9-2022

Essays on Social Mobility and Distribution

Luis A. Monroy Gómez Franco
The Graduate Center, City University of New York

Essays on Social Mobility and Distribution

by

Luis Angel Monroy Gómez Franco

A dissertation submitted to the Graduate Faculty in Economics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2022

This manuscript has been read and accepted by the Graduate Faculty in Economics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Professor Miles Corak

Date	Chair of Examining Committee

Professor Christos Giannikos

Date	Executive Officer

Professor Miles Corak

Professor Jonathan Conning

Professor Nuria Rodriguez Planas

Supervisory Committee

Abstract

Essays on Social Mobility and Distribution

by

Luis Angel Monroy Gómez Franco

Adviser: Professor Miles Corak

This dissertation consists of three chapters on the relationship between immutable characteristics of a person (also known as circumstances) and the intergenerational social mobility experienced by them, as well as their human capital accumulation. It pays particular attention to the role played by the economic resources of the household of origin in producing differences in the economic outcomes across persons.

In the first chapter, I show that the decomposition of intergenerational persistence indicators into their structural and positional components offers a clearer understanding of the determinants of the heterogeneity in subnational mobility rates. The crucial element for the separate analysis of positional and structural mobility is the use of regionally defined instead of nationally defined quantiles. This constitutes a departure from the current consensus in the estimations of mobility rates at the subnational level in economics. Using the Mexican case as an example, I show that there are no significant differences across the country's regions in terms of positional mobility. This contrast with the existing results and their interpretations, particularly regarding intergenerational mobility in the south region of the country. This highlights the importance of incorporating positional measures into the battery of tools used for intranational analysis.

In the second chapter, I study the difference in intergenerational mobility patterns between

the indigenous and non-indigenous Mexican populations, analyzing the role played by the variability in skin tones inside both groups. I employ a composite sample based on two recent surveys with the same sampling frame and questionnaire design. I obtained three main results based on rank-rank regressions, transition matrices, and quantile rank regressions. First, the indigenous population faces a higher degree of intergenerational persistence and remains concentrated at the bottom of the distribution of economic resources. Secondly, the indigenous population faces less upward mobility from the bottom of the distribution but similar rates of persistence to the non-indigenous population at the top. Furthermore, there is a skin tone gradient for the indigenous and non-indigenous populations, where lighter skin tones tend to remain at a higher rank in the distribution of economic resources than darker skin tones.

In the third chapter, I propose a framework for analyzing the short and long-run effects of temporary educational disruptions on children's learning progression. The framework includes continuous parental investments, filling a gap in the literature on the Potential Pedagogical Function. Similarly, it considers the role of economic resources as part of households' attenuation capacity. I take this framework to the data to estimate the potential effects of the COVID-19 pandemic in Mexico. The potential persistent loss in learning lies between 20%-90% of learning acquired during a typical school year, depending on the effectiveness of the policies implemented in 2020 and 2021.

Acknowledgments

Every intellectual endeavor is the fruit of a collective effort, and this dissertation is not an exemption. I want to thank Professor Miles Corak for his guidance and mentorship throughout the process of writing the dissertation. His continued interest in my work and comments and questions about it helped me to develop my ideas fully and to express them forcefully in the essays that form the dissertation. I also want to thank Professors Jonathan Conning and Nuria Rodríguez-Planas for their invaluable inputs to this dissertation. I am certain that their comments, questions, and suggestions allowed me to improve each essay significantly.

I want to thank Professors Branko Milanovic and Leslie McCall for helping me find in the Stone Center on Socio-Economic Inequality a community of fellow scholars interested in economic inequality and social mobility issues. The community of the Stone Center is an example of how the academy should work: cordially, rigorously, interdisciplinarily, and always aiming to gain a better understanding of the society we live in. I want to thank Professor Janet Gornick for her fantastic work in building such a community, her support during my time in the Graduate Center, and all the teachings she imparted inside and outside the classroom.

I have been fortunate to find mentors throughout my academic career that have pushed me to be the best possible version of myself. In that sense, I want to thank Professors Eva O.

Arceo-Gómez, Raymundo M. Campos-Vázquez and Juan Carlos Moreno-Brid for all their teachings and support across the years. Moreover, more importantly, I want to thank them for their friendship and concern when times got rough. I would also like to dedicate this dissertation to the memory of Professor Jaime Ros Bosch, who taught me how to be an applied economist and always supported me in the goal of finishing my degree. I hope I can be as good a mentor to others as the four of them have been to me.

I also have been fortunate enough during this time to find colleagues that became friends and mentors. I want to thank Gastón Yalonetzky, Roberto Vélez-Grajales, Ingrid Bleynat, and Luis Baldomero-Quintana for all their support during the ups and downs of the Ph.D. I want to thank Gastón, Ingrid and Roberto, particularly, for their trust and willingness to work with me on several projects throughout these years.

Living five years as a migrant has entailed a series of challenges, but undoubtedly the presence of friends has allowed me to surpass them successfully. First, I want to thank the friends who have been fellow travelers in life for now more than 16 years: Canek, Cynthia, Jessica, and Daniel; without you, I would not have made it in New York. Even the thousands of kilometers were no barrier to your friendship and love, and I will always be grateful for it. I also thank Ana, Jorge, and Ruso for more than a decade of friendship. More than once, we have talked in the middle of the night to lend moral support when things get rough, or a deadline looms over our heads. But we have also celebrated each other's successes, and I sincerely thank you for both. I also want to thank Raúl, Aranxa, Checo, Diego, and José Carlos for all the chats we held during these years, both the academic and non-academic ones, as they were a refuge in which I could find solace.

New York City not only brought with it new challenges and experiences but also brought

with it amazing friends. I want to thank Julio, Juliana, Twisha, Philip, Marco, Majo, Sandra, and Joaquin for helping me to find a community in this marvelous albeit imposing city. All those chats over coffee or beer were indispensable inputs to recharge energies and keep going with the Ph.D. These years also taught me that one could find people with whom deep connections can be established and maintained regardless of distance and circumstances. I want to thank Paloma Villagómez-Ornelas for showing me that and being there in the good and the bad, in the happy and not-so-happy moments of this journey called Ph.D. in economics.

During my Ph.D. I also had to deal with what has become a familiar presence in academia: clinical depression. I want to thank M.D. Nadia Becerril and M.A. César Galicia for helping me recover when my brain decided to go haywire. It is not an understatement to say that I would not have finished the degree without your support. I want to dedicate this dissertation to all graduate students who have struggled with mental illness during their studies, hoping we can build a better place for future generations.

Finalmente, quiero agradecer profundamente a mi mamá, María de los Ángeles Monroy Gómez-Franco, y a mis tías Celia y Concepción por todo el apoyo, cariño, amor y ternura que me han prodigado desde que tengo memoria. Sin ustedes no sería lo que soy, y me queda claro que han trabajado sin descanso para darme todas las oportunidades que ustedes no tuvieron. Espero poder honrar esos sacrificios con mi trabajo, y les dedico a ustedes y a mis abuelos, Celia y Crispín, está disertación.

Contents

List of Tables

List of Figures

Chapter 1

Regional comparisons of intergenerational social mobility: the importance of positional mobility.

1.1 Introduction

One of the most active areas of economics research on intergenerational social mobility is the study of intranational heterogeneity in the patterns and determinants of mobility. However, existing research[1] has left aside the analysis of intranational differences in positional mobility across a country's regions. In this paper, I contribute to filling this void by providing the first set of strictly positional measures of regional intergenerational mobility for Mexico. In contrast with previous estimations of regional patterns of intergenerational mobility (Vélez-Grajales et al., 2018b; Delajara and Graña, 2018; Monroy-Gómez-Franco and Vélez-Grajales, 2021; Delajara et al., 2022), I do not find differences in rank persistence

[1] See Chetty et al. (2015); Heidrich (2017); Delajara and Graña (2018); Corak (2019); Connolly et al. (2019); Eriksen and Munk (2020); Deutscher and Mazumder (2020); Monroy-Gómez-Franco and Vélez-Grajales (2021); Delajara et al. (2022); Acciari et al. (2022) for recent examples of this literature.

and aggregate persistence measures for the Mexican regions. The reason for the difference in results is that previous estimates employed as reference group the national distribution, which leads to a mixing of the structural and positional mobility patterns, as shown by Deutscher and Mazumder (2021).

Following Markandya (1982, 1984), let the marginal distribution of life outcome y at time t be defined as $g(y_{i,t})$, the marginal distribution of the same life outcome at time $t-1$ as $g(y_{i,t-1})$ and the corresponding joint distribution as $f(g(y_{i,t-1}), g(y_{i,t}))$. Then, it is possible to define *structural mobility* as the changes in the marginal distributions between t and $t-1$ maintaining constant the relative position of each i. In contrast, *positional or exchange mobility* is concerned with the changes in the position occupied by person i between t and $t-1$, holding constant the marginal distributions. In other words, *structural mobility* refers to the changes in the marginal distributions between two generations, while *positional mobility* refers to the changes in the copula that links the positions for the same individual across both marginal distributions.

The joint analysis of both mobility concepts has been the focus of most economics research[2], particularly that arising from traditional economic models of intergenerational mobility such as Becker and Tomes (1979, 1986) and Becker et al. (2018). However, for this literature, positional mobility is only a side product of the different intergenerational dynamics in absolute terms of the households in the population, and it is not relevant by itself. This is a direct result of the assumption that a person's utility function is defined only in terms of absolute consumption, which implies that there are no effects associated with being located at a specific part of the distribution of economic resources.

[2]See Fields and Ok (1999); Jäntti and Jenkins (2015); Durlauf et al. (2022) for surveys of the economics literature on intergenerational social mobility.

The existing research on intranational regional differences in intergenerational mobility patterns at the regional level has followed a similar approach. The seminal work by Chetty et al. (2014, 2015) introduced the use of nationally defined ranks as the dimension of analysis upon which the different regional patterns of intergenerational persistence are to be compared. Although originally motivated by the data restrictions faced by the authors, the approach is also consistent with a hybrid interpretation of intergenerational mobility in which the concepts of structural and positional mobility are mixed. On the one hand, changes in the position of a household from a particular region in the national distribution between two generations imply changes in the absolute level of resources accessible to that household, thus implying structural mobility. On the other hand, it also captures persistence (or the lack thereof) at a specific part of the national distribution, thus implying positional mobility with respect to the national distribution, but not necessarily the regional distribution.

The main limitation of this approach is that, by definition, it limits the possibility of identifying if the main driver of the observed intergenerational mobility patterns is structural or positional mobility. In this paper, I seek to contribute to the existing literature on regional differences in social mobility by showing that focusing on positional mobility can provide a deeper understanding of the regional differences. This implies opening a research agenda that has not been explored so far in the literature while simultaneously complementing the existing findings for the Mexican case.

The rest of the paper proceeds as follows. In the following section, I discuss the importance of analyzing positional intergenerational mobility at the subnational scale, emphasizing how it complements the existing literature on regional differences in mobility. Next, I describe the different measures of intergenerational mobility used in the paper, emphasizing their

positional interpretation. In the fourth section of the paper, I describe the main characteristics of the database employed, as well as the construction process of the outcome variable. The dataset can be considered an example of the type of data set employed for social mobility analysis in contexts where panel databases are non-existent or non-intergenerational. After that, I present the results for positional mobility across Mexican regions, showing how they complement the existing results and provide a better insight into the determinants of regional differences in mobility. Finally, I discuss some final remarks and future avenues of work.

1.2 Positional mobility and regional intergenerational mobility

As mentioned above, positional intergenerational mobility refers to the changes in the position occupied by a person in the distribution of a particular outcome with respect to the one occupied by her predecessor in the corresponding distribution, assuming that the marginal distribution of the outcome variable remains constant between the two cohorts. The relevance of this concept of social mobility arises from two growing bodies of economics literature: the analysis of positional concerns in terms of welfare and the political economy of social mobility and elite circulation.

1.2.1 Positional concerns: the microeconomics of positional mobility

A growing body of literature finds that individuals care not only about their absolute level of consumption or income, but also about their position with respect to different reference groups[3]. The existence of positional preferences has been identified in societies across the world, from European (Clark and Senik, 2010; Garratt et al., 2016; Mujcic and Frijters, 2012), to Latin American (Esposito and Villaseñor, 2019; Castilla, 2012; Kuegler, 2009),

[3]See Weiss and Fershtman (1998); Heffetz and Frank (2011) and Clark et al. (2008) for surveys of this literature.

North American (Luttmer, 2005; Japaridze and Sayour, 2021), African (Kingdon and Knight, 2007; Akay et al., 2014; Lentz, 2017) and Asian societies (Clark et al., 2022; Knight et al., 2009; Carlsson et al., 2008; Carlsson and Qin, 2010). Similarly, research identifies that positional concerns are present in multiple domains and that they are particularly strong for goods in which comparisons are socially invited, such as income, wealth, education, and several types of durable consumption goods (cars, houses, among others) (Birdal and Ongan, 2016; Bogaerts and Pandelaere, 2013; Carlsson et al., 2007; Murphy and Weinhardt, 2020).

These findings have a direct implication for the literature on intergenerational social mobility, as they suggest that persons base their decisions considering their implications on their relative standing with respect to other members of society. In the framework proposed by Becker et al. (2018), this implies that parental investments are made not only to improve the absolute income of the descendants but to guarantee persistence at or mobility towards the upper echelons of the distribution of economic resources in society. In other words, parental investments are made with a specific result in terms of *positional mobility* in mind[4]. Furthermore, this implication is strengthened when considering the existence of last place aversion (Kuziemko et al., 2014), as it implies that parents will seek to avoid the scenario where their descendants end at the bottom of the distribution.

The second implication is the need to identify the relevant distribution (or reference group) upon which individuals evaluate their positional mobility experience. As Deutscher and Mazumder (2021) show, using the national distribution to compare mobility patterns at the regional level ends up producing measures of mobility that do not adequately capture the positional mobility inside each region. However, this would only be a problem for positional

[4]For a theoretical discussion on how positional mobility can affect the utility of an individual, see Markandya (1982, 1984).

analysis under the traditional framework of mobility analysis (i.e. Becker et al. (2018); Becker and Tomes (1986)) if individuals use as a reference group the regional and not the national distribution. If that is not the case, the current practice of using the national distribution as reference group would be a correct way of measuring positional mobility, as Chetty et al. (2015) and Acciari et al. (2022) argue.

The literature on reference groups has relied principally on experimental methods to identify the relevant comparison group in different dimensions. For example, Falk and Knell (2004); Lentz (2017); Chang (2013) and Pérez-Asenjo (2011) show that individuals are more likely to compare themselves with those that are perceived to be similar in characteristics such as education level, age, and gender. Moreover, Mageli et al. (2022); Lentz (2017) and Celse (2018) show that positional concerns arise when the comparison occurs against a member of society outside the close networks of the person. An unfavorable comparison against a similar but unfamiliar member of society produces a loss of welfare in the respondents. There is no effect when the comparison is made against a member of the person's network. However, Carlsson et al. (2008); Knight et al. (2009); Kingdon and Knight (2007); Davis and Wu (2014) and Bhuiyan (2018) show that positional concerns increase when the reference group is part of the same ethnic group or region as the person. These results imply that for the analysis of regional differences in positional mobility, the relevant reference group from the point of view of the observation unit is the regional distribution of the outcome variable, not the national one. This provides further evidence in support of the argument made by Deutscher and Mazumder (2021) regarding the interpretation of the existing results based on nationally defined ranks.

In the specific case of Mexico, Castilla (2012) and Esposito and Villaseñor (2019) show that positional concerns are present in the realms of income and education. Moreover, Esposito

and Villaseñor (2019) show that, at least in the realm of education, the relevant reference group is the region the person inhabits, not the national population. Both results support the approach of this paper of analyzing regional differences in intergenerational mobility from a positional point of view, using the regional distributions of economic resources as reference groups.

1.2.2 The political economy of social mobility and elite persistence.

In contrast with the literature on positional concerns that focuses on individual welfare, the political economy literature on elite persistence and social highlights the relevance of positional mobility from an institutional perspective. Although less abundant in economics than the literature on positional concerns, this literature highlights the institutional mechanisms through specific groups in society persist at specific points of the distribution of economic resources and, consequently, diminish the rate of intergenerational positional mobility in society.

A strand of this literature has focused on the theoretical analysis of the different mechanisms through which elites can capture institutions and implement economic policies that guarantee their persistence at the top of the distribution of economic resources. The seminal work in this area corresponds to Acemoglu and Robinson (2008). They emphasize under which conditions transformations of the formal institutional environment do not modify the distribution of economic resources in society due to the countervailing influence of *de facto* political power. In other words, they highlight how those at the top of the economic distribution can derail any change in the formal institutions that might reduce their persistence at the top by making use of the influence they have on the command of resources and, consequently, in their distribution. Albertus and Menaldo (2014)) provide empirical support to this hypothesis. The authors show that the relative political power of elites during transitions to democracy impacts

the level of inequality observed in the country once democracy is fully in place. In societies where economic elites had a solid political position during the transition to democracy, the implementation of democratic institutions had a smaller effect on the distribution of economic resources than in the societies where the elites were weakened at the moment of the transition.

These results have implications regarding positional social mobility. In the case of elites, they have a clear incentive to diminish the accession of "non-elite" members of society to the top of the distribution, as that would imply their displacement and a reduction in their political power. Consequently, less upward positional mobility is observed in societies where elites retain more political power. This would be particularly the case concerning transitions to and from the top of the distribution of economic resources, which are the positions occupied by the elite. Figueroa (2008); Rahman Khan (2012) and Bavaro and Patriarca (2022) discuss several of the different mechanisms through which elite members can curtail positional mobility. In particular, Figueroa (2008); Palma (2020) and Bavaro and Patriarca (2022) highlight the role of institutional arrangements that disincentivize the participation of challengers in the markets that provide membership to the elite. For example, Bavaro and Patriarca (2022) analyze the role of referrals in restricting the number of competitors for positions and how they benefit the members of the already existing network of elite members.

Acemoglu et al. (2017), derive another implication concerning the relationship between democracy (and inclusive institutions) and positional mobility. They show that positional mobility can be, under certain conditions regarding the distribution of perceived mobility, a destabilizing force for a democratic regime. For example, if the median voter expects to move upwards in the distribution and become part of the elite, she will prefer an institutional setup that guarantees that power remains in the hands of the elite in the future over a democratic arrangement. In contrast, when the median voter expects a political takeover by the elite

in the future, she is incentivized to maintain democracy if her perceived mobility does not guarantee her elite membership

Both implications highlight the importance of analyzing positional mobility patterns to understand better the institutional arrangement regarding the distribution of resources in a society. This is consistent with the emphasis of stratification economics on the hierarchical organization of society and how members of each echelon will deploy different resources to defend their position or climb in the social hierarchy (Darity Jr., 2022). Similarly, it is consistent with the sociological literature on Effectively Maintained Inequality (Lucas, 2017), which highlights how elite members modify the institutional landscape once a particular educational level has reached saturation to maintain privileged access to the subsequent educational levels.

The existing evidence for Mexico highlights the regional or local character of the relationship between the capacity of elites to modify the institutional order to guarantee their persistence at the top. Garfias (2018) finds that during the Great Depression, local elites that experienced a negative shock in their resources due to the fall in commodity prices were more likely to face expropriations by the state and see their lands redistributed among the agricultural workers than local elites not engaged in the production of commodities. Thus, the particular regional conditions are crucial for the capacity of elites to maintain their position, providing another reason to focus on positional mobility using the regional distribution of economic resources as the reference group.

1.3 Measures of positional intergenerational mobility at the regional scale

The basis of positional mobility measures is the construction of a ranking based on the origin and current distribution of the outcome variable of interest and the corresponding copula that links both distributions. However, as Deutscher and Mazumder (2021) remark, a critical factor in determining if the concept of mobility being measured is positional mobility or something else is how the cut-off points of the rankings are defined. To see this, Deutscher and Mazumder (2021) define $F(\cdot)$ as a strongly positional mobility measure if it fulfills the following condition for the current $(y_{i,t})$ and origin $(y_{i,t-1})$ distributions of the outcome variable

$$F(g(y_{i,t-1}), h(y_{i,t})) = F(y_{i,t-1}, y_{i,t}) \; \forall g \text{ and } h \text{ monotonic and increasing} \tag{1.1}$$

In words, $F(\cdot)$ can only be considered a strongly positional measure of intergenerational mobility if changes in the marginal distributions of the outcome variable do not produce a change in the measured mobility. Thus, it only captures the changes in the copula joining both marginal distributions.

Consider then the case of intranational regions, the regional distributions of the outcome variable, and its distribution at the national level. The national distribution is the joint distribution of the regional ones. If the objective is to measure positional mobility at the national level, then measures of mobility defined over the national quantiles will fulfill equation 1.1. However, if the objective is to measure mobility at the regional level, that is no longer the case. The reason is that changes in the regional marginal distributions lead to changes in the position occupied by the origin or current household in the corresponding national distribution, even when those changes do not affect the copula of the regional distributions.

Thus, using nationally defined ranks to assess mobility patterns at the regional level, as Chetty et al. (2015) and subsequent literature do, does not fulfill the condition in equation 1.1. Consequently, those results cannot be interpreted as positional mobility but instead as a hybrid measure of positional and structural mobility.

The solution to this problem is to rank the members of each region separately. In other words, to produce rankings based on the regional and not the national distribution. By doing that, changes in the regional marginal distribution will not affect the regional copula linking the positions between the origin and current distributions. This means that mobility measures calculated using the regional distributions and ranks as support will fulfill the condition in equation 1.1, allowing them to measure positional mobility correctly. Moreover, and as explained in the previous section, defining the regional distribution as the reference group upon which ranks are calculated is consistent with the findings on positional concerns and political economy that deal with positional mobility.

Besides this decision on the reference population, I will rely exclusively on mobility measures based on rankings in the following analysis. Specifically, I will employ the rank-rank persistence coefficient, transition matrices, and several indexes that collapse the information in a transition matrix into a synthetic indicator As Nybom and Stuhler (2017) show, this type of measures of intergenerational mobility is more robust to life-cycle bias than traditional measures of intergenerational persistence, such as the intergenerational income elasticity. Moreover, it allows me to compare the findings of this paper with those of the previously existing literature for Mexico, specifically with those from Delajara and Graña (2018); Monroy-Gómez-Franco and Corak (2019); Monroy-Gómez-Franco and Vélez-Grajales (2021) and Delajara et al. (2022). In order to compare the effects of choosing the regional distribution instead of the national one, I will estimate the same battery of indicators using both rankings.

That is, rankings constructed with regional thresholds and rankings constructed using the national thresholds.

The rank-rank persistence coefficient measures the average degree of persistence at a specific rank of the outcome distribution. Since the work by Chetty et al. (2014, 2015) it has become a common measure to compare mobility and persistence patterns across regions of a country, as it summarizes the persistence levels implied by the copula linking the origin and current distributions (see, for example, (Connolly et al., 2019; Corak, 2019; Heidrich, 2017; Acciari et al., 2022). Defining $R_{i,t}$ as the rank of individual i in the relevant distribution of the outcome at time t, and $R_{i,t-1}$ as the rank of the same individual in the outcome distribution corresponding to time $t-1$, the rank-rank persistence coefficient is the β coefficient of the following regression.

$$R_{it} = \alpha + \beta R_{i,t-1} + u_{i,t} \tag{1.2}$$

As Hertz (2008) shows, the persistence coefficient can be decomposed into two elements: one corresponding to the share of the persistence observed within the different groups that compose the population, and another corresponding to the persistence element arising from the differences between said groups. Formally, define $\hat{\pi_r}$ as the share of the total population that inhabits in region r, $\hat{\beta}_r$ as the estimate of the national rank persistence coefficient among members of region r, $\hat{\gamma}$ the region-size-weighted between group regression coefficient, the estimate variance of the origin rank R_{t-1} among members of region r as $\hat{\sigma}^2_{R_r,t-1}$, the estimate of the variance in the origin rank at the national level as $\hat{\sigma}^2_{R,t-1}$, and the estimates of the regional and national means of the current and origin rank as $\bar{R}_{t,r}$, $\bar{R}_{t-1,r}$ and $\bar{R}_t$, $\bar{R}_{t-1}$. Then, as Hertz (2008) shows, the coefficient of equation 1.2 estimated for the whole national sample can be exactly decomposed as follows

$$\hat{\beta} = \sum_{r=1}^{5} \hat{\pi}_r \left(\hat{\beta}_r \frac{\hat{\sigma}^2_{y_{r,t-1}}}{\hat{\sigma}^2_{y_{t-1}}} \right) + \hat{\gamma} \frac{\sum_{r=1}^{5} \hat{\pi}_r (\bar{y}_{t-1,r} - \bar{y}_{t-1})^2}{\hat{\sigma}^2_{y_{t-1}}} \tag{1.3}$$

The first term represents the within-region component, while the second represents the between-group component of such persistence. This decomposition can be interpreted in terms of structural and positional intergenerational mobility for the case of regional analyses on intergenerational persistence. If the marginal distributions of the different regions are equal, the between-region component should not contribute anything to the persistence at the national level, i.e $\hat{\beta}$ is determined only by positional mobility. If that is not the case, then $\hat{\beta}$ is at least partially determined by differences in structural mobility across the different regions. Notice that if ranks are defined using only the regional distribution as a reference, the second component should be equal to zero.

A limitation of the rank-rank persistence coefficient is that it cannot capture non-linearities in the persistence rate across different ranks of the origin distribution of the outcome variable. This limitation can be supplemented by employing transition matrices, which characterize the persistence rate at different points of the distribution of the outcome variable at the origin. Transition probabilities are defined as the conditional probability that a person with origin in rank j $(O = j)$ reaches rank z in adulthood $(C = z)$, where the maximum rank is k, which conventionally is set to be $k = 5$ (in other words, the ranks correspond to quintiles). Formally, this probability $P(C = z|O = j)$ is defined as a function of the population with origin in rank j (N_j) and the share population with origin in rank j and in rank z in adulthood $(N_{z,j})$:

$$\Omega_{z|j} = P(C = z|O = j) = \frac{N_{z,j}}{N_j} \text{ for } j = 1...k \text{ and } z = 1...k \tag{1.4}$$

These transition probabilities can be arranged in a matrix to characterize the mobility

patterns at different parts of the relevant distribution. This is the transition matrix, $M_{o,c}$, which, assuming $k = 5$, is formally defined as follows:

$$M_{o,c} \equiv \begin{bmatrix} \Omega_{1|1} & \dots & \Omega_{5|1} \\ \vdots & \vdots & \vdots \\ \Omega_{1|5} & \dots & \Omega_{5|5} \end{bmatrix} \tag{1.5}$$

From the information in the transition matrix, it is possible to construct indexes that summarize both the degree of fluidity and immobility in the society from which the transition matrix was calculated. Among these indexes, In this paper, I consider the Prais-Shorrocks index (Prais, 1955; Shorrocks, 1978), and the immobility index (Symeonaki and Stamatopoulou, 2020).

The Prais-Shorrocks index is a synthetic measure that captures the degree of mobility observed in the transition matrix $M_{o,c}$ as a function of the deviations from the rank persistence depicted by the main diagonal of $M_{o,c}$. Defining $\text{tr}(M_{o,c})$ as the trace of matrix $M_{o,c}$, it is possible to define the Prais-Shorrocks index, $PS_{M_{o.c}}$, as follows:

$$PS_{M_{o.c}} = \frac{1}{k-1}\big(k - \text{tr}(M_{o,c})\big) \tag{1.6}$$

If $PS_{M_{o.c}} = 0$, it implies that there are no deviations from the main diagonal, so there is complete persistence at the origin positions. In contrast $PS_{M_{o.c}} = 1$, implies that the transition probabilities corresponding to each origin rank are the same. In other words, the probability of reaching any rank in the transition matrix is independent from the rank of origin.

The immobility index ($IM_{M_{o.c}}$), as defined by Symeonaki and Stamatopoulou (2020), is

the share that rank persistence represents out of the total number of ranks considered in the transition matrix. This ratio is the rate at which a member of society is expected to persist at any given rank. Formally this is

$$IM_{M_{o.c}} = \frac{\text{tr}(M_{o,c})}{k} \tag{1.7}$$

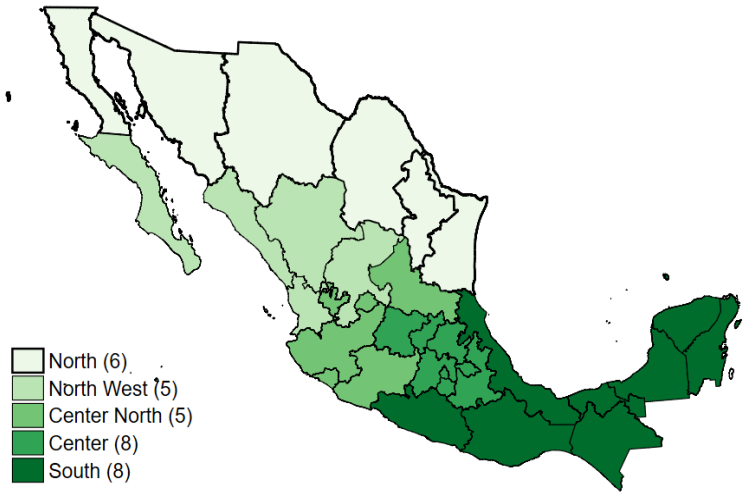
As in the case of the rank-rank coefficient, I estimate the transition matrices and the corresponding indexes using both nationally and regionally defined ranks in order provide a comparison between the positional mobility measure and the measure commonly used in the literature (see, for example Monroy-Gómez-Franco and Corak (2019); Monroy-Gómez-Franco and Vélez-Grajales (2021) and Delajara et al. (2022)).

1.4 Data

Although panel data represents the ideal data for intergenerational social mobility analysis, such datasets are not usually available in middle and low-income countries. Mexico is one of the countries in this situation, as there are no panel datasets that capture intergenerational information. An alternative is the use of retrospective surveys that recover information about the conditions of the household of origin of the interviewee. In this paper, I employ a database with these characteristics, the ESRU-EMOVI Social Mobility in Mexico Survey 2017 (EMOVI-2017). The EMOVI-2017 is a probabilistic survey representative of the Mexican non-institutionalized population between 24 and 65 years old. The survey includes a rich set of questions on the conditions of the household inhabited by the person when 14 years old and the conditions of the currently inhabited household. In particular, the survey recovers information on the types of durable goods, household appliances, and services available in both households. Similarly, the survey recovers information on the educational attainment and occupation of both the respondents and the parents.

EMOVI-2017 is the first survey with retrospective information with statistical representativeness at national and regional levels. This characteristic enables the comparative analysis of intergenerational mobility patterns between the country's regions. However, the regional representativeness of the EMOVI-2017 sample refers to a supra-state regional division of the country that divides the country into five regions constituted by different states. The regionalization of the country was based on the shared economic characteristics of the different states inside each region in 2017, adding geographical contiguity to this criteria (Centro de Estudios Espinosa Yglesias, 2019). This sample characteristic limits my capacity to disaggregate the analysis to perform state-level comparisons. Figure 1 shows the composition of the regions in the sample. The list of states that form each region are shown in Appendix A.

Figure 1.1. Geographical distribution of the different regions



It is important to note that the regional representativeness of the survey refers to the current region inhabited by the respondent, although it has information on the region of origin of the respondent. Due to the sample size, it is not possible to analyze in depth the determinants of the selection-into-migration at the regional scale, nor to study the impacts of

internal migration on social mobility. As my interest is to analyze the patterns of intra-regional mobility across the country's different regions, I opt only to consider the population that, in 2017, remained in their region of origin. For this reason, my results should be interpreted as concerning the population that never migrated outside their region of origin.

The use of retrospective data for intergenerational analysis poses several. A first concern is the presence of recall bias. Recall bias arises from incorrect answers by the respondent regarding the living conditions of the household in which they lived when they were 14 years old. The questionnaire of EMOVI-2017 seeks to attenuate this bias in two ways. Firstly, the questions on the availability of household appliances and services use as the reference point for the respondent the situation when the person was 14 years old. There is ample evidence from neurosciences and memory research that strongly suggests that the events and situations that occurred during adolescence tend to be remembered more precisely than those that occurred at other stages in life[5]. By situating the reference point at this life stage, the survey seeks to improve the quality of the information recovered through the retrospective questions. A second strategy employed by EMOVI 2017 to reduce the effects of recall bias is that the questions on the durable goods and services available in the origin household only ask about access or ownership and do not require a full description or valuation of them. This reduces the demand on the memory of the respondent.

Another concern is life-cycle bias, which can exist if the interviewee's age and that of her parents at the reference point are very different. To attenuate this bias, I restrict my analysis to the cohort between 30 and 60 years old in 2017. This cohort would also represent individuals who are at their prime working age. As shown in Table 1.1, parents of this cohort

[5]On this subject see, among others, Janssen and Murre (2008); Murre et al. (2013); Kilford et al. (2016); Hartshorne et al. (2018).

had, on average, a very similar age at the reference point compared to the respondents in 2017. This helps to attenuate the concerns about life cycle bias affecting my estimations. Table 1.1 also shows two well-known facts about Mexican history during the last 40 years: an increase in the average educational attainment as the public education system expanded and an increase in the urbanization rate across all regions (see de la Torre and Vélez-Grajales (2016) for a further discussion on this facts). By focusing only on one cohort, I can abstract from the particular effects of this historical development on intergenerational social mobility that was not experienced by older cohorts.

As the questions about the goods and services were available in the origin and current households only recover ownership information, it is impossible to estimate a monetary valuation of the household wealth or income at both points. However, the information about ownership and access to the different goods and services implies the existence of a latent variable for both households that represents the household's economic resources and economic status. This latent variable can be estimated through a household asset index that summarizes the implicit information regarding the household's economic status in the multiple binary variables that describe the access and ownership of particular goods and services. This approach to the estimation of economic resources is frequently used in development economics (see, for example, Filmer and Pritchett (2001); Filmer and Scott (2012); McKenzie (2005); Poirier et al. (2020); Wendelspiess-Chávez-Juárez (2015)). It also has been previously used to estimate social mobility patterns (Torche, 2015; Campos-Vázquez and Medina-Cortina, 2019; Vélez-Grajales et al., 2018b; Monroy-Gómez-Franco and Corak, 2019; Monroy-Gómez-Franco and Vélez-Grajales, 2021; Delajara et al., 2022).

Table 2.6 describes the goods and services considered to construct both the origin and current household economic resources indexes. As all the assets considered are binary variables,

Table 1.1. Descriptive statistics
(30-60 years old sample)

	National	North	North West	Centre North	Centre	South
Age of the respondent	41.75	41.86	41.97	41.66	41.69	41.79
(Mean)	(0.112)	(0.196)	(0.190)	(0.269)	(0.234)	(0.186)
Age of parents at reference	41.88	41.40	42.98	42.47	41.36	42.42
(Mean)	(0.108)	(0.228)	(0.356)	(0.311)	(0.227)	(0.234)
Women	0.540	0.512	0.528	0.540	0.547	0.551
(Proportion of the population)	(0.009)	(0.016)	(0.014)	(0.012)	(0.017)	(0.019)
Indigenous	0.125	0.042	0.039	0.052	0.096	0.299
(Proportion of the population)	(0.011)	(0.007)	(0.008)	(0.010)	(0.011)	(0.031)
Urban community of origin	0.665	0.836	0.467	0.626	0.773	0.447
(Proportion of the population)	(0.017)	(0.019)	(0.055)	(0.047)	(0.024)	(0.037)
Urban community at the interview	0.864	0.949	0.784	0.626	0.918	0.742
(Proportion of the population)	(0.013)	(0.012)	(0.041)	(0.047)	(0.020)	(0.034)
Average school years the parents	4.40	5.06	4.19	4.23	5.05	3.03
(Mean)	(0.098)	(0.130)	(0.275)	(0.252)	(0.239)	(0.191)
Average school years of respondent	9.80	10.31	9.83	9.30	10.41	8.74
(Mean)	(0.096)	(0.221)	(0.221)	(0.226)	(0.239)	(0.222)
Regional population	1	0.162	0.066	0.142	0.392	0.238
(Proportion of national population)		(0.016)	(0.010)	(0.017)	(0.033)	(0.020)

Note: Standard errors in parentheses. Sample weights employed. The North region consists of Baja California, Sonora, Chihuahua, Coahuila, Nuevo León and Tamaulipas; North West consists of Baja California Sur, Sinaloa, Nayarit, Durango and Zacatecas; the Center North region is form by Jalisco, Aguascalientes, Colima, Michoacán and San Luis Potosí; the Center region is formed by Guanajuato, Querétaro, Hidalgo, Estado de México, ; Mexico City, Morelos, Tlaxcala, and Puebla; the South region is formed by Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán and Quintana Roo. The reference point corresponds to the moment when the respondent was 14 years old. Indigenous population is identified as those with at least one parent who spoke an indigenous tongue. Urban communities are defined as those communities with more than 1,500 inhabitants.

the most suitable methodology to estimate the latent variable representing the household's economic status is multiple correspondence analysis (MCA). In contrast with other methodologies such as Principal Component Analysis (PCA), MCA relies on the relative frequencies of the different variables instead of using euclidean distances to estimate the latent variable representing the household's economic status[6]. Tables B1 and B2 show the coordinates

[6]For a more extensive discussion of the method see Monroy-Gómez-Franco (2022)

Table 1.2. Goods and services employed to construct the economic resources index

Household asset/service	Household at 14 years old	Current Household	Household asset/service	Household at 14 years old	Current Household
Potable water	✓	✓	Water heater	✓	✓
Stove	✓	✓	Domestic service	✓	✓
Electricity	✓	✓	Cellphone		✓
Refrigerator	✓	✓	Toaster	✓	✓
Washing machine	✓	✓	Owner inhabited dwelling	✓	✓
Landline	✓	✓	Tablet		✓
Computer	✓	✓	Vacuum cleaner	✓	✓
DVD Player / Cassette recorder	✓	✓	Video-game console	✓	✓
Microwave	✓	✓	Car	✓	✓
Television	✓	✓	Dirt floor	✓	✓
Tractor		✓	Toilette facilities	✓	
Cable TV	✓	✓	Overcrowded household	✓	✓
Owner of another dwelling	✓	✓	Bank account	✓	✓
Owner of commercial premises	✓	✓	Credit card	✓	✓
Non-agricultural land		✓	Internet		✓

assigned to each good or service of the origin and current index and their contribution to their explained inertia. This allows to identify which goods and services play a larger role in differentiating the economic status of households.

To produce robust results, I generate three series of rankings for the origin and current distribution of economic resources, dividing the distribution into 50, 20, and 10 quantiles for the estimation of the rank-rank persistence coefficient and into quintiles for the transition matrix-based indicators.

1.5 Positional mobility across Mexican regions

Table 1.3 shows the rank-rank persistence coefficient estimated over nationally and regionally defined ranks. As mentioned, the coefficient estimated using the national ranks corresponds to a hybrid measure of positional and structural mobility, while the one estimated on the regional ranks corresponds to a positional mobility measure.

The results for the hybrid measure are in line with previous findings in the literature (Delajara and Graña, 2018; Vélez-Grajales et al., 2018b; Monroy-Gómez-Franco and Corak, 2019; Monroy-Gómez-Franco and Vélez-Grajales, 2021; Delajara et al., 2022): the south region has the highest degree of rank persistence in the country, with the center of the country as the following region. On the other extreme, the most mobile region is the northeast. This ordering is robust to using more coarse partitions of the sample for the estimation.

However, the positional mobility measure results indicate that this ordering is a product of the dynamics of the regions in terms of structural mobility and not those of positional mobility. In contrast with the results of the hybrid measure, the results from the positional mobility measure indicate that, with the exemption of the northwest region, the rest of the country's regions exhibit a very similar rate of intergenerational rank persistence. This challenges the findings from the existing literature, which suggested that intergenerational mobility in the south region was substantially lower than in the rest of Mexico. My results show that the south has the same intergenerational positional persistence as the north or center regions. Thus, it cannot be considered a region "less fluid" than the others. In that sense, the hybrid approach seems to be heavily influenced by the very different macrodynamics of the south of Mexico, as initially suggested by Monroy-Gómez-Franco and Corak (2019).

Besides the last element, there are two other results. The first is that, similarly to the hybrid mobility measure, the northwest region has the lowest degree of positional persistence. This suggests that the positional mobility dynamics of the region dominate the structural ones and, in that sense, play a more prominent role in determining the rank persistence than in other regions. The second result is that the levels of positional persistence observed across all regions are high, even considering the northwest case. This suggests a pattern of elite persistence similar to the one identified for Peru by Figueroa (2008) and for Chile by Palma (2020).

Table 1.3. Rank-Rank persistent coefficient by region
(Percentiles defined over regional and national distributions)

Region	Type of measure	50 quintiles	20 quintiles	10 quintiles
North	Positional	0.507 (0.027)	0.534 (0.025)	0.537 (0.026)
	Hybrid	0.507 (0.024)	0.508 (0.023)	0.502 (0.022)
Northwest	Positional	0.416 (0.028)	0.435 (0.032)	0.433 (0.031)
	Hybrid	0.431 (0.033)	0.433 (0.035)	0.432 (0.036)
Center north	Positional	0.511 (0.027)	0.555 (0.026)	0.544 (0.027)
	Hybrid	0.495 (0.026)	0.497 (0.026)	0.496 (0.027)
Center	Positional	0.564 (0.029)	0.594 (0.029)	0.591 (0.029)
	Hybrid	0.609 (0.029)	0.613 (0.028)	0.608 (0.029)
South	Positional	0.559 (0.026)	0.533 (0.025)	0.547 (0.025=
	Hybrid	0.633 (0.029)	0.639 (0.030)	0.644 (0.030)

Note: Clustered standard errors in parentheses. Cluster unit is the primary sampling unit. Positional row corresponds to estimations based on regionally defined quintiles, while hybrid corresponds to the nationally defined quintiles. Sample weights employed. The North region consists of Baja California, Sonora, Chihuahua, Coahuila, Nuevo León and Tamaulipas; Northwest consists of Baja California Sur, Sinaloa, Nayarit, Durango and Zacatecas; the Center North region is form by Jalisco, Aguascalientes, Colima, Michoacán and San Luis Potosí; the Center region is formed by Guanajuato, Querétaro, Hidalgo, Estado de México, ; Mexico City, Morelos, Tlaxcala, and Puebla; the South region is formed by Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán and Quintana Roo. The corresponding regresions are presented in tables C1-C6 of the appendix and they only consider individuals who remained in the same region they were when 14 years old.

The previous results suggest that Mexico is more homogenous than previously thought regarding positional mobility and that the existing literature are mixing the structural and positional mobility dynamics. To further investigate this issue, table 1.4 shows the decomposition of the national rank-persistence coefficient in its between and within components, following the exact decomposition proposed by Hertz (2008) and presented in equation 1.3. The results shown in table 1.4 correspond to the estimation of equation 1.2 using the ranking of 50 quantiles. As mentioned, the within region component corresponds to positional mobility of the region, whereas the between regions component correspond to the differences in structural mobility between regions.

Table 1.4. Decomposition of the national intergenerational persistence by region (Coefficients scaled by population share)

Region	Within-group regression	Between-group regression	Total	Positional component (share of total)	Structural component (share of total)
North	0.069	0.017	0.086	0.798	0.202
Northwest	0.025	-0.001	0.024	1.049	-0.049
Center worth	0.063	0.004	0.067	0.947	0.053
Center	0.207	0.018	0.225	0.918	0.082
South	0.120	0.105	0.225	0.531	0.469
National	0.483	0.144	0.627	0.771	0.229

Sample weights employed. The North region consists of Baja California, Sonora, Chihuahua, Coahuila, Nuevo León and Tamaulipas; Northwest consists of Baja California Sur, Sinaloa, Nayarit, Durango and Zacatecas; the Center North region is form by Jalisco, Aguascalientes, Colima, Michoacán and San Luis Potosí; the Center region is formed by Guanajuato, Querétaro, Hidalgo, Estado de México, ; Mexico City, Morelos, Tlaxcala, and Puebla; the South region is formed by Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán and Quintana Roo. The regressions only consider individuals who remained in the same region they were when 14 years old and are estimated using the national ranks. Results from the regr .

The decomposition results suggest that the major contributor to the persistence rates observed in Mexico is the degree of positional persistence in the country and not the differences in structural mobility between regions. Furthermore, the positional mobility component in all regions is the principal contributor to rank persistence, highlighting the need to study it separately from the structural mobility one. The regions where positional persistence plays

a higher role are the northwest, the center north, and the country's center. This implies that in those regions, structural mobility has not been a relevant force in determining the mobility patterns of those regions. In contrast, the south is where structural mobility has played a more significant influence, representing 46% of the total rank persistence in the region.

Table 1.5 shows the results for the Shorrocks-Prais and immobility indexes estimated for the transition matrices constructed using the national (hybrid) and the regional ranks (positional). As in the case of the intergenerational rank persistence, the results for the hybrid concept of mobility imply that the south is the region with the highest level of persistence. In contrast, the positional mobility measure rejects this result, suggesting a level of persistence similar to the rest of the country.

An advantage of analyzing the full transition matrices is that they allow to identify the presence of non-linearities in the intergenerational persistence rate across the different ranks of the origin distribution of economic resources. Figure 1.2 shows the persistence probabilities at $O = 1$ and $O = 5$. In other words, the transition probabilities are at the origin distribution's extreme quintiles. Panels a and b show the persistence probabilities for the case of the matrices constructed using the regional quintiles, while panels c and d show the corresponding probabilities for the matrices constructed using the national ranks

The figure provides more information regarding the divergence between the results from the positional and hybrid mobility measures. Concerning positional mobility, panels a and b of figure 1.2 show that persistence at the top and the bottom of the regional distributions is relatively similar across the country. Moreover, they also indicate that persistence at the top is higher than at the bottom. The latter also occurs for the hybrid mobility measure. With the exemption of the south region, the persistence at the top is higher than at the bottom

Table 1.5. Mobility measures based on transition matrices
(Percentiles defined over regional distributions)

Region	Type of measure	$PS_{M_{o.c}}$	$IM_{M_{o.c}}$
National	Positional	0.760	0.392
North	Positional	0.824	0.341
	Hybrid	0.844	0.325
North West	Positional	0.877	0.298
	Hybrid	0.848	0.322
Center North	Positional	0.819	0.345
	Hybrid	0.827	0.339
Center	Positional	0.772	0.382
	Hybrid	0.781	0.375
South	Positional	0.814	0.349
	Hybrid	0.701	0.439

Note: Clustered standard errors in parentheses. Cluster unit is the primary sampling unit. Positional row corresponds to estimations based on regionally defined quintiles, while hybrid corresponds to the nationally defined quintiles. Sample weights employed. The North region consists of Baja California, Sonora, Chihuahua, Coahuila, Nuevo León and Tamaulipas; Northwest consists of Baja California Sur, Sinaloa, Nayarit, Durango and Zacatecas; the Center North region is form by Jalisco, Aguascalientes, Colima, Michoacán and San Luis Potosí; the Center region is formed by Guanajuato, Querétaro, Hidalgo, Estado de México, ; Mexico City, Morelos, Tlaxcala, and Puebla; the South region is formed by Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán and Quintana Roo. The corresponding transition matrices are presented in tables D1 to D10 of the appendix and they only consider individuals who remained in the same region they were when 14 years old.

for all other regions.

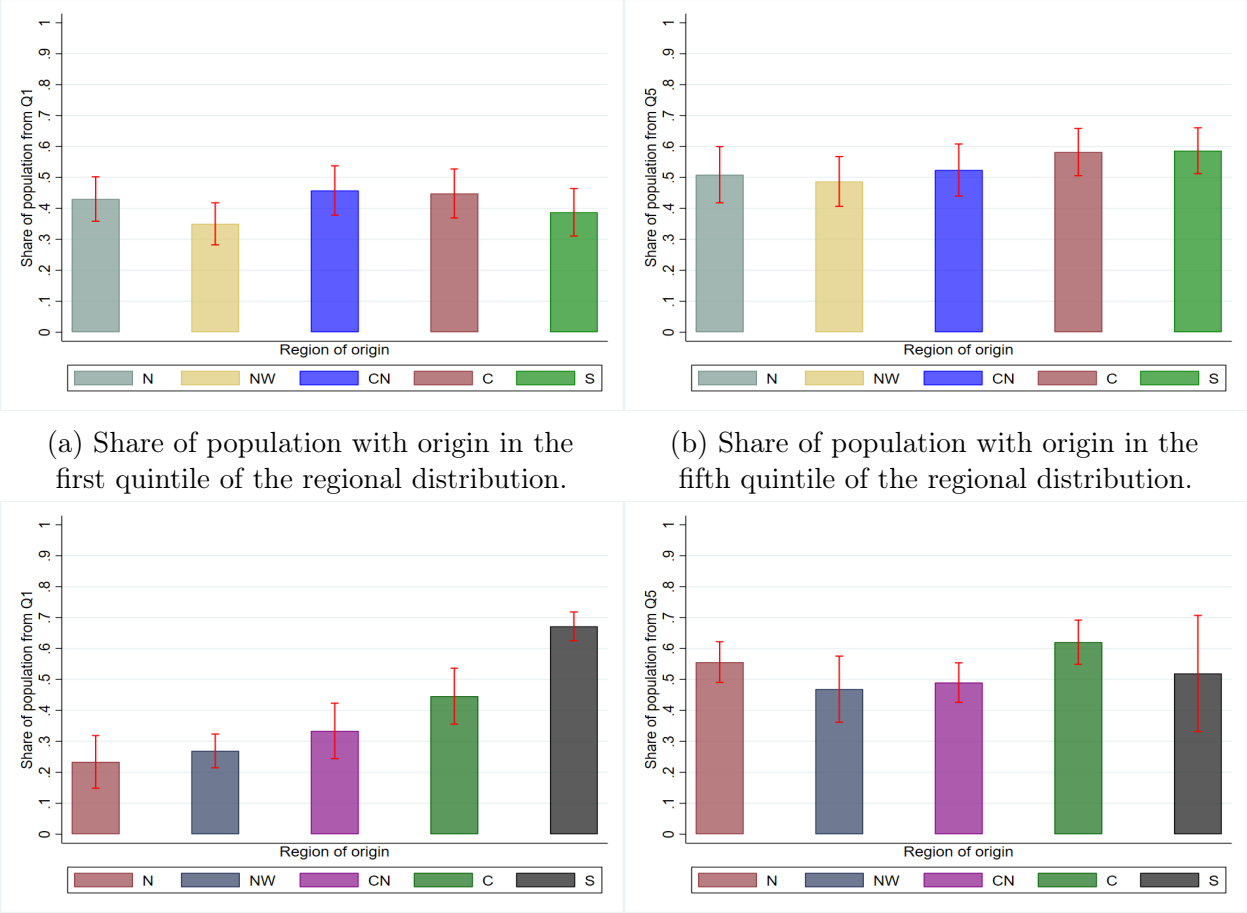
The comparison between the four panels of figure 1.2 highlights that the considerable persistence rate at the bottom of the south region's distribution of economic resources is not the fruit of a higher degree of positional immobility in that point of the distribution than in the rest of the country. In fact, the point estimate of positional mobility from the bottom for the south region is the second lowest in the country. This implies that the divergence of the south in terms of intergenerational economic mobility with respect to the rest of Mexico is associated with the lack of convergence of the region in terms of economic conditions. As Esquivel (1999); Sakikawa (2012) and German-Soto et al. (2020) show, convergence in terms of output stopped since the early 1980s. Consequently, most of the south region's sample has been exposed to this particular growth environment during most of their lives.

These results provide a more complex image of the Mexican regional landscape than the one produced by the existing literature. While Delajara et al. (2022) suggest that the south suffers a double burden (low intergenerational mobility and precarious economic conditions), my results show that, internally, the region has the same positional mobility rates as the rest of the country. Thus, it is not possible to assert that the region is by itself less mobile or more rigidly stratified than the other regions. This is in line with the evidence by Hausmann et al. (2021) that suggests that the causes of the precarious living conditions are in the aggregate performance of the region's economy and not in some characteristic of the region's society.

1.6 Final Remarks

The existing literature on intranational differences in intergenerational social mobility shows that differences in social mobility exist not only between countries but also within them

Figure 1.2. Persistence at the bottom and top quintiles of origin distribution



(a) Share of population with origin in the first quintile of the regional distribution.

(b) Share of population with origin in the fifth quintile of the regional distribution.

(c) Share of population with origin in the first quintile of the national distribution.

(d) Share of population with origin in the fifth quintile of the national distribution.

Note: Linearized standard errors shown in red. Sample weights employed. The persistence probabilities correspond to the sahre of the population with origin in quintile o who remained at the same quintile. The North region consists of Baja California, Sonora, Chihuahua, Coahuila, Nuevo León and Tamaulipas; North West consists of Baja California Sur, Sinaloa, Nayarit, Durango and Zacatecas; the Center North region is form by Jalisco, Aguascalientes, Colima, Michoacán and San Luis Potosí; the Center region is formed by Guanajuato, Querétaro, Hidalgo, Estado de México, ; Mexico City, Morelos, Tlaxcala, and Puebla; the South region is formed by Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán and Quintana Roo.

(Chetty et al., 2015; Heidrich, 2017; Connolly et al., 2019; Corak, 2019; Monroy-Gómez-Franco and Corak, 2019; Monroy-Gómez-Franco and Vélez-Grajales, 2021; Acciari et al., 2022; Delajara et al., 2022). The next phase in that research program is to understand that heterogeneity's determinants better. The first step in that direction, particularly relevant for

cases in which intergenerational panel datasets are not existing, is decomposing the observed intergenerational mobility rates into their structural and positional components.

The decomposition of intergenerational mobility rates into the structural and positional components allows the identification of the major contributor to the persistence patterns. The structural component corresponds to the differences in aggregate dynamics of the intranational regions. In contrast, the positional component consists of the differences in the probability that the inhabitants of a region exchange positions in the distribution between two generations. Therefore, properly accruing the influence of each component is crucial not only to gain a better understanding of the phenomenon but also to design the most appropriate policies to increase intergenerational mobility in a country..

In this paper, I use the Mexican case to illustrate the relevance of this decomposition for interpreting the differences in mobility rates at the subnational level. My results show that the higher intergenerational persistence rate observed for the south region of the country, previously documented by the literature (Vélez-Grajales et al., 2018b; Monroy-Gómez-Franco and Corak, 2019; Monroy-Gómez-Franco and Vélez-Grajales, 2021; Delajara et al., 2022) is a product of the large gap between the region and the rest of the country in terms of aggregate income. No significant differences are observed when estimating the positional mobility rate across Mexican regions. This implies that the results from the previous literature are primarily determined by the structural mobility component, whereas all the country's regions exhibit a similar rate of social fluidity.

The results of this paper shows the importance of decomposing the mobility rates in their structural and positional components, as it allows a better understanding of the main drivers of the intranational persistence patterns. This opens a new avenue of research in this

area, as it allows the exploration of each component separately, and consequently, a finer analysis of the determinants of the mobility patterns. Similarly, it allows to tie the political economy and the positional preferences literatures with a more robust empirical approach for the analysis of the hypotheses of those literatures in the realm of intergenerational mobility.

The results of this paper show the importance of decomposing the mobility rates in their structural and positional components, as it allows a better understanding of the main drivers of the intranational persistence patterns. This opens a new avenue of research in this area, as it allows the exploration of each component separately and, consequently, a more nuanced analysis of the determinants of the mobility patterns. Similarly, it allows tying the political economy and the positional preferences literatures with a more robust empirical approach for analyzing the hypotheses of those literatures in intergenerational mobility.

However, several limitations of this paper are relevant to the literature as a whole. First, the limited regionalization of the database prevents a deeper analysis of the difference in positional mobility across the Mexican states. This is relevant for other countries, as the primary determinant of the regional representativeness of a survey is the sample size and design, while themselves being determined by the funding available for the survey. Secondly, although the use of retrospective data allows circumventing the lack of intergenerational panel datasets, it also limits the type of mobility being assessed. Using an index of economic resources allows to identify mobility patterns in a dimension conceptually similar to the permanent income of the observed unit, but it limits the comparability with other studies that rely on income data.

Chapter 2

Shades of social mobility: Colorism, ethnic origin and intergenerational social mobility.

2.1 Introduction

As Roth (2016); Telles (2012); Sen and Wasow (2016) and Solís et al. (2019) argue, notions about "race" are usually constructed from different sets of immutable characteristics shared by a particular group of society. The set of characteristics and the role played by each in constructing the notion of race varies across societies and is heavily contingent on their historical development. However, as Dixon and Telles (2017) , skin tone is one of the few characteristics used across the globe to construct social classifications, establishing a hierarchy across the different tonalities. This hierarchization of skin tones with lighter skin tones at the top, also known as colorism (Dixon and Telles, 2017),), has become an area of active research in recent years as part of the efforts to understand the institutional under-

pinnings of economic inequality between and within the different groups that compose society[1].

However, an unexplored aspect is how colorism has a differentiated impact across societal groups with different ethnic backgrounds. In this paper, I contribute to filling this gap by studying the difference in intergenerational mobility patterns between the indigenous and non-indigenous Mexican populations, analyzing the role played by the variability in skin tones inside each group. For this analysis, I employ a composite sample based on two recent surveys with the same sampling frame and questionnaire design. I obtain three main results based on rank-rank regressions, transition matrices, and quantile rank regressions. First, the indigenous population faces a higher degree of intergenerational persistence in the economic dimension and remains concentrated at the bottom of the distribution of economic resources. Secondly, the indigenous population faces less upward mobility from the bottom of the distribution but similar rates of persistence to that of the non-indigenous population at the top. Furthermore, there is a skin tone gradient for the indigenous and non-indigenous populations, where lighter skin tones tend to remain at a higher rank in the distribution of economic resources than darker skin tones. Finally, the mobility gap between the indigenous and non-indigenous populations is larger than the differences in the patterns of social mobility by skin tone observed inside each group.

This paper contributes to several literatures in economics. First, it contributes to the literature on intra-national differences in intergenerational mobility by providing evidence of substantial differences in intergenerational persistence rates across groups of different ethnic origins. Secondly, it contributes to the economic analysis of colorism by providing evidence that the skin tone differences translate into economic outcome differences of an intergenera-

[1]These institutional underpinnings form the basis of the intergroup inequalities and hierarchies that are present in society. In other words, they constitute the stratification regime of the society, following Darity Jr. (2022).

tional dimension and that they are present across ethnic groups. Thirdly, it contributes to the nascent literature on stratification economics by highlighting how immutable characteristics such as skin tone and ethnic origin are used to construct a stratification structure in the society that generates differences in economic outcomes between and within social groups.

In the next section, I discuss the results of research on the effects of colorism on economic outcomes, highlighting the differences in the conceptualization of the role of skin tone in American and Mexican society. In the third section, I present the data employed in the analysis, describing in detail the characteristics of the composite sample and the construction of the indicator of economic resources used in the analysis. Next, I discuss the methods employed to characterize and analyze mobility patterns, considering how to capture inter-ethnic and intra-ethnic differences. In the fifth section of the paper, I present the results of this analysis, highlighting the differences in the information provided from the rank-rank regressions and the non-linearities identified in the transition matrices and quantile regressions. Finally, the paper concludes with some final remarks regarding the limitations of the analysis and possible areas of future inquiry.

2.2 The analysis of colorism in economics

Most economics research on colorism has focused on studying the US case. This focus, as Dixon and Telles (2017) point out, imposes a notion of colorism independent from that of racism, as the latter is associated more commonly in the US with the origin of the person. For this reason, the economics literature has focused on analyzing differences in the skin tone among members of the same "race." Specifically, the literature has focused on how colorism produces differences in the labor outcomes of African Americans and among the immigrant

populations in the US.

In the case of African-Americans, Goldsmith et al. (2006, 2007) and Reece (2021) find that the race wage gap between African Americans and whites diminishes as the lightness of the skin tone of the African American worker increases. Darity Jr. et al. (2002) find the same pattern among the Latino population. Devaraj and Patel (2017) find a negative association between darker skin tones and the probability of being self-employed. The authors also find that a darker skin tone is associated with a lower labor income, conditional on being self-employed. Similarly, Diette et al. (2015) found that those with a darker skin tone tend to suffer larger adverse effects on their mental health due to unemployment. However, as Akee and Yuksel (2012) and Kreisman and Rangel (2015) indicate, between 1980 and 2000, the wage gap between darker and lighter tone women decreased, while for men, it remained constant.

Although the effects of colorism on the intergenerational mobility patterns of the American Population have not been analyzed, there is a sizable literature on the differences in intergenerational mobility in income between African Americans and Whites in the US. The results from this literature point out that African Americans experience less upward mobility in income than Whites (Chetty et al., 2019; Bhattacharya and Mazumder, 2011)) and, in general, more intergenerational persistence. Davis and Mazumder (2020) show that this result holds for all the regions in the US. As Hertz (2008) shows, the contribution of this racial gap in social mobility represents 40% of the total intergenerational persistence observed in the US. When considering the Latino population, both Chetty et al. (2019) and Davis and Mazumder (2020) find that their social mobility patterns are more similar to those of the White than to those of the African American population.

This paper contributes to this literature by providing evidence of the differences in social mobility patterns between the ethnic groups of the same country. By analyzing differences in intergenerational mobility patterns by ethnic origin and skin tone, I provide evidence regarding the determinants of intragroup differences in social mobility, something that is explored at length by Davis and Mazumder (2020) and Chetty et al. (2019) for the US case. Furthermore, I find evidence that, as in the US, in Mexico, the ethnic group that has been found to face more disadvantages across different markets also tends to remain across generations at a lower point of the distribution of economic resources than the other ethnic groups.

In the case of the immigrant population, Hersch (2008) finds that a darker skin tone is associated with a lower wage among the immigrant population. However, Rosenblum et al. (2016) find that the case of Latin American migrants drives this result. Moreover, the negative association between darker skin tones and lower wages is only observed among Latin American immigrants, not immigrants from other regions. Furthermore, Dávila et al. (2011) analyze the specific case of Mexican immigrants, finding that those with lighter skin color are more likely to insert themselves in white-collar occupations than those with darker skin tones. Finally, Han (2020) analyses the intragenerational occupational mobility of immigrant workers, comparing their current occupational status with the one they had in their country of origin. The author finds that all immigrants suffer a penalization in occupational terms, as they are currently employed in occupations with a lower status than the one they had in their country of origin. However, the gap is a function of the skin tone of immigrants, being the smallest for light-skinned immigrants.

However, in the Latin American context, skin tone plays a different role in the construction of group hierarchies. As Telles (2012); Telles and Paschel (2014) and Dixon and Telles (2017) point out, the ideas of race and skin tone overlap, making skin tone one of the key immutable

characteristics upon which group inequalities are constructed. In contrast to the US, most Latin American stratification regimes did not rely on formal institutions to differentiate across skin tones but were sustained through informal mechanisms. The absence of formal stratification institutions based on skin tone can be attributed to the role of the "Mestizaje ideology."

This ideology, as Segato (2010) points out, hinges on the notion that the majority of the population is the result of the mixing between the Spaniards and indigenous inhabitants. For that reason, formal systems that discriminate based on skin tone lack legitimacy. As Dixon and Telles (2017) document, this ideology formed the basis of the nation-building effort in the continent during the XX century. Saldívar (2014) and Navarrete (2016) show that Mexico was no exception. In the Mexican case, this ideological framework allowed society to interpret the differences between the indigenous and the non-indigenous populations as arising from cultural differences and not from a difference in treatment based on the ethnic origin of the person (Saldívar, 2014; Telles and Torche, 2018; Villarreal, 2010).

Both tenets, that skin tone differences did not play a role in generating differences in economic outcomes and that the inequalities between indigenous and non-indigenous were due to cultural differences, have been recently questioned in the social sciences literature in general and economics in particular [2]. González de Alba (2010); Castañeda-Navarrete (2013) and Canedo (2018) find that poverty incidence affects the indigenous population disproportionately. For example, Canedo (2018) finds that, in 2014, 67% of the indigenous population who lived in urban communities were multidimensionally poor, while the incidence

[2]It is worth mentioning that economics is a late comer to this issue in the Mexican case. Research from political science by Aguilar (2011, 2013) had identified that voters prefer candidates with lighter skin tones, holding all else constant. Similarly, sociological research by Flores and Telles (2012); Martínez Casas et al. (2014) and Villarreal (2010) identified that a darker skin tone is associated with lower educational attainment and occupational status. These findings have been confirmed by recent work in both areas by Solís and Güémez (2020); Blanco-Bosco (2020) and Campos-Vazquez and Rivas-Herrera (2021)

among the non-indigenous urban population was 40%. Among the rural population, 81% of the indigenous rural population is poor, while only 54% of the non-indigenous population is.

The higher poverty rates reflect a series of systematic disadvantages faced by the indigenous population. For example, Serván-Mori et al. (2022) show that nearly 33% of the gap in access to maternal medical care between the indigenous and the non-indigenous groups cannot be explained by observable factors, pointing to systemic discrimination against the indigenous population. Similarly, Canedo (2019); Cano-Urbina and Mason (2016), and Aguilar-Rodriguez et al. (2018) find a systematic wage penalty associated with speaking an indigenous tongue. In a more recent study, Arceo-Gómez and Torres (2021) disentangle the effect of speaking an indigenous tongue and identifying as a member of an indigenous group. The authors find that after controlling for self-selection into identification as indigenous, there is a wage penalty to the indigenous identity of nearly 8% of the wage. Except for Canedo (2019), the other studies find that although observable characteristics such as the education level explain most of the gap, about 20%-30% remains unexplained.

This paper contributes to the literature by providing the first set of estimates on intergenerational social mobility in economic resources for the indigenous population in Mexico. Using two surveys with a shared sample frame and questionnaire design, I overcome the sample size restriction that prevented past efforts to analyze the subject. Specifically, I estimate transition matrices and intergenerational rank coefficients for the indigenous population comparable to those of the non-indigenous population. My results suggest that the disadvantages faced by the indigenous population translate into substantially higher rates of intergenerational persistence at the bottom of the distribution of economic resources.

On the second issue, the existence of systematic differences in economic outcomes associ-

ated with differences in the skin tone of the population, experimental research has shown that dark-skinned women tend to receive fewer job callbacks than their light-skinned peers, while this is not observed in men (Arceo-Gómez and Campos-Vázquez, 2014). Similarly, Arceo-Gómez and Campos-Vázquez (2019) show that employers who advertised jobs using discriminatory language (based on appearance and gender) had a higher callback rate but reduced their callback rate for women with a dark skin tone. Furthermore, Campos-Vázquez (2021) documents the presence of this *preference for whiteness* (Goldsmith et al., 2007) in the escort service market as light-skinned escorts charge higher prices than their peers with a different tonality. Similarly, Reeskens and Velasco-Aguilar (2021) find that light-skinned Mexicans have higher educational achievement and work in higher-ranking occupations than the rest of the population. Finally, Hernández-Trillo and Martínez-Gutiérrez (2022) show in a field experiment that dark-skinned persons face more barriers than light-skinned persons when applying for commercial credits.

Campos-Vázquez and Medina-Cortina (2019); Rojas-Lomelín et al. (2019) and Monroy-Gómez-Franco and Vélez-Grajales (2021) find that dark-skinned individuals experience higher intergenerational persistence at the bottom of the distribution of economic resources across all regions of the country. Regarding the intra-regional variability of this result, Baldomero-Quintana et al. (2022) find evidence that dark-skinned individuals inhabit more frequently in the vicinity of localities that in colonial times were segregated. They find lower amounts of human capital in those areas and higher intergenerational persistence. Both findings suggest institutional persistence since colonial times affecting the dark-skinned population. Campos-Vázquez and Medina-Cortina (2018) show that teenagers have interiorized these lower prospects regarding economic outcomes. They find that expectations about success in life fall among dark-skinned female teenagers when the prevalence of light-skinned Mexicans among the elite is made salient.

In two recent papers, Monroy-Gómez-Franco et al. (2022) and Plassot et al. (2022) analyze the role played by skin tone variability in the level of inequality of opportunity in Mexico[3]. In both studies, the authors find that skin tone and indigeneity significantly contribute determining inequality of opportunity, each contributing to it separately. This suggests that colorism in Mexico does not overlap completely (Telles, 2012) with the traditional ethnic divide between indigenous and non-indigenous populations.

By exploring this hypothesis regarding intergenerational mobility patterns in economic resources, this paper contributes to two different kinds of literature. On the one hand, it contributes to the nascent literature on stratification economics by analyzing how intergroup inequalities defined upon an immutable characteristic (skin tone) can produce substantial inequalities among the members of different ethnic groups. On the other hand, it contributes to the literature focused on understanding how inequality of opportunity is constructed, and the role played by ethnic origin and skin tone as circumstances that affect a person's life outcomes.

2.3 Data

I employ as the primary data source for this paper a composite sample constructed by pooling the information from the samples of two nationwide household surveys explicitly designed for the study of intergenerational social mobility in Mexico: the 2017 ESRU Social Mobility Survey of the *Centro de Estudios Espinosa Yglesias* and the 2016 Module of Intergenerational Social Mobility of the *Instituto Nacional de Estadistica y Geografía*. Each survey's sample

[3]Inequality of opportunity is defined as the share of observed inequality in economic resources explained by differences in circumstances (Ferreira and Gignoux, 2011; Ferreira et al., 2011). Circumstances are defined as variables outside the control of the individual that affect the outcomes of a person (Roemer, 1998).

is representative of the Mexican non-institutionalized population between 25 to 64 years old, consisting of 17,665 observations in the first and 25,634 observations in the second. Pooling together both surveys allows obtaining a large sample size to study the differences in the patterns of intergenerational economic mobility between Mexico's indigenous and non-indigenous populations. It is worth mentioning that Delajara et al. (2022) previously used this pooled sample to estimate intergenerational social mobility patterns at the state level.

The pooling of the samples is possible because the surveys share both questionnaire and sample designs. The questionnaires of the surveys share a battery of retrospective questions on the availability of durable goods and access to public utilities and financial services. In both surveys, the reference point of the retrospective questions corresponds to the moment the interviewee was 14 years old. Moreover, and crucially for my interests, the surveys use the same strategy to retrieve information on the skin tone of the respondent and their ethnic belonging. In the case of skin tone, the surveys ask the person to compare their perceived skin tone with a series of tones in a color palette of 11 different tones. The reference color palette corresponds to the one Telles, ed (2014) designed to capture the variability of skin tonalities in Latin America[4]. In the case of ethnic belonging, both surveys ask if the respondent's parents spoke an indigenous tongue. In the case of MMSI 2016, this is supplemented by a set of self-reported measures regarding ethnicity. Throughout the rest of this paper, I will refer to the population with at least one parent that spoke an indigenous tongue as the *indigenous population*, unless otherwise stated.

[4]There is a growing literature comparing the different scales designed for the measurement of skin tones in surveys. Campbell et al. (2020) show that regardless of the scale adopted if the measurement is taken by the interviewer, her characteristics will influence the tonality assigned to the person. Moreover, Gordon et al. (2022) show that skin tone measurement across different scales is consistent. The authors also show a positive correlation between the colorimeter "objective" measures and those from tonality scales such as PERLA. An advantage of the PERLA scale is that it allows capturing a higher variability of lighter skin tones than, for example, the Masey-Martin scale (Massey and Martin, 2003).

Table 2.1 shows the relative population shares and general characteristics of the samples from each survey and the characteristics of the composite sample. The samples register the same average parental and respondent years of schooling, sex structure, and population share with parents who spoke an indigenous tongue. There is a minor difference in the urban population size across samples, as the EMOVI 2017 characterizes the Mexican population as more urban both in the reference moment and in the present than the MMSI 2016. Table 2.2 shows the same exercise but for the case of the 11 tones of the PERLA skin tone palette. Again, the samples reflect a very similar population composition regarding skin tone distribution. In order to maximize the sample size of each tonality and following previous research on the distribution of skin tones in Mexico, I opt to collapse the 11 tones into three groups of skin tones. Skin tone group 1 (or $c = 1$) corresponds to PERLA tones 1-3, skin tone group 2 (or $c = 2$) corresponds to PERLA tones 4-6 and skin tone group 3 ($c = 3$) corresponds to PERLA tones 7-11[5]. The population share of the resulting categories is shown in the lower part of table 2.2.

The first result of the analysis is to produce the skin tone distribution of the indigenous and non-indigenous Mexican population, which figure 2.1 shows. A clear element is that the skin tone distribution of the indigenous population tilts towards the darker tones in the palette, but that does not mean there is no variability in skin tones among this population. The figure shows that the variability in skin tones among the indigenous and non-indigenous populations is relatively similar.

An essential element to consider is the sample composition in terms of the number of observations corresponding to each subgroup of interest. As the indigenous population in Mexico represents 10% of the total population, this is particularly important as there is a

[5]See Campos-Vázquez and Medina-Cortina (2019); Monroy-Gómez-Franco and Vélez-Grajales (2021)

Table 2.1. Descriptive statistics

	Composite Sample	MMSI 2016	EMOVI 2017
Sex	0.527 (0.003)	0.529 (0.005)	0.526 (0.004)
Urban community (current household)	0.831 (0.010)	0.790 (0.012)	0.874 (0.012)
Urban community (household of origin)	0.587 (0.013)	0.515 (0.009)	0.666 (0.015)
Average parental years of schooling	4.63 (0.057)	4.60 (0.064)	4.665 (0.096)
Average respondent years of schooling	9.72 (0.060)	9.650 (0.072)	9.805 (0.095)
At least one parent speaks an indigenous tongue	0.137 (0.007)	0.149 (0.007)	0.124 (0.011)
Respondent speaks indigenous tongue		0.071 (0.006)	
Indigenous self-identification		0.140 (0.005)	
Sample size	43,299	25,634	17,665

Note: Sample weights employed. Communities with more than 2,500 inhabitants are categorized as urban, both for the origin and the current household. The population with at least one parent that spoke an indigenous tongue is considered the indigenous population unless otherwise is stated. Indigenous self-identification refers to the population that, when asked about how they classify themselves, answers as belonging to one of the Mexican indigenous groups.

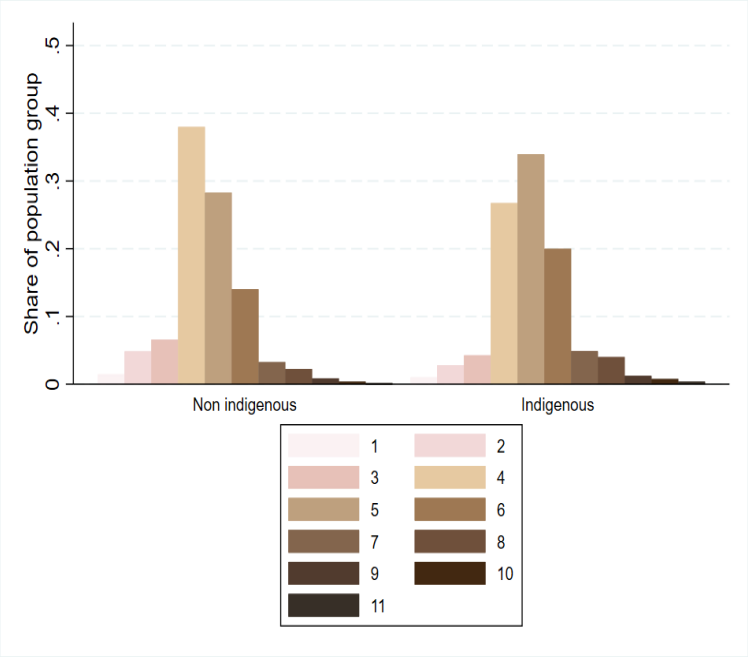
credible threat of not having enough observations for performing a robust statistical analysis across the three skin tone groups. Table 2.3 shows the composition of the composite sample in terms of ethnic origin and skin tone. The smallest group corresponds to light-skinned indigenous Mexicans, for which there are only 523 observations. Although it represents 10% of the total of light-skinned individuals in the country, the sample size does impose

Table 2.2. Skin tone Descriptive statistics

	Composite Sample	MMSI 2016	EMOVI 2017
PERLA Scale			
Tone 1	0.014	0.020	0.007
	(0.001)	(0.001)	(0.001)
Tone 2	0.046	0.049	0.042
	(0.003)	(0.002)	(0.005)
Tone 3	0.063	0.052	0.075
	(0.003)	(0.002)	(0.005)
Tone 4	0.364	0.373	0.355
	(0.005)	(0.005)	(0.008)
Tone 5	0.291	0.300	0.280
	(0.004)	(0.004)	(0.005)
Tone 6	0.148	0.130	0.167
	(0.003)	(0.003)	(0.006)
Tone 7	0.035	0.028	0.043
	(0.002)	(0.002)	(0.003)
Tone 8	0.025	0.030	0.020
	(0.001)	(0.002)	(0.002)
Tone 9	0.009	0.010	0.008
	(0.001)	(0.001)	(0.001)
Tone 10	0.004	0.005	0.003
	(0.0004)	(0.001)	(0.0004)
Tone 11	0.001	0.002	0.001
	(0.0002)	(0.0003)	(0.0004)
Aggregate tone groups			
Tone Group 1	0.123	0.121	0.124
	(0.005)	(0.003)	(0.009)
Tone Group 2	0.803	0.804	0.802
	(0.005)	(0.004)	(0.009)
Tone Group 3	0.074	0.075	0.074
	(0.003)	(0.003)	(0.004)
Sample size	43,299	25,634	17,665

Note: Sample weights employed. Skin tone 1 refers to the lightest skin tone in the PERLA tone palette, whereas skin tone 11 corresponds to the darkest one. For the palette, see figure A1 in the appendix.

Figure 2.1. Skin tone composition of the indigenous and non-indigenous Mexican populations



Note: Sample weights employed

some restrictions in analytical terms. In particular, it prevents me from analyzing this subpopulation separately from the national distribution, as the small sample size would lead to very noisy estimates. In other words, I can estimate changes in nationally defined ranks for this population but not the mobility patterns of the members of this subgroup with respect to each other (i.e., relative mobility patterns within the group).

As mentioned, the surveys recover information on the household conditions inhabited by the respondents at 14 years of age through a series of retrospective questions. Following the findings from neuroscience and psychology research on memory, the surveys situate the reference point for the retrospective questions during the teenage years. As Janssen and Murre (2008); Murre et al. (2013); Kilford et al. (2016); Hartshorne et al. (2018) show, the teenage years represent a critical period for the formation of memories and the acquisition of new cognitive skills. Centering the reference point of the retrospective questions at 14 years old al-

Table 2.3. Sample distribution across ethnic origin and skin tone group

	Non-indigenous		Indigenous	
	Total observations	Percent of skin tone group	Total observations	Percent of skin tone group
Skin tone Group 1	4,460	89.50%	523	10.50%
Skin tone Group 2	29,445	82.98%	5,012	17.02%
Skin tone Group 3	2,506	71.87%	705	28.13%

Note: Data from the composite sample. No sample weights employed. Skin tone group 1 corresponds to PERLA tones 1-3, group 2 to PERLA tones 4-6 and group 3 to PERLA tones 7-11.

lows to gear this fact to recover more salient and precise descriptions of the household of origin.

A second element in the design of the surveys aimed at diminishing the loss of precision in the description of the household of origin due to recall bias is that the surveys only ask about ownership and access to durable goods, public utilities, and financial services. Consequently, the respondent is not asked about the value or specific characteristics of the goods and services, imposing a smaller burden on her memory. The downside of this approach is that it is impossible to obtain a monetary value for the origin or the current household conditions. An alternative is to use the implicit information regarding the economic resources of the household present in the combination of goods and services possessed or to which the household has access. This information is useful for constructing a latent variable through a data reduction technique, such as Multiple Correspondence Analysis (MCA) or Principal Components Analysis (PCA). This approach to capturing economic status has been used in development economics in the study of the income distribution (Filmer and Pritchett, 2001; **?**; Filmer and Scott, 2012; Poirier et al., 2020), inequality of opportunity (Wendelspiess-Chávez-Juárez, 2015; Vélez-Grajales et al., 2018a; Monroy-Gómez-Franco et al., 2022; Plassot et al., 2022), and intergenerational social mobility (Monroy-Gómez-Franco and

Vélez-Grajales, 2021; Campos-Vázquez and Medina-Cortina, 2019; Torche, 2015; Delajara et al., 2022).

2.3.1 Construction of the indexes of economic resources

Tables 2.4 and 2.5 indicate the relative frequency of the goods and services in MMSI 2016 and EMOVI 2017. These household assets are the basis for constructing an index summarizing the economic resources available to both the origin and current households. Both tables show that the relative frequency of most assets is very similar across samples, allowing the construction of an asset index with a subset of these assets for the composite sample. Table 2.6 shows the goods and services selected to construct the current and origin household index.

All the goods and services in table 2.6 are binary variables. This characteristic restricts the method suitable to capture the latent variable of economic resources through the asset indexes. Specifically, I cannot employ methods that rely on euclidian distances to construct the weights assigned to each asset, such as PCA. Thus, for the construction of the economic resources indexes, I employ MCA, which is a method that relies on the relative frequencies of the different variables for the construction of the latent variable representing the economic resources of a household. This method was applied by Monroy-Gómez-Franco and Vélez-Grajales (2021) to analyze social mobility in Mexico.[6]. Considering that the sample consists of individuals between 35 to 64 years old, I construct a separate index for each 10-year cohort in the sample (25-34, 35-44, 45-54, 55-64). This makes MCA capture the underlying latent variable for each cohort separately, avoiding biasing the estimation by comparing members of different cohorts to each other.

[6]See Appendix B for an in-depth description of the methodology employed to construct the economic resources index.

Table 2.4. Relative frequency of ownership of services and goods across samples (Origin household)

Asset	EMOVI 2017	MMSI 2016	Composite sample	Asset	EMOVI 2017	MMSI 2016	Composite sample
Bank account	0.113 (0.002)	0.126 (0.002)	0.120 (0.002)	Microwave	0.502 (0.004)	0.510 (0.003)	0.506 (0.002)
Credit card	0.111 (0.002)	0.104 (0.002)	0.107 (0.002)	Refrigerator	0.583 (0.004)	0.548 (0.003)	0.565 (0.002)
Cattle	0.143 (0.003)	0.191 (0.002)	0.168 (0.002)	Washing machine	0.357 (0.004)	0.359 (0.003)	0.358 (0.002)
Electricity in household	0.851 (0.003)	0.816 (0.002)	0.833 (0.002)	Potable water access inside the household	0.672 (0.004)	0.576 (0.003)	0.623 (0.002)
Owned inhabited dwelling	0.771 (0.003)	0.809 (0.002)	0.790 (0.002)	TV set	0.660 (0.004)	0.666 (0.003)	0.663 (0.002)
Land for agricultural uses	0.171 (0.003)	0.250 (0.003)	0.211 (0.002)	Stove	0.669 (0.004)	0.689 (0.003)	0.679 (0.002)
Tractor	0.030 (0.001)	0.016 (0.001)	0.023 (0.001)	Toaster	0.079 (0.002)	0.116 (0.002)	0.098 (0.001)
Work animals	0.132 (0.003)	0.187 (0.002)	0.160 (0.002)	Car	0.235 (0.003)	0.227 (0.003)	0.231 (0.002)
Land for non-agricultural uses	0.071 (0.002)	0.068 (0.002)	0.069 (0.001)	Another dwelling	0.046 (0.002)	0.058 (0.001)	0.050 (0.001)
Commercial venue	0.056 (0.002)	0.044 (0.001)	0.050 (0.001)	Landline	0.247 (0.003)	0.238 (0.001)	0.242 (0.002)

Note: Sample weights employed.

A concern in intergenerational social mobility analysis is the effect of life cycle bias in estimating the intergenerational persistence rate. Similarly, and specific to the Mexican case, Monroy-Gómez-Franco and Vélez-Grajales (2021) showed that the distribution of skin tones is not homogenous across the national territory, introducing a possible source of bias due to the concentration of the darker skin tones in the south of the country, which is also the region where living conditions are more precarious compared to the rest of the country. To attenuate these two possible sources of bias, I regress the asset indexes constructed through

Table 2.5. Relative frequency of ownership of services and goods across samples (Current household)

	EMOVI 2017	MMSI 2016	Composite Sample		EMOVI 2017	MMSI 2016	Composite Sample
Bank account	0.277 (0.003)	0.356 (0.003)	0.317 (0.002)	TV Set	0.856 (0.003)	0.719 (0.003)	0.786 (0.002)
Credit Card	0.324 (0.004)	0.346 (0.003)	0.335 (0.002)	Stove	0.939 (0.002)	0.916 (0.002)	0.927 (0.001)
Cattle	0.037 (0.001)	0.064 (0.002)	0.051 (0.001)	Toaster	0.204 (0.003)	0.212 (0.003)	0.208 (0.002)
Electricity in household	0.992 (0.001)	0.993 (0.001)	0.993 (0.001)	Car	0.450 (0.004)	0.430 (0.003)	0.440 (0.002)
Owner of inhabited dwelling	0.585 (0.004)	0.524 (0.003)	0.554 (0.002)	Boiler	0.566 (0.004)	0.469 (0.003)	0.516 (0.002)
Land for agricultural uses	0.044 (0.002)	0.088 (0.002)	0.067 (0.001)	Cable TV	0.500 (0.004)	0.546 (0.003)	0.524 (0.002)
Tractor	0.009 (0.001)	0.019 (0.001)	0.014 (0.001)	Dwelling floor material different from dirt	0.974 (0.001)	0.977 (0.001)	0.975 (0.001)
Work animals	0.031 (0.001)	0.038 (0.001)	0.035 (0.001)	Cellphone	0.854 (0.003)	0.913 (0.002)	0.884 (0.002)
Land for non-agricultural uses	0.026 (0.001)	0.063 (0.002)	0.045 (0.001)	Computer	0.319 (0.004)	0.381 (0.003)	0.350 (0.002)
Commercial venue	0.096 (0.002)	0.033 (0.001)	0.064 (0.001)	Domestic service	0.106 (0.002)	0.057 (0.001)	0.081 (0.001)
Owned another dwelling	0.044 (0.002)	0.115 (0.002)	0.080 (0.001)	Internet	0.414 (0.004)	0.458 (0.003)	0.436 (0.002)
Microwave	0.502 (0.004)	0.510 (0.003)	0.506 (0.002)	Tablet	0.224 (0.003)	0.259 (0.003)	0.242 (0.002)
Refrigerator	0.910 (0.002)	0.886 (0.002)	0.898 (0.001)	VHS/DVD	0.439 (0.004)	0.467 (0.003)	0.454 (0.002)
Washing machine	0.766 (0.003)	0.748 (0.003)	0.757 (0.002)	Videogame console	0.186 (0.003)	0.158 (0.002)	0.172 (0.002)
Potable water access inside the household	0.908 (0.002)	0.937 (0.002)	0.923 (0.001)	Landline	0.381 (0.004)	0.410 (0.003)	0.396 (0.001)

Note: Sample weights employed.

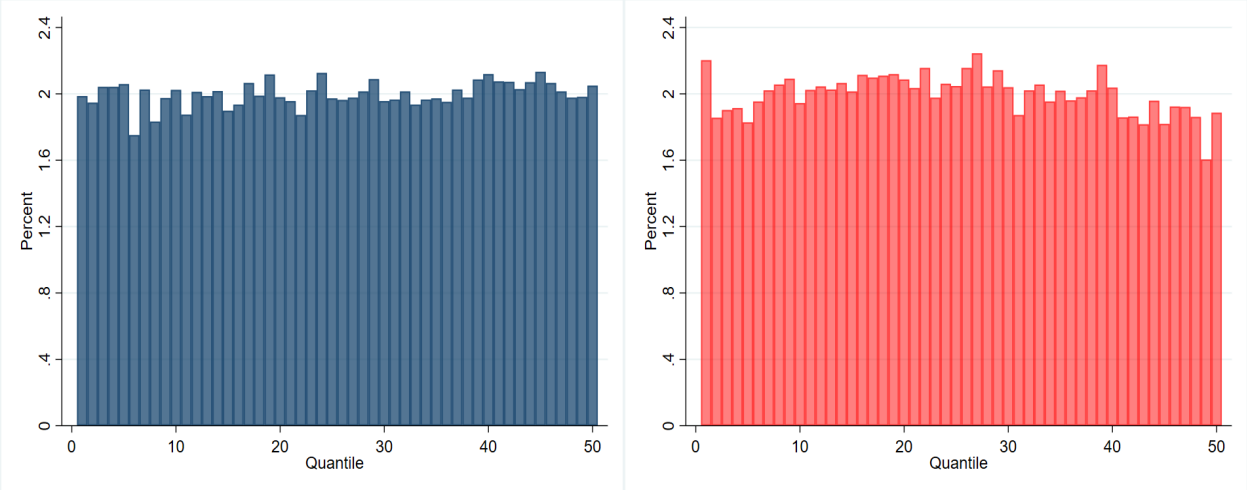
MCA on a series of regional dummies, the age and squared age of the respondent. I will use the residual of these regressions as the asset indexes for the social mobility analysis.

Table 2.6. Goods and services employed to construct the asset index

Household asset/service	Household at 14 years old	Current Household	Household asset/service	Household at 14 years old	Current Household
Overcrowded household	✓	✓	Bank account	✓	✓
Credit Card	✓	✓	Electricity	✓	✓
Landline	✓	✓	Cellphone		✓
Toaster	✓	✓	Car	✓	✓
Stove	✓	✓	Refrigerator	✓	✓
Drinking water	✓	✓	Tablet		✓
Washing machine	✓	✓	T.V. Set	✓	✓
DVD Player / Cassette recorder		✓	Video-game console		✓
Cable T.V.		✓	Owner of commercial venue	✓	✓
Microwave		✓	Domestic service		✓
Tractor		✓	Owner of another dwelling	✓	✓
Computer		✓	Owner of non-agricultural lands		✓
Owner of inhabited dwelling		✓	Tractor		✓
Internet		✓	Water heater	✓	✓

Furthermore, Nybom and Stuhler (2017) show that rank-based measures of intergenerational persistence, and consequently, social mobility, are less affected by life-cycle bias than other measures. Thus, I use the residuals from the abovementioned regression to construct a ranking of the distribution of economic resources of the origin and current household. In order to avoid the presence of empty cells in the distribution, I divide said ranking into 50 quantiles. As figure 2.2 shows, this allows me to obtain a ranking with 50 units of similar size, diminishing the effect of ties in the ranking process. As a result, the current and the origin rankings show a consistent composition of the ranks.

Figure 2.2. Distribution of economic resources indexes



(a) Origin household economic resources index (b) Current household economic resources index

Note: Survey weights are employed.
Source:Composite EMOVI-MMSI dataset.

This process has a significant implication for interpreting the measures of intergenerational social mobility described in the following section. As mentioned above, I construct the ranks of the distributions of economic resources conditional on the respondent's age and region of origin, using the national population as the reference population. As a result of these transformations, estimating rank-based measures of social mobility on this data leads to measures of change in positions of the national distribution abstracting from systemic regional differences. Following Deutscher and Mazumder (2021) the measures are not strictly positional, as changes in the group mean of the outcome variable for one group between one period and the next lead to changes in the rank-persistence coefficient of all groups. Together with the sample size for the light-skinned indigenous Mexicans, this prevents me from analyzing if there exists regional heterogeneity in the incidence of colorism across groups of different ethnic origins. However, an upside of this approach is that it allows disentangling the effect of the region of origin from those of the ethnic origin and skin tone, allowing for a more precise estimation.

2.4 Methods

The values of the economic resources index are not directly interpretable, as they only represent a latent variable capturing the overall economic status of the household. However, they have an ordinal character, which allows me to rank individuals according to the amount of resources available in their current and origin household. Consequently, I can analyze the patterns of intergenerational rank persistence of the indigenous and non-indigenous Mexicans by their skin tone. To do so, I estimate a series of rank-rank regressions in which the dependent variable is the rank of the current household in the current distribution of the economic resources (R_{it}) and the regressor is the rank of the origin household in the corresponding distribution ($R_{i,t-1}$). This approach is similar to that of Chetty et al. (2015, 2019). Formally, the regression equation is

$$R_{it} = \alpha + \beta R_{i,t-1} + \epsilon_i \tag{2.1}$$

In this case, the β coefficient corresponds to the degree of intergenerational rank persistence. I follow Goldsmith et al. (2006, 2007)work on the effects of colorism on labor market outcomes and adapt it to the context of rank-rank regressions. Initially, I estimate a modified version of equation 1 that adds a binary variable that identifies if a person is indigenous or not ($Ind_i = 1$ if i is indigenous), and this variable interacts with the rank of origin. This specification is akin to the specification used by Chetty et al. (2019) to estimate differences in social mobility between whites and blacks in the US. Formally this is:

$$R_{it} = \alpha + bR_{i,t-1} + \phi Ind_i + \delta(Ind_i \times R_{i,t-1}) + v_i \tag{2.2}$$

The coefficient δ captures the differences in intergenerational persistence between the indigenous and non-indigenous Mexican populations. Similarly, ϕ captures the differences in the rank to which each population converges. Chetty et al. (2015) also provide a different interpretation of this coefficient, pointing out that it will also represent the difference in the absolute mobility starting from the origin quantile between the indigenous and non-indigenous groups.

A third model opens up the categories of indigenous and non-indigenous differentiating according to the person's skin tone. This allows capturing any intragroup differences by skin color, allowing to capture if differences in skin tone mediate the differences by ethnic origin. Let $IndST_i^c = 1$ if i is indigenous with skin tone from group c, and zero otherwise. Similarly, define $NIndST_i^c = 1$ if i is non-indigenous with a skin tone from group c. The reference group for the estimation is the set of non-indigenous people with a skin tone among the lightest group ($c = 1$). Then, the model proposed by Goldsmith et al. (2006, 2007) adapted for rank-rank regressions is the following

$$\begin{aligned} R_{it} = \alpha + BR_{i,t-1} + \sum_{c=1}^{3} \Phi_c IndST_i^c + \sum_{c=2}^{3} \Gamma_c NIndST_i^c \\ + \sum_{c=1}^{3} \Theta_c \Big(IndST_i^c \times R_{i,t-1} \Big) + \sum_{c=2}^{3} \eta_c \Big(NIndST_i^c \times R_{i,t-1} \Big) + u_i \end{aligned} \tag{2.3}$$

The persistence rate coefficient provides a summary measure of intergenerational rank transmission. However, it does not allow to check if there are non-linearities in it nor to

identify the underlying pattern behind the differences in rank persistence rates between ethnic and skin tone groups.

I address this limitation by estimating the transition matrices for each ethnic-skin tone group defined. Transition matrices are $k \times k$ matrices in which each element represents the probability that an individual with origin in quantile o is currently at quantile p. Formally, let N_o^p be the population with origin in quantile o and currently in quantile p, and N_o be the population with origin in quantile o. The transition probability between quantiles o and p, $P_{[p|o]}$ is defined as follows

$$P_{[p|o]} = \frac{N_o^p}{N_o} \tag{2.4}$$

The transition probabilities can be grouped in a transition matrix. The rows of the transition matrix M represent the quantile of origin, while the columns order the probabilities by quintile of the household in the current economic resources distribution. Dividing the origin and current distributions into quintiles gives the following 5×5 transition matrix:

$$M_{o,p} \equiv \begin{bmatrix} P_{1|1} & \dots & P_{5|1} \\ \vdots & \vdots & \vdots \\ P_{1|5} & \dots & P_{5|5} \end{bmatrix} \tag{2.5}$$

Thus, by construction, transition matrices provide information about the rank persistence rate at different points of the distribution of economic resources at the origin. The cut-off points for each quintile are defined considering the national distributions, considering indigenous and non-indigenous populations together. The common and exogenous benchmarks

allow for a direct comparison of the transition rates at different distribution points between ethnic/skin tone groups. This is the strategy Chetty et al. (2019) followed to compare persistence rates between African-Americans and Whites in the US and by Monroy-Gómez-Franco and Vélez-Grajales (2021) to compare across skin tone and regional groups in Mexico.

To obtain a finer disaggregation, I estimate the non-parametric expected conditional quantile by decile of origin. A limitation of both transition matrices and the non-parametric estimation is that they require a substantial sample size to provide robust results. This is particularly problematic when dealing with relatively small subsamples such as the skin tone subgroups among the indigenous population.

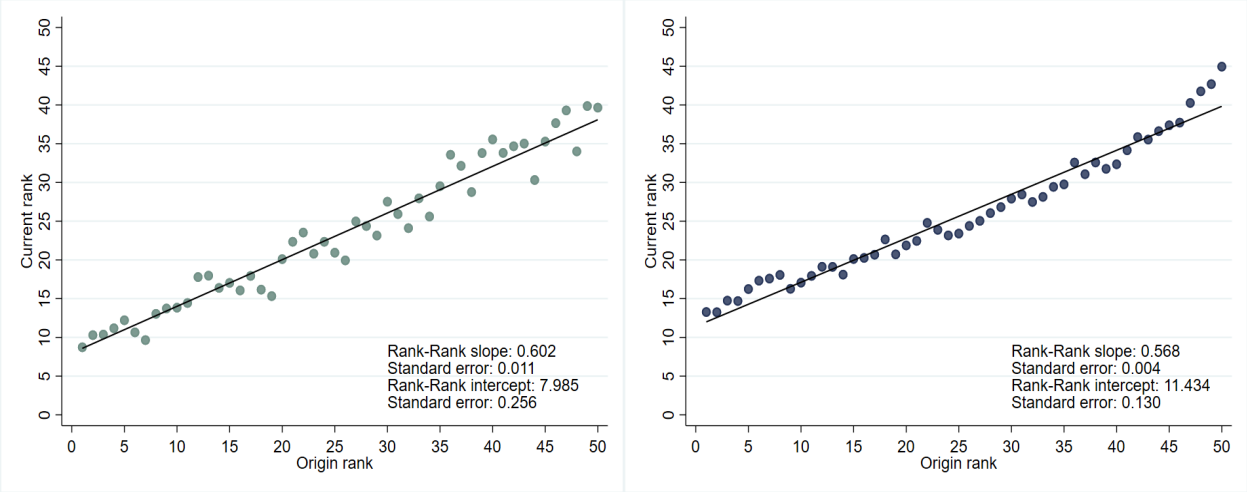
2.5 Results

2.5.1 Rank-rank regressions

A first step in the analysis is to establish if a linear approximation to the relationship between the origin and the current household rank is adequate. Figure 2.3 shows the binned scatter plots between both ranks for the indigenous and the non-indigenous populations and the associated regression line. As can be seen, a linear regression approximates the behavior of the data well enough, although, in the case of the non-indigenous population, the behavior at the top of the distribution seems to become partially nonlinear. This is something also observed by Chetty et al. (2015) for the US data.

Table 2.7 shows the results of the estimation of equations one to three. The estimated persistence rate in column 1 corresponds to the unconditional persistence rate for the Mexican population between 25 and 64 years old. The estimated value, 0.583, is similar to the values

Figure 2.3. Binned scatter plots



(a) Indigenous population (b) Non-indigenous population

Note: Survey weights are employed. Bins defined to include one quintile out of 50.
Source: Composite EMOVI-MMSI dataset.

estimated in previous exercises by Monroy-Gómez-Franco and Vélez-Grajales (2021) (point estimate 0.598, SE: 0.006) using ESRU-EMOVI 2017 exclusively, Delajara and Graña (2018) (point estimate 0.54, SE: 0.01) using a previous wave of ESRU-EMOVI corresponding to 2011 and by Delajara et al. (2022)) (point estimate of 0.62, SE: 0.006) using a similar version of the composite sample employed in this paper. This serves as a validation test for the pooling of MMSI-2016 and the EMOVI-ESRU 2017.

Column 2 of Table 2.7 shows the estimate of the differences in intergenerational rank persistence between the indigenous and non-indigenous populations in Mexico without disaggregating by skin tone. The positive estimate of the interaction coefficient, albeit not statistically significant, suggests that the indigenous population in Mexico experiences more intergenerational rank persistence than the non-indigenous population. Moreover, the negative estimate of the coefficient on indigenous status implies that the indigenous population converges to a lower rank in the national distribution than the non-indigenous population.

Table 2.7. Rank-Rank regressions, composite sample

Current rank	(1)	(2)	(3)
Origin rank	0.583***	0.568***	0.601***
	(0.009)	(0.010)	(0.021)
Indigenous		-3.450***	
		(0.600)	
Indigenous × Origin rank		0.0343	
		(0.0227)	
Indigenous, skin tone group 1			-4.557***
			(1.640)
Indigenous, skin tone group 2			-4.188***
			(0.914)
Indigenous, skin tone group 3			-6.796***
			(1.252)
Non-indigenous, skin tone group 2			-1.071*
			(0.607)
Non-indigenous, skin tone group 3			-2.268**
			(0.951)
Indigenous, skin tone group 1			0.112**
× Origin rank			(0.048)
Indigenous, skin tone group 2			-0.019
× Origin rank			(0.032)
Indigenous, skin tone group 3			-0.019
× Origin rank			(0.065)
Non-indigenous, skin tone group 2			-0.043**
× Origin rank			(0.019)
Non-indigenous, skin tone group 3			-0.100***
× Origin rank			(0.035)
Intercept	10.65***	11.43***	12.64***
	(0.282)	(0.312)	(0.686)
Observations	42,986	42,343	42,343
R-squared	0.340	0.347	0.353

Note: Sample weights are employed in all estimations. Standard errors clustered by primary sample unit of the survey. The indigenous population corresponds to the population with at least a parent that speaks an indigenous tongue. Skin tone group 1 corresponds to PERLA tones 1-3, group 2 to PERLA skin tones 4-6 and group 3 to PERLA skin tones 7-11. Ranks correspond to 50 quantiles of the origin and current distribution of the index of economic resources.

This result is consistent with the findings of Rojas-Lomelín et al. (2019).

When the categories of indigenous and non-indigenous are disaggregated by the person's skin tone (column 3), the estimate of the coefficient corresponding to the intercept of each ethnic-skin tone group is negative in all cases, and except for the non-indigenous population with a skin tone in group 2, it is statistically significant at the 95% level. This implies that, compared to the non-indigenous light-skinned population, the rest of the Mexican population is expected to converge to a lower rank in the distribution of economic resources[7].

However, the story is more complex regarding the estimate of the differences in the intergenerational rank persistence rate. Except for the light-skinned indigenous population (indigenous, skin tone group one), the point estimate of the interaction term between the ethnic-skin tone group variables and the rank of origin is negative, although only statistically significant for the non-indigenous groups. This would imply that the light-skinned indigenous population experiences a higher intergenerational rank persistence than the light-skinned non-indigenous, while the rest of the population experiences lower persistence rates.

The next step is to compare the estimates across the different ethnic-skin tone groups. Table 8 shows the linear combination tests corresponding to these comparisons. In the case of the indigenous population, the point estimates of the intercept term are increasing as the skin tonality becomes lighter. However, only the difference between the intercept for the indigenous population with tones 4-6 and tones 7-11 is statistically different from zero at a 95% confidence level. This is evidence that lighter-skinned Mexican indigenous people converge to a higher rank in the distribution of economic resources than their peers with

[7]These results are also observed if a self-identification criterion is employed instead of defining the indigenous population through a linguistic criterion. As only the MMSI 2016 contains this information, the estimations shown in table C1 of the appendix only use this sample, not the composite sample.

Table 2.8. Differences in rank persistence and intercept across groups (Linear combination test)

Hypothesis	Intercept coefficients (A)		Effect on persistence coefficient (B)	
	Coefficient difference	Clustered standard error	Coefficient difference	Clustered standard error
$Ind_{1-3} - Ind_{4-6}$	-0.368	1.487	0.131	0.047
$Ind_{1-3} - Ind_{7-11}$	2.239	1.796	0.131	0.074
$Ind_{4-6} - Ind_{7-11}$	2.608	1.036	0.001	0.062
$NonInd_{4-6} - NonInd_{7-11}$	1.197	0.783	0.057	0.029
$Ind_{1-3} - NoInd_{4-6}$	-3.486	1.499	0.155	0.045
$Ind_{4-6} - NoInd_{4-6}$	-3.117	0.640	0.024	0.025
$Ind_{7-11} - NoInd_{4-6}$	-5.725	0.983	0.024	0.059
$Ind_{1-3} - NoInd_{7-11}$	-2.289	1.618	0.212	0.054
$Ind_{4-6} - NoInd_{7-11}$	-1.920	0.966	0.081	0.036
$Ind_{7-11} - NoInd_{7-11}$	-4.528	1.218	0.081	0.063

Note: Survey weights employed. The difference coefficient corresponds to the linear combination of the corresponding coefficients from the estimation of model 3 of table 2.7.

darker skin tones. In the case of the non-indigenous population, the difference between the groups with skin tones in groups 2 and 3 is not statistically significant.

Another comparison of interest is between groups with the same skin tone but different ethnic ascription. Table 8 also shows these comparisons. In this cross-ethnic group comparison, the indigenous group systematically has a smaller intercept than the non-indigenous, and the difference is statistically significant at the 95% confidence level. Furthermore, the intercept corresponding to the indigenous groups with lighter skin tones is also systematically smaller than that of non-indigenous groups with darker skin tones. However, in the case of the persistence rate, the pattern reverses, and the indigenous groups tend to have a higher

rank persistence rate than the non-indigenous groups. And when comparing indigenous groups, the lighter skin tone group has a higher persistence rate than the other two tonality groups.

These results start to delineate a stratification regime in which, on average, the non-indigenous population achieves higher echelons of the economic resources distributions than those reached by the indigenous population. Moreover, the evidence suggests that inside both ethnic groups, persons with a lighter skin tone tend to achieve higher ranks than those with darker skin tones. This is in line with the results from Monroy-Gómez-Franco et al. (2022) which suggest that skin-tone stratification operates as a layer of stratification that generates inequality of opportunity, but it is not the main driver of inequality in Mexican society. In this case, the evidence suggests that stratification by ethnic status generates larger differences in expected rank persistence than skin-tone stratification, but the latter plays a role in the stratification within both ethnic groups.

2.5.2 Transition matrices and non-linearities

Tables 2.9 and 2.10 show the transition matrices for Mexico's non-indigenous and indigenous populations. Both transition matrices are constructed using the same distribution to define the quintile cut-offs (the national distribution), allowing to compare the transition probabilities at each distribution point directly. It is worth noting that the matrices show the unadjusted probabilities of transition, and as such, they are not bi-stochastic transition matrices. Both transition matrices have a significant persistence at the top of the distribution. Namely, more than one in two of those in the fifth quintile of the distribution at 14 years old remained there in adulthood. However, in the case of the bottom part of the distribution, the persistence among the indigenous population is substantially higher. Specifically, 58 out

of 100 in a household ranked in that position at 14 years old remained there in adulthood. The corresponding odds for the non-indigenous population are 40 out of 100.

Table 2.9. Transition matrix
(Non-indigenous population)

	Current Q1	Current Q2	Current Q3	Current Q4	Current Q5
Origin Q1	0.408 (0.013)	0.278 (0.005)	0.177 (0.010)	0.096 (0.001)	0.041 (0.002)
Origin Q2	0.262 (0.001)	0.283 (0.001)	0.240 (0.002)	0.143 (0.002)	0.071 (0.001)
Origin Q3	0.151 (0.001)	0.241 (0.002)	0.258 (0.001)	0.232 (0.001)	0.119 (0.001)
Origin Q4	0.071 (0.001)	0.158 (0.001)	0.225 (0.001)	0.309 (0.001)	0.237 (0.001)
Origin Q5	0.024 (0.001)	0.055 (0.001)	0.122 (0.001)	0.250 (0.002)	0.549 (0.002)

The off-diagonal elements also indicate a substantial difference in the mobility patterns of the indigenous and the non-indigenous population. In the case of the non-indigenous population, the upward mobility transition probabilities (upper triangle of the transition matrix) are larger than those of the indigenous population. In contrast, the downward transition probabilities (lower triangle of the transition matrix) of the indigenous population are higher than those of the non-indigenous population. This sheds light on the similar rates of intergenerational persistence between the indigenous and the non-indigenous population. Although they persist at similar rates, the indigenous population persists more frequently at the bottom of the distribution than at the top, while the reverse is true for the non-indigenous population. The similar mobility rates also have different implications: the indigenous population experiences more downward mobility than the non-indigenous one, while the

latter experiences upward mobility more frequently.

Table 2.10. Transition matrix
(Indigenous population)

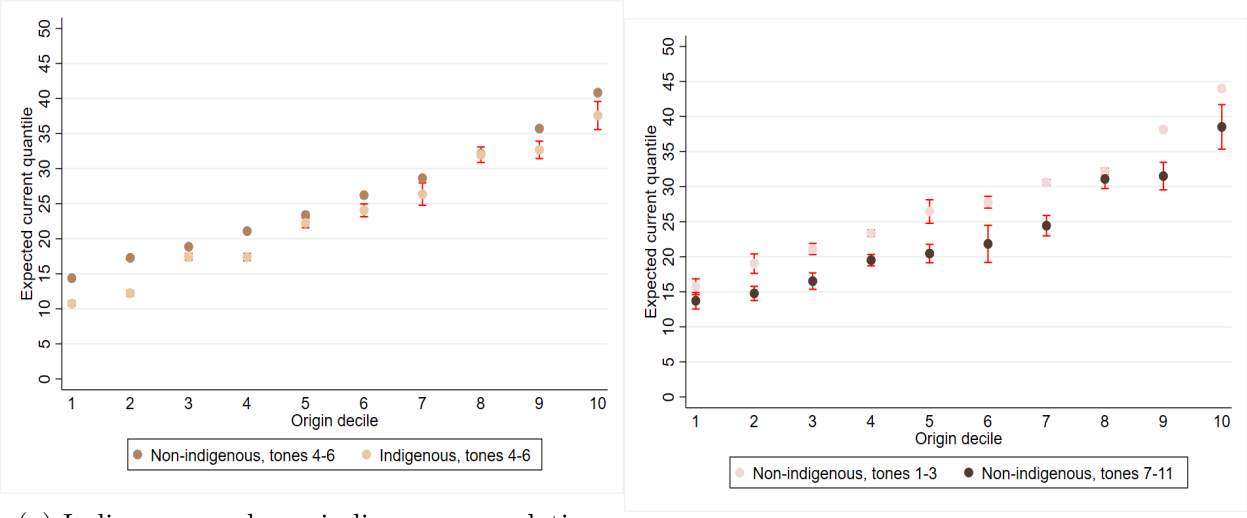
	Current Q1	Current Q2	Current Q3	Current Q4	Current Q5
Origin Q1	0.600 (0.003)	0.212 (0.001)	0.118 (0.001)	0.052 (0.001)	0.018 (0.001)
Origin Q2	0.366 (0.004)	0.290 (0.003)	0.194 (0.002)	0.102 (0.001)	0.049 (0.001)
Origin Q3	0.201 (0.001)	0.251 (0.01)	0.247 (0.002)	0.196 (0.003)	0.097 (0.002)
Origin Q4	0.117 (0.001)	0.139 (0.002)	0.219 (0.002)	0.264 (0.001)	0.261 (0.002)
Origin Q5	0.076 (0.002)	0.078 (0.001)	0.156 (0.001)	0.196 (0.003)	0.495 (0.001)

Given the differences in mobility patterns across different distribution segments, it is worth analyzing how they map once the ethnic categories are disaggregated by skin color. A first, non-parametric approach is to estimate the expected rank conditional on the decile of the distribution of economic resources of origin and the person's skin color. A limitation of this approach is that it requires a substantial sample size for each ethnicity-skin tone group to derive relatively precise estimates. Given the skin tone distribution of the indigenous population, this limits the comparison between the different skin tone groups of the indigenous population.

Panel 2.4a of figure 2.4 shows the expected rank for the indigenous and non-indigenous populations with a skin tone of group 2. In all cases, the expected rank of the non-indigenous population lies above that of the indigenous. However, the distance between the expected ranks below the median is larger than above it. This pattern indicates a higher persistence

for the indigenous than non-indigenous at the bottom of the distribution. In contrast, Panel 2.4b shows the expected ranks for the non-indigenous populations with skin tone in groups 1 and 3. The predicted ranks for the population with a skin tone in group 1 lie above those corresponding to skin tone group 3. The distance between the ranks increases for those above the median while remaining constant among those below it. This suggests a mobility pattern that below the median is similar for both groups, but above the median favors more upward mobility than those with a lighter skin tone.

Figure 2.4. Expected quantile per decile of origin



(a) Indigenous and non-indigenous population, tones 4-6

(b) Non-indigenous tones 1-3 and 7-11

Note: Survey weights are employed. Tones refer to PERLA skin tones.
Source: Composite EMOVI-MMSI dataset.

An alternative to this approach is to estimate a series of quantile regressions at different distribution points. Table 2.11 shows the estimation results of this process. Specifically, it shows the estimates corresponding to the estimation of equation 3 at the 20th, 50th, 70th, and 90th percentile of the 50 quantiles distribution of economic resources. In other words, it corresponds to a series of quantile rank regressions. The results suggest a very similar pattern to

the one observed in Table 2.7 in terms of the order of the ethnic-skin tone groups. Namely, in both the indigenous and non-indigenous groups, the highest intercept corresponds to the lightest skin tone group and the lowest to the darkest skin tone group. Furthermore, as before, the intercepts corresponding to the non-indigenous groups lie above those of the indigenous groups.

Table 2.11. Quantile rank-rank regressions

Current rank	20^{th}	50^{th}	70^{th}	90^{th}
Origin rank	0.737***	0.742***	0.595***	0.306***
	(0.024)	(0.024)	(0.021)	(0.025)
Indigenous, skin tone group 1	-1.026	-5.507***	-5.745**	-4.004**
	(0.733)	(1.384)	(2.235)	(2.028)
Indigenous, skin tone group 2	0.735	-5.930***	-8.209***	-8.222***
	(0.491)	(0.912)	(1.213)	(1.503)
Indigenous, skin tone group 3	1.180*	-6.618***	-12.92***	-13.85**
	(0.662)	(1.292)	(1.736)	(5.789)
Non-indigenous, skin tone group 2	0.895*	-1.083	-2.709***	-3.211***
	(0.482)	(0.858)	(1.016)	(1.209)
Non-indigenous, skin tone group 3	1.415**	-3.387***	-5.297***	-6.813***
	(0.633)	(1.189)	(1.302)	(1.961)
Indigenous, skin tone group 1	-0.037	0.138***	0.167***	0.105**
× Origin rank	(0.071)	(0.037)	(0.055)	(0.051)
Indigenous, skin tone group 2	-0.302***	0.030	0.155***	0.194***
× Origin rank	(0.037)	(0.035))	(0.034)	(0.042)
Indigenous, skin tone group 3	-0.443***	-0.050	0.226**	0.319
× Origin rank	(0.075)	(0.088)	(0.094)	(0.203)
Non-indigenous, skin tone group 2	-0.158***	-0.046*	0.030	0.067**
× Origin rank	(0.026)	(0.026)	(0.024)	(0.027)
Non-indigenous, skin tone group 3	-0.325***	-0.322***	0.054	0.149***
× Origin rank	(0.036)	(0.048)	(0.038)	(0.052)
Intercept	-0.474	9.387***	19.46***	34.72***
	(0.456)	(0.790)	(0.933)	(1.107)
Pseudo R^2	0.1963	0.2487	0.2354	0.1574
Observations	42,343	42,343	42,343	42,343

Note: Sample weights are employed in all estimations. Robust standard errors. The indigenous population corresponds to the population with at least a parent that speaks an indigenous tongue. Skin tone group 1 corresponds to PERLA tones 1-3, group 2 to PERLA skin tones 4-6 and group 3 to PERLA skin tones 7-11. Ranks correspond to 50 quantiles of the origin and current distribution of the index of economic resources. Q10 corresponds to the 20th percentile, Q25 to the median, Q35 to the 70th percentile and Q45 to the 90th percentile of the 50 quantiles distribution.

However, there are differences in the magnitude of the intercept gaps across skin tone groups at different distribution points. For the non-indigenous group, the penalization in the intercept (and thus on the expected rank) associated with having a darker skin tone increases as one moves upward in the current economic resources distribution quantiles. This confirms the patterns identified in the non-parametric estimations. In contrast, for the indigenous group, the intercept penalization remains constant between the 70th and 90th percentile for those with an intermediate skin tone. However, it increases for those with the darkest skin tone. This suggests that regardless of the ethnic group, as one moves upward in the distribution of the economic resources, the weight of colorism increases. While Monroy-Gómez-Franco et al. (2022) provided some suggestive evidence of this pattern, they could not indicate if this occurred across ethnic groups. The evidence in this paper suggests that colorism is present and shapes the patterns of social mobility of both the indigenous and non-indigenous Mexican population, compounding in the case of the former the effects of the exclusion documented in the literature cited at the beginning of this paper.

2.6 Robustness analysis

Two elements might affect the robustness of my results. Firstly, it is possible that dividing the sample into 50 quantiles produces quantiles with too few observations in them, leading to more noisy estimates and increasing the possible influence of outliers affecting the result. A second issue is that the wide age span of the survey respondents (25 to 64 years old) might introduce a cohort bias into the estimated coefficients.

To address the first issue, I estimate equations 2.2 and 2.3 using an alternative ranking of the origin and current household economic resources indexes based on deciles instead of 50 quantiles. This increases the number of observations per quantile, reducing the influence of

possible outliers. As table 2.12 shows, this change does not affect my main results. The results are consistent with the primary estimations, showing high persistence for both light-skinned populations and a gradient by skin tonality in the intercepts. The gradient, as in the main estimates, shows that the indigenous population converges to a lower point of the national distribution than the non-indigenous population and that the darker skin tones converge to a lower point than lighter skin tones. These results provide evidence that the results from the main estimations are not driven by the size of the cells constructed.

To address the possibility of the results being the fruit of the pooling of individuals from different cohorts, I estimate equation 2.3 for each 10 year cohort separately using the ranks defined over 50 quantiles. Table 2.13 shows that for all cohorts, the estimate of the persistence rate for the non-indigenous light-skinned population is above 0.5. This implies a high degree of intergenerational persistence, regardless of the cohort to which the individual belongs. Furthermore, the pattern observed for the intercepts in the main estimation is also present when considering each cohort separately. Namely, lighter skin tones converge to a higher point of the distribution than darker skin tones, and in all cases, the indigenous populations converge to a lower part of the distribution than the non-indigenous groups.

However, it is important to note that except for the oldest and youngest cohorts, the differences in the persistence rate across population groups are not statistically significant. This suggests that these two cohorts drive the differences in this parameter across groups in the main estimation. Together with the previous results on the intercept coefficients, this suggests a stratification pattern in which the indigenous and non-indigenous populations converge to different parts of the distribution of the economic resources, with the indigenous population disadvantaged. Moreover, the results also confirm the finding of the main estimations regarding the role skin tones played within each ethnic group. In both, lighter skin tones have

a higher expected rank than darker skin tones, which implies that both groups are stratified by skin tone.

2.7 Final Remarks

The results of this paper indicate that there exist large differences between Mexico's indigenous and non-indigenous populations regarding the social mobility patterns they experience. The results from the rank-rank regressions suggest that the indigenous population experiences more intergenerational rank persistence than the non-indigenous. Moreover, the results suggest that the indigenous population is concentrated at the bottom of the distribution of economic resources. The transition matrices for each group indicate that, in the case of the indigenous population, they experience lower upward mobility rates than the non-indigenous when starting below the median of the distribution of the economic resources. However, both groups experience similar mobility rates in the upper parts of the distribution.

These results, however, are more complex once the indigenous and non-indigenous population groups are partitioned according to their skin tone. The evidence suggests that there is a skin tone gradient in both populations. In both cases, lighter skin tones tend to experience more mobility and be concentrated at upper ranks of the distribution than darker skin tones. This evidence suggests that colorism in Mexican society does not overlap with stratification by ethnic origin, but instead, it crisscrosses it. This suggests a more complex stratification regime than what the literature on socioeconomic differences by skin tone in Mexico suggested. The results also suggest that the differences in labor and credit market outcomes by skin tone from the existing literature can be producing persistent differences with intergenerational effects.

Table 2.12. Rank-Rank regressions using deciles, composite sample

Current rank	(1)	(2)	(3)
Origin rank	0.575***	0.561***	0.590***
	(0.008)	(0.009)	(0.021)
Indigenous		-0.688***	
		(0.125)	
Indigenous × Origin rank		0.029	
		(0.022)	
Indigenous, skin tone group 1			-1.001***
			(0.341)
Indigenous, skin tone group 2			-0.832***
			(0.192)
Indigenous, skin tone group 3			-1.359***
			(0.253)
Non-indigenous, skin tone group 2			-0.222*
			(0.132)
Non-indigenous, skin tone group 3			-0.457**
			(0.205)
Indigenous, skin tone group 1			0.124***
× Origin rank			(0.048)
Indigenous, skin tone group 2			-0.022
× Origin rank			(0.031)
Indigenous, skin tone group 3			-0.021
× Origin rank			(0.061)
Non-indigenous, skin tone group 2			-0.039**
× Origin rank			(0.020)
Non-indigenous, skin tone group 3			-0.095***
× Origin rank			(0.035)
Intercept	2.337***	2.499***	2.750***
	(0.058)	(0.065)	(0.145)
Observations	42,986	42,343	42,343
R-squared	0.332	0.339	0.345

Note: Sample weights are employed in all estimations. Standard errors clustered by primary sample unit of the survey. The indigenous population corresponds to the population with at least a parent that speaks an indigenous tongue. Skin tone group 1 corresponds to PERLA tones 1-3, group 2 to PERLA skin tones 4-6 and group 3 to PERLA skin tones 7-11. Ranks correspond to 50 quantiles of the origin and current distribution of the index of economic resources.

Table 2.13. Rank-Rank regressions by cohort, composite sample

Current rank	25-34	35-44	45-54	55-64
Origin rank	0.699***	0.603***	0.502***	0.553***
	(0.0470)	(0.0219)	(0.0370)	(0.0605)
Indigenous, skin tone group 1	-1.546	-4.814**	-1.443	-9.689*
	(2.910)	(2.168)	(4.483)	(5.623)
Indigenous, skin tone group 2	-3.588*	-3.179***	-6.190***	-5.301*
	(2.111)	(1.037)	(1.805)	(2.922)
Indigenous, skin tone group 3	-5.712**	-7.002***	-6.803*	-8.645**
	(2.307)	(1.398)	(4.080)	(4.014)
Non-indigenous, skin tone group 2	0.513	-0.875	-3.018**	-1.031
	(1.849)	(0.844)	(1.352)	(2.148)
Non-indigenous, skin tone group 3	-1.591	-2.484**	-5.176**	3.870
	(2.227)	(1.176)	(2.042)	(3.884)
Indigenous, skin tone group 1	0.130*	0.056	0.031	0.326*
× Origin rank	(0.077)	(0.078)	(0.127)	(0.165)
Indigenous, skin tone group 2	-0.012	-0.025	0.007	-0.089
× Origin rank	(0.059)	(0.038)	(0.064)	(0.103))
Indigenous, skin tone group 3	-0.012	-0.025	0.007	-0.089
× Origin rank	(0.059)	(0.038)	(0.064)	(0.103)
Non-indigenous, skin tone group 2	-0.109**	-0.026	0.017	-0.112**
× Origin rank	(0.047)	(0.023)	(0.039)	(0.055)
Non-indigenous, skin tone group 3	-0.136**	-0.081*	0.013	-0.327**
× Origin rank	(0.068)	(0.042)	(0.074)	(0.127)
Intercept	10.03***	12.05***	15.63***	15.64***
	(1.737)	(0.804)	(1.296)	(2.342)
Observations	8,044	22,542	7,830	3,927
R-squared	0.414	0.368	0.298	0.250

Note: Sample weights are employed in all estimations. Standard errors clustered by primary sample unit of the survey. Ranks correspond to 50 quantiles of the origin and current distribution of the index of economic resources.

Two direct implications can be derived from these results. The first one concerns the study of differences in intergenerational mobility patterns across different ethnic groups of a particular country. As I have shown in this paper, to obtain a better knowledge of stratification patterns and how inequalities consolidate through time, it is necessary to study differences both between and within ethnic groups. Incorporating this more complex view of heterogeneity into the study of intergenerational economic mobility is the next step in the literature. It requires will require changes in the type of data that is collected as well as in how ethnic groups are defined and conceptualized. The latter requires a more nuanced discussion of the relevant cleavages for each country in terms of the construction of ethnic identities and how said construction may hide some key elements, as is the case in Mexico concerning skin tone.

The second implication concerns the stratification regime in Mexico. The evidence I provide in this paper shows that skin tone serves as an organizing social criterion for distributive purposes in Mexico. Moreover, it shows that it acts independently from ethnic origin, as skin tone differences in the patterns of social mobility are observed among indigenous and non-indigenous Mexicans. This directly contradicts the narrative of the "Mestizaje myth," as it shows that although there were no formal institutions through which discrimination by skin color was implemented in Mexico in the XXth century, there are informal institutions that produce systematic economic inequalities among individuals of different skin tones. If the idea of a non-racist society in which skin color plays no role in life as everyone either shares the same mixed origin or has integrated into the common core of "Mexicanness" can be considered the "Mexican dream," the evidence of this paper shows that idea is not sustained by reality. Ethnic origin and skin tone do play a role in how far a person moves upward or downward in the distribution of economic resources.

A clear avenue of extension of this analysis is to consider how other social stratifiers interact with the ethnic and skin tone gradient. In particular, how gender interacts with both dimensions, similar to the recent work on labor market outcomes by Paul et al. (2022) for the US case. The results from Arceo-Gómez and Campos-Vázquez (2014, 2019) and Campos-Vázquez and Medina-Cortina (2018) indicate a gendered gradient in how skin tone serves as a variable upon which inequalities are constructed. This is also something that has been observed in ethnographic work on the subject (Krozer and Urrutia-Gómez, 2021). Similarly, the results by Torche (2015) indicate that there a substantial differences in social mobility between men and women.

Secondly, it is necessary to analyze the underlying factors behind the clear differences in intergenerational social mobility patterns by ethnic origin and skin tone. The work by Baldomero-Quintana et al. (2022) suggest that a deeper exploration of the persistent consequences of the colonial legacy in Mexico might help to understand better the patterns detected in this paper. However, more work in this vein needs to be conducted.

Chapter 3

Modelling the learning impacts of educational disruptions in the short and long run

3.1 Introduction

A classic problem for policymakers is the lack of information on the long-run effects of any shock. This is particularly true for processes in which their outcomes are observed only after long periods, an example of which is the learning process. This lack of information makes the implementation of measures to attenuate the effects of the shock more complex to the policymaker, as neither the magnitude nor the characteristics of those impacted are apparent at the time of the shock. In this paper, I integrate and expand previous work on learning progression and educational disruptions by Kaffenberger and Pritchett (2020b, 2021) and Neidhöfer et al. (2021) into a unified model of learning progression, parental investments and instruction interruptions. This integration allows me to disaggregate the impacts of an instructional disruption into its immediate effects and its cumulative effects of learning,

highlighting the role played by parental resources (and its distribution) in attenuating those effects. Furthermore, this approach allows to expand the calculation of the costs of an educational disruption proposed by Neidhöfer et al. (2021) to consider the immediate costs and the cumulative ones.

However, in contrast with Kaffenberger and Pritchett (2020b, 2021) and Neidhöfer et al. (2021), I model parental interventions as constantly occurring throughout the learning trajectory of children following recent research on the economics of parenting (Doepke et al., 2019). The integrated framework also allows me to analyze how different policy interventions can interact with parental interventions to attenuate the effects of the disruptions. Thus, the model provides a framework for the policymaker to opportunely assess the effects of a shock and the policy alternatives available to her.

The model presented in this paper represents an alternative to the traditional approach to modeling learning acquisition in economics, represented by Cunha et al. (2010). Although both models focus on learning accumulation instead of schooling, they take very different approaches to the subject. In the case of Cunha et al. (2010),the focus of the model is at the level of the individual, requiring vast information regarding non-cognitive and cognitive skills in the relevant population in order to be calibrated for the estimation of optimal interventions. This reduces the approach's applicability to contexts where data is more scarce, or it is only available at an aggregated level. In contrast, the model presented here requires less information regarding the individual skills while still being able to replicate the observed patterns in learning accumulation and being suitable to inform about the impacts of different interventions. These characteristics make it relevant for the case of developing countries where data constraints limit the application of formal modeling to inform the policymaking process about the long-run effects of interventions.

A growing empirical literature has documented the medium and long-run effects of an instructional disruption on educational attainment, standardized test scores and labor market outcomes. The papers in this literature find that an instructional disruption translates into lower educational attainment, scores or earnings in the absence of private or public attenuation measures. These effects are observed both when the cause of the instructional interruption is a major societal disruption (for example, the Second World War (Ichino and Winter-Ebmer, 2004) , the Chinese Cultural Revolution (Meng and Zhao, 2021), or a natural disaster(Andrabi et al., 2021; Sacerdote, 2012) and less socially disruptive causes, like a teacher strike (Jaume and Willén, 2019; Belot and Webbink, 2010) or an abnormal frequency of snowing days (Marcotte and Hemelt, 2008).

Due to data restrictions, the studies that focus on the effects of educational disruptions after five years of the event tend to rely on formal measures of schooling, such as enrollment and years of education as a proxy of the human capital or learning stock of the children affected by the shock (see, for instance, Ichino and Winter-Ebmer (2004); Belot and Webbink (2010); Meng and Zhao (2021)). A recent series of studies on the relationship between learning acquisition and schooling progression questions this assumption. Evidence from Indonesia (Beatty et al., 2021), Rwanda (Crawfurd, 2021), multiple African countries (Pritchett and Sandefur, 2020; Kaffenberger and Pritchett, 2020a) and Pakistan (Bau et al., 2021) indicates that the increase in the average number of school years in those countries has not been matched by an increase in the stock of learning of children as measured by standardized tests. Thus, this evidence suggests that the impacts of a transitory instructional disruption on the human capital of those affected are not necessarily captured in the schooling dimension, and thus they should be modeled in the learning dimension or in outcomes linked to the stock of learning, such as labor earnings as Jaume and Willén (2019) do.

The decoupling between learning and school progression is also present in the analyses by Andrabi et al. (2021), and Sacerdote (2012) on the role of responses to natural disasters on the persistence of the effects of an instructional interruption. Andrabi et al. (2021) analyze the case of the 2005 earthquake in Kashmir, Pakistan, while Sacerdote (2012) studies the effects of Hurricane Katrina in New Orleans, USA. In both cases, the authors study the effects of the natural disaster on learning (test scores) and a measure of formal educational progression (school enrollment). However, the type of response by the public sector and households differs. In Kashmir, households affected by the earthquake received a cash transfer that represented, on average, 150% of their monthly consumption. Children affected by the hurricane relocated to school districts with better quality schools regarding their average score in standardized tests in the New Orleans case. In both studies, the authors do not find any effect of the disruption and compensatory measure in their measure of formal schooling, but they find an effect in the children's performance in standardized tests. Sacerdote (2012) finds that after an initial decline, four years after the relocation due to the hurricane, the test scores of the children who changed residence improved regarding their pre-Katrina score compared with those who did not move. In contrast, Andrabi et al. (2021) find that the score of those affected by the earthquake remained below those unaffected four years after the earthquake.

The results by Andrabi et al. (2021); Sacerdote (2012) highlight the importance of considering the effects of an instructional interruption in terms of learning and not only schooling. They also show that the short-run impacts are not necessarily representative of the long-run ones, which are influenced by the responses by households and policymakers. The rest of the paper goes as follows. In the following section, I describe the formal model used to characterize how the instructional process affects learning acquisition. Next, I describe the

methodology employed to measure the short-run losses in learning due to an instructional disruption and how these losses can be introduced into the learning acquisition model to estimate the cumulative effects of the disruption. In the fourth section of this paper, I describe the data, and the calibration of the model used to estimate the potential losses caused by COVID-19 in Mexico. Afterwards, I present the application results and conclude with some final remarks and future avenues of research.

3.2 Instructional processes and learning acquisition.

Modeling the impacts of an instruction disruption on learning requires a model of how educational instruction affects learning acquisition during an educational career. Identifying the impact of instruction on learning has been the subject of interest of a vast literature on Educational Production Functions. (EPF). EPF, as their name indicates, consider learning as the output of a multi-input production process, establishing the marginal contribution of each of the inputs to the learning acquired by the student[1].

The most common formulations of the EPF consider four groups of variables as determining the stock of learning (A_{it}) that a child i has accumulated up to moment t: family background factors (B_{it}), peer factors (P_{it}), school-related factors (S_{it}), and the innate affinities and abilities of each child (I_{it}). The stock of each of these factors determines the stock of learning of person i up to moment t. Formally, this can be expressed in a general form as follows:

[1]The origin of the Educational Production Functions literature is the Coleman Report (Coleman et al., 1966) commissioned by the U.S. Congress to identify the association between different elements of a student background and her performance in school. For a review of the early literature see Bowles (1970); Hanushek (1979) and for more recent studies Hanushek (2002, 2010); Todd and Wolpin (2003)

$$A_i(t) = f(B_i(t), P_i(t), S_i(t), I_i(t)) \tag{3.1}$$

From equation 1 it is possible to derive the formulation corresponding to the learning gains occurred between two periods. Equation 1 implies that any gains in learning between t and $t-1$ occur due to a change in the input stocks. Let Δ represent the change in a given variable between t and $t-1$, this can be formally expressed as follows:

$$\Delta A_i = f\Big(B_{i,t}, P_{i,t}, S_{i,t}, N_{i,t}\Big) - f\Big(B_{i,t-1}, P_{i,t-1}, S_{i,t-1}, N_{i,t-1}\Big) \tag{3.2}$$

From equation 1 we know that the learning observed in the previous period ($a_{i,t-1}$) is a sufficient statistic of the history of values of the inputs, transformed according to the learning technology $f(\cdot)$ (Todd and Wolpin, 2003). Thus, the level of learning observed at time t, namely $a_{i,t}$ can be expressed as the sum of the current values of the inputs ($B_{it}, P_{it}, S_{it}, I_{it}$), plus the learning observed in the previous period $a_{i,t-1}$, which represents the learning stock up until period $t-1$ This is the value-added formulation of an Educational Production Function.

$$a_{i,t} = f(B_{it}, P_{it}, S_{it}, I_{it}) + a_{i,t-1} \tag{3.3}$$

A limitation of EPFs is that they do not provide an explicit model of the mechanisms behind learning accumulation. Specifically, they do not specify the different mechanisms through which each input influences learning acquisition, nor how those mechanisms might change in the face of a temporary shock. A way to address this limitation of EPF is through a structured model of learning accumulation that explicitly represents these mechanisms. The Potential Pedagogical Function proposed by Kaffenberger and Pritchett (2021, 2020b) is such a model.

The Potential Pedagogical Function (PPF) models the interplay between the expected process of learning accumulation implied in school curricula and the preconditions of the students that interact with it. By explicitly modeling the relationships between the different inputs of the learning production process, the PPF allows the analysis of the effects of changes in those relationships and the level of the inputs on the learning accumulation process. This property makes it an ideal starting point for the construction of a model of the effects of instructional interruptions.

Following Kaffenberger and Pritchett (2020b, 2021), the first element of the PPF is the learning profile (LP) implied by the school curriculum of a specific school grade. The learning profile represents the learning gains (in this case, in terms of points in a standardized test) that a student is expected to acquire by attending school grade G, and it is a function of the learning stock accumulated by students up until the end of grade G-1. It is defined by a coverage width (w), the median of prerequisite knowledge (π^G), the minimum and maximum expected learning gains (z_{min}, z_{max}), and the learning gradient in school grade G (v). The educational authorities define all these elements in the curriculum for a specific school grade.

The coverage width of school grade G (w) represents the domain upon which the learning

profile is effective. In other words, it is the set of values of the learning stock upon which the attendance to school grade G will lead to students learning something new. The width is centered at a specific value of previous knowledge (π^G), which represents the learning stock at the end of school grade G-1 of the median student in the school system. The minimum z_{min} (maximum z_{max}) expected learning gains correspond to the learning gains that a student with the minimum (maximum) amount of previous knowledge on the subject will obtain during school grade G. The gradient indicates the gains above the minimum learning gains that a student with a learning stock above the minimum will gain during the school grade G. It is defined by the combination of the width, minimum, and maximum learning gains define the learning gradient of the school grade (v) as follows:

$$v = \frac{z_{max} - z_{min}}{w} \tag{3.4}$$

All these elements, along with the learning stock of students at G-1, $A_{i,G-1}$, define the learning profile, and consequently the expected learning gains according to the curricula during grade G ($a^G_{i,G}$) as follows:

$$LP(w, z, v, \pi^G) = a_{i,G} = z_{min} + v\Big(A_{i,G-1} - \Big(\pi^G - \frac{w}{2}\Big)\Big) \tag{3.5}$$

Equation 3.5 is equivalent to an EPF in its value-added form (equation 3.3) from the policymaker's point of view that designed the curricula of school grade G. However; it is not

necessarily equal to the observed EPF. As the evidence on the decoupling between schooling learning gains summarized in the previous section indicates, the EPF implied in the curricula of school grade G is not necessarily the one that is effective for all students attending that school grade.

The PPF formalizes this empirical insight by establishing that only the children with a stock of knowledge in the domain of the learning profile (equation 5) will gain any learning from attending school grade G. From the definition of the learning profile for grade G we know that its domain is the set $A_{i,G-1} \in \left[\pi^G - \frac{w}{2}, \pi^G + \frac{w}{2}\right]$, which implies that children with a stock of knowledge $A_{i,G-1} < \pi^G - \frac{w}{2}$ or $A_{i,G-1} > \pi^G + \frac{w}{2}$ will not learn anything while attending school grade G. Informally, this means that students without the prerequisites for course G will not learn anything if they were to attend and complete grade G, while those with a knowledge level above that of the contents of grade G will also not learn anything from attending grade G.

With these elements, it is possible to define the PPF as a function from the learning stocks of children at G-1 to the effective learning gains at the end of G ($F : A_{i,G-1} \rightarrow a_{i,G}$) with the following form.

$$PPF(LP(\cdot), A_{i,G-1}) = a_{i,G} = \begin{cases} 0 & \text{if } A_{i,G-1} < \pi^G - \frac{w}{2} \\ z_{min} + v(A_{i,G-1} - (\pi^G - \frac{w}{2})) & \text{if } \pi^G - \frac{w}{2} \leq A_{i,G-1} \leq \pi^G + \frac{w}{2} \\ 0 & \text{if } A_{i,G-1} > \pi^G + \frac{w}{2} \end{cases} \tag{3.6}$$

The PPF describes the effective EPF of school grade G in which the learning gains experienced by a student during a school grade depend on the learning stock accumulated up to the beginning of that grade. The non-linear gains allow the model to reproduce the decoupling between school and learning progression observed in the empirical data.

So far, this model only characterizes the learning gains from a school year. In order to replicate the process of learning accumulation through a school career, it is necessary to include the progression in the center of the learning profiles of different school grades. Kaffenberger and Pritchett (2021, 2020b) propose doing so by including a "shift" parameter ρ that represents the expected learning gains of the median student according to the school curricula. Formally, this implies that from school grade G to school grade $G+1$, the boundaries of the width of the learning profile increase according to the following equations:

$$(\pi^G + \rho) - \frac{w}{2} = \pi^{G+1} - \frac{w}{2} \tag{3.7}$$

$$(\pi^G + \rho) + \frac{w}{2} = \pi^{G+1} + \frac{w}{2} \tag{3.8}$$

By modeling learning progression in this way, the PPF proposed by Kaffenberger and Pritchett (2021, 2020b) implies that the only input relevant to learning accumulation is schooling. In the context of EPFs, Todd and Wolpin (2003) suggest that the score of a standardized test at the end of a school grade is a sufficient statistic of the learning attained during the school grade. That is, the test score $a_{i,G}$ summarizes all the relevant information on the educational inputs invested during school grade G and the previous stock of learning

of student i. This property, however, no longer holds in the case of a structured model such as the PPF. That is because a structured model seeks to explicitly model the mechanisms through which each input influences the outcome variable. This implies that non-modeled mechanisms are assumed to be non-existent. Thus, the description of learning accumulation in the PPF, as proposed by Kaffenberger and Pritchett (2021, 2020b), ignores the role of parental investments during a child's academic career, contradicting much of the literature on parenting and human capital investments (Doepke et al., 2019).

I supplement this omission in the basic PPF model by explicitly modeling parental investments in their children's human capital, subject to the resource constraint of the parental household in the spirit of Ben-Porath (1967); Leibowitz (1974) and Todd and Wolpin (2003). Specifically, I model parental investments as a compensatory measure to guarantee that a child remains in the domain of the learning profile of the school grade she is attending, thus replicating the empirical finding that parental investments are complements to schooling (Heckman and Mosso, 2014).

In this framework, the parents' objective is to guarantee that their child has a knowledge stock sufficiently high enough to be in the domain of the learning profile of the school grade in which the child is enrolled. Their restriction is the amount of educational and economic resources available to them.

This restriction can be expressed as the weighted sum of the educational and economic resources of the household. Let e_i be the educational attainment in school years of the parents of child i, and w_i the economic resources in the household of child i. Then, the total amount of educational resources of household i can be expressed in relative terms as the ratio between the observed attainment e_i and the maximum attainment in the population,

$\max(e)$. Similarly, the total amount of economic resources in the household expressed in relative terms is the ratio of the difference between the maximum amount of resources $\max(w)$ and the amount of household i , w_i and the range of the economic resource distribution, $\max(w) - \min(w)$. Then, the total amount of resources in household i is defined as follows.

$$\beta_i = \theta \frac{e_i}{\max(e)} + (1-\theta)\left[1 - \frac{|\max(w) - w_i|}{|\max(w) - \min(w)|}\right] \tag{3.9}$$

Equation 3.9 is similar to what Neidhöfer et al. (2021) consider the attenuation capacity of households. However, a major difference between equation 3.9 and Neidhöfer et al. (2021) is that the former considers both educational and economic resources, following the literature on the determinants of parental investments (Heckman and Mosso, 2014; Prix and Erola, 2017). In contrast, Neidhöfer et al. (2021) only consider the role played by educational resources.

Next, define $A^*_{i,G-1}$ as the stock of learning of child i before the beginning of grade G. From equation 1, this stock is a function of schooling, parental investments, peers and innate skills. Assuming that peer effects occur in the environments where schooling and parental investments take place and that innate ability is predetermined, the reduced form of $A^*_{i,G-1}$ is given by:

$$A^*_{i,G-1} = f(B_{i,t-1}, S_{i,t-1}) \tag{3.10}$$

Given that parents seek to minimize the distance between the knowledge stock of their

children and the median knowledge stock in the domain of the learning profile of grade G, the optimal investment they have to perform is equivalent to the distance between π^G and the stock of knowledge at the end of grade $G-1$, $A_{i,G-1}$ penalized by the number of resources available to the household. Thus, the stock of knowledge at the beginning of grade G will be given by:

$$A^*_{i,G-1} = A_{i,G-1} + \beta|\pi^G - A_{i,G-1}| \tag{3.11}$$

Redefining equation 8 to be in terms of the knowledge stock before the start of grade G, $A^*_{i,G-1}$, gives the following PPF, in which both the role of schooling and parental investments are explicitly modeled:

$$PPF(LP(\cdot), A^*_{i,G-1}) = a_{i,G} = \begin{cases} 0 & \text{if } A^*_{i,G-1} < \pi^G - \frac{w}{2} \\ z_{min} + v(A^*_{i,G-1} - (\pi^G - \frac{w}{2})) & \text{if } \pi^G - \frac{w}{2} \leq A^*_{i,G-1} \leq \pi^G + \frac{w}{2} \\ 0 & \text{if } A^*_{i,G-1} > \pi^G + \frac{w}{2} \end{cases} \tag{3.12}$$

Finally, the accumulation of the learning for child i is given by the following equation.

$$A_{i,G} = A^*_{i,G-1} + a_{i,G} \tag{3.13}$$

The last parameter in the model is the dropout rate. Following Kaffenberger and Pritchett (2021, 2020b), the dropout rate is modeled as a function of the knowledge stock of the children. Specifically, the assumption is that the lower tail of the distribution of the knowledge stock drops out of school at a rate consistent with the observed dropout rates in the educational system. Let ϕ^G be the grade-specific cut-off value of the knowledge stock distribution below which students drop out., then the dropout function will be defined as

$$q(A_i) = \begin{cases} 1 & \text{if } A_i \leq \phi^G \\ 0 & \text{if } A_i > \phi^G \end{cases} \tag{3.14}$$

Together, equations 3.8, 3.9, 3.11, 3.12, 3.13 and 3.14 form the learning progression model used in the rest of the paper.

3.3 Estimating the costs of an instructional disruption

In general terms, an educational disruption can have two types of costs: immediate and cumulative costs. The first set corresponds to the share of a school year lost because in-person instruction was suspended due to the occurrence of an external event (the underlying cause of the disruption). This share is equivalent to the share of the learning of a school year lost due to the in-person class suspension. The cumulative costs are those associated with the lags in the learning progression that a student might experience at further stages of her academic

career due to the disruption of in-person instruction. This second set of costs occurs due to the immediate costs that might displace the student from the domain of the learning profile of the school grade she is attending. This would imply that the student fails to obtain the expected learning gains from that course and starts to lag behind the learning profile of the subsequent school grades.

In the following subsections, I delve into the characterization of both types of costs.

3.3.1 Estimating the immediate learning cost

To estimate the immediate costs, I generalize the approach proposed by Neidhöfer et al. (2021). This approach considers three significant components of the immediate costs of a learning disruption faced by a child. The first two components are defined at the lowest geographical level of aggregation for which data is available. The third allows translating the costs from the aggregate level to the individual level.

The first component is the gross cost of the disruption, O. These costs are the sum of the share of the school year during which in-person classes did not occur, $\frac{d}{D}$, and the days lost due to the incidence of the underlying cause of the instructional interruption, expressed as a share of the school year. This second component is the product of the days lost due to the underlying cause of the disruption, λ, and the probability of being affected by it, Ω. Thus, the expected value of the gross costs of the disruption is given by:

$$E(O) = \frac{d}{D} + \frac{\lambda \times \Omega}{D} \tag{3.15}$$

The second component of the immediate costs is the government's attenuation measures to diminish the impact of the instructional interruption Z. Following Neidhöfer et al. (2021), I model this as the product between their effectiveness in substituting a day of in-person instruction γ, and the probability of access to these attenuation measures, θ. I express these measures in terms of the share of the school year in which in-person instruction did not occur. The following equation gives the expected value of these public attenuation measures:

$$E(Z) = \frac{d}{D}(\gamma \times \theta) \tag{3.16}$$

With the gross immediate costs and the public attenuation measures, it is possible to construct the expected value of the net immediate costs faced by each household, k_r as follows

$$E(k_r) = E(O) - E(Z) = \frac{d + \lambda \times \Omega \ - \ d(\gamma \times \theta)}{D} \tag{3.17}$$

The immediate net costs are assumed to be homogeneous across the region of analysis. What produces differences at the individual level is the attenuation capacity of households. This is the household's capacity to diminish the effect of the immediate net costs on the children's learning process. Neidhöfer et al. (2021) define it as a negative function of the educational resources of the parents of the children. However, their formulation leaves aside

the role of economic resources as enabling households to diminish the effects of a shock on a student's learning process. Following the formulation of total resources of a household in equation 3.9, in equation 3.18, I define the attenuation capacity of the household as a negative function of the total resources in the household:

$$1 - \beta_i = 1 - \left(\theta \frac{e_i}{\max(e)} + (1 - \theta) \left[1 - \frac{|\max(w) - w_i|}{|\max(w) - \min(w)|} \right] \right) \tag{3.18}$$

The effective learning cost experienced by child i, C_i, will depend on the attenuation capacity of her household and the immediate net costs experienced in the region. I define the effective learning cost as the share of the learning gains from a school grade lost due to instructional disruption. Formally the expected value of the effective learning costs is given by :

$$E[C_i] = (1 - \beta) \times E(k_r) = (1 - \beta) \times \frac{d + \lambda \times \Omega \ - \ d(\gamma \times \theta)}{D} \tag{3.19}$$

Equation 3.19 represents the immediate loss of learning experienced by children i in terms of the share of learning from a school grade that the students do not obtain due to the instructional disruption. This cost is included in the model of learning accumulation developed in the previous section to calculate the cumulative cost of the disruption. This calculation is explained in the following subsection.

3.3.2 Estimating the cumulative learning costs of an instructional disruption

The first step to estimate the cumulative effects of a transitory instructional disruption, is to introduce the immediate costs into the model of learning accumulation. Assuming that the transitory shock occurred when child i attended school grade G, the accumulation of learning described in equation 12 will be affected by costs $E[C_i]$. This means that the stock of learning at the end of grade G will be given by:

$$A_{i,G} = A^*_{i,G-1} + (1 - E[C_i])\, a_{i,G} \tag{3.20}$$

Using equation 3.20 as the accumulation rule at grade G and afterwards employing the model of learning progression defined by equations 3.8, 3.9, 3.11, 3.12, 3.13 and 3.14 for the rest of the school grade allows simulating the learning trajectory under the effects of the instructional shock. If the end of the learning trajectory is grade T, then the stock of learning resulting from this trajectory for child i will be $\widehat{A_{i,T}}$.This is the stock of learning under the effects of the interruption. The model defined by equations 3.8, 3.9, 3.11, 3.12, 3.13 and 3.14 can then be used to construct a counterfactual trajectory in which the shock never took place. The resulting learning stock is denoted by $A_{i,T}$.

Similarly, it is possible to define a counterfactual trajectory that considers the learning progression expected by the educational system. Equation 3.8 define this progression through the parameter ρ, which defines the learning gain for each school grade implicit in the progression of the school curricula. Thus, for T school grades the stock of learning accumulated at the end of this learning progression will be defined by

$$\bar{A_{i,T}} = T\rho \tag{3.21}$$

With these three learning stocks, it is possible to estimate the cumulative learning costs of an instructional disruption. The effective cumulative learning cost (ECC) is estimated as the difference between the learning stock at T of the trajectory without a shock, $A_{i,\ T}$, and the learning stock of the trajectory that considers the effect of the shock, $\widehat{A_{i,\ T}}$. This difference is expressed in terms of the learning progression from one grade to the next one implied in the school system curricula ρ. Thus, the difference is measured as the years of learning progression lost due to the shock with respect to a scenario without a shock. Formally,

$$ECC = \frac{\widehat{A_{i,T}} - A_{i,T}}{\rho} \tag{3.22}$$

It is also possible to estimate the cumulative cost with respect to the median progression implied by the school system curricula (SCC). The SCC is the difference between the learning stock at T of the trajectory that considers the effect of the instructional shock and the learning stock accumulated under the learning progression of the system given by ρ. As before, the difference is expressed in terms of ρ, which allows interpreting the result as the years of learning progression lost due to the shock with respect to the expected progression of the system. Formally, this is

$$SCC = \frac{\widehat{A_{i,T}} - \bar{A_{i,T}}}{\rho} \tag{3.23}$$

In the following section, I provide an empirical illustration of the model for estimating the immediate and cumulative costs of the instructional disruption in Mexico due to the COVID-19 Pandemic. The empirical illustration seeks to highlight the role of the additions to the basic model and the robustness of the estimations to changes in several key parameters.

3.4 An application: Mexican children during the COVID-19 Pandemic

3.4.1 Data and Calibration of the Learning Accumulation Model.

The first step in bringing the model described by equations 3.8, 3.9, 3.11, 3.12, 3.13 and 3.14 to produce an estimate of the potential learning effects of the COVID-19 pandemic in Mexico is to calibrate the learning progression model to the observed data on learning acquisition in the country. For this exercise and following previous exercises (Kaffenberger and Pritchett, 2020b, 2021), I use as a measure of learning acquisition the math scores of a nationally implemented standardized test. For this exercise, I use the average test score and standard deviation of the 2017 math component of the *National Plan for the Evaluation of Learnings* (Programa Nacional para la Evaluación de los Aprendizajes, PLANEA in Spanish) for students in ninth grade. This test has information at both the national and regional level, allowing me to calibrate a PPF for each one of the regions of the country[2]. In addition, due

[2]Given the data restrictions from the *Espinosa Rugarcia Social Mobility in Mexico Survey 2017* I am restricted to only consider five aggregate regions of the country composed in the following way: the North

to institutional changes in the country, the test from 2017 is the most recent information on math learning for students in ninth grade.

The model has four parameters to be calibrated: ρ, z_{min}, z_{max} and w. For the value of w, I follow Kaffenberger and Pritchett (2020b), who assume that the learning profile covers all students enrolled in the system at the beginning of their educational trajectory. This assumption, although extreme, implies overestimating the coverage of the learning profile, thus implying an underestimation of the impacts of an instructional disruption. This leaves three parameters to be calibrated: ρ, z_{min}, and z_{max}

In the Mexican context, the parameter ρ is defined at the national level, as the federal ministry of education determines the school curricula for primary and secondary education. Hence, to obtain its value, I calibrate a PPF at the national model by selecting the values of ρ, z_{min}, and z_{max} that produce a learning trajectory that at the end of ninth grade produces the values of the average score and standard deviation as close as possible to those observed in the standardized test. With this value of ρ, I calibrate the values of, z_{min}, and z_{max} for each region using the same methodology as before: selecting the values of z_{min}, and z_{max} that produce a learning trajectory that in ninth grade replicates as close as possible the observed average regional mean score and the standard deviation of the regional scores. Table 1 shows the calibrated parameter values and the comparison between the observed scores and those derived from the calibrated models.

To calculate the economic and educational resources of the household (equation 3.9), I use

region consists of Baja California, Sonora, Chihuahua, Coahuila, Nuevo León and Tamaulipas; North West consists of Baja California Sur, Sinaloa, Nayarit, Durango and Zacatecas; the Center North region is form by Jalisco, Aguascalientes, Colima, Michoacán and San Luis Potosí; the Center region is formed by Guanajuato, Querétaro, Hidalgo, Estado de México, ; Mexico City, Morelos, Tlaxcala, and Puebla; the South region is formed by Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán and Quintana Roo.

Table 3.1. Parameter values for the Mexican case

	North	North West	Center North	Center	South
Average math score in PLANEA 2017 (Total sample)	492	494	513	511	483
Average math score (Model)	483.824 (0.5367)	492.5117 (0.7484	497.2913 (0.4891)	503.016 (0.4660)	466.338 (1.2817)
w	153	153	153	153	153
$z_{\min}$	37.7	37.7	38.4	38.4	36.5
$z_{\max}$	71.63	71.63	72.96	72.96	69.35
ρ	54	54	54	54	54
v	0.2218	0.2218	0.2259	0.2259	0.2147
Initial distribution	N(0,20)	N(0,20)	N(0,20)	N(0,20)	N(0,20)

the data from the *Espinosa Rugarcia Social Mobility in Mexico Survey 2017* (ESRU-EMOVI 2017). This survey is designed to measure social mobility in Mexico and its regions and consequently has extensive information about the Mexican population's living conditions and parental characteristics. Furthermore, the survey is representative at the national and regional level for the population between 25 and 64 years old, recovering information about the conditions of the households they inhabited when 14 years old. However, I am only interested in the information of the youngest cohort (25 to 30 years old) as the conditions of their households of origin are closest to the characteristics of the current Mexican households. Consequently, the sample employed diminishes from 17,665 to 2,474 observations.

It is important to note that although the conditions of these households are not necessarily equivalent to the conditions of current Mexican households, they represent a sample of households of which ample information about conditions of origin is available and that does not suffer from co-residence bias. Table 2 shows the sample's composition in terms of sex, indigenous status, urban residence, and the years of school attainment of the respondent and both parents.

Table 3.2. Descriptive Statistics of ESRU-EMOVI 2017.

Variable	National	North	North West	Center North	Center	South
Years of school of interviewee	11.8453	11.4828	12.2753	11.6716	12.5643	10.9050
(Regional mean)	(0.1291)	(0.2009)	(0.2578)	(0.2989)	(0.2396)	(0.3074)
Years of school of the father	7.5955	8.1953	6.7974	7.4127	8.8788	5.4875
(Regional mean)	(0.2115)	(0.2568)	(0.3377)	(0.3669)	(0.4235)	(0.3393)
Years of school of the mother	7.2009	8.1953	7.1173	7.2323	8.2746	5.0322
(Regional mean)	(0.1896)	(0.2473)	(0.3714)	(0.3350)	(0.4236)	(0.3308)
Female population	0.5219	0.5307	0.5139	0.5217	0.5160	0.5284
(Share of regional population)	(0.0141)	(0.0242)	(0.0321)	(0.0352)	(0.0273)	(0.0264)
Urban community of origin	0.7297	0.8955	0.6048	0.7135	0.8374	0.4953
(Share of regional population	(0.0327)	(0.0244)	(0.0566)	(0.0449)	(0.0298)	(0.0411)
Indigenous population	0.1085	0.0525	0.0254	0.0641	0.0656	0.2618
(Share of regional population)	(0.0123)	(0.0161)	(0.0146)	(0.0169)	(0.0149)	(0.0326)
Regional population		0.1588	0.0673	0.1310	0.4137	0.2292
(Share of national population)		(0.0184)	(0.0098)	(0.0163)	(0.0361)	(0.0213)

Notes: Data from ESRU-EMOVI 2017 for respondents between 25 and 30 years old. Standard errors in parenthesis.

To measure the educational resources in the household, I use the average years of education of both parents or those of the present parent in single-parent households. In this, my measure of educational resources is the same as that employed by Neidhöfer et al. (2021). In the case of economic resources, I use a household asset index to measure them. Household asset indices have been employed in the development literature for analysis regarding the distribution of economic resources when other variables such as income are not available (see Filmer and Pritchett (2001); McKenzie (2003); Wittenberg and Leibbrandt (2017); Poirier et al. (2020)). In this case, as ESRU-EMOVI 2017 only records the ownership of an asset, I employ Multiple Correspondence Analysis (MCA) in the construction of the index. MCA uses relative frequencies across the binary variables considered to identify an underlying structure across the observations in terms of ownership of the assets. This underlying structure can produce a rank of the individuals in the sample according to the amount and type of assets they possess. Table 3 shows the assets employed in the construction of the index. With this data on educational and economic resources, I calculate equation 3.9 for all individuals in the sample.

Table 3.3. Binary variables for the parental household asset index.

The household has access to the water supply	The household has a washing machine
The household has an oven	The household has a landline telephone
The household has a television	The household has a computer
The household has a refrigerator	The household has a VHS
The household has a microwave	The household has cable television
The household owns a water heater	The household owns a vacuum cleaner
A member of the household owned the housing facilities inhabited	A member of the household owns a car
A member of the household has a bank account.	A member of the household owns a credit card.
The household hires a domestic worker.	

3.4.2 Calculation of the immediate costs

The first step to calculate the immediate learning costs of the COVID-19 pandemic is to calculate the net costs as defined by equation 15 above. The net costs are composed of the direct costs COVID-19 in terms of days of in-person instruction lost and the effect of the attenuation measures deployed by the government. Moreover, the direct costs of COVID-19 consist of the school days under remote instruction, d, plus the expected value of the days lost due to the incidence and mortality of COVID-19. τ_r, both expressed as a share of the days in a school year (190 in the Mexican case). Formally, this is

$$E\left[O_r\right] = \frac{d + \tau_r}{D} \tag{3.24}$$

In the case of the COVID-19 pandemic, the interruption of in-person instruction lasted the last third of the 2019-2020 school year (60 days) and the whole 2020-2021 school year. Thus, we calculate equation 22 for two school years, where in the first one $d = 60$ and in the second one $d = 190$. Let $P_r\left(s = 1\right)$ be the probability of a household member being sick of

COVID-19 and $P_r(m=1)$ be the probability that a household member died of COVID-19 for inhabitants of region r and define $\tau^s = 5$ as the number of school days lost due to sickness and $\tau^m = 15$ the school days lost due to the mourning of the death of a household member. Then the expected value of the days lost due to the incidence and mortality of COVID-19 will be given by equation 3.25 for inhabitants of region r

$$\tau_r = \tau^s \times P_r(s=1) + \tau^m \times P_r(m=1) \tag{3.25}$$

In which the values for τ^m and τ^s are taken from Neidhöfer et al. (2021). The probabilities $P_r(s=1)$ and $P_r(m=1)$ are calculated with the data for the cumulative incidence and cumulative mortality of COVID-19 in region r from the official government records on the spread of COVID-19 in Mexico available at `https://datos.covid-19.conacyt.mx/`. I calculate the probabilities for each school year affected by the pandemic.

The second component of the net costs is the effect of the attenuation measures deployed by the government to reduce the impact of the instructional displacement on learning ($E(Z)$ in equation 3.16). In the case of the COVID-19 pandemic, the public attenuation measures taken by the government were two. The first was a program of remote instruction implemented through televised classes in which the course syllabus materials were covered (Ramírez-Raymundo et al., 2021). This was to act in tandem with the second component, which was the displacement of all instruction to a remote instruction model through internet-based platforms and telephone contact between parents, children and teachers (Diario Oficial de la Federación, 2020). However, the government did not expand the coverage of the internet,

t.v. or telephone services.

A difficulty in modeling these two interventions is that there is no data on their effectiveness as substitutes for one day of in-person instruction. This limitation forces me to make different assumptions regarding their effectiveness. Let α be the share of in-person instruction that both interventions are able to substitute jointly and define ϕ and δ as the effectiveness parameter of each intervention (δ for the televised transmissions, ϕ for the remote instruction model) such that $\phi + \delta = \alpha$. I simulate three cases $\alpha = 0$, $\alpha = 0.5$, and $\alpha = 1$. As stated in equation 3.16, this substitutability capacity has to be weighted by the probability of access to the interventions by the relevant population, in this case, the population of region r. The relevant probabilities are, in this case, the probability that a household has access to a digital television, γ_r, the joint probability that a household has a computational device and internet access, κ_r, and the probability that the student remained in contact with their teachers λ. Thus, the formal expression of these attenuation policies is in equation 3.26.

$$E\left[Z_r\right] = \frac{d\left[(\delta \times \gamma_r) + (\psi \times \kappa_r \times j)\right]}{D} \tag{3.26}$$

I calculate the probabilities γ_r and κ_r using information from the Mexican Population Census from 2020. At the same time, I calculate the probability λ with the information from the ENCOVID-ED, a special telephone survey designed to gather information on the conditions under which remote learning took place (Instituto Nacional de Estadística y Geografía, 2021).

Substituting equations 3.24 and 3.26 into equation 3.19 provides the equation that I use to calculate the immediate costs of the pandemic as a share of a school year of learning. Formally, this is :

$$E\left[C_i\right] = (1 - \beta_i) \times \frac{d + \tau_r - d\left[(\delta \times \gamma_r) + (\psi \times \kappa_r \times \lambda)\right]}{D} \tag{3.27}$$

Table 3.4 shows the average immediate costs of the pandemic at the national level under the three scenarios considered for the parameter α. In the best case scenario, when $\alpha = 1$, the average immediate cost at the national level represents a third of a school year of learning. In the worst case scenario, when $\alpha = 0$, the costs represents a 0.9583 of a school year, where the potential maximum is of 1.3 years of learning. A first pattern observed in the data is that the assumption made about the remote instructional model's effectiveness in substituting in-person instruction plays an important role in the size and distribution of the cost. The larger the assumed effectiveness, the larger is the regional difference in the average immediate learning cost. For instance, if the remote instruction learning model is assumed to be a perfect substitute for in-person learning ($\alpha = 1$), the average immediate costs in the South represents 1.79 times the cost in the Center. On the other hand, if the substitutability is assumed to be null between those two modes of instruction ($\alpha = 0$), the cost in the South represents 1.18 times the cost in the Center.

This pattern is a consequence of the interventions' effectiveness and the inequalities in the probability of access to them. As Campos-Vázquez and Monroy-Gómez-Franco (2016) and **?** show, the South region of the country is the region with lower social mobility rates in absolute terms and with higher levels of multidimensional poverty in Mexico. Consequently,

Table 3.4. Average effective immediate cost in each scenario.

Region	Average effective immediate cost ($\alpha = 1$)	Average effective immediate cost ($\alpha = 0.5$)	Average effective immediate cost ($\alpha = 0$)
National	0.2964	0.6273	0.9583
	(0.0052)	(0.0066)	(0.0085)
North	0.2208	0.5695	0.9183
	(0.0022)	(0.0056)	(0.0091)
North West	0.2957	0.6349	0.9741
	(0.0047)	(0.0102)	(0.0156)
Center North	0.2619	0.6029	0.9440
	(0.0039)	(0.0090)	(0.0141)
Center	0.2490	0.5789	0.9089
	(0.0038)	(0.0089)	(0.0140)
South	0.4469	0.7598	1.0726
	(0.0063)	(0.0108)	(0.0152)

Note: Authors' calculations corresponding to equation 3.27. The effective immediate cost corresponds to the share of a school year of learning lost due to the transition to remote learning. Clustered standard errors in parenthesis. The clusters are defined at the primary sampling unit.

the probability of access to the policy interventions is the smallest in the south of the country compared to other regions (see table A.3 in the appendix). Thus, as the effectiveness of a policy intervention increases, the greater the effect inequalities in access to them have in increasing or decreasing the cost of the instructional interruption. In other words, interventions that are highly effective in attenuating the cost of the instructional interruption increase the opportunity cost of not having access to them. This highlights an important factor in policy: even if the policy implemented is a perfect substitute for in-person instruction, gaps in access to the policy have to be closed for the policy to attenuate the learning cost of the disruption.

Table 3.4 also highlights another factor: the attenuation capacity of households can, on average, diminish the immediate cost of the disruption by 26% of the potential maximum. This is calculated as the share of the potential maximum effective immediate cost (1.3 of a school year of learning in this case) that the difference between the maximum possible loss and the average effective cost when $\alpha = 0$ represents. The average maximum attenuation

occurs in the Center region, where on average, households can reduce effective immediate cost by 30%, whereas in the South, the private resources of the household only can attenuate 17% of the cost. This highlights the complementarity between public and private attenuation measures and the limits of a response to an instructional disruption that relies exclusively upon the households' resources.

Before proceeding to the estimation of the cumulative costs, an important question is how relevant is the inclusion of economic resources in the attenuation capacity of the household when estimating the immediate costs of an instructional cost. Table 5 shows the results of assigning different weights to the role of economic resources in the household's total resources (equation 8). I consider three scenarios: $\theta = 1$, which is the scenario Neidhöfer et al. (2021) consider, $\theta = 0.5$ which is my preferred estimation and $\theta = 0$ which is a scenario where the only determinant of household resources are economic resources. Table 3.5 shows the results of this exercise for the different values of α.

Table 3.5. Effect of different compositions of household resources
(National average effective immediate cost)

	$\theta = 1$	$\theta = 0.5$	$\theta = 0$
$\alpha = 1$	0.2827	0.5979	0.9130
	(0.0055)	(0.0080)	(0.0112)
$\alpha = 0.5$	0.2964	0.6273	0.9583
	(0.0052)	(0.0066)	(0.0084)
$\alpha = 0$	0.3101	0.6568	1.0035
	(0.0052)	(0.0061)	(0.0076)

Author's calculations. $\theta = 1$ corresponds to the case where household resources are composed exclusively of educational resources, while $\theta = 0$ represents the case where household resources are composed only of economic resources. $\alpha = 1$ corresponds to the case where the remote instruction is a perfect substitute of in person classes, $\alpha = 0$ is the case where remote instruction has no effect on learning. The effective immediate cost corresponds to the share of a school year of learning lost due to the transition to remote learning.

Table 3.5 shows that considering only the educational resources of the household as relevant ($\theta = 1$) produces the highest attenuation capacity and, consequently, the lowest effective immediate cost. In contrast, when only the economic resources of the household are considered relevant ($\theta = 0$), the result is the highest effective immediate costs. This suggests that although both types of resources are related, they do not entirely overlap. In particular, it suggests that the lower bound of the educational resources is higher than that of economic resources. This is consistent with the evidence regarding the evolution across generations of both educational attainment and economic resources in Mexico. Whereas there has been convergence in educational attainment, the same is not true regarding income or other types of economic resources. (de la Torre and Vélez-Grajales, 2016). Thus, it seems a better approach to include both types of resources when possible.

For the estimation of the cumulative costs, I consider only the specification that considers ($\theta = 0.5$)

3.4.3 Calculation of the cumulative costs

To estimate the potential cumulative costs of the COVID-19 pandemic, I simulate the process of learning accumulation in which the immediate costs estimated in equation 3.27 affects the learning acquired during sixth and seventh grade (sixth grade of *primaria* and first grade of *secundaria* in the terminology of the Mexican education system) and progresses up to ninth grade.

These simulations are done using the calibrated PPF for the Mexican case, varying the magnitude of the costs introduced in the sixth and seventh grades according to the student-specific costs estimated in the previous subsection (following equation 3.20). The

counterfactual learning stock for student i will be the mean value of the distribution produced by this simulated learning progression up to ninth grade. Then, the cumulative costs with respect a trajectory without pandemic are calculated following equation 3.22 and with respect to the school system expected trajectory using equation 3.23.

Table 3.6. Average cumulative cost with respect to baseline scenario under different assumptions

Region	Without household compensating investment			With household compensating investment		
	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 0$	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 0$
National	0.9870	1.4479	2.0505	0.1871	0.4711	0.9040
	(0.0170)	(0.0199)	(0.0178)	(0.0087)	(0.0129)	(0.0158)
North	0.8043	1.2391	1.8422	0.0773	0.1608	0.9178
	(0.0097)	(0.0099)	(0.0097)	(0.0024)	(0.0019)	(0.0183)
North West	1.0624	1.4034	2.0195	0.2154	0.4065	0.7787
	(0.0133)	(0.0177)	(0.0147)	(0.0069)	(0.0064)	(0.0154)
Center North	0.7950	1.1504	1.7660	0.1803	0.3949	0.7575
	(0.0089)	(0.0138)	(0.0253)	(0.0042)	(0.0026)	(0.0076)
Center	0.7946	1.2542	1.8789	0.1270	0.4003	0.7601
	(0.0099)	(0.0109)	(0.0099)	(0.0017)	(0.0036)	(0.0089)
South	1.5217	2.0948	2.6479	0.4622	0.8597	1.2577
	(0.0256)	(0.0254)	(0.0239)	(0.0108)	(0.0212)	(0.0354)

Author's calculations. The first three columns refer to the estimations of the cumulative costs with a PPF without parental compensation (that is a model without equation 3.11). The second three columns refer to the model that includes equation 3.11. $\alpha = 1$ corresponds to the case where the remote instruction is a perfect substitute of in person classes, $\alpha = 0$ is the case where remote instruction has no effect on learning. The effective immediate cost corresponds to the share of a school year of learning lost due to the transition to remote learning. Jackknife standard errors in parenthesis.

Table 3.7 shows two sets of results for the average cumulative costs (as defined by equation 20) under different assumptions regarding the effectiveness of the remote instruction policy adopted. The first set corresponds to a model of learning accumulation without an active role for parents in investing in their children to attenuate the gap with respect to the learning progression of the school syllabi. This is the original learning accumulation model as proposed by Kaffenberger and Pritchett (2020, 2021) without equation 20. The second set of results includes equation 18 and, as a result, consider the case in which parents invest in their

children's learning to minimise the gap with respect to the school system learning progression. This comparison allows me to examine how sensitive is the model to the inclusion of an active parental investment function, which would bring the PPF model closer to the results from the parenting literature (Doepke et al., 2019).

In the first scenario, that without parental investments, the national average of the expected cumulative cost of the instructional disruption supposed by the COVID-19 pandemic ranges between the loss of one year of learning (if $\alpha = 1$, the best-case scenario) and two years of learning (if $\alpha = 0$, the worst-case scenario) with respect to a learning progression without a shock. In the second scenario, which considers the role of parental investments throughout the learning progression, the values are 0.1871 of a school year of learning for the best-case scenario and 0.9040 of a school year of learning for the worst-case scenario. In the two scenarios, these national averages hide substantial regional heterogeneity, and in particular, the south of the country is expected to be the most affected region regardless of the scenario considered.

As expected, the cumulative cost is smaller in the model that considers parental investments throughout a child's learning trajectory. However, the role of parental investments in mitigating the widening of the initial gap caused by the instructional disruption is more apparent when comparing the relationship between immediate and cumulative costs between both models. For example, under the assumption that no parental investment occurs, when $\alpha = 1$, the cumulative costs are 3.32 times the immediate costs. Similarly, when $\alpha = 0$, the average cumulative cost at the national level is 2.1397 times the immediate cost. In contrast, when parents are assumed to actively try to attenuate the disruption throughout the children's learning progression, the cumulative costs are smaller than the immediate cost. In this case, when $\alpha = 1$ the national average cumulative cost represent s 0.6312 times the

Table 3.7. Average cumulative cost with respect to intended learning progression under different assumptions

Region	Without household compensating investment			With household compensating investment		
	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 0$	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 0$
National	0.9162	1.3771	1.9797	0.1163	0.4003	0.8332
	(0.0315)	(0.0340)	(0.0319)	(0.0252)	(0.0289)	(0.0335)
North	0.8433	1.2782	1.8812	0.0383	0.1998	0.9568
	(0.0060)	(0.0056)	(0.0048)	(0.0096)	(0.0098)	(0.0281)
North West	0.9418	1.2829	1.8989	0.0948	0.2859	0.6581
	(0.0108)	(0.0078)	(0.0071)	(0.0202)	(0.0197)	(0.0288)
Center North	0.5860	0.9413	1.5569	0.0288	0.1859	0.5484
	(0.0056)	(0.0065)	(0.0253)	(0.0124)	(0.0106)	(0.0163)
Center	0.4795	0.9391	1.5639	0.1881	0.0852	0.4450
	(0.0038)	(0.0055)	(0.0036)	(0.0083)	(0.0118)	(0.0173)
South	1.8858	2.0948	2.6479	0.8263	1.2238	1.6218
	(0.0058)	(0.0045)	(0.0048)	(0.0340)	(0.0446)	(0.0587)

Author's calculations. The first three columns refer to the estimations of the cumulative costs with a PPF without parental compensation (that is a model without equation 3.11). The second three columns refer to the model that includes equation 3.11. $\alpha = 1$ corresponds to the case where the remote instruction is a perfect substitute of in person classes, $\alpha = 0$ is the case where remote instruction has no effect on learning. The effective immediate cost corresponds to the share of a school year of learning lost due to the transition to remote learning. Jackknife standard errors in parenthesis.

national average immediate cost, whereas when $\alpha = 0$ the national average cumulative cost represents 0.9433 times the national average immediate cost.

However, it is essential to highlight that there are persistent costs of temporary instructional disruption even when parental investments are included in the model. This result is consistent with the empirical literature reviewed in the introduction and highlights the role of policy interventions in preventing this persistence. As my results indicate, parental investments can prevent the cumulative costs from increasing with respect to the immediate cost, but they are not enough to completely attenuate the cumulative effect of the temporal disruption. This is particularly true for families with fewer resources and, thus, a lower investment capacity in their children.

Table 3.8. Inequality in the magnitude of the cumulative cost
(Gini coefficient)

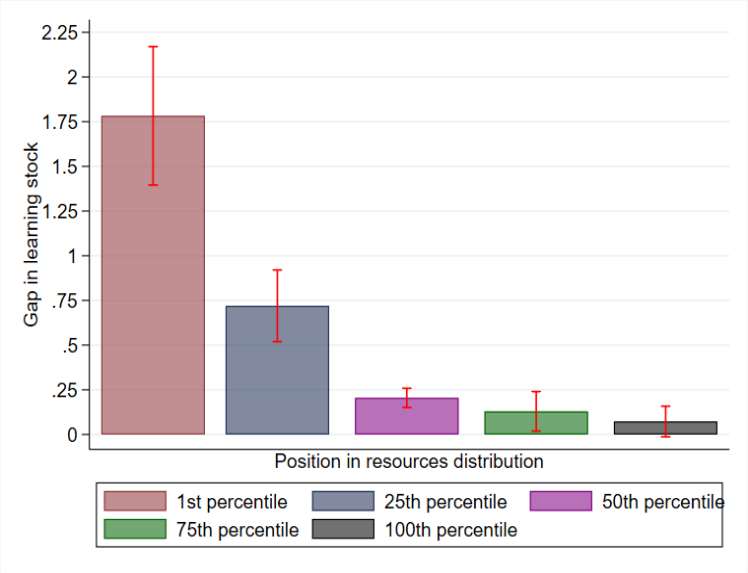
Region	Without household compensating investment			With household compensating investment		
	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 0$	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 0$
National	0.1935	0.1529	0.1015	0.3877	0.2896	0.1828
	(0.0046)	(0.0037)	(0.0026)	(0.0057)	(0.0065)	(0.0046)
North	0.1061	0.0884	0.0575	0.2863	0.1719	0.1869
	(0.0065)	(0.0052)	(0.0035)	(0.0121)	(0.0077)	(0.0098)
North West	0.0820	0.0805	0.0495	0.2187	0.1095	0.1168
	(0.0052)	(0.0044)	(0.0028)	(0.0101)	(0.0059)	(0.0051)
Center North	0.0879	0.0783	0.0555	0.1852	0.0817	0.0884
	(0.0055)	(0.0046)	(0.0028)	(0.0094)	(0.0036)	(0.0039)
Center	0.0992	0.0644	0.0402	0.1291	0.0773	0.0906
	(0.0068)	(0.0046)	(0.0027)	(0.0066)	(0.0049)	(0.0051)
South	0.1179	0.0846	0.0662	0.1629	0.1724	0.1894
	(0.0055)	(0.0037)	(0.0028)	(0.0066)	(0.0049)	(0.0051)

Author's calculations. The first three columns refer to the estimations of the cumulative costs with a PPF without parental compensation (that is a model without equation 3.11). The second three columns refer to the model that includes equation 3.11. $\alpha = 1$ corresponds to the case where the remote instruction is a perfect substitute of in person classes, $\alpha = 0$ is the case where remote instruction has no effect on learning. The effective immediate cost corresponds to the share of a school year of learning lost due to the transition to remote learning. Jackknife standard errors in parenthesis.

To analyse the effect of varying the assumptions on parental compensating investments and the efficacy of public policy, table 3.8 shows the Gini coefficient of the distribution of cumulative costs under the scenarios analysed above. As expected, the inclusion of parental investments throughout the learning progression results in a more unequal distribution of the cumulative costs than when this factor is not considered. The reason is that when this factor is excluded, the inequality in household resources only affects the immediate costs, after which it stops playing a role in the model. However, when parental investments are included, the inequality in their distribution translates directly into inequality in the cumulative costs experienced by children. The unequal impacts of the COVID-19 class suspension can be seen in figure 1. As can be seen, those who lie below the median of the economic and educational resources joint distribution are the ones who suffer the most persistent effect in terms of magnitude. In contrast, those at the top of the distri-

bution can recover from the short-run shock and do not experience a persistent negative effect.

Figure 3.1. Distribution of impact across percentiles of resource distribution



The results from my preferred specification (that which includes parental attenuation) suggest that as a result of the temporary instructional disruption, the knowledge stock of the Mexican children who experienced the shock at sixth grade will be, on average, between 0.1871 and 0.9040 school years of learning when they reach ninth grade. In other words, in the best-case scenario, at the end of their primary education cycle, the cohort who experienced the shock in sixth grade will have the learning stock equivalent to a student who missed two months of schooling in a scenario without the pandemic. In the worst-case scenario, they will have the learning stock of a child who missed almost an entire year of in-person instruction. However, this average loss in years of learning masks substantial regional heterogeneity. For the children from the Center region, the expected loss of years of learning lies between 10% (a month of in-person instruction) and 76% of a school year of learning progression (7 months of instruction). In contrast, the expected cost for the children from the South region lies between 40% of a school year of learning (about four months of

instruction) and a school year and a quarter of learning. These results already consider the role of parental effort in mitigating the ripple effects of the temporary disruption, highlighting the importance of complementary public policy aimed at reinforcing that mitigation effort.

3.5 Final Remarks

This paper's first set of implications has to do with the recent literature on models of instructional disruptions and their costs (Kaffenberger and Pritchett, 2020b, 2021; Kaffenberger, 2021; Neidhöfer et al., 2021). Firstly, this literature has sought to explicitly model the mechanism through which instructional disruptions affect learning acquisition both in the short and the long run. However, as these papers come from two different strands of the literature (development economics and education studies), their modeling strategies have remained isolated from each other. In this paper, I show that the insights from both types of models can be put together in a cohesive framework that matches the empirical findings that initially motivated the literature. Secondly, in this paper, I show that it is possible to integrate continuous parental investments into this framework, supplementing a gap in the literature and closing the gap between learning progression models based on the Potential Pedagogical Function and more traditional economic models. The inclusion of this factor does not prevent the model from replicating the persistence effect of temporary shocks into the learning acquisition process. Furthermore, the model highlights that the explicit inclusion of economic resources as part of the resources of the household for attenuating the impacts of a temporary shock provides information not captured by the educational resources, which has been the only variable used so far in the literature (Neidhöfer et al., 2021).

The second set of implications refers to the empirical application of the model. The

results suggest that the instructional disruption caused by the COVID-19 pandemic will cause a persistent learning loss with respect to the counterfactual without the disruption in the Mexican case. The magnitude of the loss varies by region and depends on the performance of the remote learning model implemented as a substitute for in-person classes. If the remote instruction model were a perfect substitute, the average loss in the country for the cohort affected would be equivalent in terms of learning gains lost to missing a month of in-person classes. However, if the remote instruction model failed, the loss would be, on average, equivalent to missing a whole year of learning. These average results are worse in the south of the country, a region that is already behind the rest of the country in learning progression and scholarly attainment. Given that these results already consider the effort performed by households in mitigating this learning loss, the policy implications are clear. In order to minimize the effect of the pandemic shock in the future stock of learning of the affected children, it is necessary to implement a remedial education policy geared towards the students who inhabit households with less resources.

Appendices

Appendix A

Chapter 1

A.1 Regionalization of the sample

Table A1. Regional composition

Region	States
North	Baja California, Sonora, Chihuahua, Coahuila, Nuevo León and Tamaulipas
North West	Baja California Sur, Sinaloa, Nayarit, Durango and Zacatecas
Center North	Jalisco, Aguascalientes, Colima, Michoacán and San Luis Potosí
Center	Guanajuato, Querétaro, Hidalgo, Estado de México, Mexico City, Morelos, Tlaxcala, and Puebla
South	Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán and Quintana Roo

A.2 Construction of the economic resources index

Table B1. Components of the origin household's economic resources index

Good/service	Value	Coordinates	Explained inertia	Good/service	Value	Coordinates	Explained inertia
Overcrowded household	0	1.207	0.022	VHS/DVD player	0	-0.595	0.001
	1	-0.726	0.013		1	2.344	0.026
Access to potable water	0	-1.640	0.036	Microwave	0	-0.322	0.004
	1	0.812	0.018		1	3.135	0.037
Stove	0	-1.662	0.036	Cable TV	0	-0.260	0.002
	1	0.816	0.018		1	2.976	0.028
Electricity	0	-2.233	0.027	Owners of another dwelling	0	-0.094	0.000
	1	0.350	0.004		1	2.368	0.009
TV set	0	-1.589	0.027	Owners of commercial venue	0	-0.123	0.001
	1	0.805	0.004		1	2.266	0.011
Refrigerator	0	-1.471	0.037	Automobile	0	-0.573	0.010
	1	0.805	0.028		1	2.080	0.037
Washing machine	0	-0.935	0.023	Bank account	0	-0.213	0.002
	1	1.905	0.028		1	3.309	0.027
Landline	0	-0.710	0.016	Credit Card	0	-0.236	0.002
	1	2.415	0.053		1	3.569	0.032
Computer	0	-0.178	0.016	Water heater	0	-0.909	0.022
	1	2.415	0.053		1	1.749	0.042
WC in house premises	0	-1.443	0.035	Owners of inhabited dwelling	0	-0.189	0.001
	1	1.025	0.025		1	0.053	0.001
Toaster	0	-0.264	0.003	Non-dirt house floor	0	-1.905	0.034
	1	3.598	0.035		1	0.576	0.010
Domestic service	0	-0.199	0.001	Vacuum	0	-0.168	0.001
	1	2.631	0.019		1	4.229	0.027
Videogame console	0	-0.244	0.002				
	1	3.317	0.030				

Table B2. Components of the current household's economic resources index

Good/service	Value	Coordinates	Explained inertia	Good/service	Value	Coordinates	Explained inertia
Overcrowded household	0	0.375	0.004	VHS/DVD player	0	-1.082	0.022
	1	-1.101	0.011		1	1.356	0.028
Access to potable water	0	-2.308	0.018	Microwave	0	-1.434	0.036
	1	0.246	0.002		1	1.442	0.036
Stove	0	-2.932	0.019	Cable TV	0	-1.157	0.023
	1	0.200	0.001		1	1.185	0.024
Electricity	0	-1.837	0.001	Owners of another dwelling	0	-0.118	0.000
	1	0.012	0.000		1	2.513	0.010
Tractor	0	-0.010	0.001	Owners of commercial venue	0	-0.098	0.000
	1	0.927	0.001		1	1.897	0.006
Refrigerator	0	-2.726	0.024	Owner of agricultural lands	0	-0.134	0.001
	1	0.286	0.003		1	1.460	0.006
Washing machine	0	-2.186	0.038	Bank account	0	-0.577	0.009
	1	0.649	0.011		1	2.110	0.033
Landline	0	-1.069	0.024	Credit Card	0	-0.503	0.007
	1	1.733	0.040		1	2.546	0.037
Computer	0	-1.004	0.024	Water heater	0	-1.305	0.026
	1	2.109	0.049		1	1.029	0.020
Internet access	0	-1.285	0.033	Owners of inhabited dwelling	0	-0.395	0.001
	1	1.824	0.047		1	0.078	0.000
Toaster	0	-0.623	0.011	Dirt house floor	0	0.066	0.000
	1	2.497	0.043		1	-2.496	0.006
Domestic service	0	-0.247	0.002	Vacuum	0	-0.429	0.006
	1	2.135	0.016		1	3.003	0.039
Automobile	0	-1.125	0.023	Tablet	0	-0.686	0.013
	1	1.276	0.026		1	2.445	0.045
Videogame console	0	-0.244	0.002	Cellphone	0	-2.183	0.023
	1	3.317	0.030		1	0.349	0.004

A.3 Regression tables

Table C1. Rank-Rank regressions
(Ranks defined over the national distribution, 50 quantiles)

Region	North	North-west	Center-north	Center	South
Rank of origin	0.507	0.431	0.495	0.609	0.633
	(0.0237)	(0.0330)	(0.0260)	(0.0286)	(0.0295)
Intercept	14.70	15.63	13.04	11.52	5.977
	(0.984)	(1.113)	(0.767)	(1.081)	(0.588)
Observations	1,830	1,355	1,886	3,403	2,616
R-squared	0.269	0.200	0.303	0.345	0.412

Notes: Clustered standard errors. Sample weights employed. Ranks correspond to 50 quantiles of the corresponding national distribution. The North region consists of Baja California, Sonora, Chihuahua, Coahuila, Nuevo León and Tamaulipas; North West consists of Baja California Sur, Sinaloa, Nayarit, Durango and Zacatecas; the Center North region is form by Jalisco, Aguascalientes, Colima, Michoacán and San Luis Potosí; the Center region is formed by Guanajuato, Querétaro, Hidalgo, Estado de México, Mexico City, Morelos, Tlaxcala, and Puebla; the South region is formed by Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán and Quintana Roo. The regressions only consider individuals who remained in the same region they were when 14 years old.

Table C2. Rank-Rank regressions
(Ranks defined over the national distribution, 20 quantiles)

Region	North	North-west	Center-north	Center	South
Rank of origin	0.508	0.433	0.497	0.613	0.639
	(0.0229)	(0.0354)	(0.0260)	(0.0278)	(0.0303)
Constant	6.011	6.373	5.348	4.654	2.421
	(0.388)	(0.470)	(0.315)	(0.431)	(0.243)
Observations	1,830	1,355	1,886	3,403	2,616
R-squared	0.270	0.200	0.302	0.347	0.410

Notes: Clustered standard errors. Sample weights employed. Ranks correspond to 20 quantiles of the corresponding national distribution. The North region consists of Baja California, Sonora, Chihuahua, Coahuila, Nuevo León and Tamaulipas; North West consists of Baja California Sur, Sinaloa, Nayarit, Durango and Zacatecas; the Center North region is form by Jalisco, Aguascalientes, Colima, Michoacán and San Luis Potosí; the Center region is formed by Guanajuato, Querétaro, Hidalgo, Estado de México, Mexico City, Morelos, Tlaxcala, and Puebla; the South region is formed by Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán and Quintana Roo. The regressions only consider individuals who remained in the same region they were when 14 years old.

Table C3. Rank-Rank regressions
(Ranks defined over the national distribution, 10 quantiles)

Region	North	North-west	Center-north	Center	South
Rank of origin 0.502	0.432	0.496	0.608	0.644	
	(0.0223)	(0.0363)	(0.0266)	(0.0298)	(0.0304)
Constant	3.169	3.335	2.811	2.450	1.286
	(0.199)	(0.237)	(0.170)	(0.230)	(0.127)
Observations	1,830	1,355	1,886	3,403	2,616
R-squared	0.261	0.197	0.297	0.342	0.408

Notes: Clustered standard errors. Sample weights employed. Ranks correspond to 10 quantiles of the corresponding national distribution. The North region consists of Baja California, Sonora, Chihuahua, Coahuila, Nuevo León and Tamaulipas; North West consists of Baja California Sur, Sinaloa, Nayarit, Durango and Zacatecas; the Center North region is form by Jalisco, Aguascalientes, Colima, Michoacán and San Luis Potosí; the Center region is formed by Guanajuato, Querétaro, Hidalgo, Estado de México, Mexico City, Morelos, Tlaxcala, and Puebla; the South region is formed by Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán and Quintana Roo. The regressions only consider individuals who remained in the same region they were when 14 years old.

Table C4. Rank-Rank regressions
(Ranks defined over the regional distributions, 50 quantiles)

Region	North	North-west	Center-north	Center	South
Rank of origin 0.507	0.416	0.512	0.564	0.559	
	(0.0266)	(0.0285)	(0.0274)	(0.0290)	(0.0263)
Constant	1.481	1.748	1.465	1.307	1.328
	(0.109)	(0.129)	(0.0842)	(0.110)	(0.108)
Observations	1,830	1,355	1,886	3,403	2,616
R-squared	0.257	0.173	0.262	0.317	0.314

Notes: Clustered standard errors. Sample weights employed. Ranks correspond to 50 quantiles of the corresponding regional distribution. The North region consists of Baja California, Sonora, Chihuahua, Coahuila, Nuevo León and Tamaulipas; North West consists of Baja California Sur, Sinaloa, Nayarit, Durango and Zacatecas; the Center North region is form by Jalisco, Aguascalientes, Colima, Michoacán and San Luis Potosí; the Center region is formed by Guanajuato, Querétaro, Hidalgo, Estado de México, Mexico City, Morelos, Tlaxcala, and Puebla; the South region is formed by Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán and Quintana Roo. The regressions only consider individuals who remained in the same region they were when 14 years old.

Table C5. Rank-Rank regressions
(Ranks defined over the regional distributions, 20 quantiles)

Region	North	North-west	Center-north	Center	South
Rank of origin	0.534	0.435	0.555	0.594	0.533
	(0.0254)	(0.0318)	(0.0257)	(0.0292)	(0.0249)
Constant	4.899	5.937	4.669	4.275	5.101
	(0.390)	(0.498)	(0.296)	(0.398)	(0.377)
Observations	1,830	1,355	1,886	3,403	2,616
R-squared	0.286	0.192	0.308	0.355	0.326

Notes: Clustered standard errors. Sample weights employed. Ranks correspond to 20 quantiles of the corresponding regional distribution. The North region consists of Baja California, Sonora, Chihuahua, Coahuila, Nuevo León and Tamaulipas; North West consists of Baja California Sur, Sinaloa, Nayarit, Durango and Zacatecas; the Center North region is form by Jalisco, Aguascalientes, Colima, Michoacán and San Luis Potosí; the Center region is formed by Guanajuato, Querétaro, Hidalgo, Estado de México, Mexico City, Morelos, Tlaxcala, and Puebla; the South region is formed by Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán and Quintana Roo. The regressions only consider individuals who remained in the same region they were when 14 years old.

Table C6. Rank-Rank regressions
(Ranks defined over the regional distributions, 10 quantiles)

Region	North	North-west	Center-north	Center	South
Rank of origin 0.537	0.433	0.544	0.591	0.547	
	(0.0260)	(0.0312)	(0.0269)	(0.0292)	(0.0252)
Constant	2.546	3.121	2.507	2.250	2.563
	(0.202)	(0.255)	(0.165)	(0.203)	(0.201)
Observations	1,830	1,355	1,886	3,403	2,616
R-squared	0.287	0.190	0.297	0.349	0.326

Notes: Clustered standard errors. Sample weights employed. Ranks correspond to 10 quantiles of the corresponding regional distribution. The North region consists of Baja California, Sonora, Chihuahua, Coahuila, Nuevo León and Tamaulipas; North West consists of Baja California Sur, Sinaloa, Nayarit, Durango and Zacatecas; the Center North region is form by Jalisco, Aguascalientes, Colima, Michoacán and San Luis Potosí; the Center region is formed by Guanajuato, Querétaro, Hidalgo, Estado de México, Mexico City, Morelos, Tlaxcala, and Puebla; the South region is formed by Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán and Quintana Roo. The regressions only consider individuals who remained in the same region they were when 14 years old.

A.4 Transition matrices

A.4.1 Transition matrices, Nationally defined quantiles

Table D1. North region transition matrix
(Quintiles defined over the national distribution)

	Q1 (Bottom)	Q2	Q3	Q4	Q5 (Top)
Q1 (Bottom)	0.234	0.322	0.264	0.117	0.064
	(0.044)	(0.039)	(0.038)	(0.020)	(0.023)
Q2	0.202	0.237	0.224	0.231	0.106
	(0.031)	(0.041)	(0.025)	(0.035)	(0.017)
Q3	0.127	0.249	0.270	0.203	0.151
	(0.022)	(0.024)	(0.024)	(0.016)	(0.026)
Q4	0.032	0.144	0.253	0.327	0.244
	(0.010)	(0.024)	(0.034)	(0.020)	(0.036)
Q5 (Top)	0.005	0.031	0.123	0.284	0.556
	(0.002)	(0.009)	(0.019)	(0.024)	(0.033)

Note: Rows correspond to the origin quantile, columns to current quintile. Sample weights employed. The North (N) region consists of Baja California, Sonora, Chihuahua, Coahuila, Nuevo León and Tamaulipas. Quintiles defined over the national distribution.

Table D2. Northwest region transition matrix
(Quintiles defined over the national distribution)

	Q1 (Bottom)	Q2	Q3	Q4	Q5 (Top)
Q1 (Bottom)	0.269	0.336	0.223	0.121	0.050
	(0.028)	(0.032)	(0.030)	(0.026)	(0.020)
Q2	0.182	0.320	0.247	0.171	0.080
	(0.030)	(0.024)	(0.044)	(0.036)	(0.023)
Q3	0.111	0.290	0.265	0.224	0.111
	(0.020)	(0.026)	(0.028)	(0.031)	(0.023)
Q4	0.015	0.155	0.252	0.286	0.291
	(0.010)	(0.032)	(0.045)	(0.050)	(0.040)
Q5 (Top)	0.068	0.079	0.179	0.206	0.469
	(0.028)	(0.020)	(0.034)	(0.028)	(0.055)

Note: Rows correspond to the origin quantile, columns to current quintile. Sample weights employed. Clustered standard errors. Cluster unit is the primary sampling unit. North West (NW) consists of Baja California Sur, Sinaloa, Nayarit, Durango and Zacatecas. Quintiles defined over the national distribution.

Table D3. Center-north region transition matrix
(Quintiles defined over the national distribution)

	Q1 (Bottom)	Q2	Q3	Q4	Q5 (Top)
Q1 (Bottom)	0.334	0.322	0.270	0.258	0.104
	(0.046)	(0.039)	(0.033)	(0.031)	(0.030)
Q2	0.188	0.358	0.288	0.114	0.052
	(0.028)	(0.032)	(0.036)	(0.028)	(0.016)
Q3	0.077	0.355	0.264	0.228	0.076
	(0.020)	(0.033)	(0.031)	(0.027)	(0.014)
Q4	0.025	0.180	0.385	0.248	0.161
	(0.009)	(0.021)	(0.034)	(0.027)	(0.027)
Q5 (Top)	0.016	0.047	0.180	0.278	0.490
	(0.003)	(0.010)	(0.027)	(0.019)	(0.033)

Note: Rows correspond to the origin quantile, columns to current quintile. Sample weights employed. Clustered standard errors. Cluster unit is the primary sampling unit. Center North (CN) region is form by Jalisco, Aguascalientes, Colima, Michoacán and San Luis Potosí. Quintiles defined over the national distribution.

Table D4. Center region transition matrix
(Quintiles defined over the national distribution)

	Q1 (Bottom)	Q2	Q3	Q4	Q5 (Top)
Q1 (Bottom)	0.446	0.273	0.180	0.070	0.032
	(0.046)	(0.045)	(0.030)	(0.018)	(0.016)
Q2	0.243	0.220	0.258	0.158	0.121
	(0.032)	(0.023)	(0.050)	(0.030)	(0.022)
Q3	0.156	0.225	0.244	0.249	0.126
	(0.039)	(0.025)	(0.026)	(0.036)	(0.022)
Q4	0.047	0.134	0.225	0.345	0.249
	(0.017)	(0.022)	(0.021)	(0.033)	(0.025)
Q5 (Top)	0.016	0.032	0.070	0.261	0.620
	(0.006)	(0.008)	(0.015)	(0.030)	(0.037)

Note: Rows correspond to the origin quantile, columns to current quintile. Sample weights employed. Clustered standard errors. Cluster unit is the primary sampling unit. Center region is formed by Guanajuato, Querétaro, Hidalgo, Estado de México, Mexico City, Morelos, Tlaxcala, and Puebla. Quintiles defined over the national distribution.

Table D5. South region transition matrix
(Quintiles defined over the national distribution)

	Q1 (Bottom)	Q2	Q3	Q4	Q5 (Top)
Q1 (Bottom)	0.672	0.218	0.078	0.027	0.006
	(0.024)	(0.018)	(0.012)	(0.006)	(0.003)
Q2	0.434	0.341	0.135	0.058	0.031
	(0.044)	(0.032)	(0.018)	(0.012)	(0.011)
Q3	0.143	0.348	0.291	0.188	0.030
	(0.022)	(0.034)	(0.033)	(0.033)	(0.008)
Q4	0.095	0.189	0.194	0.371	0.150
	(0.040)	(0.031)	(0.026)	(0.038)	(0.024)
Q5 (Top)	0.026	0.062	0.194	0.198	0.519
	(0.017)	(0.067)	(0.057)	(0.052)	(0.096)

Note: Rows correspond to the origin quantile, columns to current quintile. Sample weights employed. Clustered standard errors. Cluster unit is the primary sampling unit. The South region is formed by Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán and Quintana Roo. Quintiles defined over the national distribution.

A.4.2 Transition matrices, regionally defined quantiles

Table D6. North region transition matrix
(Quintiles defined over the regional distribution)

	Q1 (Bottom)	Q2	Q3	Q4	Q5 (Top)
Q1 (Bottom)	0.430	0.246	0.160	0.102	0.062
	(0.037)	(0.028)	(0.025)	(0.021)	(0.014)
Q2	0.313	0.242	0.222	0.161	0.062
	(0.025)	(0.024)	(0.026)	(0.025)	(0.019)
Q3	0.127	0.233	0.247	0.186	0.163
	(0.027)	(0.032)	(0.019)	(0.021)	(0.039)
Q4	0.072	0.193	0.251	0.277	0.206
	(0.016)	(0.028)	(0.021)	(0.030)	(0.028)
Q5 (Top)	0.013	0.082	0.121	0.275	0.509
	(0.005)	(0.019)	(0.026)	(0.027)	(0.047)

Note: Rows correspond to the origin quantile, columns to current quintile. Sample weights employed. The North (N) region consists of Baja California, Sonora, Chihuahua, Coahuila, Nuevo León and Tamaulipas. Quintiles defined over the regional distribution.

Table D7. Northwest region transition matrix
(Quintiles defined over the regional distribution)

	Q1 (Bottom)	Q2	Q3	Q4	Q5 (Top)
Q1 (Bottom)	0.350	0.263	0.198	0.133	0.055
	(0.035)	(0.019)	(0.028)	(0.029)	(0.022)
Q2	0.284	0.226	0.185	0.206	0.099
	(0.038)	(0.027)	(0.037)	(0.038)	(0.028)
Q3	0.214	0.254	0.206	0.211	0.116
	(0.028)	(0.034)	(0.028)	(0.027)	(027)
Q4	0.095	0.179	0.266	0.223	0.237
	(0.019)	(0.027)	(0.043)	(0.024)	(0.033)
Q5 (Top)	0.067	0.068	0.150	0.229	0.487
	(0.019)	(0.019)	(0.025)	(0.025)	(0.041)

Note: Rows correspond to the origin quantile, columns to current quintile. Sample weights employed. Clustered standard errors. Cluster unit is the primary sampling unit. North West (NW) consists of Baja California Sur, Sinaloa, Nayarit, Durango and Zacatecas. Quintiles defined over the regional distribution.

Table D8. Center-north region transition matrix
(Quintiles defined over the regional distribution)

	Q1 (Bottom)	Q2	Q3	Q4	Q5 (Top)
Q1 (Bottom)	0.458	0.233	0.168	0.100	0.041
	(0.041)	(0.030)	(0.027)	(0.023)	(0.014)
Q2	0.262	0.263	0.218	0.169	0.088
	(0.030)	(0.031)	(0.029)	(0.024)	(0.016)
Q3	0.148	0.272	0.250	0.228	0.102
	(0.026)	(0.034)	(0.036)	(0.025)	(0.018)
Q4	0.101	0.167	0.257	0.230	0.246
	(0.022)	(0.033)	(0.049)	(0.026)	(0.041)
Q5 (Top)	0.029	0.069	0.105	0.272	0.524
	(0.010)	(0.019)	(0.025)	(0.025)	(0.043)

Note: Rows correspond to the origin quantile, columns to current quintile. Sample weights employed. Clustered standard errors. Cluster unit is the primary sampling unit. Center North (CN) region is form by Jalisco, Aguascalientes, Colima, Michoacán and San Luis Potosí. Quintiles defined over the regional distribution.

Table D9. Center region transition matrix
(Quintiles defined over the regional distribution)

	Q1 (Bottom)	Q2	Q3	Q4	Q5 (Top)
Q1 (Bottom)	0.448	0.282	0.167	0.065	0.049
	(0.041)	(0.037)	(0.016)	(0.014)	(0.017)
Q2	0.294	0.291	0.222	0.129	0.064
	(0.037)	(0.038)	(0.027)	(0.032)	(0.016)
Q3	0.172	0.211	0.246	0.246	0.125
	(0.041)	(0.024)	(0.024)	(0.039)	(0.019)
Q4	0.068	0.170	0.235	0.343	0.184
	(0.015)	(0.026)	(0.021)	(0.026)	(0.023)
Q5 (Top)	0.018	0.049	0.134	0.216	0.582
	(0.007)	(0.017)	(0.017)	(0.026)	(0.039)

Note: Rows correspond to the origin quantile, columns to current quintile. Sample weights employed. Clustered standard errors. Cluster unit is the primary sampling unit. Center region is formed by Guanajuato, Querétaro, Hidalgo, Estado de México, Mexico City, Morelos, Tlaxcala, and Puebla. Quintiles defined over the regional distribution.

Table D10. South region transition matrix
(Quintiles defined over the regional distribution)

	Q1 (Bottom)	Q2	Q3	Q4	Q5 (Top)
Q1 (Bottom)	0.387	0.300	0.146	0.132	0.035
	(0.040)	(0.028)	(0.021)	(0.240)	(0.010)
Q2	0.344	0.246	0.238	0.124	0.048
	(0.039)	(0.027)	(0.035)	(0.016)	(0.011)
Q3	0.182	0.260	0.236	0.210	0.113
	(0.037)	(0.028)	(0.027)	(0.026)	(0.019)
Q4	0.069	0.150	0.265	0.290	0.226
	(0.015)	(0.023)	(0.027)	(0.025)	(0.027)
Q5 (Top)	0.015	0.040	0.112	0.246	0.586
	(0.007)	(0.012)	(0.016)	(0.030)	(0.038)

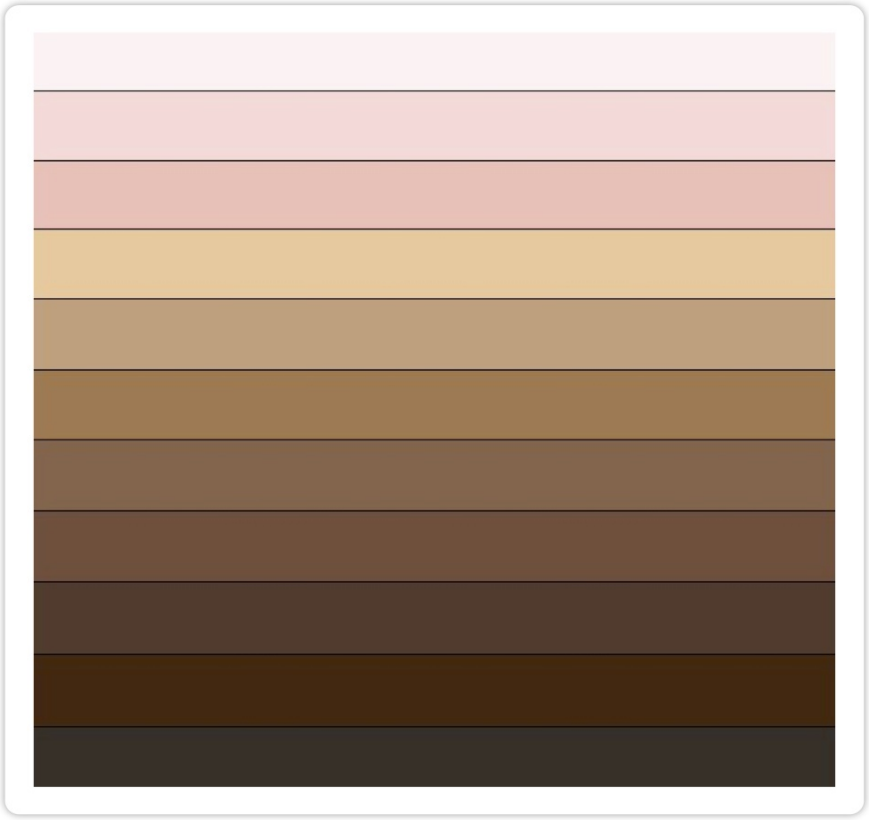
Note: Rows correspond to the origin quantile, columns to current quintile. Sample weights employed. Clustered standard errors. Cluster unit is the primary sampling unit. The South region is formed by Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán and Quintana Roo. Quintiles defined over the regional distribution.

Appendix B

Chapter 2

B.1 PERLA color palette

Figure A1. PERLA skin tone palette



Note: Source Telles, ed (2014)

B.2 Construction of economic resources index

Construction of economic resources index In the absence of information about parental and current income, it is necessary to use another type of information to estimate the amount of economic resources available in the origin and current household. An alternative is to use information regarding the types of durable goods and services each household has access to, as a larger number of assets owned would imply a higher income level than households with fewer goods and services. This approach implies the existence of a latent variable that represents the household's economic status that can be recovered from the information regarding the assets and services owned by the household. This problem is akin to a dimensionality problem, in which the objective is to reduce the information from the variables representing the set

of assets and services into one variable that captures the implicit economic status of the household. I employ Multiple Correspondence Analysis (MCA) to achieve this goal.

In contrast with Principal Component Analysis, MCA does not require the variables to be continuous nor capable of being summarized by their mean and variance. This makes it an ideal method for computing an asset index based on binary variables that indicate the ownership of a series of durable goods. From Greenacre and Hastie (1987), the first step of the MCA algorithm is to construct an indicator matrix in which each row corresponds to the ownership profile of an individual while the columns are the categories of each binary variable representing the ownership of the different goods. Next denote this matrix by Z. From Z it is possible to construct the Burt matrix, C, as $C = Z^T Z$. Following Nenadic and Greenacre (2007), all the elements $c_{i.j}$ of C are divided by the grand total of the matrix: $N = \sum_{i,j} c_{i,j}$, and the resulting matrix is the correspondence matrix, P, formed by $p_{i,j}$ elements. The row totals r_i of this matrix are the "masses" of each category. These masses are employed in the construction of the standardized residuals $\mathbf{S} = \{\frac{p_{i,j} - r_i r_j}{\sqrt{r_i r_j}}\}$.

Performing the eigendecomposition of $\mathbf{S}$ it is possible to find the eigenvectors $V = \{v_{i,s}\}$ and the eigenvalues λ_s that solve the problem $\mathbf{S} = V \Lambda V^T$, in which Λ is a diagonal matrix with the eigenvalues in the main diagonal. This algorithm produces the eigenvalues and eigenvectors that maximize the correlation across the different assets, while maximizing the dispersion of the coordinates implicit in the eigenvectors (Greenacre and Hastie, 1987; Kohn, 2012). The coordinates for each category are derived from $f_{i,s} = \alpha_{i,s} \lambda_s = \frac{v_{i,s}}{\sqrt{r_i}} \lambda_s$. Multiplying the ownership profile of each individuals by the vector of coordinates produces the row score, which is the latent variable representing the index of economic resources. The square value of the eigenvalues is the principal inertia, namely the amount of dispersion replicated by the generated coordinates.

Tables B1-B8 show the mass, the contribution to the first principal inertia, and the coordinates for each category of the binary variables representing goods and services.

Table B1. Components of the origin household index, composite sample. (Cohort 25-34)

Household asset or service	Value	Mass	Share of total explained inertia	Coordinates
Overcrowded household	0	0.034	0.027	0.944
	1	0.038	0.024	-0.843
Bank account	0	0.060	0.013	-0.445
	1	0.012	0.064	2.251
Credit card	0	0.059	0.014	-0.474
	1	0.012	0.066	2.250
Electricity	0	0.004	0.047	-3.384
	1	0.068	0.002	0.178
Fridge	0	0.019	0.081	-2.085
	1	0.053	0.029	0.741
Washing machine	0	0.034	0.059	-1.351
	1	0.037	0.054	1.234
Potable water	0	0.018	0.058	-1.859
	1	0.054	0.019	0.606
T.V. set	0	0.011	0.072	-2.435
	1	0.060	0.014	0.464
Stove	0	0.014	0.075	-2.303
	1	0.058	0.018	0.550
Toaster	0	0.063	0.006	-0.322
	1	0.009	0.047	2.382
Car	0	0.050	0.023	-0.691
	1	0.021	0.055	1.644
Another dwelling	0	0.067	0.001	-0.129
	1	0.005	0.018	1.864
Commercial venue	0	0.067	0.001	-0.115
	1	0.004	0.014	1.810
Landline	0	0.047	0.034	-0.874
	1	0.025	0.064	1.651

Table B2. Components of the origin household index, composite sample. (Cohort 35-44)

Household asset or service	Value	Mass	Share of total explained inertia	Coordinates
Overcrowded household	0	0.022	0.032	1.280
	1	0.049	0.014	-0.571
Bank account	0	0.063	0.008	-0.342
	1	0.008	0.059	2.566
Credit card	0	0.064	0.006	-0.303
	1	0.007	0.057	2.770
Electricity	0	0.011	0.055	-2.127
	1	0.060	0.010	0.396
Fridge	0	0.032	0.064	-1.422
	1	0.040	0.050	1.124
Washing machine	0	0.047	0.034	-0.877
	1	0.024	0.067	1.699
Potable water	0	0.028	0.053	-1.388
	1	0.043	0.035	0.912
T.V. set	0	0.023	0.067	-1.684
	1	0.049	0.031	0.785
Stove	0	0.022	0.067	-1.720
	1	0.050	0.030	0.758
Toaster	0	0.064	0.006	-0.316
	1	0.007	0.054	2.726
Car	0	0.056	0.015	-0.527
	1	0.016	0.054	1.898
Another dwelling	0	0.068	0.001	-0.121
	1	0.004	0.019	2.202
Commercial venue	0	0.068	0.001	-0.104
	1	0.003	0.016	2.231
Landline	0	0.055	0.021	-0.629
	1	0.016	0.073	2.173

Table B3. Components of the origin household index, composite sample. (Cohort 45-54)

Household asset or service	Value	Mass	Share of total explained inertia	Coordinates
Overcrowded household	0	0.016	0.029	1.406
	1	0.055	0.009	-0.421
Bank account	0	0.066	0.005	-0.260
	1	0.006	0.055	3.024
Credit card	0	0.068	0.003	-0.197
	1	0.004	0.052	3.577
Electricity	0	0.021	0.054	-1.556
	1	0.050	0.023	0.663
Fridge	0	0.044	0.046	-1.029
	1	0.028	0.073	1.627
Washing machine	0	0.056	0.021	-0.628
	1	0.015	0.080	2.345
Potable water	0	0.035	0.049	-1.177
	1	0.036	0.048	1.146
T.V. set	0	0.038	0.050	-1.143
	1	0.033	0.058	1.314
Stove	0	0.033	0.055	-1.266
	1	0.038	0.048	1.110
Toaster	0	0.068	0.003	-0.196
	1	0.004	0.047	3.538
Car	0	0.061	0.010	-0.420
	1	0.011	0.059	2.399
Another dwelling	0	0.069	0.001	-0.089
	1	0.003	0.015	2.316
Commercial venue	0	0.069	0.001	-0.109
	1	0.003	0.019	2.565
Landline	0	0.061	0.013	-0.471
	1	0.010	0.076	2.754

Table B4. Components of the origin household index, composite sample. (Cohort 55-64)

Household asset or service	Value	Mass	Share of total explained inertia	Coordinates
Overcrowded household	0	0.016	0.018	1.155
	1	0.056	0.005	-0.321
Bank account	0	0.067	0.004	-0.247
	1	0.004	0.069	3.957
Credit card	0	0.068	0.003	-0.205
	1	0.003	0.064	4.331
Electricity	0	0.026	0.046	-1.299
	1	0.046	0.026	0.773
Fridge	0	0.049	0.035	-0.840
	1	0.022	0.078	1.887
Washing machine	0	0.061	0.014	-0.496
	1	0.010	0.087	3.006
Potable water	0	0.039	0.038	-0.991
	1	0.033	0.045	1.161
T.V. set	0	0.046	0.036	-0.883
	1	0.025	0.068	1.644
Stove	0	0.041	0.042	-1.000
	1	0.030	0.058	1.382
Toaster	0	0.069	0.002	-0.176
	1	0.002	0.061	4.885
Car	0	0.062	0.008	-0.377
	1	0.009	0.058	2.627
Another dwelling	0	0.070	0.000	-0.083
	1	0.002	0.018	3.013
Commercial venue	0	0.069	0.001	-0.115
	1	0.003	0.025	3.051
Landline	0	0.065	0.009	-0.373
	1	0.007	0.080	3.481

Table B5. Components of the current household index, composite sample (25-34)

Asset	Value	Mass	Inertial explained	coord	Asset	Value	Mass	Inertial explained	coord
Overcrowded dwelling	0	0.024	0.008	0.577	Domestic service	0	0.034	0.002	-0.226
	1	0.013	0.015	-1.090		1	0.003	0.019	2.424
Bank Account	0	0.025	0.013	-0.712	internet	0	0.021	0.040	-1.389
	1	0.012	0.027	1.549		1	0.016	0.051	1.754
Credit Card	0	0.024	0.011	-0.669	Boiler	0	0.018	0.028	-1.261
	1	0.013	0.021	1.268		1	0.019	0.026	1.190
Electricity	0	0.000	0.004	-2.780	Table	0	0.027	0.016	-0.764
	1	0.037	0.000	0.024		1	0.010	0.044	2.148
Dwelling	0	0.025	0.001	-0.054	VHS or DVD player	0	0.018	0.023	-1.157
	1	0.012	0.001	0.113		1	0.019	0.022	1.105
Tractor	0	0.037	0	-0.010	Videogame	0	0.029	0.011	-0.621
	1	0.000	0.001	1.201		1	0.008	0.041	2.318
Agricultural land	0	0.036	0	-0.075	Landline	0	0.024	0.022	-0.969
	1	0.001	0.001	0.579		1	0.013	0.042	1.816
Shop	0	0.035	0	-0.073	Car	0	0.022	0.020	-0.958
	1	0.002	0.004	1.231		1	0.015	0.030	1.456
Another dwelling	0	0.035	0.000	-0.075	Cable TV	0	0.017	0.025	-1.239
	1	0.002	0.005	1.231		1	0.020	0.021	1.059
Microwave	0	0.018	0.038	-1.466	Cellphone	0	0.003	0.023	-2.556
	1	0.019	0.037	1.424		1	0.034	0.002	0.258
Refrigerator	0	0.004	0.040	-2.969	Computer	0	0.024	0.027	-1.072
	1	0.033	0.005	0.368		1	0.013	0.051	1.981
Washing machine	0	0.009	0.043	-2.133	TV set	0	0.007	0.021	-1.801
	1	0.028	0.015	0.724		1	0.030	0.005	0.400
Potable water	0	0.003	0.019	-2.362	Stove	0	0.003	0.032	-3.162
	1	0.034	0.002	0.222		1	0.034	0.002	0.240
Toaster	0	0.030	0.009	-0.564					
	1	0.007	0.037	2.301					

Table B6. Components of the current household index, composite sample (35-44)

Asset	Value	Mass	Inertial explained	coord	Asset	Value	Mass	Inertial explained	coord
Overcrowded dwelling	0	0.025	0.006	0.520	Domestic service	0	0.034	0.001	-0.205
	1	0.012	0.014	-1.095		1	0.003	0.019	2.678
Bank Account	0	0.024	0.016	-0.830	internet	0	0.020	0.041	-1.400
	1	0.013	0.032	1.599		1	0.017	0.050	1.712
Credit Card	0	0.024	0.014	-0.787	Boiler	0	0.019	0.027	-1.227
	1	0.037	0.026	1.420		1	0.018	0.028	1.249
Electricity	0	0.000	0.004	-3.097	Tablet	0	0.027	0.015	-0.741
	1	0.037	0.000	0.021		1	0.010	0.042	2.109
Dwelling	0	0.016	0.001	-0.189	VHS or DVD player	0	0.020	0.019	-0.999
	1	0.021	0.001	0.145		1	0.017	0.021	1.133
Tractor	0	0.036	0.000	-0.021	Videogame	0	0.030	0.009	-0.561
	1	0.001	0.001	1.279		1	0.007	0.042	2.514
Agricultural land	0	0.035	0.000	-0.055	Landline	0	0.023	0.025	-1.055
	1	0.002	0.003	0.986		1	0.014	0.040	1.658
Shop	0	0.035	0.000	-0.089	Car	0	0.020	0.024	-1.114
	1	0.002	0.005	1.606		1	0.017	0.029	1.319
Another dwelling	0	0.033	0.001	-0.188	Cable TV	0	0.017	0.021	-1.123
	1	0.004	0.011	1.725		1	0.020	0.018	0.982
Microwave	0	0.018	0.036	-1.415	Cellphone	0	0.004	0.023	-2.491
	1	0.019	0.035	1.392		1	0.034	0.002	0.263
Refrigerator	0	0.004	0.035	-2.875	Computer	0	0.023	0.030	-1.145
	1	0.004	0.004	0.337		1	0.014	0.051	1.919
Washing machine	0	0.009	0.040	-2.116	TV set	0	0.009	0.017	-1.423
	1	0.028	0.013	0.669		1	0.028	0.005	0.444
Potable water	0	0.003	0.015	-2.312	Stove	0	0.003	0.030	-3.089
	1	0.034	0.001	0.182		1	0.034	0.002	0.252
Toaster	0	0.029	0.011	-0.610					
	1	0.008	0.040	2.299					

Table B7. Components of the current household index, composite sample (45-54)

Asset	Value	Mass	Inertial explained	coord	Asset	Value	Mass	Inertial explained	coord
Overcrowded dwelling	0	0.029	0.003	0.315	Domestic service	0	0.034	0.002	-0.218
	1	0.003	0.018	-1.116		1	0.003	0.019	2.460
Bank Account	0	0.026	0.014	-0.733	internet	0	0.021	0.038	-1.342
	1	0.011	0.032	1.728		1	0.016	0.049	1.738
Credit Card	0	0.026	0.013	-0.704	Boiler	0	0.017	0.026	-1.272
	1	0.011	0.030	1.667		1	0.020	0.022	1.053
Electricity	0	0.000	0.006	-3.468	Tablet	0	0.030	0.012	-0.629
	1	0.037	0.000	0.027		1	0.007	0.048	2.556
Dwelling	0	0.010	0.002	-0.323	VHS or DVD player	0	0.023	0.019	-0.928
	1	0.027	0.001	0.124		1	0.014	0.029	1.469
Tractor	0	0.036	0.000	-0.024	Videogame	0	0.032	0.005	-0.410
	1	0.001	0.002	1.323		1	0.005	0.038	2.874
Agricultural land	0	0.035	0.000	-0.064	Landline	0	0.020	0.029	-1.220
	1	0.017	0.004	1.144		1	0.015	0.035	1.439
Shop	0	0.034	0.001	-0.147	Car	0	0.020	0.025	-1.131
	1	0.003	0.008	1.538		1	0.017	0.030	1.353
Another dwelling	0	0.034	0.001	-0.203	Cable TV	0	0.018	0.022	-1.140
	1	0.003	0.014	1.946		1	0.019	0.021	1.067
Microwave	0	0.018	0.037	-1.448	Cellphone	0	0.005	0.024	-2.165
	1	0.019	0.036	1.399		1	0.032	0.004	0.350
Refrigerator	0	0.003	0.032	-2.871	Computer	0	0.025	0.026	-1.033
	1	0.034	0.003	0.289		1	0.012	0.052	2.067
Washing machine	0	0.009	0.041	-2.162	TV set	0	0.008	0.018	-1.534
	1	0.028	0.013	0.666		1	0.029	0.005	0.439
Potable water	0	0.002	0.015	-2.354	Stove	0	0.003	0.030	-3.095
	1	0.035	0.001	0.169		1	0.034	0.002	0.233
Toaster	0	0.029	0.012	-0.644					
	1	0.008	0.043	2.342					

Table B8. Components of the current household index, composite sample (55-64)

Asset	Value	Mass	Inertial explained	coord	Asset	Value	Mass	Inertial explained	coord
Overcrowded dwelling	0	0.029	0.002	0.267	Domestic service	0	0.033	0.004	-0.339
	1	0.008	0.007	-0.905		1	0.004	0.026	2.452
Bank Account	0	0.029	0.009	-0.562	internet	0	0.024	0.029	-1.108
	1	0.008	0.030	1.889		1	0.013	0.054	2.057
Credit Card	0	0.027	0.011	-0.656	Boiler	0	0.016	0.024	-1.265
	1	0.037	0.026	1.458		1	0.021	0.018	0.935
Electricity	0	0.000	0.003	-1.769	Tablet	0	0.031	0.011	-0.606
	1	0.037	0.000	0.013		1	0.006	0.054	2.899
Dwelling	0	0.008	0.003	-0.532	VHS or DVD player	0	0.024	0.021	-0.955
	1	0.029	0.001	0.146		1	0.013	0.040	1.830
Tractor	0	0.037	0.000	-0.012	Videogame	0	0.033	0.004	-0.342
	1	0.000	0.002	0.958		1	0.004	0.037	3.199
Agricultural land	0	0.036	0.000	-0.040	Landline	0	0.021	0.028	-1.185
	1	0.001	0.005	1.572		1	0.016	0.035	1.493
Shop	0	0.033	0.001	-0.168	Car	0	0.021	0.021	-1.035
	1	0.004	0.008	1.480		1	0.016	0.029	1.397
Another dwelling	0	0.035	0.001	-0.112	Cable TV	0	0.021	0.024	-1.109
	1	0.002	0.011	2.399		1	0.016	0.030	1.396
Microwave	0	0.018	0.036	-1.429	Cellphone	0	0.009	0.025	-1.673
	1	0.019	0.035	1.384		1	0.028	0.009	0.561
Refrigerator	0	0.003	0.028	-2.596	Computer	0	0.028	0.020	-0.848
	1	0.034	0.003	0.250		1	0.009	0.057	2.466
Washing machine	0	0.009	0.039	-2.031	TV set	0	0.006	0.020	-1.830
	1	0.028	0.013	0.663		1	0.031	0.004	0.349
Potable water	0	0.003	0.015	-1.966	Stove	0	0.002	0.022	-2.573
	1	0.034	0.001	0.192		1	0.035	0.002	0.180
Toaster	0	0.029	0.013	-0.675					
	1	0.008	0.047	2.478					

B.3 Alternative definition of indigenous group.

Table C1. Rank-rank regression, alternative definition

Current rank	(1)	(2)	(3)
Origin rank	0.592***	0.575***	0.588***
	(0.009)	(0.009)	(0.020)
Indigenous		-3.273***	
		(0.609)	
Indigenous × Origin rank		-0.012	
		(0.025)	
Indigenous, skin tone group 1			-5.103***
			(1.570)
Indigenous, skin tone group 2			-3.643***
			(0.848)
Indigenous, skin tone group 3			-5.372***
			(1.292)
Non-indigenous, skin tone group 2			-0.379
			(0.702)
Non-indigenous, skin tone group 3			-3.747***
			(1.092)
Indigenous, skin tone group 1 × Origin rank			0.070
			0.077
Indigenous, skin tone group 2 × Origin rank			-0.030
			(0.032)
Indigenous, skin tone group 3 × Origin rank			-0.069
			(0.063)
Non-indigenous, skin tone group 2 × Origin rank			-0.022
			(0.021)
Non-indigenous, skin tone group 3 × Origin rank			-0.024
			(0.045)
Intercept	10.62***	11.53***	12.25***
	(0.284)	(0.289)	(0.681)
Observations	25,408	25,408	25,408
R-squared	0.366	0.372	0.377

Note: Sample weights are employed in all estimations. Standard errors clustered by primary sample unit of the survey. The indigenous population corresponds to the population that self-identifies as member of an indigenous group. Skin tone group 1 corresponds to PERLA tones 1-3, group 2 to PERLA skin tones 4-6 and group 3 to PERLA skin tones 7-11. Ranks correspond to 50 quantiles of the origin and current distribution of the index of economic resources.

Bibliography

Acciari, Paolo, Alberto Polo, and Giovanni L. Violante, "And Yet It Moves: Intergenerational Mobility in Italy," *American Economic Journal: Applied Economics*, July 2022, *14* (3), 118–63.

Acemoglu, Daron and James A. Robinson, "Persistence of Power, Elites, and Institutions," *American Economic Review*, March 2008, *98* (1), 267–93.

_ **, Georgy Egorov, and Konstantin Sonin**, "Social Mobility and Stability of Democracy: Reevaluating De Tocqueville," *The Quarterly Journal of Economics*, 11 2017, *133* (2), 1041–1105.

Aguilar-Rodriguez, Adriana, Alfonso Miranda, and Yu Zhu, "Decomposing the language pay gap among the indigenous ethnic minorities of Mexico: is it all down to observables?," *Economics Bulletin*, 2018, *38* (2), 689–695.

Aguilar, Rosario, "Social and political consequences of stereotypes related to racial phenotypes in Mexico," Working Paper 230, Centro de Investigación y Docencia Económicas 2011.

_ , "Los tonos de los desafíos democráticos: el color de la piel y la raza en México," *Política y Gobierno*, 2013, pp. 25–57.

Akay, Alpaslan, Lisa Andersson, Peter Martinsson, and Haileselassie Medhin, "Positional Concerns among the Poor: Does Reference Group Matter? Evidence from Survey Experiments," *Journal of African Economies*, 07 2014, *23* (5), 673–699.

Akee, Randall and Mutlu Yuksel, "The Decreasing Effect of Skin Tone on Women's Full-Time Employment," *ILR Review*, 2012, *65* (2), 398–426.

Albertus, Michael and Victor Menaldo, "Gaming Democracy: Elite Dominance during Transition and the Prospects for Redistribution," *British Journal of Political Science*, 2014, *44* (3), 575–603.

Andrabi, Tahir, Benjamin Daniels, and Jishnu Das, "Human Capital Accumulation and Disasters: Evidence from the Pakistan Earthquake of 2005," *Journal of Human Resources*, 2021.

Arceo-Gómez, Eva O. and Pedro J. Torres, "Brechas salariales por autoidentificación indígena y rasgos lingüísticos en México," *Sobre México Temas de Economía*, jul. 2021, (3), 129–161.

_ **and Raymundo M. Campos-Vázquez**, "Race and Marriage in the Labor Market: A Discrimination Correspondence Study in a Developing Country," *American Economic Review*, 2014, *104* (5), 376–380.

_ **and** _ , "Double Discrimination: Is Discrimination in Job Ads Accompanied by Discrimination in Callbacks?," *Journal of Economics, Race, and Policy*, Dec 2019, *2* (4), 257–268.

Baldomero-Quintana, Luis, Enrique de la Rosa, and L. Guillermo Woo-Mora, "Infrastructures of Race? Colonial Indigenous Zoning and Contemporaneous Urban Segregation," Working Paper, William & Mary Department of Economics June 2022.

Bau, Natalie, Jishnu Das, and Andres Yi Chang, "New evidence on learning trajectories in a low-income setting," *International Journal of Educational Development*, 2021, *84*, 102430.

Bavaro, Michele and Fabrizio Patriarca, "Referrals, intergenerational mobility and human capital accumulation," *Economic Modelling*, 2022, *110*, 105811.

Beatty, Amanda, Emilie Berkhout, Luhur Bima, Menno Pradhan, and Daniel Suryadarma, "Schooling progress, learning reversal: Indonesia's learning profiles between 2000 and 2014," *International Journal of Educational Development*, 2021, *85*, 102436.

Becker, Gary S. and Nigel Tomes, "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy*, 1979, *87* (6), 1153–1189.

_ **and** _ , "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics*, 1986, *4* (3), S1–S39.

_ **, Scott Duke Kominers, Kevin Murphy, and Jörg L. Spenkuch**, "A Theory of Intergenerational Mobility," *Journal of Political Economy*, 2018, *126* (S1), S7–S25.

Belot, Michèle and Dinand Webbink, "Do Teacher Strikes Harm Educational Attainment of Students?," *Labour*, 2010, *24* (4), 391–406.

Ben-Porath, Yoram, "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 1967, *75* (4), 352–365.

Bhattacharya, Debopam and Bhaskar Mazumder, "A Nonparametric Analysis of Black-White Differences in Intergenerational Income Mobility in the United States," *Quantitative Economics*, 2011, *2* (3), 335–379.

Bhuiyan, Muhammad Faress, "Life Satisfaction and Economic Position Relative to Neighbors: Perceptions Versus Reality," *Journal of Happiness Studies*, October 2018, *19* (7), 1935–1964.

Birdal, Murat and T. Ongan, "Why Do We Care About Having More than Others? Socioeconomic Determinants of Positional Concerns in Different Domains," *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, March 2016, *126* (2), 727–738.

Blanco-Bosco, Emilio E., "Discriminación étnico-racial y oportunidades educativas en México," *Sociológica*, 12 2020, *35*, 139 – 180.

Bogaerts, Tess and Mario Pandelaere, "Less is more: Why some domains are more positional than others," *Journal of Economic Psychology*, 2013, *39*, 225–236.

Bowles, Samuel, *Towards an Educational Production Function*, NBER,

Campbell, Mary E., Verna M. Keith, Vanessa Gonlin, and Adrienne R. Carter-Sowell, "Is a Picture Worth A Thousand Words? An Experiment Comparing Observer-Based Skin Tone Measures," *Race and Social Problems*, 2020, *12* (2), 242–245.

Campos-Vazquez, Raymundo M. and Carolina Rivas-Herrera, "The Color of Electoral Success: Estimating the Effect of Skin Tone on Winning Elections in Mexico," *Social Science Quarterly*, 2021, *102* (2), 844–864.

Campos-Vázquez, Raymundo M., "The higher price of whiter skin: an analysis of escort services," *Applied Economics Letters*, 2021, *28* (1), 1–4.

_ **and Eduardo Medina-Cortina**, "Identidad Social y Estereotipos por Color de Piel. Aspiraciones y Desempeño en Jóvenes Mexicanos," *El Trimestre Económico*, 2018, *85* (1), 53–79.

_ **and** _ , "Skin Color and Social Mobility: Evidence from Mexico," *Demography*, 2019, *56* (1), 77–113.

_ **and Luis Monroy-Gómez-Franco**, "¿El crecimiento económico reduce la pobreza en México?," *Revista de Economí Mexicana. Anuario UNAM*, 2016, *1* (1), 140–185.

Canedo, Ana P., "Analyzing Multidimensional Poverty Estimates in Mexico From an Ethnic Perspective: A Policy Tool for Bridging the Indigenous Gap," *Poverty & Public Policy*, 2018, *10* (4), 543–563.

_ , "Labor Market Discrimination Against Indigenous Peoples in Mexico: A Decomposition Analysis of Wage Differentials," *Iberoamericana. Nordic Journal of Latin American and Caribbean Studies*, 2019, *48* (1), 12–27.

Cano-Urbina, Javier and Patrick L. Mason, "Acculturation and the labor market in Mexico," *IZA Journal of Labor Policy*, December 2016, *5* (1), 1–29.

Carlsson, Fredrik and Ping Qin, "It is better to be the head of a chicken than the tail of a phoenix: Concern for relative standing in rural China," *The Journal of Socio-Economics*, 2010, *39* (2), 180–186.

_ , Gautam Gupta, and Olof Johansson-Stenman, "Keeping up with the Vaishyas? Caste and relative standing in India," *Oxford Economic Papers*, 05 2008, *61* (1), 52–73.

_ , Olof Johansson-Stenman, and Peter Martinsson, "Do You Enjoy Having More than Others? Survey Evidence of Positional Goods," *Economica*, 2007, *74* (296), 586–598.

Castañeda-Navarrete, Jennifer, "Poverty Dynamics in Mexico, 2002-2005. An Ethnicity Approach," *Brazilian Review of Econometrics*, 2013, *33* (1).

Castilla, Carolina, "Subjective well-being and reference-dependence: Insights from Mexico," *The Journal of Economic Inequality*, June 2012, *10* (2), 219–238.

Celse, Jérémy, "Do You Enjoy Having More Than Others or More Than Another? Exploring the Relationship Between Relative Concerns and the Size of the Reference Group," *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, August 2018, *138* (3), 1089–1118.

Centro de Estudios Espinosa Yglesias, "Consideraciones para la definición de las regiones de la Encuesta ESRU de Movilidad Social en México 2017," Technical Note, Centro de Estudios Espinosa Yglesias 2019.

Chang, Wen-Chun, "Climbing up the Social Ladders: Identity, Relative Income, and Subjective Well-being," *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, August 2013, *113* (1), 513–535.

Chetty, Raj, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter, "Race and Economic Opportunity in the United States: an Intergenerational Perspective*," *The Quarterly Journal of Economics*, 12 2019, *135* (2), 711–783.

_ , _ , Patrick Kline, and Emmanuel Saez, "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," *The Quarterly Journal of Economics*, 2015, *129* (4), 1553–1623.

_ , _ , _ , _ , and Nicholas Turner, "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility," *American Economic Review: Papers and Proceedings*, 2014, *104* (5), 141–147.

Clark, Andrew E. and Claudia Senik, "Who Compares to Whom? The Anatomy of Income Comparisons in Europe," *The Economic Journal*, 2010, *120* (544), 573–594.

_ , _ , **and Katsunori Yamada**, "The Joneses in Japan: income comparisons and financial satisfaction," *The Japanese Economic Review*, April 2022, *73* (2), 351–372.

_ , **Paul Frijters, and Michael A. Shields**, "Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles," *Journal of Economic Literature*, March 2008, *46* (1), 95–144.

Coleman, James, Ernest Campbell, Carol Hobson, James McPartland, Alexander Mood, Frederick Weinfeld, and Robert York., "Equality of Educational Opportunity," Technical Report, Washington D.C. 1966.

Connolly, Marie, Miles Corak, and Catherine Haeck, "Intergenerational Mobility Between and Within Canada and the United States," *Journal of Labor Economics*, 2019, *37* (S2), S595–S641.

Corak, Miles, "The Canadian Geography of Intergenerational Income Mobility," *The Economic Journal*, 2019, *130* (631), 2134–2174.

Crawfurd, Lee, "Accounting for repetition and dropout in contemporaneous cross-section learning profiles: Evidence from Rwanda," *International Journal of Educational Development*, 2021, *85*, 102443.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 2010, *78* (3), 883–931.

Darity Jr., William A., "Position and Possessions: Stratification Economics and Intergroup Inequality," *Journal of Economic Literature*, June 2022, *60* (2), 400–426.

Darity Jr., William, Darrick Hamilton, and Jason Dietrich, "Passing on blackness: Latinos, race, and earnings in the USA," *Applied Economics Letters*, 2002, *9* (13), 847–853.

Davis, Jonathan and Bhashkar Mazumder, "Racial and Ethnic Differences in the Geography of Intergenerational Mobility," Working Paper, SSRN 2020.

Davis, Lewis and Stephen Wu, "Social Comparisons and Life Satisfaction Across Racial and Ethnic Groups: The Effects of Status, Information and Solidarity," *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, July 2014, *117* (3), 849–869.

de la Torre, Rodolfo and Roberto Vélez-Grajales, *Informe sobre Desarrollo Humano. México 2016: Desigualdad y Movilidad*, United Nations Development Program, 2016.

Delajara, Marcelo and Dositeo Graña, "Intergenerational Social Mobility in Mexico and its Regions Results from Rank-Rank Regressions," *Sobre México. Revista de Economía*, 2018, *4* (1), 22–37.

_ , **Raymundo M. Campos-Vazquez, and Roberto Velez-Grajales**, "The regional geography of social mobility in Mexico," *Regional Studies*, 2022, *56* (5), 839–852.

Deutscher, Nathan and Bhashkar Mazumder, "Intergenerational mobility across Australia and the stability of regional estimates," *Labour Economics*, 2020, *66*, 101861.

_ **and** _ , "Measuring Intergenerational Income Mobility: A Synthesis of Approaches," Working Paper 39, Stone Center on Socio-Economic Inequality June 2021.

Devaraj, Srikant and Pankaj C. Patel, "Skin Tone and Self-Employment: Is there an Intra-Group Variation among Blacks?," *The Review of Black Political Economy*, 2017, *44* (1-2), 137–166.

Diario Oficial de la Federación, "Acuerdo 26/12/20 por el que se establecen las orientaciones pedagógicas y los criterios para la evaluación del aprendizaje para la educación preescolar, primaria y secundaria en el periodo de contingencia sanitaria generada por el virus SARS-CoV2 (COVID-19) para el ciclo escolar 2020-2021.," 12 2020. `https://www.dof.gob.mx/nota_detalle.php?codigo=5608934&fecha=28/12/2020`.

Diette, Timothy M., Arthur H. Goldsmith, Darrick Hamilton, and William Darity Jr., "Skin Shade Stratification and the Psychological Cost of Unemployment: Is there a Gradient for Black Females?," *The Review of Black Political Economy*, 2015, *42* (1-2), 155–177.

Dixon, Angela R. and Edward E. Telles, "Skin Color and Colorism: Global Research, Concepts, and Measurement," *Annual Review of Sociology*, 2017, *43* (1), 405–424.

Doepke, Matthias, Giuseppe Sorrenti, and Fabrizio Zilibotti, "The Economics of Parenting," *Annual Review of Economics*, 2019, *11* (1), 55–84.

Durlauf, Steven N., Andros Kourtellos, and Chih Ming Tan, "The Great Gatsby Curve," Working Paper 29761, National Bureau of Economic Research February 2022.

Dávila, Alberto, Marie T. Mora, and Sue K. Stockly, "Does Mestizaje Matter in the US? Economic Stratification of Mexican Immigrants," *American Economic Review*, May 2011, *101* (3), 593–97.

Eriksen, Jesper and Martin D. Munk, "The geography of intergenerational mobility — Danish evidence," *Economics Letters*, 2020, *189*, 109024.

Esposito, Lucio and Adrián Villaseñor, "Relative Deprivation and School Enrolment. Evidence from Mexico," *Review of Income and Wealth*, 2019, *65* (1), 166–186.

Esquivel, Gerardo, "Convergencia regional en México, 1940-1995," *El Trimestre Económico*, 1999, *66* (264), 725–761.

Falk, Armin and Markus Knell, "Choosing the Joneses: Endogenous Goals and Reference Standards," *The Scandinavian Journal of Economics*, 2004, *106* (3), 417–435.

Ferreira, Francisco H. G. and Jéremié Gignoux, "The Measurement of Inequality of Opportunity: Theory and an Application to Latin America," *Review of Income and Wealth*, 2011, *57* (4), 622–657.

_ , _ , and Meltem Aran, "Measuring Inequality of Opportunity with Imperfect Data: The Case of Turkey," *Journal of Economic Inequality*, 2011, *9* (4), 651–680.

Fields, Gary S. and Efe A. Ok, "The Measurement of Income Mobility: An Introduction to the Literature," in Jacques Silber, ed., *Handbook of Income Inequality Measurement*, Vol. 71 of *Recent Economic Thought Series*, Springer, 1999, pp. 557–598.

Figueroa, Adolfo, "Competition and circulation of economic elites: Theory and application to the case of Peru," *The Quarterly Review of Economics and Finance*, 2008, *48* (2), 263–273.

Filmer, Deon and Kinnon Scott, "Assessing Asset Indices," *Demography*, 2012, *49* (1), 359–392.

_ and Lant H. Pritchett, "Estimating Wealth Effects without Expenditure Data-or Tears: An Application to Educational Enrollments in States of India," *Demography*, 2001, *38* (1), 115–132.

Flores, René and Edward Telles, "Social Stratification in Mexico Disentangling Color, Ethnicity and Class," *American Sociological Review*, 2012, *77* (3), 486–494.

Garfias, Francisco, "Elite Competition and State Capacity Development: Theory and Evidence from Post-Revolutionary Mexico," *American Political Science Review*, 2018, *112* (2), 339–357.

Garratt, Elisabeth, Tarani Chandola, Kingsley Purdam, and Alex Wood, "The interactive role of income (material position) and income rank (psychosocial position) in psychological distress: a 9-year longitudinal study of 30,000 UK parents," *Social Psychiatry and Psychiatric Epidemiology*, 10 2016, *51.*

German-Soto, Vicente, Reyna Rodríguez-Pérez, and Alexandra Gallegos-Morales, "Exposure to globalization and regional convergence in Mexico," *Estudios Económicos*, Jul. 2020, *35* (2), 267–295.

Goldsmith, Arthur H., Darrick Hamilton, and William Darity Jr., "Shades of Discrimination: Skin Tone and Wages," *The American Economic Review*, 2006, *96* (2), 242–245.

_ , _ , and _ , "From Dark to Light: Skin Color and Wages among African-Americans," *The Journal of Human Resources*, 2007, *42* (4), 701–738.

González de Alba, Iván G., "Poverty in Mexico from an Ethnic Perspective," *Journal of Human Development and Capabilities*, 2010, *11* (3), 449–465.

Gordon, Rachel A., Amelia R. Branigan, Mariya A. Khan, and Johanna G. Nunez, "Measuring Skin Color: Consistency, Comparability, and Meaningfulness of Rating Scale Scores and Handheld Device Readings," *Journal of Survey Statistics and Methodology*, 01 2022, *10* (2), 337–364.

Greenacre, Michael J. and Trevor Hastie, "The Geometric Interpretation of Correspondence Analysis," *Journal of the American Statistical Association*, 1987, *82* (398), 437–447.

Han, JooHee, "Does Skin Tone Matter? Immigrant Mobility in the U.S. Labor Market," *Demography*, 03 2020, *57* (2), 705–726.

Hanushek, Eric A., "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *The Journal of Human Resources*, 1979, *14* (3), 351–388.

_ , "Publicly provided education," in A. J. Auerbach and M. Feldstein, eds., *Handbook of Public Economics*, Vol. 4 of *Handbook of Public Economics*, Elsevier, 2002, chapter 30, pp. 2045–2141.

_ , "Education Production Functions: Developed Country Evidence," in Penelope Peterson, Eva Baker, and Barry McGaw, eds., *International Encyclopedia of Education (Third Edition)*, third edition ed., Oxford: Elsevier, 2010, pp. 407–411.

Hartshorne, Joshua K., Joshua B. Tenenbaum, and Steven Pinker, "A critical period for second language acquisition: Evidence from 2/3 million English speakers," *Cognition*, 2018, *177*, 263–277.

Hausmann, Ricardo, Carlo Pietrobelli, and Miguel Angel Santos, "Place-specific determinants of income gaps: New sub-national evidence from Mexico," *Journal of Business Research*, 2021, *131*, 782–792.

Heckman, James J. and Stefano Mosso, "The Economics of Human Development and Social Mobility," *Annual Review of Economics*, 2014, *6* (1), 689–733.

Heffetz, Ori and Robert H. Frank, "Preferences for Status: Evidence and Economic Implications," in Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, eds., *Handbook of Social Economics*, Vol. 1, North-Holland, 2011, pp. 69–91.

Heidrich, Stefanie, "Intergenerational mobility in Sweden: a regional perspective," *Journal of Population Economics*, October 2017, *30* (4), 1241–1280.

Hernández-Trillo, Fausto and Ana Martínez-Gutiérrez, "The Dark Road to Credit Applications: The Small-Business Case of Mexico," *Journal of Financial Services Research*, 2022, *62* (1-2), 1–25.

Hersch, Joni, "Profiling the New Immigrant Worker: The Effects of Skin Color and Height," *Journal of Labor Economics*, 2008, *26* (2), 345–386.

Hertz, Tom, "A group-specific measure of intergenerational persistence," *Economics Letters*, 2008, *100* (3), 415–417.

Ichino, Andrea and Rudolf Winter-Ebmer, "The Long-Run Educational Cost of World War II," *Journal of Labor Economics*, 2004, *22* (1), 57–87.

Instituto Nacional de Estadística y Geografía, "Encuesta para la Medición del Impacto del COVID-19 en la Educación. Presentación de resultados," Technical Report, Instituto Nacional de Estadística y Geografía, Aguascalientes 2021.

Janssen, Steve M. J. and Jaap M. J. Murre, "Reminiscence bump in autobiographical memory: Unexplained by novelty, emotionality, valence, or importance of personal events," *The Quarterly Journal of Experimental Psychology*, 2008, *61* (12), 1847–1860.

Japaridze, Irakli and Nagham Sayour, "Dying from envy: The role of inequality," *Health Economics*, 2021, *30* (6), 1374–1392.

Jaume, David and Alexander Willén, "The Long-Run Effects of Teacher Strikes: Evidence from Argentina," *Journal of Labor Economics*, 2019, *37* (4), 1097–1139.

Jäntti, Markus and Stephen P. Jenkins, "Income Mobility," in Anthony B. Atkinson and François Bourguignon, eds., *Handbook of Income Distribution*, Vol. 2, Elsevier, 2015, pp. 807–935.

Kaffenberger, Michelle, "Modelling the long-run learning impact of the Covid-19 learning shock: Actions to (more than) mitigate loss," *International Journal of Educational Development*, 2021, *81*, 102326.

_ **and Lant Pritchett**, "Aiming higher: Learning profiles and gender equality in 10 low- and middle-income countries," *International Journal of Educational Development*, 2020, *79*, 102272.

_ **and** _ , "Failing to Plan? Estimating the Impact of Achieving Schooling Goals on Cohort Learning," Working Paper 20083, Research on Improving Systems of Education 2020.

_ **and** _ , "A structured model of the dynamics of student learning in developing countries, with applications to policy," *International Journal of Educational Development*, 2021, *82*, 102371.

Kilford, Emma J., Emily Garrett, and Sarah-Jayne Blakemore, "The development of social cognition in adolescence: An integrated perspective," *Neuroscience Biobehavioral Reviews*, 2016, *70*, 106–120.

Kingdon, Geeta Gandhi and John Knight, "Community, comparisons and subjective well-being in a divided society," *Journal of Economic Behavior Organization*, 2007, *64* (1), 69–90.

Knight, John, Lina Song, and Ramani Gunatilaka, "Subjective well-being and its determinants in rural China," *China Economic Review*, 2009, *20* (4), 635–649. Symposium on Health Economics Issues in China.

Kohn, Jennifer L., "What is Health? A Multiple Correspondence Health Index," *Eastern Economic Journal*, 2012, *38* (2), 223–250.

Kreisman, Daniel and Marcos A. Rangel, "On the Blurring of the Color Line: Wages and Employment for Black Males of Different Skin Tones," *The Review of Economics and Statistics*, 03 2015, *97* (1), 1–13.

Krozer, Alice and Andrea Urrutia-Gómez, "Not in the eyes of the beholder: Racialisation, Whiteness and Beauty Standards in Mexico," Working Paper 5, El Colegio de México: Project on Racial and Ethnic Discrimination in Mexico 2021.

Kuegler, Alice, "A curse of comparison? : Evidence on reference groups for relative income concerns," Technical Report, World Bank 2009.

Kuziemko, Ilyana, Ryan W. Buell, Taly Reich, and Michael I. Norton, " "Last-Place Aversion": Evidence and Redistributive Implications *," *The Quarterly Journal of Economics*, 01 2014, *129* (1), 105–149.

Leibowitz, Arleen, "Home Investments in Children," *Journal of Political Economy*, 1974, *82* (2), S111–S131.

Lentz, Erin C., "Keeping Up with the Neighbors? Reference Groups in Ghana," *Economic Development and Cultural Change*, 2017, *66* (1), 91–112.

Lucas, Samuel R., "An Archaeology of Effectively Maintained Inequality Theory," *American Behavioral Scientist*, 2017, *61* (1), 8–29.

Luttmer, Erzo F. P., "Neighbors as Negatives: Relative Earnings and Well-Being*," *The Quarterly Journal of Economics*, 08 2005, *120* (3), 963–1002.

Mageli, Ingvild, Andrea Mannberg, and Eirik Eriksen Heen, "With whom, and about what, do we compete for social status? Effects of social closeness and relevance of reference groups for positional concerns," *Journal of Behavioral and Experimental Economics*, 2022, *98*, 101867.

Marcotte, Dave E. and Steven W. Hemelt, "Unscheduled School Closings and Student Performance," *Education Finance and Policy*, 07 2008, *3* (3), 316–338.

Markandya, Anil, "Intergenerational exchange mobility and economic welfare," *European Economic Review*, 1982, *17* (3), 307–324.

_ , "The Welfare Measurement of Changes in Economic Mobility," *Economica*, 1984, *51* (204), 457–471.

Martínez Casas, Regina, Emiko Saldívar, René D. Flores, Christina A. Sue, and Edward Telles, "The Different Faces of Mestizaje: Ethnicity and Race in Mexico," in Edward Telles, ed., *Pigmentocracies: Ethnicity, Race, and Color in Latin America*, University of North Carolina Press, 2014, pp. 36–80.

Massey, Douglas S. and Jennifer A. Martin, "NIS Skin Color Scale," 2003.

McKenzie, David, "Measuring Inequality with Asset Indicators," *Journal of Population Economics*, 2005, *18* (2), 229–260.

McKenzie, David J., "How do Households Cope with Aggregate Shocks? Evidence from the Mexican Peso Crisis," *World Development*, 2003, *31* (7), 1179 – 1199.

Meng, Xin and Guochang Zhao, "The long shadow of a large scale education interruption: The intergenerational effect," *Labour Economics*, 2021, *71*, 102008.

Monroy-Gómez-Franco, Luis, "Shades of social mobility: Colorism, ethnic origin and intergenerational social mobility.," Working Paper, SocArXiv Jul 2022.

_ **and Miles Corak**, "A Land of Unequal Chances: Social Mobility Across Mexican Regions," Working Paper 2019.

_ **and Roberto Vélez-Grajales**, "Skin Tone Differences in Social Mobility in Mexico: Are We Forgetting Regional Variance?," *Journal of Economics, Race, and Policy*, 2021, *4* (4), 257–274.

_ **,** _ **, and Gastón Yalonetzky**, "Layers of Inequality: Unequal Opportunities and Skin Color in Mexico," *The Review of Black Political Economy*, 2022, *49* (3), 230–250.

Mujcic, Redzo and Paul Frijters, "Economic choices and status: measuring preferences for income rank," *Oxford Economic Papers*, 01 2012, *65* (1), 47–73.

Murphy, Richard and Felix Weinhardt, "Top of the Class: The Importance of Ordinal Rank," *The Review of Economic Studies*, 05 2020, *87* (6), 2777–2826.

Murre, Jaap M.J., Steve M.J. Janssen, Romke Rouw, and Martijn Meeter, "The rise and fall of immediate and delayed memory for verbal and visuospatial information from late childhood to late adulthood," *Acta Psychologica*, 2013, *142* (1), 96 – 107.

Navarrete, Federico, *Mexico racista. Una denuncia*, Grijalbo, 2016.

Neidhöfer, Guido, Nora Lustig, and Mariano Tommasi, "Intergenerational transmission of lockdown consequences: Prognosis of the longer-run persistence of COVID-19 in Latin America," *Journal of Economic Inequality*, 2021.

Nenadic, Oleg and Michael J. Greenacre, "Correspondence Analysis in R, with Two- and Three-dimensional Graphics: The ca Package," *Journal of Statistical Software*, 2007, *20* (3), 1–13.

Nybom, Martin and Jan Stuhler, "Biases in Standard Measures of Intergenerational Income Dependence," *Journal of Human Resources*, 2017, *52* (3), 800–825.

Palma, José Gabriel, "Why the rich always stay rich (no matter what, no matter the cost)," *CEPAL Review*, 2020, (132), 93–132.

Paul, Mark, Khaing Zaw, and William Darity Jr., "Returns in the Labor Market: A Nuanced View of Penalties at the Intersection of Race and Gender in the US," *Feminist Economics*, 2022, *28* (2), 1–31.

Plassot, Thibaut, Isidro Soloaga, and Pedro Torres, "Inequality of Opportunity in Mexico and its Regions: A Data-Driven Approach," *The Journal of Development Studies*, 2022, *0* (0), 1–17.

Poirier, Mathieu J. P., Karen A. Grépin, and Michel Grignon, "Approaches and Alternatives to the Wealth Index to Measure Socioeconomic Status Using Survey Data: A Critical Interpretive Synthesis," *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, February 2020, *148* (1), 1–46.

Prais, Sigbert, "Measuring Social Mobility," *Journal of the Royal Statistical Society. Series A (General)*, 1955, *118* (1), 56–66.

Pritchett, Lant and Justin Sandefur, "Girls' schooling and women's literacy: schooling targets alone won't reach learning goals," *International Journal of Educational Development*, 2020, *78*, 102242.

Prix, Irene and Jani Erola, "Does death really make us equal? Educational attainment and resource compensation after paternal death in Finland," *Social Science Research*, 2017, *64*, 171–183.

Pérez-Asenjo, Eduardo, "If happiness is relative, against whom do we compare ourselves? Implications for labour supply," *Journal of Population Economics*, October 2011, *24* (4), 1411–1442.

Rahman Khan, Shamus, "The Sociology of Elites," *Annual Review of Sociology*, 2012, *38* (1), 361–377.

Ramírez-Raymundo, Rodolfo, José Gutiérrez-García, and Magdalena Rodríguez de la Huerta, "Orientaciones para apoyar el estudio en casa de niñas, niños y adolescentes. Educación preescolar, primaria y secundaria," Technical Report, Secretaría de Educación Pública, Mexico City 2021.

Reece, Robert L., "The Gender of Colorism: Understanding the Intersection of Skin Tone and Gender Inequality," *Journal of Economics, Race, and Policy*, March 2021, *4* (1), 47–55.

Reeskens, Tim and Rodrigo Velasco-Aguilar, "Being white is a full time job? Explaining skin tone gradients in income in Mexico," *Journal of Ethnic and Migration Studies*, 2021, *47* (1), 46–68.

Roemer, John, *Equality of Opportunity*, Harvard University Press, 1998.

Rojas-Lomelín, Marco A., Jhader Aguad-Revilla, and Judith A. Morrison, "Diversidad étnico-racial en México y su influencia en la movilidad social," Working Paper IDB-TN-1674, Inter-American Development Bank 2019.

Rosenblum, Alexis, William Darity, Jr., Angel L. Harris, and Tod G. Hamilton, "Looking through the Shades: The Effect of Skin Color on Earnings by Region of Birth and Race for Immigrants to the United States," *Sociology of Race and Ethnicity*, 2016, *2* (1), 87–105.

Roth, Wendy D., "The multiple dimensions of race," *Ethnic and Racial Studies*, 2016, *39* (8), 1310–1338.

Sacerdote, Bruce, "When the Saints Go Marching Out: Long-Term Outcomes for Student Evacuees from Hurricanes Katrina and Rita," *American Economic Journal: Applied Economics*, January 2012, *4* (1), 109–35.

Sakikawa, Kaoko, "Regional Convergence in Mexico, 1970–2005: A Panel Data Approach," *Growth and Change*, 2012, *43* (2), 252–272.

Saldívar, Emiko, "'It's Not Race, It's Culture': Untangling Racial Politics in Mexico," *Latin American and Caribbean Ethnic Studies*, 2014, *9* (1), 89–108.

Segato, Rita, "Los cauces profundos de la raza latinoamericana: una relectura del mestizaje," *Crítica y Emancipación*, 2010, *2* (3), 11–44.

Sen, Maya and Omar Wasow, "Race as a Bundle of Sticks: Designs that Estimate Effects of Seemingly Immutable Characteristics," *Annual Review of Political Science*, 2016, *19* (1), 499–522.

Serván-Mori, Edson, Clara Juárez-Ramírez, Sergio Meneses-Navarro, Ileana Heredia-Pi, Nancy Armenta-Paulino, Emanuel Orozco-Núñez, and Gustavo Nigenda, "Ethnic Disparities in Effective Coverage of Maternal Healthcare in Mexico,

2006–2018: a Decomposition Analysis," *Sexuality Research and Social Policy*, 2022, pp. 1–14.

Shorrocks, Anthony, "The measurement of mobility," *Econometrica*, 1978, *46* (5), 1013–1024.

Solís, Patricio, Alice Krozer, Carlos Arroyo Bautista, and Braulio Güémez, "Discriminación étnico-racial en México: una taxonomía de las prácticas," Working Paper 1, Proyecto sobre Discriminación Étnico Racial en México 2019.

_ **and Braulio Güémez**, "Características étnico-raciales y desigualdad de oportunidades económicas en México," *Estudios Demográficos y Urbanos*, nov. 2020, *36* (1), 255–289.

Symeonaki, Maria and Glykeria Stamatopoulou, "On the Measurement of Positive Labor Market Mobility," *SAGE Open*, 2020, *10* (3), 2158244020934489.

Telles, Edward, "The overlapping concepts of race and colour in Latin America," *Ethnic and Racial Studies*, 2012, *35* (7), 1163–1168.

_ **and Florencia Torche**, "Varieties of Indigeneity in the Americas," *Social Forces*, 10 2018, *97* (4), 1543–1570.

_ **and Tianna Paschel**, "Who Is Black, White, or Mixed Race? How Skin Color, Status, and Nation Shape Racial Classification in Latin America," *American Journal of Sociology*, 2014, *120* (3), 864–907.

_ **, ed.**, *Pigmentocracies: Ethnicity, Race and Color in Latin America.*, The University of North Carolina Press, 2014.

Todd, Petra E. and Kenneth I. Wolpin, "On the Specification and Estimation of the Production Function for Cognitive Achievement*," *The Economic Journal*, 2003, *113* (485), F3–F33.

Torche, Florencia, "Intergenerational Mobility and Gender in Mexico," *Social Forces*, 2015, *94* (2), 563–587.

Villarreal, Andrés, "Stratification by Skin Color in Contemporary Mexico," *American Sociological Review*, 2010, *75* (5), 652–678.

Vélez-Grajales, Roberto, Luis Monroy-Gómez-Franco, and Gastón Yalonetzky, "Inequality Of Opportunity In Mexico," *Journal of Income Distribution*, 2018, *27* (3-4), 143–157.

_ **, Omar Stabridis, and Enrique Minor-Campa**, "Still Looking for the Land of Opportunity. The Case of Mexico," *Sobre México. Revista de Economía*, 2018, *4* (1), 54–69.

Weiss, Yoram and Chaim Fershtman, "Social status and economic performance:: A survey," *European Economic Review*, 1998, *42* (3), 801–820.

Wendelspiess-Chávez-Juárez, Florian, "Measuring Inequality of Opportunity with Latent Variables," *Journal of Human Development and Capabilities*, 2015, *16* (1), 106–121.

Wittenberg, Martin and Murray Leibbrandt, "Measuring Inequality by Asset Indices: A General Approach with Application to South Africa," *Review of Income and Wealth*, 2017, *63* (4), 706–730.