Does It Matter Where You Grow Up? Childhood Exposure Effects in Latin America and the Caribbean*

Ercio Muñoz†

May 2022

Abstract

This paper studies whether the observed differences in intergenerational educational mobility across regions in Latin America and the Caribbean are due to the sorting of families or the effect of growing up in these different places. The analysis exploits differences in the ages of children at the time their families moved across locations, to isolate regional childhood exposure effects from sorting. The findings show a convergence rate of 3.5 percent per year of exposure between age 1 to 11, implying that children who moved at age of 1 would pick up 35 percent of the observed differences in mobility between origin and destination. These results are robust to using a specification that identifies the effect of place within households, the use of only anomalously high migration outflows, instrumenting the choice of destination with historical migration, and a combination of both approaches.

JEL-Codes: D63, I24, J13, J62, R23, Z13.

Keywords: *Educational intergenerational mobility, Geography, Latin America, Neighborhood effects.*

*I thank Jonathan Conning, Núria Rodríguez-Planas, Wim Vijverberg, Roy van der Weide, Chico Ferreira, Lucila Berniell, Dolores de la Mata and several seminar/conference participants for helpful comments. Pablo Vargas and Marcelo Gomez provided outstanding research assistance. I greatly acknowledge the financial support of CAF -development Bank of Latin America, the Center for Latin American, Caribbean and Latino Studies at CUNY, the doctoral student research grant round 16 from CUNY, and the Mario Capelloni dissertation fellowship from CUNY. The usual disclaimers apply. First version: July 2021.

†Email: emunozsaavedra@worldbank.org.

I Introduction

There is substantial variation in intergenerational mobility (IGM) in education–measured as the likelihood of children completing at least primary education when parents did not complete primary school–across provinces and districts of 24 countries in the Latin America and the Caribbean (LAC) region (see Munoz, 2021a).[1] This variation in upward mobility can be the result of sorting of families (i.e., the fact that different families choose to live in different places) or, alternatively, can be the result of the existence of causal effects of regions (i.e., the place where people live affects the level of upward mobility independently of the characteristics of the family).

In this paper, I use 21 census samples that span 11 countries in LAC to investigate how much of the variation in upward mobility can be attributed to the effect of being exposed to certain places during childhood. In particular, I use families that moved across districts/provinces and exploit variation in children's age at the time of the move to isolate sorting from regional childhood exposure effects. This empirical strategy that relies on observational data was first proposed in Chetty and Hendren (2018a) to study the effect of commuting zones on IGM in income in the US and adapted by Alesina, Hohmann, Michalopoulos, and Papaioannou (2021) to the context of educational IGM in developing countries.

I find evidence of regional childhood exposure effects as well as significant sorting-selection. I estimate a convergence rate of 3.5% per year of exposure between the ages 1 to 11, which implies that children who move at age 1 would pick up $10 \times 3.5 = 35\%$ of the observed difference in permanent residents' outcomes between their origin and destination regions. In addition, I find significant selection effects of approximately 42%, which implies that individuals who move to a region where permanent residents have 1 percentage point higher upward mobility have close to 0.42 higher mobility themselves purely due to selection. Given that exposure effects are identified under the strong assumption that the timing of

[1] See Figure A1 in the Appendix for a reproduction of the map of IGM at the district-level.

the move is unrelated to other determinants of primary completion, I also estimate these effects within households, using plausibly exogenous migration outflows, and instrumenting destinations with past migration patterns. All these exercises produce results that are qualitatively similar. Moreover, I explore potential heterogeneity by estimating childhood exposure effects by sub-populations and I explore an alternative outcome that computes upward mobility with secondary education as the level of interest. I do not find much evidence of heterogeneity and the results of the estimates using secondary education show similar patterns to the baseline exercise but with larger convergence rates.

This paper is related to three strands of the economic literature. First, it is related to the recent literature that estimates intergenerational mobility (IGM) in education within countries and documents important variation across places in developing countries (see for example, Alesina, Hohmann, Michalopoulos, & Papaioannou, 2020; Alesina et al., 2021; Asher, Novosad, & Rafkin, 2021; Munoz, 2021a, 2021b; Van der Weide, Ferreira de Souza, & Barbosa, 2020). This is an extension of the literature on intergenerational income mobility within countries initiated by Chetty, Hendren, Kline, and Saez (2014) that is mostly focused on high-income economies, and the intergenerational mobility in education at the country-level (see Torche, 2019, for a survey focused on developing countries).[2] Second, it adds to the literature that studies how the place in which children grow up matters for medium or long-term socioeconomic outcomes (see Chyn & Katz, 2021, for a recent survey about place effects). This includes studies that use random or quasi-random variation to identify the effects (see for example Chetty, Hendren, & Katz, 2016; Chyn, 2018; Damm & Dustmann, 2014) and those using observational data exploiting samples of movers (Alesina et al., 2020, 2021; Chetty & Hendren, 2018a; Deutscher, 2020; Laliberté, 2021; Ward, 2020). Third, this paper relates to the set of papers studying the different drivers of intergenerational mobility with the help of data at the sub-national level and the application of quasi-experimental methods (see for example, Card, Domnisoru, & Taylor, 2022; Derenoncourt, 2021; Sharkey

[2]Important recent contributions are Narayan et al. (2018); Van der Weide, Lakner, Gerszon Mahler, Narayan, and Ramasubbaiah (2021) documenting IGM in education at the country-level for 153 economies.

& Torrats-Espinosa, 2017).

The paper is organized as follows. Section II describes the data set. Section III discusses the empirical methods. Section IV goes over the empirical results. Section V offers some robustness exercises and additional results. Finally, section VI concludes with final remarks.

II Data

I use census data obtained from IPUMS International (Integrated Public Use Microdata Series, IPUMS, 2019), hosted at the University of Minnesota Population Center, which reports harmonized representative samples (typically 10%) of full census micro data sets for a large number of countries.

Countries. I use individual records, retrieved from 21 national censuses from 11 countries: Brazil, Colombia, Cuba, Ecuador, El Salvador, Guatemala, Jamaica, Mexico, Panama, Trinidad and Tobago, and Uruguay (see Table A1 in the Appendix for the details about years and the fraction of the data available by census sample). There is a much longer list of censuses available for Latin America and the Caribbean that can be used to estimate educational intergenerational mobility (see Munoz, 2021a), however, the main constraint that limits the number of samples being used in this study is the availability of the information necessary (which is detailed below) to analyze within-country migration as it is required in the empirical strategy.

Education. The data set contains a variable reporting educational attainment that is re-coded by IPUMS to capture educational attainment in terms of the level of schooling completed without necessarily reflecting any particular country's definition of the various levels of schooling in terms of terminology or number of years of schooling. It contains four categories: 1) Less than primary completed, 2) primary completed, 3) secondary completed, and 4) university completed. This variable applies, to the extent possible, the United Nations standard of six years of primary schooling, three years of lower secondary schooling, and three years of higher secondary schooling. In addition, there is a variable reporting years of schooling available for some samples (17 out of 21 samples).

Linking generations. The data collection is organized at the household level,[3] so it is possible to link individuals who live with their parents in the same household at the time of the interview using a variable that details the relationship between each individual and the head of the household. For simplicity, I use individuals linked to their probable father and probable mother according to the procedures used by IPUMS for family interrelationships.[4] This method of linking generations differs from the one used in Munoz (2021a) for the main analysis, which links individuals ages 14-25 to all the older relatives living in the same household, however, as mentioned in the robustness section of Munoz (2021a), both methods give indicators of mobility that are close to being perfectly correlated.

For the main analysis, I use a sample of individuals co-residing with at least one "probable" parent, I focus on individuals with age in the range 14-25 in the main analysis where completion of primary education is the outcome of interest. However, when I explore completion of secondary education as the outcome of interest, I use individuals with age in the range 18-25. These age ranges are chosen to target individuals with an age sufficiently large so they can complete the educational level of interest if progressing in line with the official age for which the educational system is designed but young enough so they are likely to still live with their parents. The effort of maximizing the co-residence rate attempts to minimize potential truncation bias in the estimates of intergenerational mobility (see Emran, Greene, & Shilpi, 2018; Francesconi & Nicoletti, 2006). Nonetheless, Munoz and Siravegna (2021) show that the bias is small for estimates of absolute mobility that use census data. In addition, there is a low level of re-ranking when these estimates are used to rank economies across time and space by level of mobility relative to the ranking obtained with estimates that use retrospective information (i.e., surveys that ask all individuals for the level of education of their parents).

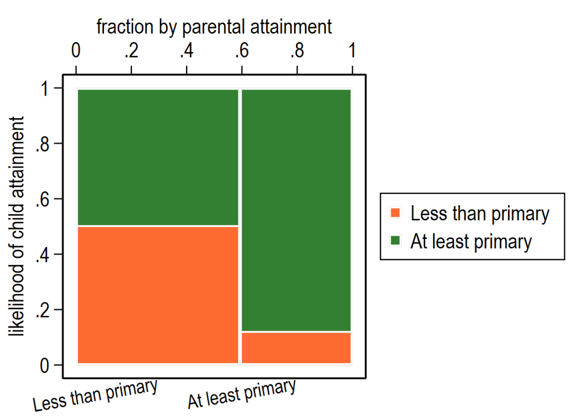
[3] I exclude several census samples for which this is not the case.

[4] More details can be found in the following link: https://usa.ipums.org/usa/chapter5/chapter5.shtml.

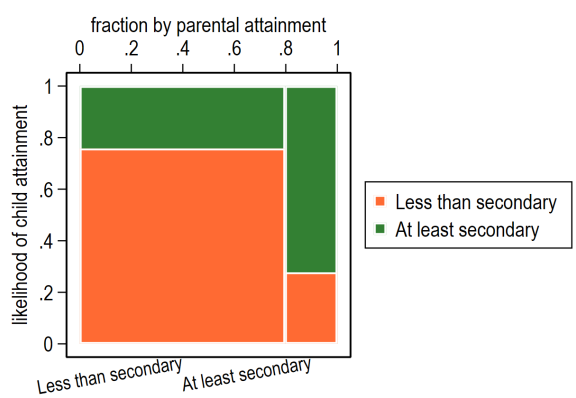
Figure 1 shows the educational attainment in terms of completion of primary (see Figure 1a) and secondary (see Figure 1b) education for young individuals and the average attainment of their linked parents. In the sample, a majority of individuals have parents that completed less than primary or less than secondary education. Furthermore, conditional on the average level of educational attainment of parents being low (less than primary or less than secondary), the likelihood of children achieving higher education than their parents is less than 50% (Figure 1a) and in the case of secondary education just a little bit more than 20% Figure 1b).

The focus on primary education as the main outcome of interest is based on the fact that primary education is an important margin for the parents in the sample as shown in Figure 1a. Furthermore, the use of secondary education reduces the sample size available for the analysis (by approximately 50%) as the rate of co-residence falls with age (see Figure A2 in the Appendix). However, as I will show later, estimates using secondary education show a consistent picture with respect to the impact of place on mobility.

Figure 1: Educational attainment of young individuals and their parents



(a) Primary completion



(b) Secondary completion

Notes: The figures are constructed using individuals aged 14-25 years (left) and 18-25 years (right). Parental educational attainment corresponds to the parents' average.

Geographic information. IPUMS reports residence at the time of the interview for at most two levels of administrative units in which the households were enumerated: 1)

Provinces, which are administrative units roughly similar to states in the US; and 2) Districts, which are fine administrative units similar to counties in the US. These variables contain the geographies for every country harmonized spatio-temporally to provide spatially consistent boundaries across samples in each country.

Migration. IPUMS International reports, when available in the original source, the administrative unit in which the individual was born, the previous location of the previous residence, and the number of years living in the current location. For some censuses, this information is all at the level of province (e.g., Brazil 2010 has only state-level information) and for others, information at the district level is also available (e.g., Cuba 2012 has municipality-level information). I use these variables to identify a sample of movers, i.e. individuals who at the time of the interview live in a place different than the one in which they were born, as well as a sample of permanent residents (or non-movers) that correspond to individuals living in the same place in which they were born. Hence, I consider the place of origin as the place where they were born, the current place as the destination, and I construct a variable that records age at move using current age minus the time they have been living in the current place.

Table 1 provides a set of summary statistics comparing the characteristics of individuals (and their families) that are non-movers versus those that are movers.[5] In terms of family characteristics, movers show slightly higher income percentile (according to average income of parents) and urban status (at destination in the case of movers). However, permanent residents have a higher share of dwelling ownership and are more likely to be couple families. In the case of educational outcomes, individuals who are permanent residents complete primary education with exactly the same likelihood but exhibit lower years of schooling than movers. Overall, the table suggests that permanent residents are relatively similar to those individuals who move (or at least they are not remarkably different).[6]

[5]A caveat of this comparison is that some characteristics such as household income and years of schooling are not available in all the surveys, hence, the sample composition may vary across rows.

[6]Table A2 in the Appendix describes the sample size by census sample distinguishing between movers

Table 1: Summary statistics for permanent residents and movers

	Permanent residents				Movers			
	Mean	Std. dev.	Median	N	Mean	Std. dev.	Median	N
Family characteristics								
Income percentile	49.72	28.89	49	8409346	52.52	28.50	54	928209
Ownership of dwelling	0.76	0.43	1	12843887	0.60	0.49	1	2242290
Urban status	0.72	0.45	1	12542856	0.81	0.39	1	2020376
Number of siblings	2.23	1.30	2	13257078	2.22	1.38	2	2341902
Couple family	0.56	0.50	1	13257078	0.40	0.49	0	2341902
Educational outcomes								
Completed primary	0.63	0.48	1	13234115	0.63	0.48	1	2324712
Years of schooling	6.16	3.68	6	8993451	6.34	3.67	6	1763570

Notes: The samples consider the characteristics of individuals age 14-25. Permanent residents are those individuals living in the same region in which they were born. Movers are those individuals living in a region different than the one where they were born. Average income percentil is computed using percentiles by census-year using the average total income of parents in the previous month or year. The sample size varies across variables due to different availability of data between censuses and also differences in the amount of missing data.

III Methods

As discussed in a recent survey by Chyn and Katz (2021), the potential influence of neighborhoods on individual outcomes can operate through contemporaneous (or situational effects) and past neighborhoods through exposure (or developmental) effects that accumulate during childhood. The econometric approach followed here is focused solely on the latter, which has the key prediction that the gains from moving to places with beneficial characteristics is larger for children who are younger at the time of the move and thus exposed for a longer period.

Given the focus on exposure, my goal is to determine how much a child's potential outcomes would improve on average if he were to grow up in a region where the permanent residents' outcomes are 1 percentage point higher. Following Chetty and Hendren (2018b), I define exposure effect at age m as the impact of spending year m of one's childhood in a region where the outcomes of permanent residents are 1 percentage point higher.

and non-movers.

To understand better the meaning of exposure effects consider the following hypothetical scenario. If I were able to randomly assign individuals to new regions d starting at age m for the rest of their childhood, a simple econometric approach to estimate the mean impact of spending year m of childhood and onward in a region where permanent residents have 1 percentage point better outcomes (here defined as ξ_m) would be:

$$y_i = \alpha_m + \xi_m \bar{y}_{db} + \epsilon_i \tag{1}$$

where α_m are age at the time of move fixed effects, $\bar{y}_{db}$ is the average outcomes of permanent residents in region d of the same cohort b as the individual i, and ϵ_i captures all other determinants of the outcomes studied. The fact that individuals were randomly assigned would ensure that the OLS estimate of ξ_m in equation 1 is consistent. These coefficients then could be used to compute the exposure effect as $\omega_m = \xi_m - \xi_{m-1}$.

However, the estimation of equation 1 in observational data will almost surely give an inconsistent estimate because there may be unobservables correlated to the outcome of interest. Hence, the empirical approach to obtain an estimate of exposure effects requires a strong assumption, which is discussed after presenting the baseline semi-parametric specification.

I use the research design proposed by Chetty and Hendren (2018a) to estimate the role of commuting zones in shaping intergenerational income mobility in the US, which was adapted to educational mobility by Alesina et al. (2020, 2021). This empirical strategy uses observational data to isolate regional childhood exposure effects from sorting and consists of two steps. First, I estimate the observed outcomes for individuals who are permanent residents, by region and birth decade. Second, I use individuals who move during childhood between regions to estimate how much of the previously estimated differences in intergenerational mobility between the place of origin and destination (for individuals in their same birth decade) is reflected in their chances of finishing primary school and how this changes with the age at the time of the move. If regions affect individual mobility, the impact should be stronger, the longer the exposure to the new region (i.e., the difference between origin and destination is reflected in their chances in a larger magnitude when the move happens in earlier ages). These two steps are explained in detail in the next subsections.

III.1 Predicted upward mobility using permanent residents

As a first step, I estimate the level of upward mobility in education (by region and birth decade) defined as the likelihood of obtaining at least a primary education for individuals whose parents did not finish primary school using a sample of permanent residents with age between 14 and 25. I define permanent residents as individuals who live at the time of the census interview in the same region (i.e., province or district) in which they were born. Hence, the indicator of upward intergenerational mobility (γ_{rb}) used corresponds to the estimation of the following conditional probability by region r and birth cohort b:

$$\gamma_{rb} = P(D_{irb}^{children} = 1 | D_{irb}^{parent} = 0) \tag{2}$$

where $D_{irb}^{children}$ is an indicator variable equal to one if individual i, living in region r, and born in decade b obtains at least primary education. Similarly, $D_{irb}^{parent} = 0$ is an indicator variable equal to one if the parents of the same individual were able to complete at least primary school.

The sample estimates of γ_{rb} are then used to construct the main variable of interest for the next step, which is defined as:

$$\Delta_{ijb} = \hat{\gamma}_{ib} - \hat{\gamma}_{jb} \tag{3}$$

This variable captures the difference in upward intergenerational mobility of permanent residents of a given birth decade b living in region i and those living in region j.

III.2 Estimating regional childhood exposure effects using movers

In the second step, I consider only individuals with age between 14 and 25 who moved between regions at age 1-20 and whose parents did not complete primary school. The idea is to estimate how the chances of finishing primary school are affected by the difference between the chances of permanent residents in their new region of residence relative to the region of origin depending on the age at which they move. This is estimated using two econometric approaches, a semi-parametric and a parametric specification.

Baseline semiparametric specification. Following Alesina et al. (2021), I estimate regional childhood exposure effects with the following specification[7]:

$$y_{ihbmod} = \alpha_{ob} + \alpha_m + \sum_{m=1}^{20} \beta_m I(m_i = m)\Delta_{odb} + \epsilon_{ihbmod} \tag{4}$$

where the subscripts refer to individual i, member of household h, born in decade b, and who moves at age m from region of origin o to region of destination d. Parameters α_{ob} are fixed effects by region of origin and birth cohort. Parameters α_m are fixed effects by age at which the individual moved to capture disruption effects. The set of interactions allow the effect of Δ_{odb}, which is the difference in upward mobility between the region of origin o and region of destination d for cohort b as defined in the previous section (see Equation 3), on the chances of obtaining at least primary school to differ by age at which individuals move.

As discussed in Chetty and Hendren (2018b), the set of β_m coefficients captures both a standard selection effect that measures how parental inputs and other determinants of children's outcomes for movers co-vary with permanent resident outcomes, and the impact of regions (ξ_m) as defined in equation 1. The key identification assumption necessary for causal interpretation is that the timing of the move is unrelated to other unobserved determinants of primary completion. In other words, families that moved may be different from those who did not, but migrant families may differ only in their timing (e.g., two families moving from

[7]A formal and detailed description of this research design can be found in Chetty and Hendren (2018a).

region i to j when the child is 2 years old and 3 years old are not systematically different). This assumption allows me to use the estimates of β_m for different ages to get the exposure effect ω_m as $\beta_m = \xi_m + \phi$, where ϕ denotes the age invariant selection effect that gets canceled out when taking the difference.

Intuitively, this specification tries to mimic in a quasi-experimental way the ideal (but infeasible) experiment where I randomly force families to move from one region (with a given IGM) to another region (with different level of IGM) at different ages to study how IGM varies by the age of move.

Parametric specification. I also estimate regional childhood exposure effects using the following more parsimonious specification:

$$
\begin{aligned}
y_{ihbmod} = \sum_{b=b_0}^{B} 1(b_i = b)(\alpha_b^1 + \alpha_b^2 \gamma_{ob}) + \sum_{m=1}^{20} \zeta_m 1(m_i = m) \\
+1(m_i < 5)(\beta_0 + (20 - m_i)\beta_1)\Delta_{odb} \\
+1(5 \leq m_i \leq 11)(\gamma_0 + (20 - m_i)\gamma_1)\Delta_{odb} \\
+1(m_i \geq 12)(\delta_0 + (20 - m_i)\delta_1)\Delta_{odb} + \epsilon_{ihbmod}
\end{aligned}
\tag{5}
$$

where now the equation imposes a piecewise linear structure, allowing the regional exposure effects to differ for pre-school years (ages 1-4), the ages (arguably more) relevant for primary school (ages 5-11), and post-primary education years (ages 12-20).

The only difference between the specifications discussed in this section (see equations 4 and 5) and the ones used in Alesina et al. (2021) is that I do not include a set of interactions between indicator variables by birth cohort and the difference in upward mobility between destination and origin. These interactions are motivated in Chetty and Hendren (2018b) to alleviate concerns about differences in measurement error across cohorts due to changes in the ability to measure parents' locations. These concerns are not present in my set up and adding them makes little difference in practice.[8]

[8]The results including this set of interactions are available in the Appendix.

Within-household identification. In addition, I estimate similar specifications adding household fixed effects such that equation 4 in the baseline semi-parametric approach becomes:

$$y_{ihbmod} = \alpha_h + \alpha_{ob} + \alpha_m + \sum_{m=1}^{20} \beta_m I(m_i = m)\Delta_{odb} + \epsilon_{ihbmod} \tag{6}$$

and similarly, equation 5 of the parametric approach becomes:

$$\begin{aligned} y_{ihbmod} = \alpha_h + \sum_{b=b_0}^{B} 1(b_i = b)(\alpha_b^1 + \alpha_b^2 \gamma_{ob}) + \sum_{m=1}^{20} \zeta_m 1(m_i = m) \\ +1(m_i < 5)(\beta_0 + (20 - m_i)\beta_1)\Delta_{odb} \\ +1(5 \leq m_i \leq 11)(\gamma_0 + (20 - m_i)\gamma_1)\Delta_{odb} \\ +1(m_i \geq 12)(\delta_0 + (20 - m_i)\delta_1)\Delta_{odb} + \epsilon_{ihbmod} \end{aligned} \tag{7}$$

In both cases, the identification now comes from comparing siblings in households with more than one individual with age between 14 and 25. This tackles concerns of endogeneity as a result of time-invariant unobservables at the household-level.

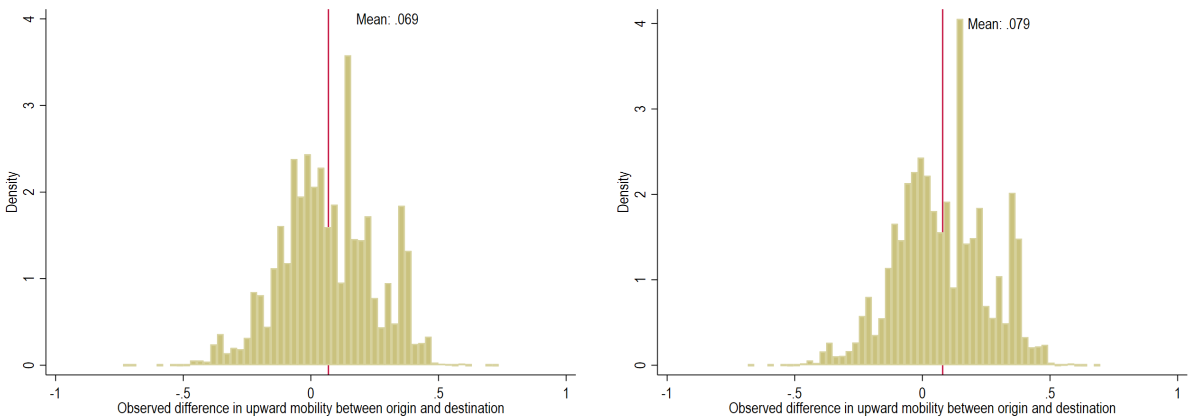
IV Regional childhood exposure effects in LAC

In this section, I present the main empirical results. First, I display the variation in upward mobility between origins and destinations faced by those who move across regions. Second, I report the results of the baseline semi-parametric estimation. Third, I report the results of the parametric estimation.

IV.1 Observed differences between origin and destination

As discussed in the previous section, the empirical strategy uses the intergenerational upward mobility of permanent residents to describe regions. In practice, this implies that the difference in this indicator between region of destination and region of origin becomes the independent variable of interest in the regressions. Figure 2 displays the distribution of this variable. On average, children moved to places with higher levels of upward mobility. This is true in a sample that consider all the movers (see Figure 2a) and a restricted sample of households with more than one individual of age in the range 14-25, which is used when I include household fixed effects in the regressions (see Figure 2b).

Figure 2: Distribution of observed differences in intergenerational upward mobility



(a) Full sample **(b)** Sample of siblings

Notes: The figure shows the observed differences in intergenerational upward mobility between origin and destination (computed using only permanent residents) for a sample of individuals who currently reside in a different place than their birth place. The sample on the right is restricted to households with more than 1 member aged 14-25 co-residing with at least one parent that did not complete primary education.

IV.2 Baseline semi-parametric estimates

In the first set of regression results, I estimate the average increase in an individual's likelihood of completing at least primary education (conditional on having parents that did not complete primary school) from moving at age m to a region (and live there onward) with 1 percentage point higher expected probability for permanent residents of the same birth decade (see equation 4).

Figure 3 plots the estimated β_m coefficients from equation 4.[9] First, the plot reveals non-

[9]Figure A2 in the Appendix reports these results with a specification that includes an interaction between the destination-origin differences in mobility and indicators variables by cohort.

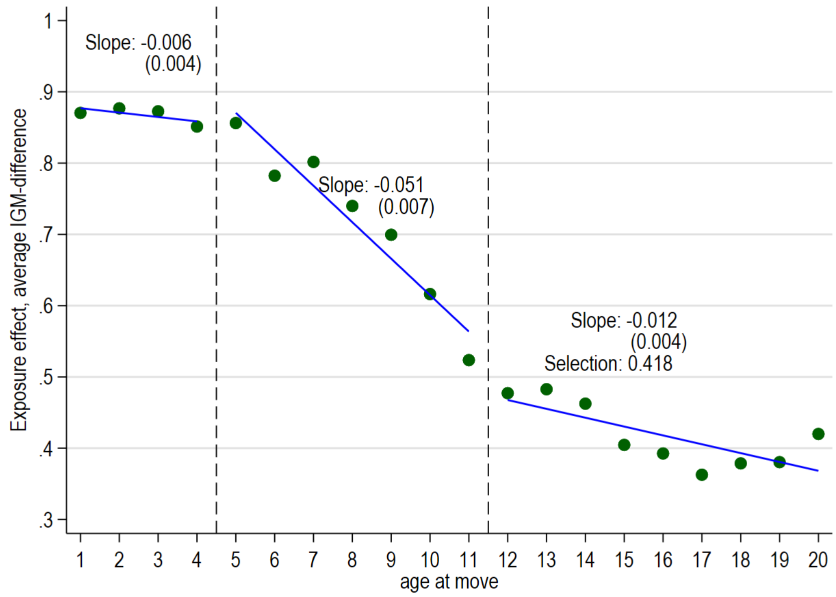
negligible selection effects that are captured by the average level (0.418) of the coefficients β_m associated to ages at move greater than 11, the age at which most individuals finish primary education if they follow the typical schedule of the educational system. This implies that individuals who move to a region where permanent residents have 1 percentage point higher upward mobility have 0.418 higher mobility themselves purely due to selection effects. Second, a flatter segment can be appreciated for very young ages that corresponds to the period before starting formal schooling and a much stepper decline is observed between ages 5 and 11. The average exposure effect is obtained as the average $\beta_m - \beta_{m-1}$ between ages 1-11 is 3.5%. This implies that a child with parents that did not complete primary school that moves at age 1 to a region with higher chances of finishing primary will pick up on average 35% of the difference between origin and destination.

These results are remarkably similar to the findings in previous studies for different contexts and measures of intergenerational mobility (see for example, Alesina et al., 2021; Chetty & Hendren, 2018a; Deutscher, 2020).

Within-household results. The previous results are based on the strong assumption that selection does not vary with age at move. An obvious concern is that this assumption may not hold in practice and for that reason Chetty and Hendren (2018a) offer a set of validation exercises to mitigate those concerns and validate the proposed research design. In line with that and with the aim of adding evidence in support of the empirical strategy, I also estimate these effects exploiting variation within households. In other words, I run a similar regression as before but including household fixed effects (see equation 6) to base the estimation on the comparison of "siblings" of different ages that had different lengths of exposure to the same places.

Figure 4 plots the estimated β_m coefficients by age at move using a restricted sample of households with more than one individual aged 14-25. Figure 4a plots the estimated coefficients obtained when the model does not include household fixed effects and Figure 4b plots the same estimated coefficients when the model includes household fixed effects and there-

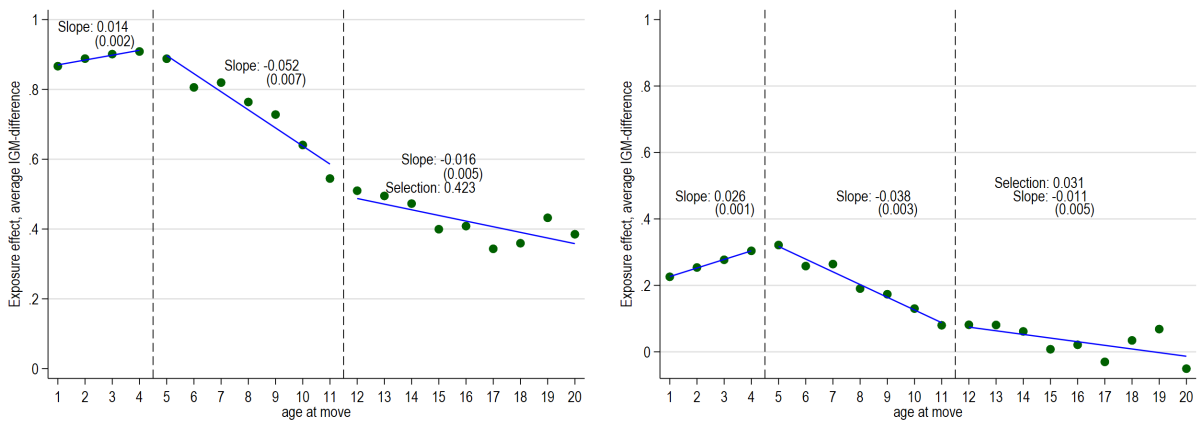
Figure 3: Regional childhood exposure effect estimates for the likelihood of completing at least primary education when parents were not able to complete this level



Notes: Estimated coefficients β_m from equation 4. These coefficients capture the expected increase of an individual's likelihood of completing at least primary school (given that their parents were not able to do so) from moving at age m to a place with 1 percentage point higher expected probability for permanent residents. They are estimated by regressing an indicator of primary school completion of those whose parents move in their childhood on the interaction of their age at move m with $\Delta_{odb} = \gamma_{db} - \gamma_{ob}$ – the difference between upward mobility for permanent residents of the same birth-decade b in the destination d versus the origin o. Controls capture: cohort and origin effects (via indicators for birth-decade interacted with origin); and disruption effects (via indicators for age at move). The sample includes individuals age 14-25 living with at least one parent that did not complete primary education and who moved before being 18 years old.

fore exploits variation within households. First, the restricted sample delivers qualitatively similar results relative to the baseline full sample. Second, the estimated effects in the model with household fixed effects are lower than the baseline estimates. The selection effect drops significantly to 0.020 (statistically indistinguishable from zero) and the estimated exposure effects are somewhat lower than the baseline results but still statistically significant. This rules out that the regional childhood exposure effects estimated in the baseline specification are driven by invariant household unobserved factors.

Figure 4: Place childhood exposure effect estimates for the likelihood of completing at least primary education when parents were not able to complete this level – observational and within family estimates



(a) Household FE sample, without HH FE

(b) With household FE

Notes: Estimated coefficients β_m from equation 4. These coefficients capture the expected increase of an individual's likelihood of completing at least primary school (given that their parents were not able to do so) from moving at age m to a place with 1 percentage point higher expected probability for permanent residents. They are estimated by regressing an indicator of primary school completion of those whose parents move in their childhood on the interaction of their age at move m with $\Delta_{odb} = \gamma_{db} - \gamma_{ob}$ – the difference between upward mobility for permanent residents of the same birth-decade b in the destination d versus the origin o. Controls capture: cohort and origin effects (via indicators for birth-decade interacted with origin); disruption effects (via indicators for age at move); and household effects in the case of the second figure (via household indicators). Both regressions use the same sample, which includes only households with at least two individuals age 14-25 living with at least one parent that did not complete primary education and who moved before being 18 years old.

IV.3 Parametric estimates

The results of the semi-parametric estimation (see Figure 3 and 4) suggest that the effects of moving at different ages evolve in a way that allows a parametrization that can simplify equations 4 and 6 to make them more parsimonious.

Table 2 reports the estimates of childhood exposure effects using the parametric approach described in the previous section (see equation 5). The results of this approach confirm the previous findings showing a strong exposure effect for ages 5-11, a flatter section between ages 1-4 (although with a statistically significant slope in the case of the within-household specification), and similarly a flatter segment after age 11.[10]

The choice of age kinks is mainly motivated by how the educational system is structured. However, a data driven approach would point out to the same kinks (see Figures A3 in the Appendix). Table A4 in the Appendix reports coefficient estimates and the respective measures of goodness of fit for different options in terms of the age kinks used in the parametric specification. Metrics such as the R-squared and the information criteria AIC or BIC indicate that the model provides practically the same fit if the kinks are moved from 4 to 5 (i.e., first segment being ages 1-5) and 11 to 10 (i.e., last segment being ages 10-20).

Table 2: Parametric estimates of regional childhood exposure effects

	(1) IGM	(2) IGM	(3) IGM
β: 1-4	0.000524	-0.0140*	-0.0247**
	(0.007)	(0.008)	(0.012)
γ: 5-11	0.0494***	0.0512***	0.0391***
	(0.004)	(0.005)	(0.006)
δ: 12-20	0.0155***	0.0201***	0.0125***
	(0.003)	(0.004)	(0.004)
R-squared	0.095	0.092	0.685
N	436792	271984	271984
Household FE	No	No, hhfe sample	Yes

Notes: The dependent variable in all specifications is an indicator that takes the value of one for children of parents without completed primary school who have completed at least primary education and zero otherwise (IGM). The independent variables comprise a linear in origin-average-IGM (calculated for the birth-cohort relevant to the individual among nonmovers) term, age-at-move indicator variables, birth-decade $\times$ destination indicators interacted with destination-minus-origin differences in upward IGM, all of which are not reported, and three linear terms for destination-minus-origin differences in the relevant-birth-cohort-nonmover average IGM for moves taking place when the child moves, ages 1-4, 5-11, and 12-18. Double clustered at the origin and at the destination region standard errors are reported in parenthesis.

[10] Table A3 in the Appendix reports results from a parametric specification that includes an interaction between the destination-origin differences in mobility and indicators by cohort.

V Robustness and additional results

In this section I run three additional exercises. First, with the aim of addressing concerns of endogeneity, I investigate whether the baseline results are robust to the use of a subset of moves that are plausibly more exogenous, to instrumenting destinations with historical migration, and to the blending of both approaches. Second, I explore potential heterogeneity by estimating regional childhood exposure effects for sub-populations. Third, I explore whether the evidence of regional childhood exposure effects obtained in the baseline specification holds if I focus on intergenerational upward mobility computed with secondary education as an alternative outcome. The motivation is that primary education as the threshold of interest may seem less policy relevant for current cohorts, although as I have shown, the shares of parents with less than primary school, as well as the children with less than primary school, in the sample under analysis are significant.

V.1 Addressing concerns of endogeneity

V.1.a Anomalous migration outflows

To alleviate concerns that time-varying factors may jointly drive household moves and children's educational investments in proportion to exposure to the region with higher mobility, I re-estimate the model using a subset of moves that are more likely to reflect plausibly exogenous moves.

For this purpose, I construct a panel data set of outflows by region-of-origin-and-year-of-move by counting the number of movers by region of origin and year of move (i.e., census year minus age at move) regardless of their destinations. Next, I use this panel data set to regress outflows on a constant and a linear trend by region of origin, compute the residuals from these regressions, and use them to sort (in ascending order) observations within each region of origin (assigning them percentile ranks). In this way, years that are assigned large percentile ranks (i.e., years with large residuals) can be considered as periods with anomalously large migration outflows.

I use the percentile ranks to identify a subset of years for each region that can be considered to be a period with large anomalous outflows (or at least larger than usual) and run the baseline parametric regression on them. Figure 5 displays the estimates of regional exposure effects for the three age segments used in the parametric specification (see equation 5).[11] I find estimates that are very stable until the percentile 50 and then become more volatile. However, despite the volatility, the estimates remain qualitatively similar with slopes that are non-statistically significant in early ages, a significant childhood exposure effect in the segment of primary school age, and smaller significant effect post primary school age. This suggests that the regional childhood exposure effects found in the baseline results are not driven by families choosing the timing of the move.

V.1.b Expected destination of moving households

To alleviate concerns that time-varying factors may jointly drive household choice of destination and children's educational investments in proportion to exposure to the region with higher mobility, I use past migration destinations from each origin to predict where moving household will settle with a "shift-share" design.

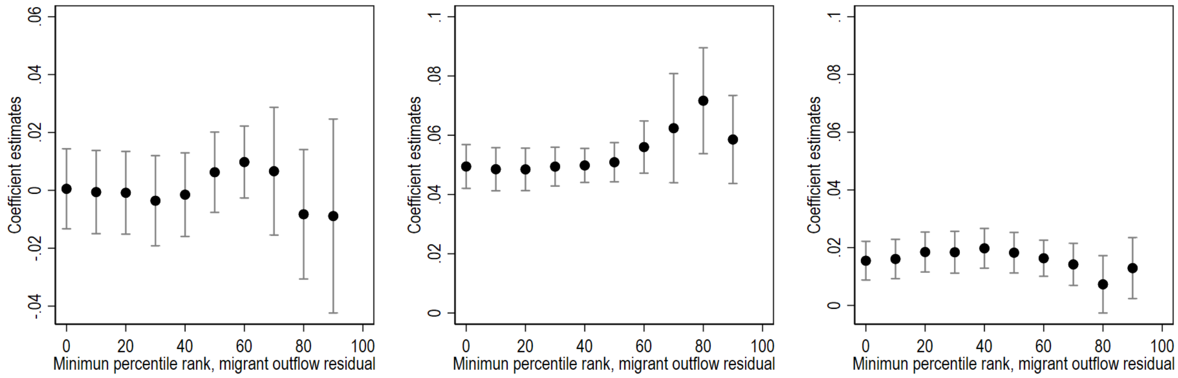
For each year of move y and origin o, I compute the share that moves to destination-d as $\sigma_{ody} = \frac{\sum_{x=T_0}^{y-10} movers_{odx}}{\sum_{d=1}^{D} \sum_{x=T_0}^{y-10} movers_{odx}}$ where D is the total number of regions in a given country and T_0 is the first year in which I observe a mover from this origin.

For individuals who move in year y from region o to region d, I compute the predicted difference in mobility for permanent residents $\hat{\Delta}_{odby}$ as the historic share-weighted analog, $\hat{\Delta}_{oby} = \sum_{d=1}^{D} \Delta_{odb} \times \sigma_{ody}$. Figure 6 plots the predicted difference against the actual difference for the sample of movers.

I use the predicted difference in upward mobility between origin and destination to in-

[11] Table A5 in the Appendix reports these estimates.

Figure 5: Regional childhood exposure effect estimates using anomalous migration outflows



(a) β: 1-4 **(b)** γ: 5-11 **(c)** δ: 12-20

Notes: The dependent variable in all specifications is an indicator that takes the value of one for children of parents without completed primary school who have completed at least primary education and zero otherwise (IGM). The independent variables comprise a linear in origin-average-IGM (calculated for the birth-cohort relevant to the individual among non-movers) term, age-at-move indicator variables, birth-decade×destination indicators interacted with destination-minus-origin differences in upward IGM, all of which are not reported, and three linear terms for destination-minus-origin differences in the relevant-birth-cohort-nonmover average IGM for moves taking place when the child moves, ages 1-4, 5-11, and 12-18. 95% confidence intervals constructed with double clustered at the origin and at the destination region standard errors are reported. Each estimate corresponds to the result of a regression using a subset of the data that consider observations in years ranked above a given threshold (i.e., anomalous migration outflows).

strument the actual difference. Table 3 reports the results of the parametric approach using this instrumental variable strategy. The table displays the reduced form, which estimates equation 5 replacing the actual observed difference between origin and destination with the predicted one, and also the results of a two state least squares estimation. The results are fairly consistent with the baseline results shown before, which suggest that the main findings are not driven by the factors that affect the choice of destinations together with educational attainment.

V.1.c Blending anomalous migration outflows to expected destinations

In this subsection I analyze whether the baseline results are robust to the use of plausibly exogenous moves and to instrumenting the choice of destinations. Figure 7 displays the re-

Figure 6: Predicted versus actual difference in IGM between origin and destinations

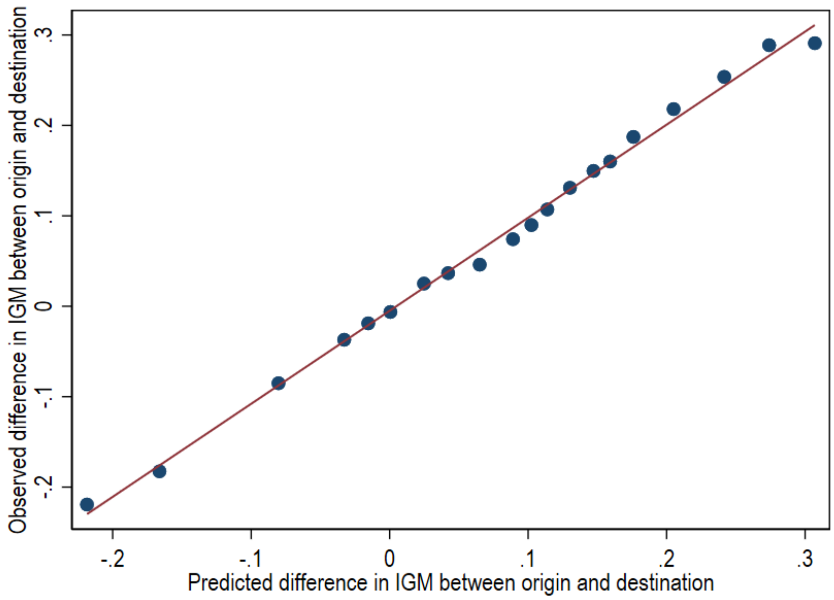


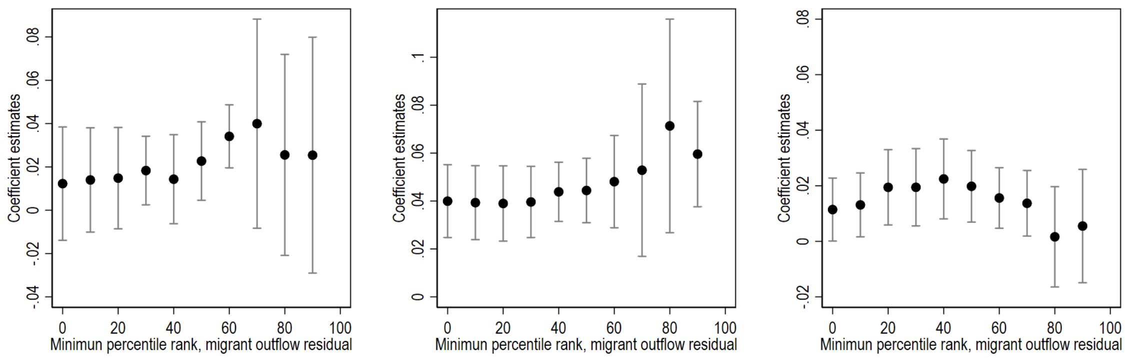
Table 3: Regional childhood exposure effects instrumenting destinations

	(1) IGM	(2) IGM	(3) IGM	(4) IGM	(5) IGM	(6) IGM
β: 1-4	0.0130	-0.0162	-0.0233	0.00661	-0.0232	-0.0208*
	(0.008)	(0.014)	(0.024)	(0.008)	(0.021)	(0.012)
γ: 5-11	0.0441***	0.0473***	0.0334***	0.0442***	0.0348***	0.0471***
	(0.008)	(0.010)	(0.007)	(0.003)	(0.007)	(0.006)
δ: 12-20	0.0104*	0.0121	0.00520	0.0107**	0.00576	0.0120*
	(0.006)	(0.007)	(0.005)	(0.005)	(0.005)	(0.006)
R-squared	0.071	0.067	0.684	0.040	0.001	0.038
N	403751	254661	254661	403216	254348	254348
Household FE	No	No, hhfe sample	Yes	No	Yes	No, hhfe sample
Estimator	OLS	OLS	OLS	2SLS	2SLS	2SLS

Notes: The dependent variable in all specifications is an indicator that takes the value of one for children of parents without completed primary school who have completed at least primary education and zero otherwise (IGM). The independent variables comprise a linear in origin-average-IGM (calculated for the birth-cohort relevant to the individual among nonmovers) term, age-at-move indicator variables, birth-decade × destination indicators interacted with destination-minus-origin differences in upward IGM, all of which are not reported, and three linear terms for destination-minus-origin differences in the relevant-birth-cohort-nonmover average IGM for moves taking place when the child moves, ages 1-4, 5-11, and 12-18. Double clustered at the origin and at the destination region standard errors are reported in parenthesis. OLS columns report reduced form estimates where destination-minus-origin differences in upward IGM are substituted by a predicted difference using historical migration. 2SLS instrument destination-minus-origin differences in upward IGM with predicted differences.

gional childhood exposure estimates applying both approaches together. The results remain consistent with the baseline findings.

Figure 7: Regional childhood exposure effect estimates using anomalous migration outflows and instrumenting destinations



(a) β (ages 1-4) **(b)** γ (ages 5-11) **(c)** δ (ages 12-20)

Notes: The dependent variable in all specifications is an indicator that takes the value of one for children of parents without completed primary school who have completed at least primary education and zero otherwise (IGM). The independent variables comprise a linear in origin-average-IGM (calculated for the birth-cohort relevant to the individual among non-movers) term, age-at-move indicator variables, birth-decade$\times$destination indicators interacted with destination-minus-origin differences in upward IGM, all of which are not reported, and three linear terms for destination-minus-origin differences in the relevant-birth-cohort-nonmover average IGM for moves taking place when the child moves, ages 1-4, 5-11, and 12-18. 95% confidence intervals constructed with double clustered at the origin and at the destination region standard errors are reported. Each estimate corresponds to the result of a regression using a subset of the data that consider observations in years ranked above a given threshold (i.e., anomalous migration outflows).

V.2 Heterogeneity

In this subsection I explore whether there is evidence of heterogeneity in the regional childhood exposure effects computed for a set of sub-populations that are identifiable in the census questionnaires. With this in mind, I estimate the following econometric specification, which is a variation of equation 5:

$$
\begin{aligned}
y_{ihbmod} = \sum_{b=b_0}^{B} 1(b_i = b)(\alpha_b^1 + \alpha_b^2 \gamma_{ob}) + \sum_{m=1}^{20} \zeta_m 1(m_i = m) + \eta^1 1(D_i = 1) \\
+1(m_i < 5)(\beta_0^{base} + (20 - m_i)\beta_1^{base})\Delta_{odb} \\
+1(5 \leq m_i \leq 11)(\gamma_0^{base} + (20 - m_i)\gamma_1^{base})\Delta_{odb} \\
+1(m_i \geq 12)(\delta_0^{base} + (20 - m_i)\delta_1^{base})\Delta_{odb} \\
+1(m_i < 5)1(D_i = 1)(\beta_0^{diff} + (20 - m_i)\beta_1^{diff})\Delta_{odb} \\
+1(5 \leq m_i \leq 11)1(D_i = 1)(\gamma_0^{diff} + (20 - m_i)\gamma_1^{diff})\Delta_{odb} \\
+1(m_i \geq 12)1(D_i = 1)(\delta_0^{diff} + (20 - m_i)\delta_1^{diff})\Delta_{odb} \\
+\epsilon_{ihbmod}
\end{aligned} \tag{8}
$$

where the parameters carry similar meaning as in equation 5 but now the specification includes a set of interactions with an indicator variable $1(D_i = 1)$, which takes the value of 1 when the individual have a given characteristic or belong to a given sub-population. I consider male/female, rural/urban, dwelling non-owner/owner, and movers facing a negative/positive difference in upward mobility between destination and origin.

Table 4 reports the results of this exercise. I find only little evidence of heterogeneity in the dimensions under analysis. In particular, childhood exposure effects in the age segment 5-11 appears to be slightly higher for those moving to urban places and slightly lower for those households that are dwelling owners. The slopes past school age appear flatter in the case of women and steeper for households that are dwelling owners.

Table 4: Parametric estimates of regional childhood exposure effects

	(1) IGM	(2) IGM	(3) IGM	(4) IGM	(5) IGM	(6) IGM	(7) IGM	(8) IGM	(9) IGM	(10) IGM	(11) IGM	(12) IGM
β: 1-4, base	-0.000785	-0.0267**	-0.0144	-0.0319*	0.0114	-0.0429*	0.00791	-0.00404	0.0106	0.0322***	0.0236	0.0465*
	(0.008)	(0.010)	(0.009)	(0.017)	(0.016)	(0.023)	(0.009)	(0.033)	(0.017)	(0.011)	(0.029)	(0.025)
γ: 5-11, base	0.0491***	0.0409***	0.0517***	0.0404***	0.0333***	0.0462***	0.0597***	0.0424***	0.0636***	0.0501***	0.0474***	0.0583***
	(0.004)	(0.006)	(0.004)	(0.007)	(0.011)	(0.009)	(0.004)	(0.007)	(0.006)	(0.007)	(0.011)	(0.007)
δ: 12-20, base	0.0288***	0.0188***	0.0350***	0.0282***	0.0253**	0.0311***	0.00301	0.00610	0.00707	0.0244***	0.0261**	0.0378***
	(0.004)	(0.004)	(0.004)	(0.010)	(0.011)	(0.009)	(0.006)	(0.006)	(0.007)	(0.009)	(0.010)	(0.013)
β: 1-4, diff	0.00580	0.0103	0.00804	0.0383**	-0.0430*	0.0284	-0.00961	-0.0260	-0.0313*	-0.0414**	-0.0730	-0.0866**
	(0.010)	(0.018)	(0.011)	(0.018)	(0.025)	(0.025)	(0.010)	(0.029)	(0.018)	(0.019)	(0.053)	(0.037)
γ: 5-11, diff	0.000200	-0.00371	-0.00182	0.0112**	0.00726	0.00708	-0.0155***	-0.00562	-0.0183***	0.00139	-0.0125	-0.00720
	(0.003)	(0.003)	(0.004)	(0.005)	(0.012)	(0.008)	(0.006)	(0.008)	(0.006)	(0.010)	(0.013)	(0.012)
δ: 12-20, diff	-0.0361***	-0.0195***	-0.0402***	-0.0144	-0.0138	-0.0125	0.0264**	0.0142	0.0286***	-0.0116	-0.0186	-0.0229*
	(0.006)	(0.004)	(0.008)	(0.010)	(0.010)	(0.009)	(0.013)	(0.011)	(0.010)	(0.011)	(0.012)	(0.013)
R-squared	0.105	0.689	0.103	0.126	0.684	0.125	0.099	0.685	0.096	0.097	0.685	0.095
N	436792	271984	271984	422994	263373	263373	433793	270266	270266	436792	271984	271984
Household FE	No	Yes	No, hhfe sample	No	Yes	No, hhfe sample	No	Yes	No, hhfe sample	No	Yes	No, hhfe sample
Base category	Male	Male	Male	Rural	Rural	Rural	Nonowner	Nonowner	Nonowner	Negative	Negative	Negative

Notes: The dependent variable in all specifications is an indicator that takes the value of one for children of parents without completed primary school who have completed at least primary education and zero otherwise (IGM). The independent variables comprise a linear in origin-average-IGM (calculated for the birth-cohort relevant to the individual among nonmovers) term, age-at-move indicator variables, birth-decade $\times$ destination indicators interacted with destination-minus-origin differences in upward IGM, all of which are not reported, and three linear terms for destination-minus-origin differences in the relevant-birth-cohort-nonmover average IGM for moves taking place when the child moves, ages 1-4, 5-11, and 12-18. Double clustered at the origin and at the destination region standard errors are reported in parenthesis.

V.3 Other outcomes

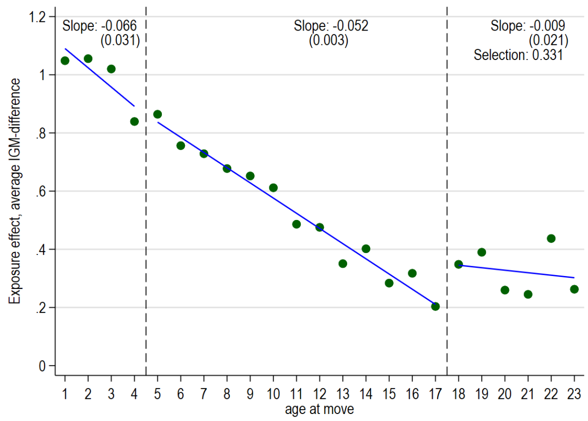
All the previous analysis concerns the likelihood of individuals completing at least primary education conditional on having parents that did not complete that level. This margin is of interest historically, as an important share of the population had not completed it in the cohorts used in the study. However, secondary education would arguably be of more interest if one were thinking about more recent cohorts. In this subsection, I estimate regional childhood exposure effects defining upward intergenerational mobility as the likelihood of completing at least secondary education conditional on having parents that did not complete such level (or conditional on parents that did not complete primary education). The main limitation of this analysis is that I need to focus on older individuals because the educational system is designed so that students finish secondary around the age of 17 and given that co-residence rates drop with age, the sample size drops down. Hence, I estimate regional childhood exposure effects using individuals with age 18-25 co-residing with at least one parent.

Figure 8 reports the estimated slopes using the semi-parametric approach (see equation 4). The exposure effects computed with individuals who moved before being 18 years old is approximately 5.5%, and the selection effect (the average of coefficients after age 17) is approximately 37%. These results imply higher convergence rates than the baseline results but also support the main finding that there is evidence of regional exposure effects in Latin America and the Caribbean.

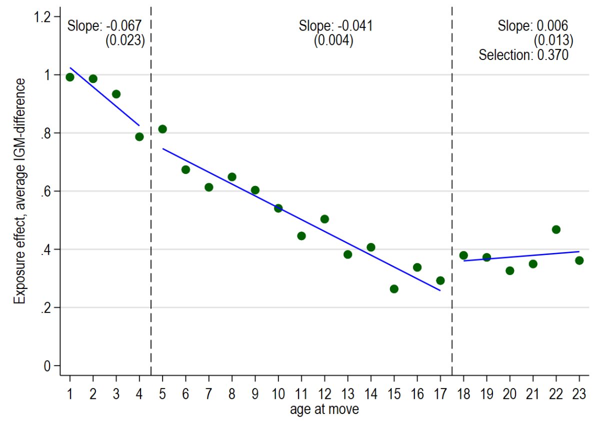
VI Final Remarks

In this paper, I show in a new setting that every additional year that a child spends in a region with higher/lower upward mobility than her birthplace influences her chances of moving up the educational ladder. I replicate the approach of Alesina et al. (2021) for Africa to provide estimates of regional childhood exposure effects in Latin America and the Caribbean, a more

Figure 8: Regional childhood exposure effect estimates for the likelihood of completing at least secondary education when parents were not able to complete this level or primary



(a) Probability of secondary given parents with less than secondary



(b) Probability of secondary given parents with less than primary

Notes: Estimated coefficients b_m from equation 4. These coefficients capture the expected increase of an individual's likelihood of completing at least secondary school (given that their parents were not able to complete primary (left) or secondary (right)) from moving at age m to a place with 1 percentage point higher expected probability for permanent residents. They are estimated by regressing an indicator of secondary school completion of those whose parents move in their childhood on the interaction of their age at move m with $\Delta_{odb} = \gamma_{db} - \gamma_{ob}$ – the difference between upward mobility for permanent residents of the same birth-decade b in the destination d versus the origin o. Controls capture: cohort and origin effects (via indicators for birth-decade interacted with origin); disruption effects (via indicators for age at move); and household effects in the case of the second figure (via household indicators). Both regressions use the same sample, which includes only households with at least two individuals age 14-25 living with at least one parent that did not complete primary education and who moved before being 18 years old.

affluent continent with less inequality, lower poverty rates, higher socioeconomic mobility, higher educational attainment, and different institutions. I find exposure effects greater than those documented in Africa using the same metric of upward mobility (i.e., the likelihood of completing at least primary education for individuals with parents who did not finish primary). This exposure rate is even higher when I consider secondary education instead of primary education as the level of focus to measure upward mobility.

This paper adds to the growing evidence documenting how the places in which children grow up can influence many socioeconomic outcomes later in life. The findings are remarkably similar to those found in other contexts that include high-income countries and the use of different measures such as intergenerational mobility in income computed with rank-rank regressions. These results suggest that place-based policies to increase the human capital accumulation of disadvantaged children that target regions left behind may be fruitful to increase the aggregate level of absolute intergenerational mobility in education in Latin America and the Caribbean. However, it is essential to note that the estimates presented in this paper speak about the levels of intergenerational mobility for birth cohorts that are already adults. Therefore, policy actions would require updated estimates that consider mobility indicators for later birth cohorts and focus on higher levels of completion such as secondary or college attendance, given the widespread adoption of compulsory schooling in the continent.

Finally, a fundamental limitation of this study is that it quantifies the average effect of exposure to regions during childhood without shedding light on the different potential mechanisms driving these effects. Hence, it leaves many open questions about the role of these mechanisms for further research.

References

Alesina, A., Hohmann, S., Michalopoulos, S., & Papaioannou, E. (2020). Religion and Educational Mobility in Africa. *NBER Working Paper*.

Alesina, A., Hohmann, S., Michalopoulos, S., & Papaioannou, E. (2021). Intergenerational Mobility in Africa. *Econometrica*, *89*(1), 1–35.

Asher, S., Novosad, P., & Rafkin, C. (2021). Intergenerational Mobility in India: Estimates from New Methods and Administrative Data. *Working Paper. Available at: http://paulnovosad.com/pdf/anr-india-mobility.pdf (accessed November 2021))*.

Card, D., Domnisoru, C., & Taylor, L. (2022). The Intergenerational Transmission of Human Capital: Evidence from the Golden Age of Upward Mobility. *Journal of Labor Economics*, *40*(S1), S39–S95.

Chetty, R., & Hendren, N. (2018a). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. *Quarterly Journal of Economics*, *133*(3), 1107–1162.

Chetty, R., & Hendren, N. (2018b). The Impacts of Neighborhoods on Intergenerational Mobility II: County-level Estimates. *The Quarterly Journal of Economics*, 1163–1228.

Chetty, R., Hendren, N., & Katz, L. F. (2016). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review*, *106*(4), 855–902.

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

Chyn, E. (2018). Moved to Opportunity: The Long-run Effects of Public Housing Demolition on Children. *American Economic Review*, *108*(10), 3028–3056.

Chyn, E., & Katz, L. F. (2021). Neighborhoods Matter: Assessing the Evidence for Place Effects. *NBER Working Paper*, *28953*.

Damm, A. P., & Dustmann, C. (2014). Does Growing Up in a High Crime Neighborhood Affect Youth Criminal Behavior? *American Economic Review*, *104*(6), 1806–1832.

Derenoncourt, E. (2021). Can You Move to Opportunity? Evidence from the Great Migration. *Forthcoming, American Economic Review*.

Deutscher, N. (2020). Place, Peers, and the Teenage Years: Long-run Neighborhood Effects in Australia. *American Economic Journal: Applied Economics*, *12*(2), 220–249.

Emran, M. S., Greene, W., & Shilpi, F. (2018). When Measure Matters: Coresidency, Truncation Bias, and Intergenerational Mobility in Developing Countries. *Journal of Human Resources*, *53*(3), 579–607.

Francesconi, M., & Nicoletti, C. (2006). Intergenerational Mobility and Sample Election in Short Panels. *Journal of Applied Econometrics*, *21*, 1265–1293.

IPUMS. (2019). *Integrated Public Use Microdata Series, International: Version 7.2 [dataset].* Minneapolis, MN: IPUMS: Minnesota Population Center.

Laliberté, J.-W. (2021). Long-term Contextual Effects in Education: Schools and Neighborhoods. *American Economic Journal: Economic Policy*, *13*(2), 336–377.

Munoz, E. (2021a). The Geography of Intergenerational Mobility in Latin America and the Caribbean. *Stone Center Working Paper*, *29*.

Munoz, E. (2021b). Intergenerational Educational Mobility within Chile. *Available at SSRN: https://ssrn.com/abstract=3969270 or http://dx.doi.org/10.2139/ssrn.3969270*.

Munoz, E., & Siravegna, M. (2021). When Measure Matters: Coresidence Bias and Intergenerational Mobility Revisited. *Available at SSRN: https://ssrn.com/abstract=3969270 or http://dx.doi.org/10.2139/ssrn.3969270*.

Narayan, A., Van der Weide, R., Cojocaru, A., Lakner, C., Redaelli, S., Gerszon Mahler, D., ... Thewissen, S. (2018). *Fair Progress?: Economic Mobility Across Generations Around the World*. The World Bank.

Sharkey, P., & Torrats-Espinosa, G. (2017). The Effect of Violent Crime on Economic Mobility. *Journal of Urban Economics*, *102*, 22–33.

Torche, F. (2019). Educational Mobility in Developing Countries. *WIDER Working Paper*(88), 1–31.

Van der Weide, R., Ferreira de Souza, P., & Barbosa, R. (2020). Intergenerational Mobility in Education in Brazil. *mimeo*.

Van der Weide, R., Lakner, C., Gerszon Mahler, D., Narayan, A., & Ramasubbaiah, R. (2021). Intergenerational Mobility Around the World. *Policy Research Working Paper*, *9707*.

Ward, Z. (2020). Internal Migration, Education, and Intergenerational Mobility: Evidence from American History. *Fortcoming at Journal of Human Resources*.

Appendices

In this Appendix I provide details on the sample construction and some additional tables and graphs.

Figure A1 shows the map of upward mobility in education at the district-level for Latin America and the Caribbean region.

Table A1 lists the country, census years and fraction of the full-count census available in the sample.

Figure A2 shows co-residence rates by age.

Table A2 describes the sample sizes by census considering individuals in the age range of interest (14-25), how many of them are considered movers and non-movers, and how many of them are used in the different regressions.

Figure A2 reports childhood exposure effects of a specification that includes an interaction between the destination-origin differences in mobility and indicators variables by cohort.

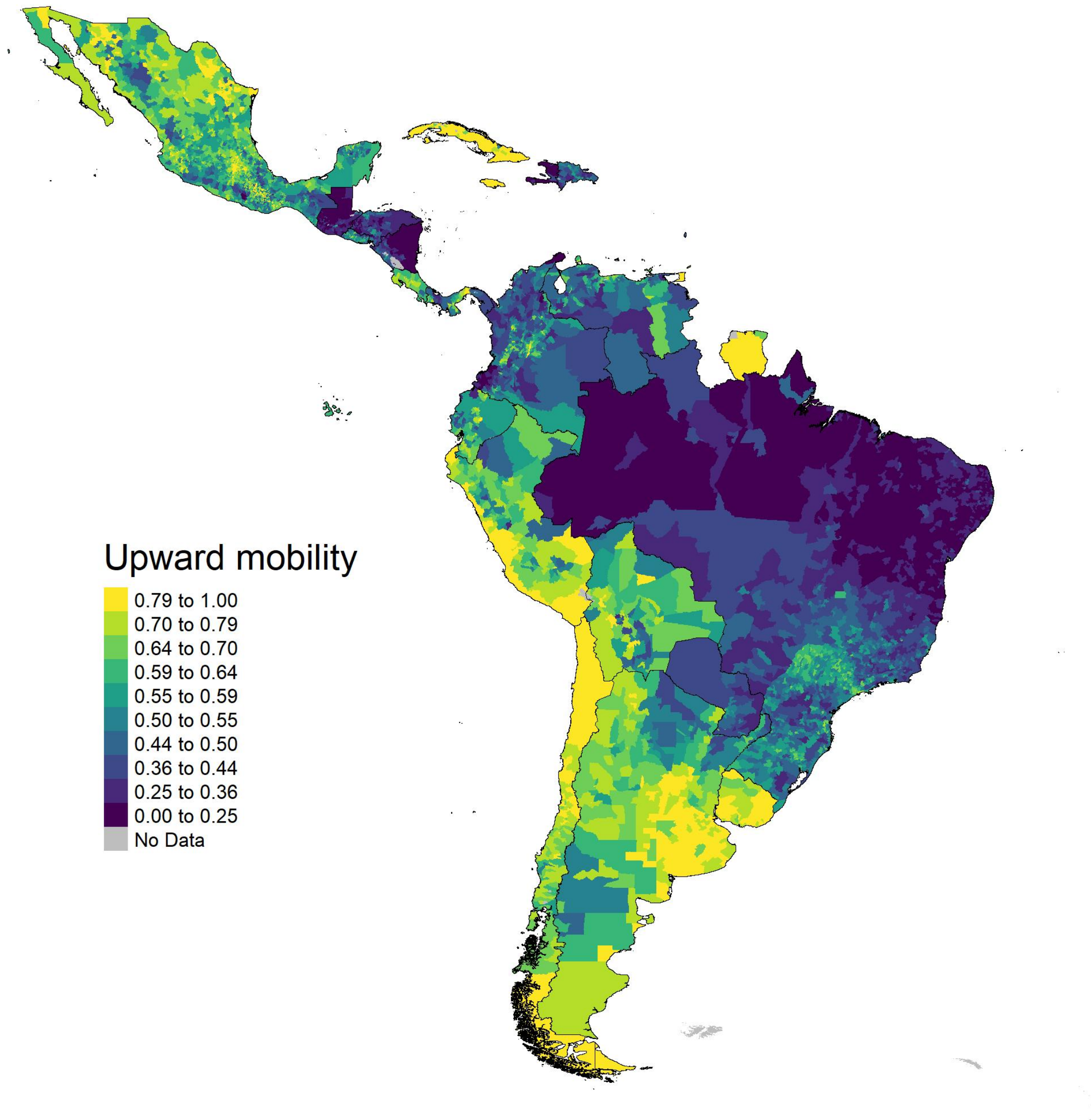
Table A3 reports the parametric estimates of a specification that includes an interaction between the destination-origin differences in mobility and indicators variables by cohort.

Figure A3 report goodness of fit (information criteria (AIC and BIC), R-squared, and adjusted R-squared) statistics for different competing model specifications.

Table A4 report the estimates for different competing model specifications.

Table A5 reports parametric estimates of regional childhood exposure effects using anomalous migrant outflows.

Figure A1: Upward Mobility in LAC

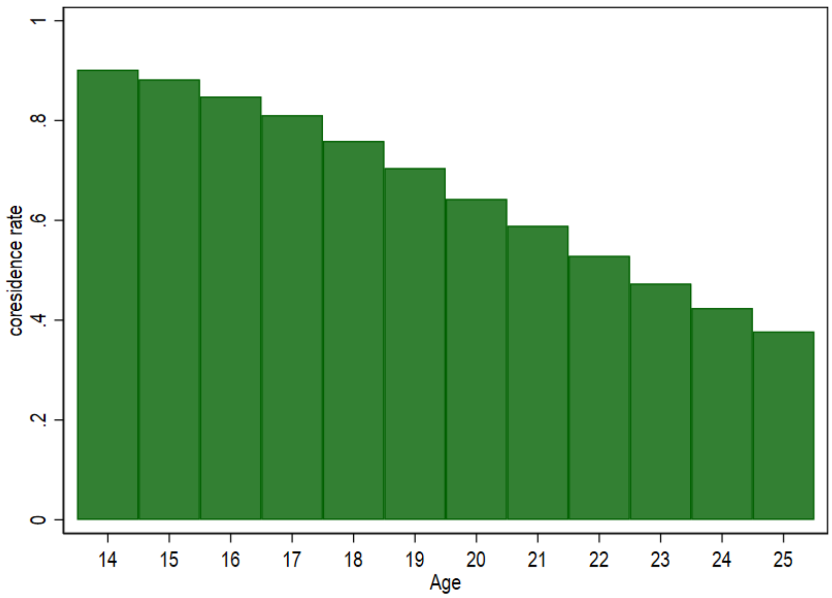


Notes: Upward mobility reflects the likelihood that children, aged 14-18, whose parents have not completed primary schooling will manage to complete at least primary education. This graph uses provinces for St. Lucia, Jamaica, Trinidad and Tobago and Suriname that do not have a finer administrative units in the data set. Source: Munoz (2021a)

Table A1: Fraction of full-count census by sample

N	Country	Census years	Fraction (%)
1	Brazil	1991, 2000, 2010	10, 10, 10
2	Colombia	1973	10
3	Cuba	2002, 2012	10, 10
4	Ecuador	1974, 1982, 2001	10, 10, 10
5	El Salvador	1992, 2007	10, 10
6	Guatemala	1981, 1994	5, 10
7	Jamaica	1982, 1991, 2001	10, 10, 10
8	Mexico	1970	1
9	Panama	1960, 1980	5, 10
10	Trinidad and Tobago	1970	10
11	Uruguay	2011	10

Figure A2: Coresidence rate by age



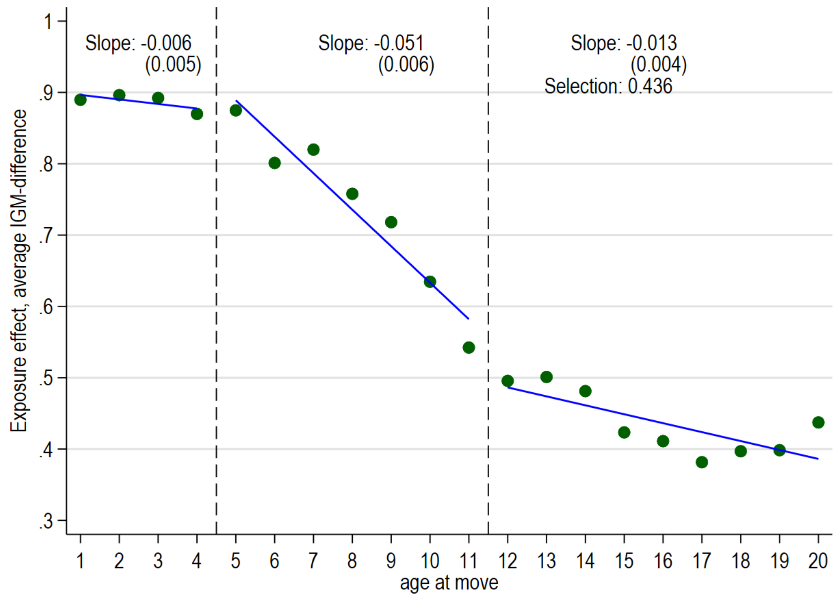
Notes: Co-residence is defined as having information of educational attainment for at least one parent. The graph is constructed using only individuals age 14-25 with information about their own educational attainment.

Table A2: Sample sizes

	All observations			Used in regressions		HH FE reg.
Country Year	All	Movers	Non-movers	Movers	Non-movers	Movers
Brazil 1991	3949560	568888	3380672	172145	1910286	118732
Brazil 2000	4814694	636011	4178683	151120	1887591	91938
Brazil 2010	4449465	496048	3953417	67934	1241344	33234
Colombia 1973	456167	83564	372603	15151	142390	9414
Cuba 2002	177974	20080	157894	40	3702	8
Cuba 2012	180743	13850	166893	8	1664	0
Ecuador 1974	162317	96782	65535	4463	25277	2773
Ecuador 1982	194579	111590	82989	3359	27476	2138
Ecuador 2001	286685	144260	142425	2251	32925	1264
El Salvador 1992	128955	22027	106928	4937	38140	3192
El Salvador 2007	130687	15362	115325	2385	41353	1319
Guatemala 1981	71827	9933	61894	2290	31695	1379
Guatemala 1994	194270	21282	172988	4522	88173	2709
Jamaica 1982	57011	19759	37252	547	4918	276
Jamaica 1991	55824	14455	41369	251	2367	129
Jamaica 2001	46514	12986	33528	53	762	14
Mexico 1970	110945	18105	92840	4202	46679	2792
Panama 1960	11005	3621	7384	217	2695	101
Panama 1980	46539	14622	31917	715	8994	463
Trinidad and Tobago 1970	16107	8907	7200	15	828	10
Uruguay 2011	57112	9770	47342	161	1785	70

Notes: This table reports the total sample size by country-year Census, and for restricted population by age and keeping only observations with information of education for children and parents.

Figure A2: Place childhood exposure effect estimates for the likelihood of completing at least primary education when parents were not able to complete this level - Specification that includes an interaction by cohort



Notes: Estimated coefficients b_m from equation 4. These coefficients capture the expected increase of an individual's likelihood of completing at least primary school (given that their parents were not able to do so) from moving at age m to a place with 1 percentage point higher expected probability for permanent residents. They are estimated by regressing an indicator of primary school completion of those whose parents move in their childhood on the interaction of their age at move m with $\Delta_{odb} = \gamma_{db} - \gamma_{ob}$ – the difference between upward mobility for permanent residents of the same birth-decade b in the destination d versus the origin o. Controls capture: cohort and origin effects (via indicators for birth-decade interacted with origin); and disruption effects (via indicators for age at move). The sample includes individuals age 14-25 living with at least one parent that did not complete primary education and who moved before being 18 years old.

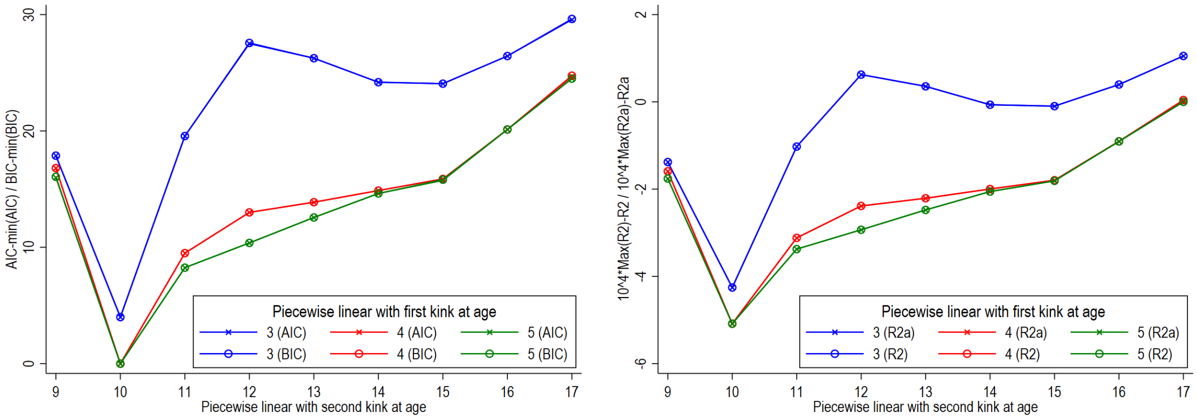
Table A3: Parametric estimates of regional childhood exposure effects - Specification that includes an interaction by cohort

	(1) IGM	(2) IGM	(3) IGM
β: 1-4	0.000653	-0.0139*	-0.0297**
	(0.007)	(0.008)	(0.013)
γ: 5-11	0.0494***	0.0511***	0.0336***
	(0.004)	(0.004)	(0.006)
δ: 12-18	0.0152***	0.0193***	0.00710**
	(0.004)	(0.004)	(0.003)
R-squared	0.095	0.092	0.685
N	436792	271984	271984
Household FE	No	No, hhfe sample	Yes

Notes: The dependent variable in all specifications is an indicator that takes the value of one for children of parents without completed primary school who have completed at least primary education and zero otherwise (IGM). The independent variables comprise a linear in origin-average-IGM (calculated for the birth-cohort relevant to the individual among nonmovers) term, age-at-move indicator variables, birth-decade $\times$ destination indicators interacted with destination-minus-origin differences in upward IGM, all of which are not reported, and three linear terms for destination-minus-origin differences in the relevant-birth-cohort-nonmover average IGM for moves taking place when the child moves, ages 1-4, 5-11, and 12-18. Double clustered at the origin and at the destination region standard errors are reported in parenthesis.

$$
\begin{aligned}
y_{ihbmod} = [\alpha_h +] \sum_{b=b_0}^{B} 1(b_i = b)(\alpha_b^1 + \alpha_b^2 \gamma_{ob}) + \sum_{m=1}^{18} \zeta_m 1(m_i = m) \\
+ \sum_{b=b_0}^{B} \kappa_b 1(b_i = b)\Delta_{odb} \\
+1(m_i < 5)(\beta_0 + (18 - m_i)\beta_1)\Delta_{odb} \\
+1(5 \leq m_i \leq 11)(\gamma_0 + (18 - m_i)\gamma_1)\Delta_{odb} \\
+1(m_i \geq 12)(\delta_0 + (18 - m_i)\delta_1)\Delta_{odb} + \epsilon_{ihbmod}
\end{aligned}
$$

Figure A3: Goodness of fit statistics by model specification



(a) Information criteria (AIC and BIC) **(b)** R-squared and adjusted R-squared

Notes: Goodness of fit statistics for different model specifications. The statistics are transformed such that lower values indicate better fit to help the readability. The dependent variable in all specifications is an indicator that takes the value of one for children of parents without completed primary school who have completed at least primary education and zero otherwise (IGM). The independent variables comprise a linear in origin-average-IGM (calculated for the birth-cohort relevant to the individual among nonmovers) term, age-at-move indicator variables, birth-decade $\times$ destination indicators interacted with destination-minus-origin differences in upward IGM, all of which are keep across competing specifications, and three linear terms for destination-minus-origin differences in the relevant-birth-cohort-nonmover average IGM for moves taking place when the child moves, age pre-primary school (1 to 3, 4, or 5), primary school ages (6 or earlier to 9 or later until 17), and post primary school ages (10 or later to 20). A constant exposure effect model as in Chetty and Hendren (2018a) was also estimated but it is not plotted as it shows poorer fit.

Table A4: Estimates and goodness of fit of different parametric specifications

	(1) IGM	(2) IGM	(3) IGM	(4) IGM	(5) IGM	(6) IGM	(7) IGM	(8) IGM	(9) IGM	(10) IGM	(11) IGM	(12) IGM	(13) IGM
β_linear	0.0080***												
	(0.001)												
β		-0.0077	-0.0077	-0.0077	-0.0077	0.0005	0.0005	0.0005	0.0005	0.0009	0.0009	0.0009	0.0009
		(0.011)	(0.011)	(0.011)	(0.011)	(0.007)	(0.007)	(0.007)	(0.007)	(0.004)	(0.004)	(0.004)	(0.004)
γ		0.0301***	0.0362***	0.0434***	0.0472***	0.0350***	0.0417***	0.0494***	0.0528***	0.0302***	0.0416***	0.0522***	0.0559***
		(0.004)	(0.004)	(0.004)	(0.004)	(0.006)	(0.005)	(0.004)	(0.004)	(0.007)	(0.006)	(0.004)	(0.004)
δ		0.0236***	0.0176***	0.0155***	0.0162***	0.0236***	0.0176***	0.0155***	0.0162***	0.0236***	0.0176***	0.0155***	0.0162***
		(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)
R-sq	0.0831	0.0946	0.0947	0.0946	0.0946	0.0946	0.0947	0.0947	0.0946	0.0946	0.0947	0.0947	0.0946
AIC	596106	590592	590576	590592	590600	590591	590574	590584	590585	590592	590574	590580	590582
BIC	596205	590724	590697	590713	590721	590723	590706	590715	590706	590735	590706	590701	590703
N	436792	436792	436792	436792	436792	436792	436792	436792	436792	436792	436792	436792	436792
age: β	NA	1-3	1-3	1-3	1-3	1-4	1-4	1-4	1-4	1-5	1-5	1-5	1-5
age: γ	NA	4-9	4-10	4-11	4-12	5-9	5-10	5-11	5-12	6-9	6-10	6-11	6-12
age: δ	NA	9-20	10-20	11-20	12-20	9-20	10-20	11-20	12-20	9-20	10-20	11-20	12-20

Notes: The dependent variable in all specifications is an indicator that takes the value of one for children of parents without completed primary school who have completed at least primary education and zero otherwise (IGM). The independent variables comprise a linear in origin-average-IGM (calculated for the birth-cohort relevant to the individual among non-movers) term, age-at-move indicator variables, birth-decade × destination indicators interacted with destination-minus-origin differences in upward IGM, all of which are not reported and are keep across competing specifications, and three linear terms for destination-minus-origin differences in the relevant-birth-cohort-nonmover average IGM for moves taking place when the child moves, early age pre primary education, typical relevant ages for primary education (approximately 6-11), and post primary education until the age 20. Each column report different options for the cutoffs defining these three periods In addition, the first column report a model where I don't distinguish between these three segments. Double clustered at the origin and at the destination region standard errors are reported in parenthesis.

Table A5: Parametric estimates of regional childhood exposure effects using anomalous migrant outflows

	(1) 100%	(2) 90%	(3) 80%	(4) 70%	(5) 60%	(6) 50%	(7) 40%	(8) 30%	(9) 20%	(10) 10%
β: 1-4	0.000524	-0.000598	-0.000849	-0.00359	-0.00150	0.00626	0.00978	0.00661	-0.00827	-0.00888
	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.006)	(0.011)	(0.011)	(0.017)
γ: 5-11	0.0494***	0.0485***	0.0485***	0.0494***	0.0498***	0.0509***	0.0560***	0.0624***	0.0716***	0.0586***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.009)	(0.009)	(0.008)
δ: 12-20	0.0155***	0.0161***	0.0185***	0.0184***	0.0198***	0.0183***	0.0163***	0.0142***	0.00730	0.0129**
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.005)	(0.005)
R-squared	0.095	0.094	0.093	0.092	0.093	0.095	0.097	0.100	0.103	0.112
N	436792	425213	411815	397561	379760	346115	306502	248518	190278	110330
Household FE	No	No	No	No	No	No	No	No	No	No

Notes: The dependent variable in all specifications is an indicator that takes the value of one for children of parents without completed primary school who have completed at least primary education and zero otherwise (IGM). The independent variables comprise a linear in origin-average-IGM (calculated for the birth-cohort relevant to the individual among nonmovers) term, age-at-move indicator variables, birth-decade $\times$ destination indicators interacted with destination-minus-origin differences in upward IGM, all of which are not reported, and three linear terms for destination-minus-origin differences in the relevant-birth-cohort-nonmover average IGM for moves taking place when the child moves, ages 1-4, 5-11, and 12-18. Double clustered at the origin and at the destination region standard errors are reported in parenthesis.