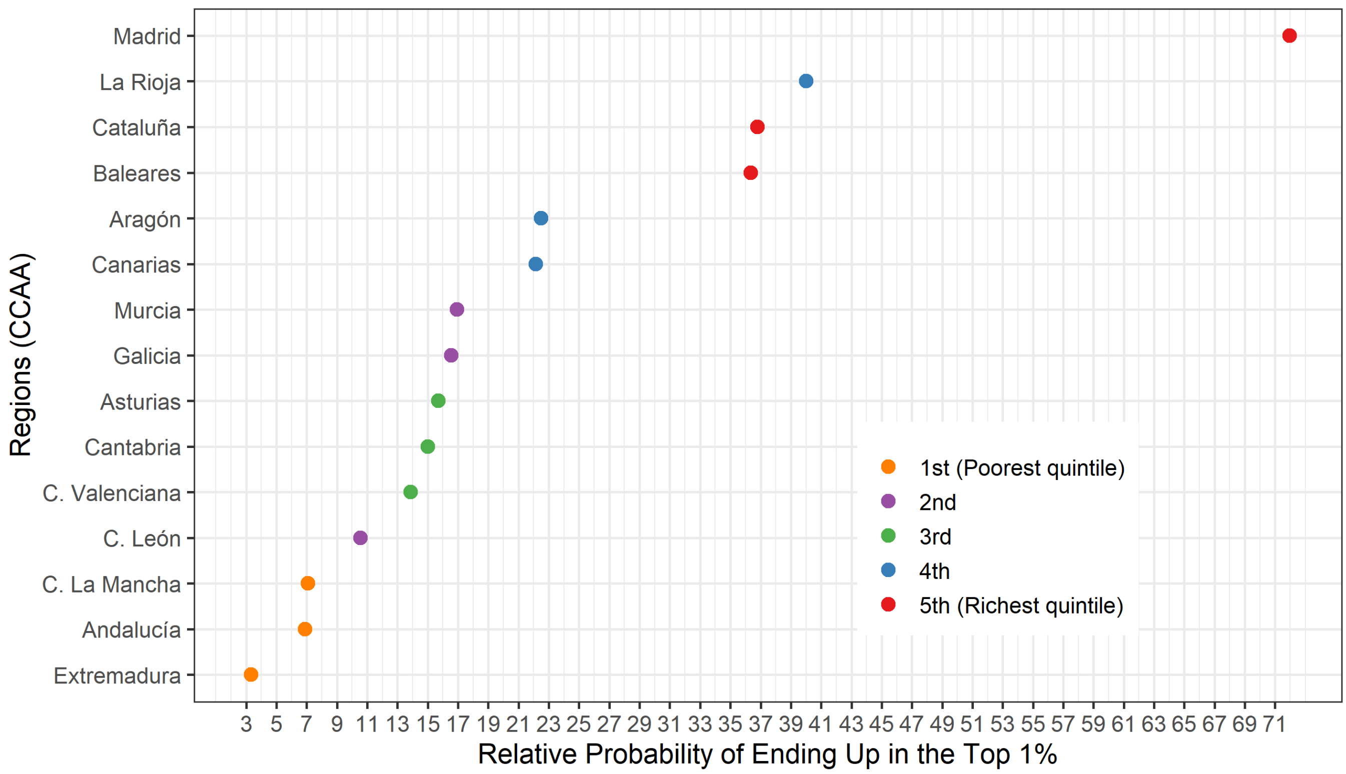
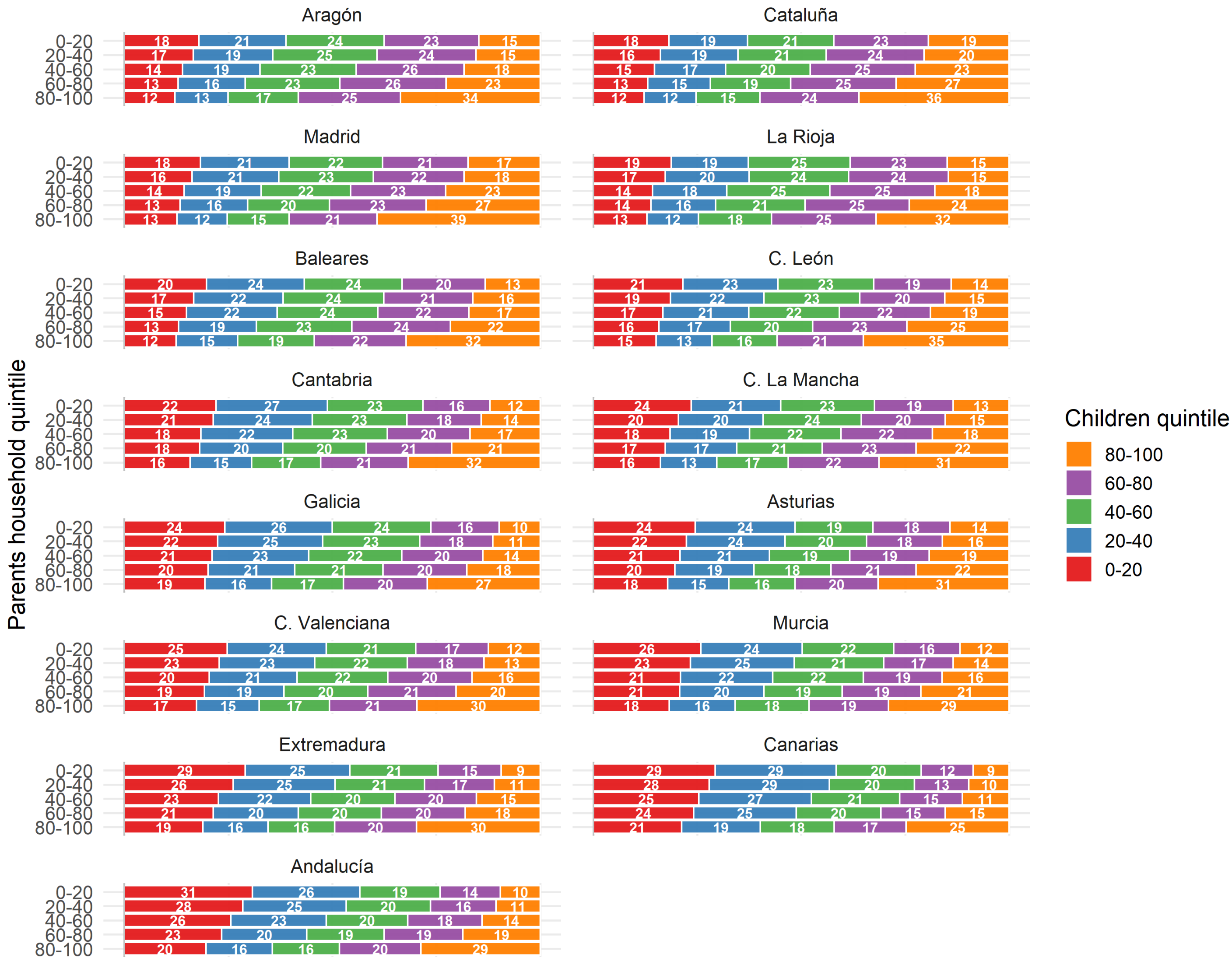
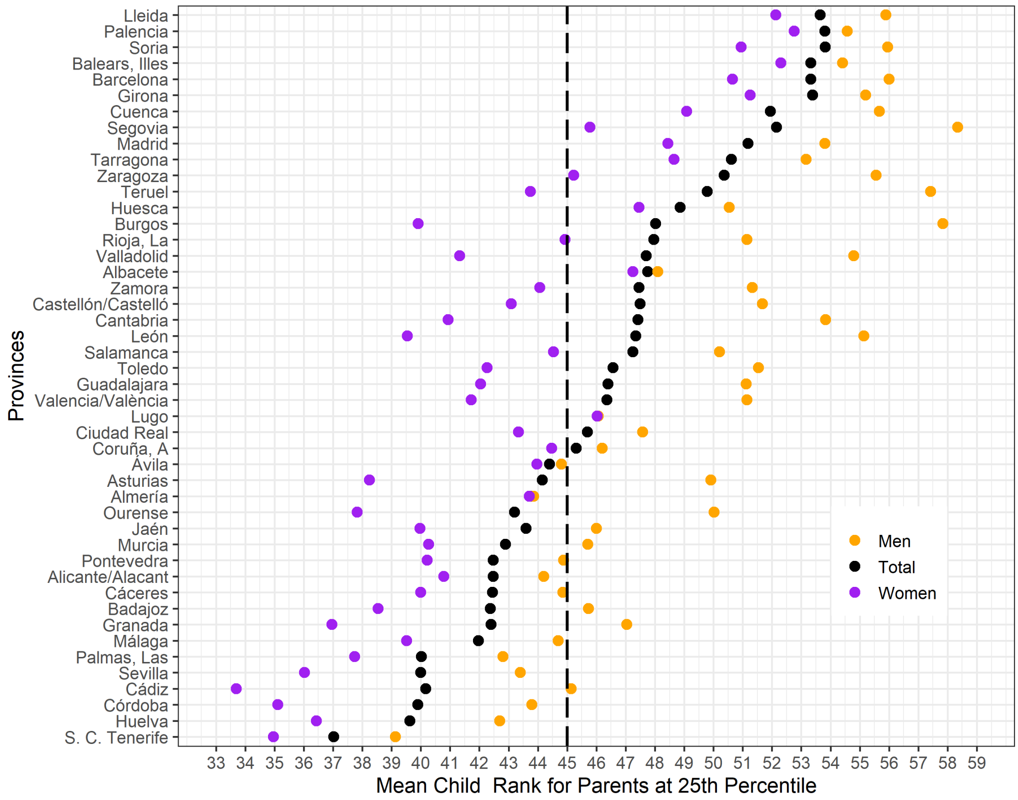
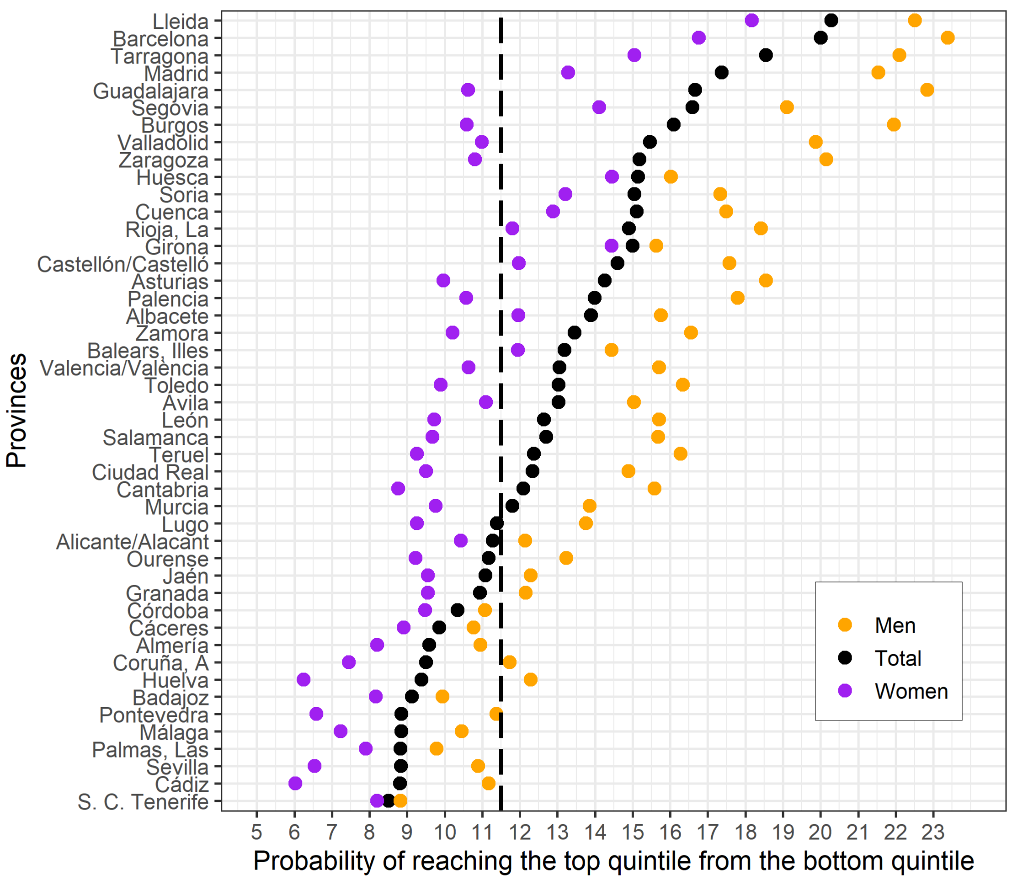


Figure A.23: Quintile Transition Matrices across Regions





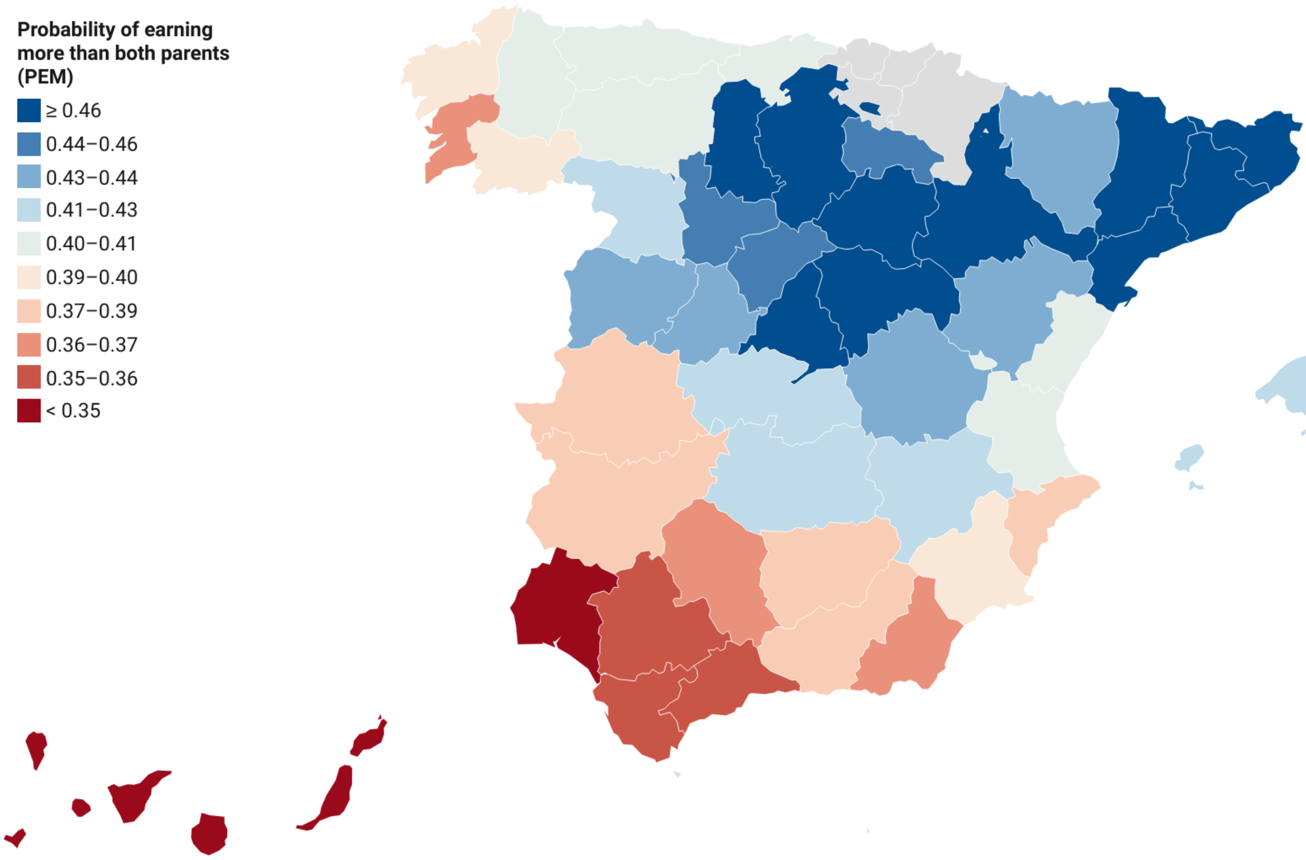
(a) Absolute Upwards Mobility (AUM) across Provinces

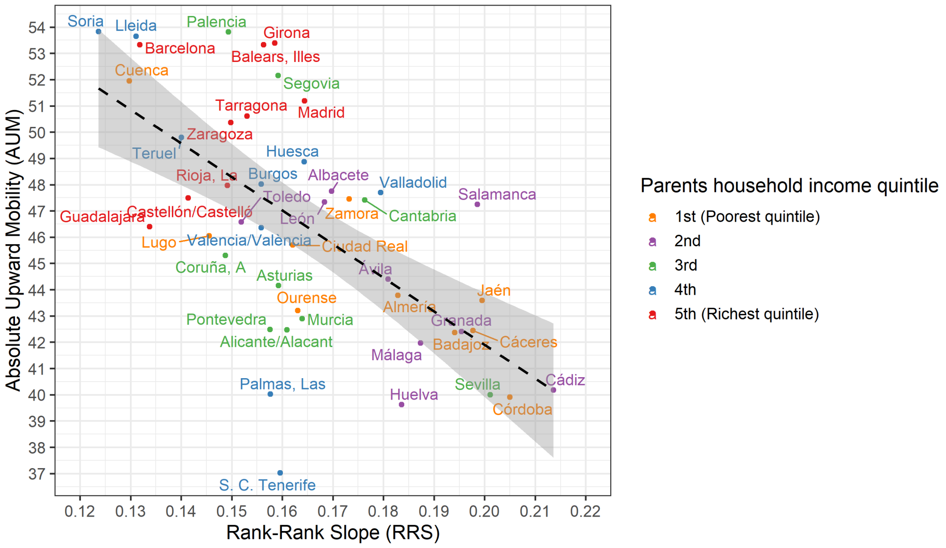


(b) P(Q5|Q1) across Provinces

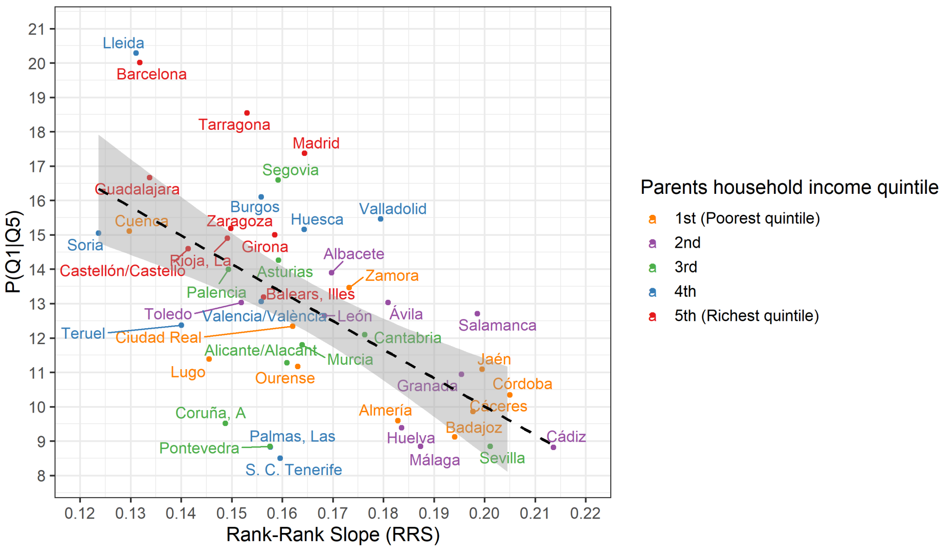
Figure A.24: Absolute Mobility Estimates Across Provinces

Figure A.25: Heatmap of the Probability of Earning more Than Parents (PEM) by Province

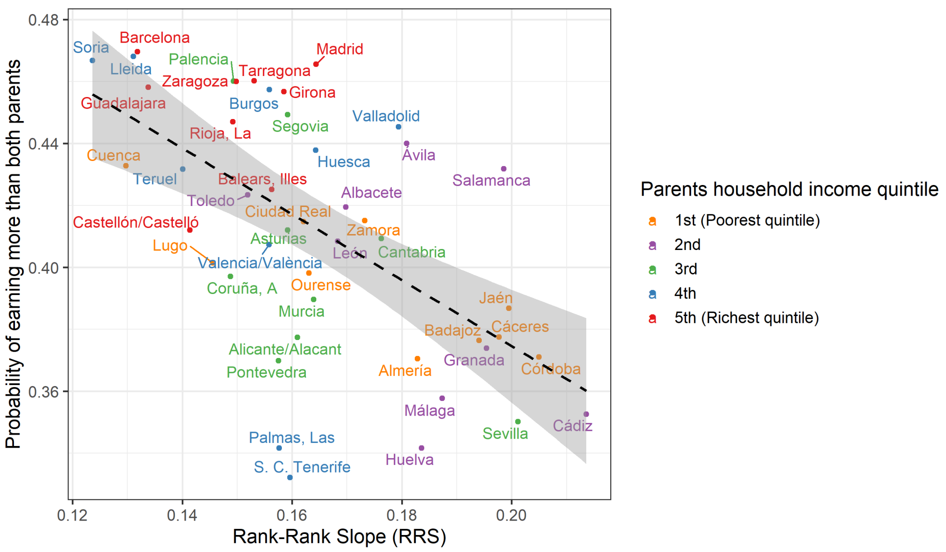




(a) AUM



(b) P(Q5|Q1)



(c) PEM

Figure A.26: The Association between relative and Absolute Mobility at the Province Level

Table A.7: Correlates of the Gender Gaps of Different Intergenerational Mobility Measures at the Province Level

Dependent variable:

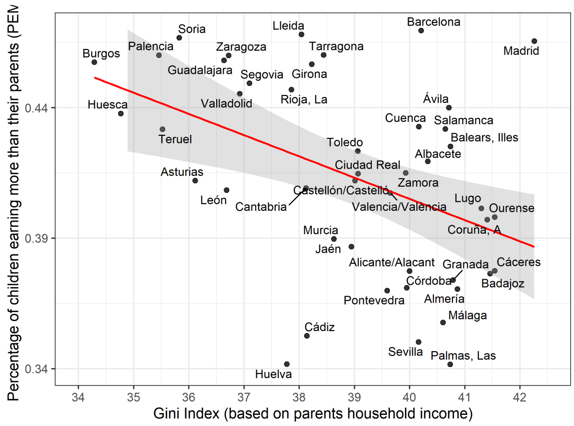
	Gap in AUM						Gap in PEM					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
AUM	−0.025						−0.001					
	(0.143)						(0.001)					
PEM		8.442						−0.024				
		(16.051)						(0.086)				
P(Q5|Q1)			0.149						−0.001			
			(0.207)						(0.001)			
RRS				15.342						0.286^{*}		
				(28.890)						(0.148)		
Parental Income					−0.00000						−0.00000	
					(0.00000)						(0.00000)	
Children Income						−0.00000						−0.00000
						(0.00000)						(0.00000)
Constant	8.006	3.372	4.920^{*}	4.320	0.072^{***}	0.068^{*}	0.111^{***}	0.077^{**}	0.075^{***}	0.019	0.072^{***}	0.068^{*}
	(6.645)	(6.642)	(2.752)	(4.805)	(0.023)	(0.035)	(0.035)	(0.036)	(0.015)	(0.025)	(0.023)	(0.035)
Observations	46	46	46	46	46	46	46	46	46	46	46	46
R^2	0.001	0.006	0.012	0.006	0.001	0.00004	0.036	0.002	0.007	0.078	0.001	0.00004

	Gap in P(Q5|Q1)					Gap in RRS					
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)
AUM	0.201^{**}					-0.002^{***}					
	(0.083)					(0.001)					
PEM		35.861^{***}					-0.209^{**}				
		(8.372)					(0.084)				
P(Q5|Q1)			0.439^{***}					-0.003^{***}			
			(0.111)					(0.001)			
RRS									0.285^{*}		
									(0.156)		
Parental Income				0.0003^{***}						-0.00000^{***}	
				(0.0001)						(0.00000)	
Children Income					0.001^{***}						-0.00001^{***}
					(0.0002)						(0.00000)
Constant	−4.335	-9.778^{***}	−0.680	−1.837	-9.908^{***}	0.143^{***}	0.124^{***}	0.078^{***}	−0.010	0.114^{***}	0.139^{***}
	(3.864)	(3.465)	(1.471)	(2.417)	(3.331)	(0.034)	(0.035)	(0.014)	(0.026)	(0.021)	(0.033)
Observations	46	46	46	46	46	46	46	46	46	46	46
R^2	0.118	0.294	0.263	0.157	0.315	0.184	0.123	0.161	0.070	0.238	0.179

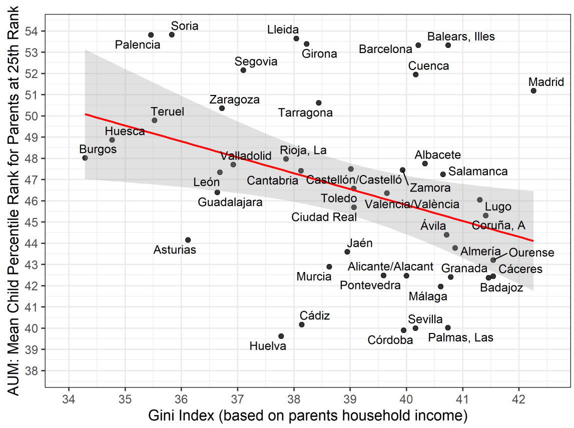
Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table A.8: The Association between Out-migration and Upward Absolute Mobility

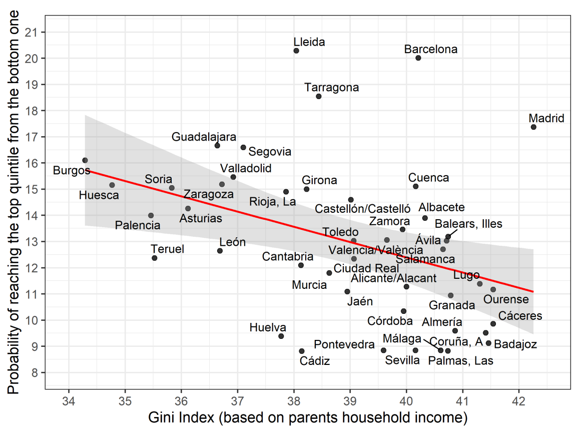
	Dependent variable:
	Child Income Rank
Parents Household Income Rank	0.190***
	(0.001)
Mover	10.635***
	(0.069)
Constant	39.568***
	(0.047)
Observations	1,479,921
R^2	0.086
Adjusted R^2	0.086
Note:	*p<0.1; **p<0.05; ***p<0.01



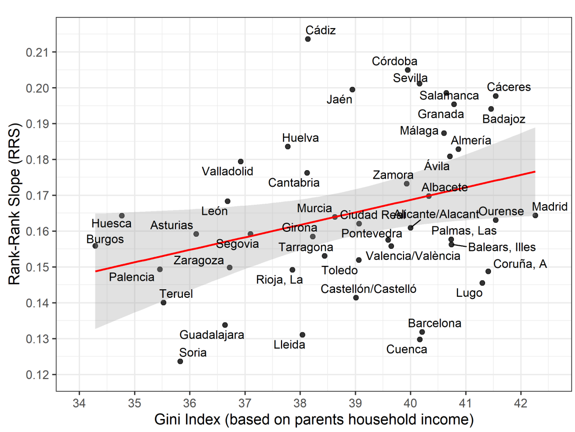
(a) AUM and Income Inequality (Gini Index)



(b) PEM and Income Inequality (Gini Index)



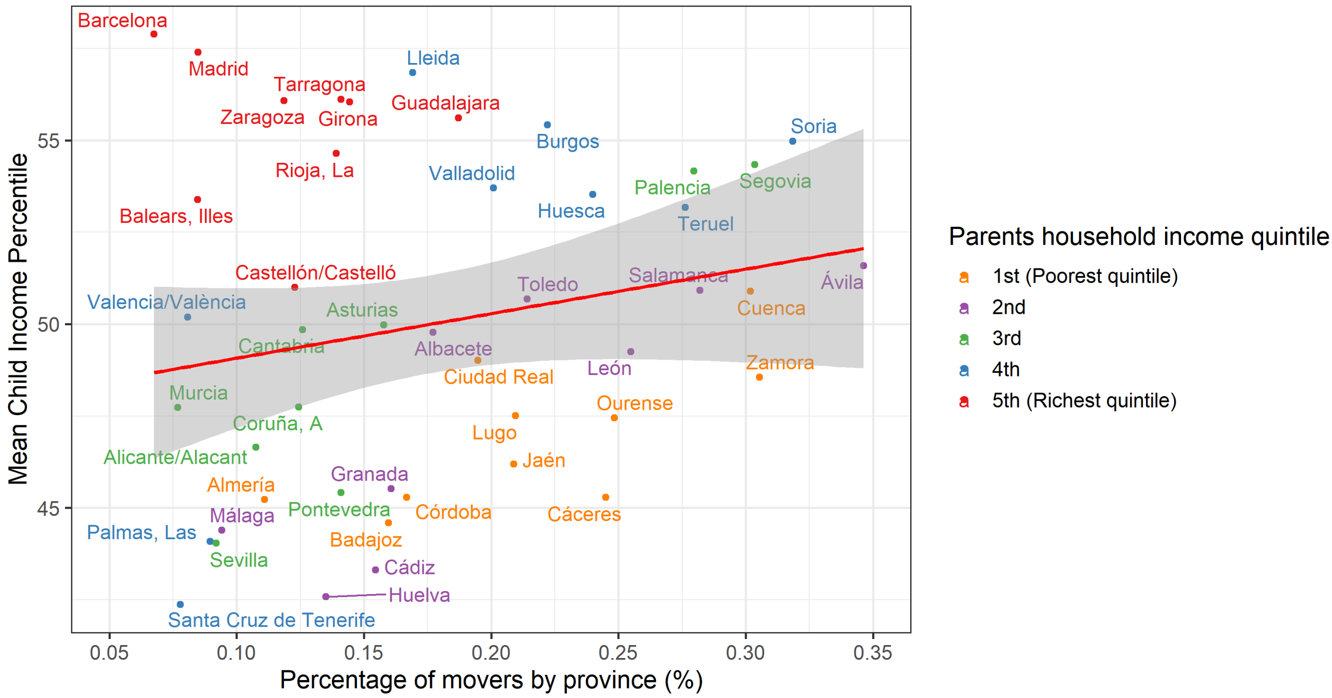
(c) P(Q5|Q1) and Income Inequality (Gini Index)

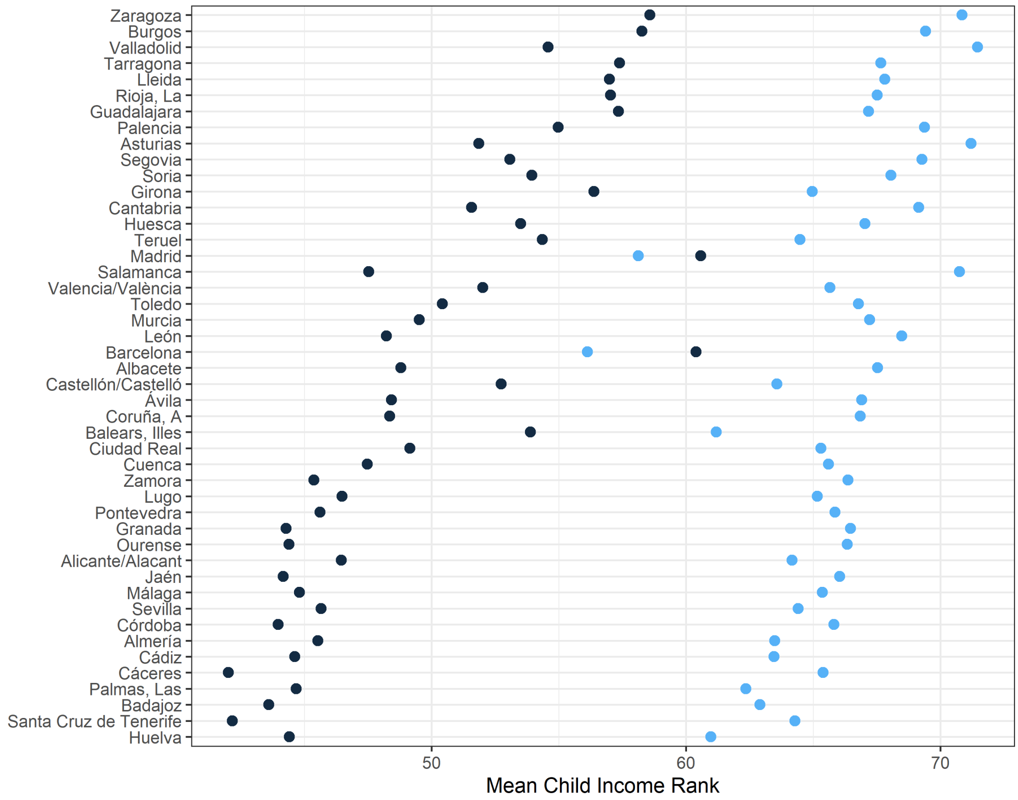


(d) RRS and Income Inequality (Gini Index)

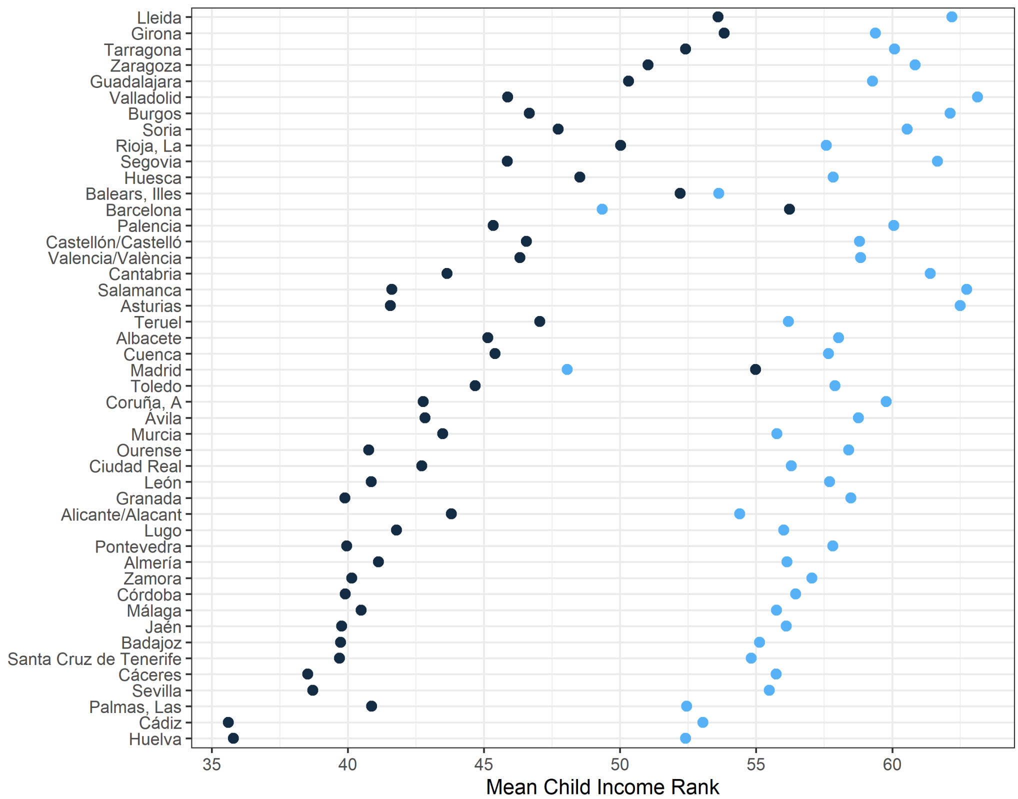
Figure A.27: The Great Gatsby Curve Across Different Intergenerational Mobility Measures

Figure A.28: The Association between Out-migration and Upward Absolute Mobility





(a) Sons: Stayers vs. Movers



(b) Daughters: Stayers vs. Movers

Figure A.29: Mean Child Percentile for Stayers and Movers at the Province Level by Parental Income Rank and Child Gender

Figure A.30: The Association between Out-migration and Upward Absolute Mobility

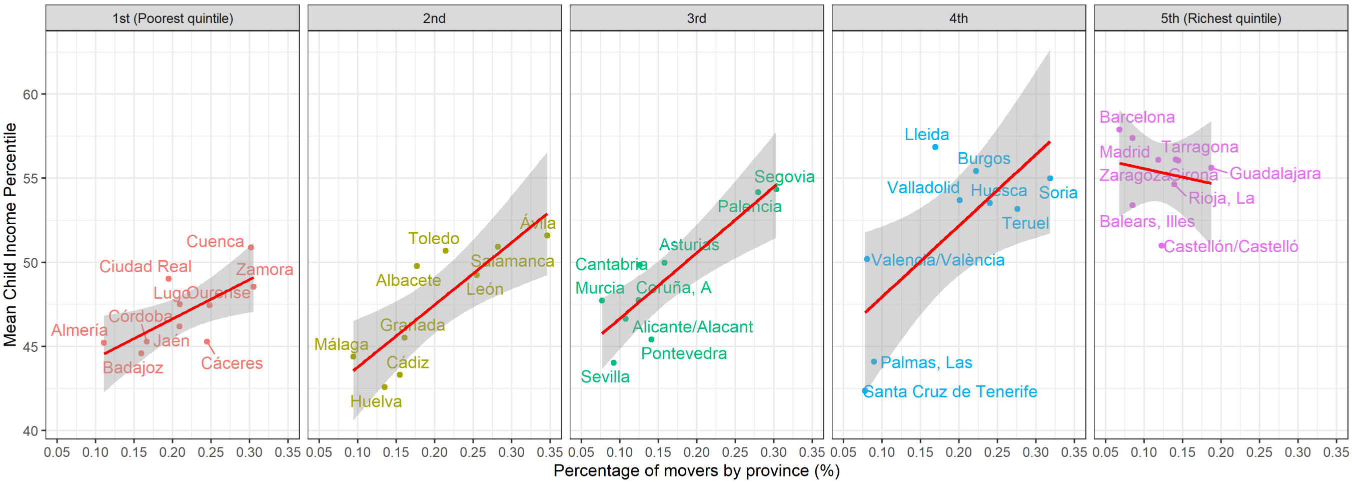


Table A.9: The Association Between Out-migration and Upward Absolute Mobility at the Province level

Dependent Variable	Quintile	Regressor	$\hat{\beta}$	S.E.	Statistic	p.value	Observations
Mean Child Income Rank	1st (Poorest quintile)	Pct. of Movers	23.37**	8.26	2.83	0.02	10
Mean Child Income Rank	2nd	Pct. of Movers	37.05**	9.53	3.89	0.01	9
Mean Child Income Rank	3rd	Pct. of Movers	39.10***	8.07	4.85	0.00	9
Mean Child Income Rank	4th	Pct. of Movers	42.34**	14.81	2.86	0.02	9
Mean Child Income Rank	5th (Richest quintile)	Pct. of Movers	-10.06	21.03	-0.48	0.65	9

Figure A.31: The Association Between Out-migration and Upward Absolute Mobility

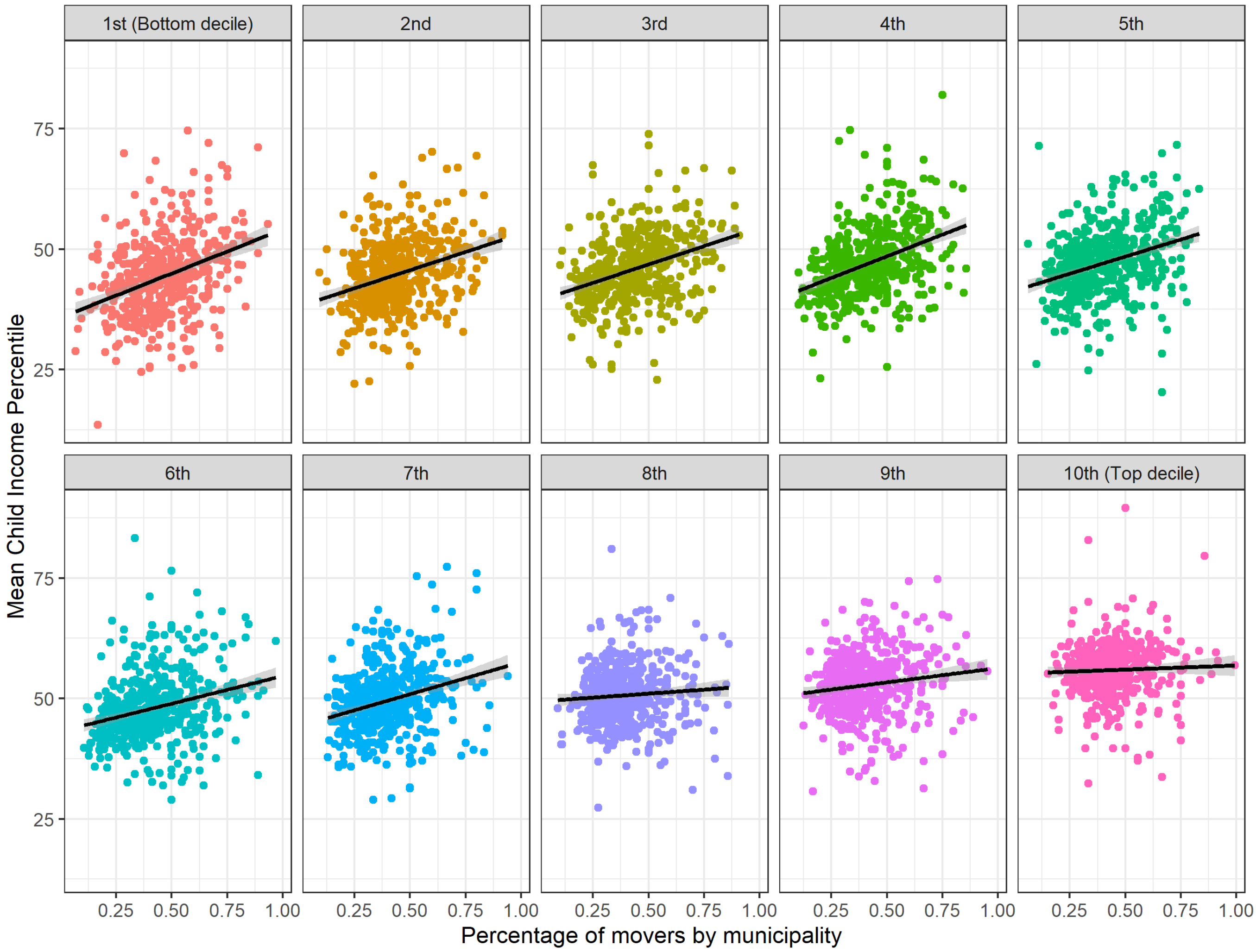


Table A.10: The Association Between Out-migration and Upward Absolute Mobility at the Municipality Level

Dependent Variable	Decile	Regressor	$\hat{\beta}$	S.E.	Statistic	p.value	Observations
Mean Child Income Rank	1st (Bottom decile)	Pct. of Movers	19.02***	1.82	10.46	0.00	504
Mean Child Income Rank	2nd	Pct. of Movers	14.60***	1.74	8.38	0.00	504
Mean Child Income Rank	3rd	Pct. of Movers	11.12***	1.61	6.90	0.00	504
Mean Child Income Rank	4th	Pct. of Movers	17.01***	1.62	10.50	0.00	504
Mean Child Income Rank	5th	Pct. of Movers	13.69***	1.67	8.17	0.00	504
Mean Child Income Rank	6th	Pct. of Movers	11.72***	1.69	6.95	0.00	504
Mean Child Income Rank	7th	Pct. of Movers	14.35***	1.80	7.97	0.00	504
Mean Child Income Rank	8th	Pct. of Movers	4.97***	1.69	2.94	0.00	504
Mean Child Income Rank	9th	Pct. of Movers	10.47***	1.74	6.03	0.00	504
Mean Child Income Rank	10th (Top decile)	Pct. of Movers	2.52	1.64	1.54	0.12	504

Table A.11: Common Support Table (Part 1)

Sibling Status	Gender	Age	Marital Status	N. of Kids	Total Sample Size	Count	% of Sample Size
Stayer	Men	30.00	Single	0.00	548	20	2.58
Stayer	Men	31.00	Married	0.00	548	1	0.13
Stayer	Men	31.00	Single	0.00	548	27	3.49
Stayer	Men	32.00	Married	0.00	548	1	0.13
Stayer	Men	32.00	Single	0.00	548	26	3.36
Stayer	Men	32.00	Single	1.00	548	1	0.13
Stayer	Men	33.00	Married	1.00	548	1	0.13
Stayer	Men	33.00	Single	0.00	548	15	1.94
Stayer	Men	34.00	Married	1.00	548	1	0.13
Stayer	Men	34.00	Married	2.00	548	2	0.26
Stayer	Men	34.00	Single	0.00	548	16	2.07
Stayer	Men	35.00	Married	1.00	548	2	0.26
Stayer	Men	35.00	Married	2.00	548	2	0.26
Stayer	Men	35.00	Single	0.00	548	10	1.29
Stayer	Men	35.00	Single	1.00	548	1	0.13
Stayer	Men	36.00	Married	0.00	548	1	0.13
Stayer	Men	36.00	Married	1.00	548	2	0.26
Stayer	Men	36.00	Single	0.00	548	11	1.42
Stayer	Men	36.00	Single	1.00	548	1	0.13
Stayer	Woman	30.00	Married	0.00	548	2	0.26
Stayer	Woman	30.00	Single	0.00	548	20	2.58
Stayer	Woman	31.00	Married	1.00	548	1	0.13
Stayer	Woman	31.00	Single	0.00	548	18	2.33
Stayer	Woman	31.00	Single	1.00	548	1	0.13
Stayer	Woman	32.00	Married	1.00	548	1	0.13
Stayer	Woman	32.00	Single	0.00	548	16	2.07
Stayer	Woman	33.00	Married	0.00	548	2	0.26
Stayer	Woman	33.00	Married	1.00	548	4	0.52

Table A.12: Common Support Table (Part 2)

Sibling Status	Gender	Age	Marital Status	N. of Kids	Total Sample Size	Count	% of Sample Size
Stayer	Woman	33.00	Married	2.00	548	1	0.13
Stayer	Woman	33.00	Single	0.00	548	13	1.68
Stayer	Woman	34.00	Married	0.00	548	1	0.13
Stayer	Woman	34.00	Married	1.00	548	2	0.26
Stayer	Woman	34.00	Married	2.00	548	1	0.13
Stayer	Woman	34.00	Single	0.00	548	16	2.07
Stayer	Woman	35.00	Married	2.00	548	1	0.13
Stayer	Woman	35.00	Single	0.00	548	14	1.81
Stayer	Woman	35.00	Single	1.00	548	2	0.26
Stayer	Woman	36.00	Married	1.00	548	2	0.26
Stayer	Woman	36.00	Married	2.00	548	3	0.39
Stayer	Woman	36.00	Single	0.00	548	10	1.29
Stayer	Woman	36.00	Single	1.00	548	2	0.26
Mover	Men	30.00	Single	0.00	548	20	2.58
Mover	Men	31.00	Married	0.00	548	1	0.13
Mover	Men	31.00	Single	0.00	548	27	3.49
Mover	Men	32.00	Married	0.00	548	1	0.13
Mover	Men	32.00	Single	0.00	548	26	3.36
Mover	Men	32.00	Single	1.00	548	1	0.13
Mover	Men	33.00	Married	1.00	548	1	0.13
Mover	Men	33.00	Single	0.00	548	15	1.94
Mover	Men	34.00	Married	1.00	548	1	0.13
Mover	Men	34.00	Married	2.00	548	2	0.26
Mover	Men	34.00	Single	0.00	548	16	2.07
Mover	Men	35.00	Married	1.00	548	2	0.26
Mover	Men	35.00	Married	2.00	548	2	0.26
Mover	Men	35.00	Single	0.00	548	10	1.29
Mover	Men	35.00	Single	1.00	548	1	0.13
Mover	Men	36.00	Married	0.00	548	1	0.13
Mover	Men	36.00	Married	1.00	548	2	0.26
Mover	Men	36.00	Single	0.00	548	11	1.42
Mover	Men	36.00	Single	1.00	548	1	0.13

Table A.13: Common Support Table (Part 3)

Sibling Status	Gender	Age	Marital Status	N. of Kids	Total Sample Size	Count	% of Sample Size
Mover	Woman	30.00	Married	0.00	548	2	0.26
Mover	Woman	30.00	Single	0.00	548	20	2.58
Mover	Woman	31.00	Married	1.00	548	1	0.13
Mover	Woman	31.00	Single	0.00	548	18	2.33
Mover	Woman	31.00	Single	1.00	548	1	0.13
Mover	Woman	32.00	Married	1.00	548	1	0.13
Mover	Woman	32.00	Single	0.00	548	16	2.07
Mover	Woman	33.00	Married	0.00	548	2	0.26
Mover	Woman	33.00	Married	1.00	548	4	0.52
Mover	Woman	33.00	Married	2.00	548	1	0.13
Mover	Woman	33.00	Single	0.00	548	13	1.68
Mover	Woman	34.00	Married	0.00	548	1	0.13
Mover	Woman	34.00	Married	1.00	548	2	0.26
Mover	Woman	34.00	Married	2.00	548	1	0.13
Mover	Woman	34.00	Single	0.00	548	16	2.07
Mover	Woman	35.00	Married	2.00	548	1	0.13
Mover	Woman	35.00	Single	0.00	548	14	1.81
Mover	Woman	35.00	Single	1.00	548	2	0.26
Mover	Woman	36.00	Married	1.00	548	2	0.26
Mover	Woman	36.00	Married	2.00	548	3	0.39
Mover	Woman	36.00	Single	0.00	548	10	1.29
Mover	Woman	36.00	Single	1.00	548	2	0.26

Table A.14: Analysis of Siblings: Observed Characteristics Balance Table

Sibling Status	Age	Gender	Marital Status	Family Income	N. of Children
Mover Sibling	32.91	0.52	0.74	31245.49	0.161
Stayer Sibling	32.91	0.52	0.74	31245.37	0.161

References

T. Abel. Cultural capital and social inequality in health. *Journal of Epidemiology & Community Health*, 62(7):e13–e13, 2008.

P. Acciari, A. Polo, and G. Violante. 'and yet, it moves': Intergenerational mobility in italy. 2019.

P. Aghion, U. Akcigit, A. Hyytinen, and O. Toivanen. The social origins of inventors. Technical report, National Bureau of Economic Research, 2017.

F. Alvaredo, L. Chancel, T. Piketty, E. Saez, and G. Zucman. *World inequality report 2018*. Belknap Press, 2018.

A. Bell, R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen. Who becomes an inventor in america? the importance of exposure to innovation. *The Quarterly Journal of Economics*, 134(2):647–713, 2019.

F. Beltran Tapia and S. de Miguel Salanova. Class, literacy and social mobility: Madrid, 1880–1905. *The History of the Family*, 26(1):149–172, 2021.

P. Bingley, L. Cappellari, and K. Tatsiramos. Family, community and long-term socio-economic inequality: Evidence from siblings and youth peers. *The Economic Journal*, 131(636):1515–1554, 2021.

S. E. Black and P. J. Devereux. Recent developments in intergenerational mobility. Technical report, National Bureau of Economic Research, 2010.

J. Blanden. How much can we learn from international comparisons of intergenerational mobility? 2009.

P. Bourdieu. *Distinction: A social critique of the judgement of taste*. Harvard university press, 1987.

P. Bourdieu. The forms of capital.(1986). *Cultural theory: An anthology*, 1:81–93, 2011.

E. Bratberg, J. Davis, B. Mazumder, M. Nybom, D. D. Schnitzlein, and K. Vaage. A comparison of intergenerational mobility curves in germany, norway, sweden, and the us. *The Scandinavian Journal of Economics*, 119(1):72–101, 2017.

M. Cervini-Plá. Exploring the sources of earnings transmission in spain. 2012.

M. Cervini-Plá. Intergenerational earnings and income mobility in s pain. *Review of Income and Wealth*, 61(4):812–828, 2015.

M. Cervini Plá and X. Ramos. Movilidad intergeneracional y emparejamiento selectivo en españa. © *Papeles de Economía Española, 2013, núm. 135, p. 217-229*, 2013.

R. Chetty and N. Hendren. The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228, 2018a.

R. Chetty and N. Hendren. The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3):1107–1162, 2018b.

R. Chetty, N. Hendren, P. Kline, and E. Saez. Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, 129(4): 1553–1623, 09 2014.

R. Chetty, N. Hendren, F. Lin, J. Majerovitz, and B. Scuderi. Childhood environment and gender gaps in adulthood. *American Economic Review*, 106(5):282–88, 2016a.

R. Chetty, M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler. The association between income and life expectancy in the united states, 2001-2014. *Jama*, 315(16): 1750–1766, 2016b.

R. Chetty, J. N. Friedman, E. Saez, N. Turner, and D. Yagan. Income segregation and intergenerational mobility across colleges in the united states. *The Quarterly Journal of Economics*, 135(3):1567–1633, 2020.

P. Chuard and V. Grassi. Switzer-land of opportunity: Intergenerational income mobility in the land of vocational education. *Available at SSRN 3662560*, 2020.

M. D. Collado, I. Ortuño-Ortín, and A. Romeu. Long-run intergenerational social mobility and the distribution of surnames. *manuscript, Universidad de Alicante*, 2012.

M. Connolly, M. Corak, and C. Haeck. Intergenerational mobility between and within canada and the united states. *Journal of Labor Economics*, 37(S2):S595–S641, 2019.

M. Corak. Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, 27(3):79–102, 2013.

M. Corak. Intergenerational mobility: what do we care about? what should we care about? *Australian Economic Review*, 53(2):230–240, 2020.

L. De Pablos Escobar and M. Gil Izquierdo. Intergenerational educational and occupational mobility in spain: does gender matter? *British Journal of Sociology of Education*, 37(5):721–742, 2016.

N. Deutscher and B. Mazumder. Intergenerational mobility across australia and the stability of regional estimates. *Labour Economics*, 66:101861, 2020.

J. Eriksen and M. D. Munk. The geography of intergenerational mobility—danish evidence. *Economics Letters*, 189:109024, 2020.

P. Fessler and M. Schürz. Private wealth across european countries: the role of income, inheritance and the welfare state. *Journal of Human Development and Capabilities*, 19(4):521–549, 2018.

M. Güell, J. V. Rodríguez Mora, and C. I. Telmer. The informational content of surnames, the evolution of intergenerational mobility, and assortative mating. *The Review of Economic Studies*, 82(2):693–735, 2015.

S. Haider and G. Solon. Life-cycle variation in the association between current and lifetime earnings. *American Economic Review*, 96(4):1308–1320, 2006.

S. Heidrich. Intergenerational mobility in sweden: a regional perspective. *Journal of Population Economics*, 30(4):1241–1280, 2017.

P. Korom. Inherited advantage: The importance of inheritance for private wealth accumulation in europe. Technical report, MPIfG Discussion Paper, 2016.

P. Matthew and D. M. Brodersen. Income inequality and health outcomes in the united states: An empirical analysis. *The Social Science Journal*, 55(4):432–442, 2018.

T. Piketty. Capital & ideology. *Harvard University Press (HUP)*, 2020.

C. Santiago-Caballero. Income inequality in central spain, 1690–1800. *Explorations in Economic History*, 48(1):83–96, 2011.

C. Santiago Caballero et al. Social mobility in nineteenth century spain: Valencia, 1841-1870. Technical report, Universidad Carlos III de Madrid. Instituto Figuerola, 2018.

G. Solon. Intergenerational income mobility in the united states. *The American Economic Review*, pages 393–408, 1992.

G. Solon. Intergenerational mobility in the labor market. In *Handbook of labor economics*, volume 3, pages 1761–1800. Elsevier, 1999.

G. Solon. Cross-country differences in intergenerational earnings mobility. *Journal of Economic Perspectives*, 16(3):59–66, 2002.

R. Van der Weide and B. Milanovic. Inequality is bad for growth of the poor (but not for that of the rich). *The World Bank Economic Review*, 32(3):507–530, 2018.