ECINEQ

Society for the Study of Economic Inequality

Working Paper Series

Unpacking inequality of opportunity in Chile: the role of birth circumstances using a Shapley decomposition

Carlos Gradín

Gabriela Zapata-Román

ECINEQ 2024 676

ECINEQ 2024 676
2024 September
www.ecineq.org

Unpacking inequality of opportunity in Chile: the role of birth circumstances using a Shapley decomposition

Carlos Gradín
Universidade de Vigo
Gabriela Zapata-Román
Universidad Central de Chile

Abstract

We investigate the role of inherited circumstances in the inequality of opportunities in the labor market in Chile. We estimate the aggregate Shapley contribution of circumstances consistently across various measures reflecting legitimate views on inequality. We argue that the decomposition of path-dependent measures does not consider the effect of circumstances on inequality through the interaction of inequality between and within types. This largely overestimates their contribution using the Gini index and underestimates it with entropy measures. Circumstances explain 27-28 percent of inequality in 2022 using three entropy measures and 17 percent with extreme sensitivity to the poorest. They also explain 36 percent of inequality with the Gini index, which is less sensitive to both extremes and primarily affected by social stratification. Despite its persistence, inequality of opportunity has mainly driven inequality since the Great Recession. We also investigated population group contributions using a decomposition based on the Recentered Influence Function.

Keyword: inequality decomposition, equality of opportunity, Shapley, RIF, Chile

JEL Cassification: D63, O15, O54

Unpacking inequality of opportunity in Chile: the role of birth circumstances using a Shapley decomposition

Carlos Gradín

(Universidade de Vigo)

Gabriela Zapata Román

(Universidad Central de Chile, University of Manchester)

July 24, 2024

Abstract:

We investigate the role of inherited circumstances in the inequality of opportunities in the labor market in Chile. We estimate the aggregate Shapley contribution of circumstances consistently across various measures reflecting legitimate views on inequality. We argue that the decomposition of path-dependent measures does not consider the effect of circumstances on inequality through the interaction of inequality between and within types. This largely overestimates their contribution using the Gini index and underestimates it with entropy measures. Circumstances explain 27-28 percent of inequality in 2022 using three entropy measures and 17 percent with extreme sensitivity to the poorest. They also explain 36 percent of inequality with the Gini index, which is less sensitive to both extremes and primarily affected by social stratification. Despite its persistence, inequality of opportunity has mainly driven inequality since the Great Recession. We also investigated population group contributions using a decomposition based on the Recentered Influence Function.

Keywords: inequality decomposition, equality of opportunity, Shapley, RIF, Chile.

JEL Classification: D63, O15, O54.

Carlos Gradín, cgradin@uvigo.gal

Gabriela Zapata Román, gabriela.zapataroman@ucentral.cl

Acknowledgments

The authors acknowledge very constructive comments from Pedro Salas-Rojo and other participants at the GEOM conference held at Bari in June 2024 and the WIDER Development Conference at Bogota in October 2022.

Carlos Gradín acknowledges Grant PID2022-137352NB-C42 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe", and ED431B2022/03 by Xunta de Galicia. Gabriela Zapata-Román acknowledges the funding of the Chilean National Research and Development Agency ANID and the *Fondecyt Postdoctorado* No 3210480.

1. INTRODUCTION

Equality of opportunity embraces the idea that an individual's economic outcomes, such as income, asset ownership, or educational achievement, and their distribution in society are not only determined by individual effort but also by the opportunity set that one starts with in life (Arneson, 1989; Cohen, 1989). In this line, Roemer (1998) developed a model that differentiates between circumstances and efforts, which jointly determine economic outcomes. To achieve equal opportunities, the distribution of economic outcomes must be necessarily independent of circumstances. Thus, the inequality generated by observable exogenous characteristics such as gender, parental income, parental education, parental occupation, or ethnic background, among others, provides a measure of inequality of opportunity (IO), quantifying the relevance of exogenous circumstances in shaping a person's outcomes. This method has been adopted, among others, by Ferreira and Gignoux (2011) and Brunori, Ferreira, & Neidhöfer (2023) for Latin American countries, Checchi et al. (2016) for Europe, and Brunori et al. (2019a) for African countries. In Latin America, circumstances account for about 23 to 35 percent of overall income inequality as measured by the Mean Log Deviation (MLD).[1]

The aim of this paper is to investigate how much of the high level of income inequality in Chile is due to family and other predetermined circumstances. The answer to this question depends on various factors, like what circumstances are considered, how we account for individual economic capacity, and how we measure inequality. In this paper, we use the most common individual characteristics determined at birth (family of origin, gender, region, ethnicity) to define different population types to calculate ex-ante inequality of opportunity as inequality explained by differences in individual market income between the various types (subgroups) that these circumstances generate. As Ferreira & Gignoux (2011) pointed out, unobserved circumstances conceptually make these results a lower bound of the actual inequality of opportunity when computed in a population. However, Brunori, Peragine, et al. (2019) remarked that the small number of household survey observations in some types may produce an upward bias in its estimation. Nevertheless, we show that this latter potential estimation bias is small in our case. We also agree with Atkinson (2015) that there are good reasons to be concerned with inequality of outcomes and not only of opportunities, even if the reduction of the latter generates greater consensus and may unravel structural social failures.

[1] The MLD, also known as M-Theil, is a member of the Entropy family of indices with $\alpha = 0$: GE_0. Estimates from Ferreira & Gignoux (2011) for Brazil, Colombia, Ecuador, Guatemala, Panama, and Peru.

Chile is an interesting case study. Since its return to democracy in 1990, Chile has shown strong economic growth, increasing living standards and drastically reducing absolute poverty. However, it has yet to successfully mitigate income inequality, remaining alongside Costa Rica as the most unequal OECD country (OECD, 2024). In this context, when trying to understand how social and economic factors have affected the distribution of opportunities in Chile, the literature has stressed the effect of the structural transformations the country has experienced over the opportunity structure (Torche & Wormald, 2004). Therefore, investigating the inequality of opportunity in Chile might help us to understand the mechanisms that produce these extreme inequalities in welfare, education, and other individual achievements (Fleurbaey & Peragine, 2013).

Regarding the measurement of inequality, the literature has shifted from initially being based on the Mean Log Deviation (since it satisfies the path-independent decomposability axiom) to a more extended use of the Gini index, as in Brunori, Peragine, et al. (2019) and subsequent literature. Most standard measures of inequality (such as entropy measures, the Gini index, and others) share the fundamental property of assessing what distributions have more inequality consistently with the comparison of their non-crossing Lorenz curves (i.e., they all verify the Pigou-Dalton principle of transfers). However, they may disagree about how the distributions compare with intersecting Lorenz curves because they emphasize distributional differences taking place at different parts of the distribution, and these may have opposite effects on inequality. The Gini index in the recent literature was introduced because it is less sensitive to extreme values within types, thus removing less inequality than other measures after smoothing incomes within types. It is important to note that it is also less sensitive to the presence of very affluent or very disadvantaged types, thus reducing inequality if instead we remove inequality between types.

Furthermore, inequality measures differ in how they decompose inequality into the contribution of inequality between and within groups. This is also important because inequality of opportunity is often implemented as inequality between population groups represented by their counterfactual (average) incomes, with groups being either types sharing the same circumstances (ex-ante approach) or tranches with the same level of effort (ex-post approach). In the context of the ex-ante approach, we argue here that the interpretation of the between-type term as the contribution of circumstances to overall inequality, usually presented as a percentage, may be misleading in the context of path-dependent indices because it does not account for the full effect of average incomes across types on inequality. With these indices, inequality within and between types does not add to overall inequality, i.e., their sum is either larger or smaller. In other words, the remaining inequality between types is different from the change in inequality after equalizing the average income of all types (i.e., removing all inequality of opportunity). This means that there is another term, an interaction of the between and within-type distributions that is, therefore, also influenced by existing inequality of opportunity and not accounted for. For example, overall inequality using some entropy measures ($\alpha > 0$) tends to be higher if the most affluent groups tend to be the most unequal. This also affects the Gini index, but the main effect is that inequality is reduced with the income stratification of groups (when they do not overlap much). This interaction term may be negative, implying an overestimation of the contribution of inequality of opportunity to overall inequality, or positive, leading to an underestimation instead.

Indeed, the primary justification for using the MLD in the early literature was that it is the only path-independent measure (Foster & Shneyerov, 2000), implying that inequality between groups is precisely equal to inequality that is gone after equalizing average income across groups. This was also followed until recently in other areas, such as the literature on global inequality and its decomposition into inequality between and within countries (e.g., Bourguignon & Morrisson, 2002). Due to this, we argue that the use of path-dependent measures, particularly the Gini index and other entropy measures or the Atkison family, may misestimate the actual importance of inequality of opportunity. However, constraining the analysis of the MLD is highly restrictive as it imposes a particular sensitivity to the extremes of the distribution that may not be shared by everybody (e.g., Brunori, Palmisano, et al., 2019). Its use makes much sense if one is more concerned with the inequality of opportunity generated by the most disadvantaged groups. Others may be more concerned with the accumulation of income among most affluent groups instead, in line with recent developments in income and wealth inequality discussions (Atkinson, 2005 and subsequent literature), in which case the use of measures like the T-Theil or the coefficient of variation would likely be more appropriate. This makes the researcher choose between the inequality measure with the desired distributive properties or the only path-dependent one.

In the approach we propose here, rather than imposing a particular view on inequality, we investigate to what extent the relevance of circumstances in shaping inequality varies with the sensitivity to different parts of the distribution exploiting the information of a battery of inequality measures (the members of the Entropy family and the Gini index) and the Lorenz curve. To overcome the critical challenge regarding the decomposability of inequality measures that are not path-independent, we adopt the Shapley approach (Chantreuil & Trannoy, 2013; Shorrocks, 2013) and estimate the contribution of inequality of opportunity as the inequality between average incomes by type and the level of inequality that would be gone after equalizing the average incomes across types. As we will discuss, this does not affect the estimates for the MLD but substantially impacts other measures, particularly the Gini index or the square of the coefficient of variation (GE_2). The Shapley approach fully accounts for the influence of between-type differences on overall inequality by attributing half of the interaction effect to between-type inequality and the other half to within-type inequality. It makes more meaningful comparisons across indices and puts the Gini and the GE_2, otherwise outliers, in line with the different measures. The Shapley decomposition provides a clear path to guarantee path independence regardless of the original decomposability properties of the index, thus widening the possibility set and allowing the study of the extent to which the relevance of inequality of opportunity depends on the observer's inequality views.

This paper aims to contribute to the current literature by analyzing the distribution of opportunities in the labor market in Chile using a battery of measures with different distributive sensitivities.[2] The Shapley approach correctly estimates the relative relevance of inequality of opportunity (conditional on a given set of circumstances) with any measure and how this varies depending on distributive views. We find that inequality of opportunity in Chile in 2022 is around 27-28 percent of net market income inequality using entropy measures, with the percentage being smaller (17 percent) with a higher sensitivity to the lower end of the distribution. The share using the Gini index, less sensitive to both extremes than other measures and where inequality is mitigated by income stratification, is 36 percent. This is higher than with MLD (27 percent) but substantially smaller than the share of between-type inequality to overall inequality (56 percent) when the interaction effect is not accounted for. We argue that it is hard to claim that inequality of opportunity represents 56 percent. In comparison, inequality within types (after removing all between-type inequality) represents an even more significant 85 percent of overall inequality. In other words, a Chilean government that managed to remove inter-type differences (based on the types here identified) would be disappointed to find out that only 15 percent of the original overall Gini inequality would have vanished. The extensive range of inequality of opportunity estimates, 17-36 percent, reflects that the relevance of inequality of opportunity depends on subjective views about inequality, something intrinsic to the measurement of inequality more broadly, even if the range is smaller 27-36 percent if we exclude an extreme sensitivity to the bottom of the distribution rarely used in the literature.

We additionally propose to enrich the so-called opportunity profile proposed by Ferreira & Gignoux (2011) by directly computing the contribution of every type (population groups defined by combinations of circumstances) to inequality of opportunity, as well as to inequality within types, based on the Recentered Influence Function of the corresponding inequality measure

[2] Earlier studies on the inequality of opportunity in Chile include Contreras et al. (2014) and Núñez & Tartakowsky (2011). Chile was also included in the sample of countries recently studied by Brunori, Ferreira, & Neidhöfer (2023).

(Gradín, 2020). A Blinder-Oaxaca type of decomposition allows us to investigate the role of the identified types in driving the inequality of opportunity, which largely drove the overall trend in inequality in Chile in the analyzed period, during and after the Great Recession.

The following section discusses the theoretical framework and methods used to measure inequality of opportunity. Later, we describe the data and discuss the results. The last section concludes.

2. EQUALITY (INEQUALITY) OF OPPORTUNITY

The general approach

Built on the critical work on social justice of political philosophers such as John Rawls and Ronald Dworkin, equality of opportunity embraces the idea that a fair society does not necessarily equalize happiness, wealth, or education. Instead, it provides its members equal access to the inputs needed to achieve the outcomes they care about (Ferreira & Peragine, 2015). Roemer (1993, 1998) states that equality of opportunity policy aims to "level the playing field" to compensate for uneven circumstances over which individuals should not be held accountable, and that affects their ability to achieve the advantages they would like to pursue.

In this literature, the selection of the circumstances has been more straightforward than the decision about what is considered an effort. "Opportunities are inherently unobservable because they are, by definition, a set of hypothetical options, some of which are exercised –and become factual– while others are not exercised and become counterfactual" (Ferreira & Peragine 2015, p. 8). For that reason, the measurement method is 'indirect' in that it measures how opportunities such as the characteristics of the family of origin (parental education and occupation) and other given personal circumstances (place of birth, gender, or ethnicity) affect the outcome of interest.

The main approaches to measuring inequality of opportunity are the 'ex-ante' (Van de Gaer 1993) and the 'ex-post' (Roemer, 1993, 1998), reflecting the compensation and reward principles. They decompose total inequality into an "ethically acceptable" component resulting from the differential effort and an "ethically unacceptable" part resulting from unequal opportunities expressed by exogenous circumstances. If the initial conditions are compensated before any effort is made, we refer to the 'ex-ante' approach related to the compensation principle. People should not have different outcomes just because they face different circumstances of origin. If the compensation occurs after the efforts have been made, we refer to the 'ex-post' approach, which is related to the reward principle. People who exercised the same level of effort –or made the same choices– should achieve the same outputs.

In the 'ex-ante' approach, an inequality index applied to the new –counterfactual– distribution of averages by type would reflect the inequality of opportunity since all effort inequalities have been removed. Likewise, in the 'ex-post' approach, an inequality index applied to this counterfactual distribution that contains average by tranches of effort should reflect the fair inequality.

The assumptions about the distribution of effort needed in the 'ex-post' approach make the estimates less robust than those obtained with the 'ex-ante' approach, which only needs variables of circumstance to be implemented (Roemer and Trannoy, 2016). Therefore, it is the most popular approach used in numerous countries (Brunori, 2016), and we adopt it here.

Summarizing, two income vectors are needed to estimate 'ex-ante' inequality of opportunity:

$$y = \{y_1, \ldots\ldots, y_n\} \in \mathbb{R}_+^N$$

$$y_b = \{\mu_1 1_{N_1} \ldots, \mu_i 1_{N_i} \ \ldots\ldots, \mu_n 1_{N_n}\} \in \mathbb{R}_+^N$$

Where y represents the overall income distribution with mean μ of a population with n types, y_i is the income distribution of type i with mean μ_i, and y_b represents the counterfactual distribution that eliminates within-type inequality (leaving only the inequality of opportunity) by giving everyone the average income of their type, where 1_{N_i} is a 1-vector of size N_i.

Although the counterfactual distribution has also been estimated fully parametrically (Bourguignon et al., 2007; Ferreira & Gignoux, 2011), the most straightforward and intuitive non-parametric method divides the population by types based on a given set of circumstances, computing the average income for each type. There is a risk of downward bias if critical circumstances are omitted and upward bias due to overfitting if the resulting types involve small subsamples (Brunori, Ferreira, & Neidhöfer, 2023; Brunori, Peragine, et al., 2019). For this reason, the most recent research has followed these studies and adopted data-driven methods to optimally partition the population and predict average incomes, such as *Conditional Inference Trees* or their extension in *Conditional Inference Random Forests.* This approach may be convenient in some contexts, like when comparing countries with varying sample sizes and heterogeneous social structures. However, the underlying classifications by type from these trees vary across samples and may not correspond with prevailing normative views about what determines opportunities in a specific country. For this reason, in the main results of this country study, we define ad-hoc circumstances that are relevant in the country and widely used in international analyses based on normative considerations, as we think they are the most meaningful in this context. We investigate the risk of upward bias due to overfitting by considering various levels of aggregation of types, introducing circumstances in a sequence. This potential bias is small as long as we account for the predominant role of two critical circumstances (parental education and gender) that tend to produce larger types.

The other circumstances, which may result in smaller types, only make a modest contribution. Thus, the results presented with the non-parametric method are robust to the risk of overfitting, likely influenced by the relatively large samples in the Chilean context. Nevertheless, we also compute conditional inference trees and random forests, leading to a similar conclusion. The risk of downward bias remains as we cannot control for unobservable circumstances.

Inequality of opportunity is then obtained by applying an inequality index $I(\cdot)$ to the distribution y_b: $I(y_b)$. This is usually called the *absolute* measure of inequality of opportunity. Using Checchi & Peragine's (2010) notation, this is usually expressed as a share of the total observed inequality $I(y)$, indicating the relevance of circumstances to explain inequality in outcomes, and called a *relative* measure of inequality of opportunity:

$$IO^b = \frac{I(y_b)}{I(y)}$$

Selection of the inequality index

The early literature used the mean logarithmic deviation (MLD, also known as M-Theil, or entropy measures with $\alpha = 0$, GE_0) because it satisfies the 'path-independent decomposability' axiom, allowing total inequality to be decomposed as the exact sum of inequality due to effort (within-type inequality) and inequality of opportunities (between-type inequality).[3]

Most common inequality measures such as the Gini index, the Generalized Entropy family, the Atkinson family, and others have in common that they verify the defining feature of an inequality measure, i.e., that inequality declines after small progressive transfers and increases after small regressive transfers that do not change the average income (the Pigou-Dalton Principle of Transfers), consistently with the comparison of non-intersecting Lorenz curves (Atkinson, 1970).[4] However, they differ in how much the impact of such transfers on inequality varies across the distribution, allowing them to differentially quantify the exact distributional change depending on what incomes are affected.[5] This is particularly relevant for the direction of inequality whenever Lorenz curves intersect and the magnitude of an inequality changes in the presence of Lorenz dominance. For example, the MLD is known to be more sensitive to the bottom of the distribution

[3] See, for example, Checchi et al. (2010); Checchi & Peragine (2010); Ferreira & Gignoux (2011); Singh (2012)). For a technical discussion of path dependency and inequality measures, see Foster & Shneyerov (2000), and for a full discussion of the decomposability of inequality measures, see Chakravarty (2009).

[4] Alongside symmetry, replication invariance, and scale invariance.

[5] For example, see how the marginal increase in population at each percentile of the income distribution impacts inequality as measured by various indices using the Recentered Influence Function (Gradín, 2020). Also, see Figure 6a below for the relative contribution of population groups based on their relative income.

than Theil (L-Theil or GE_1) or GE_2 (or 1/2 CV^2), which, however, are more sensitive to the upper tail. The higher the parameter α, the more sensitive the Generalized Entropy (GE_α) index will be to the upper part of the distribution. For instance, the Gini index is known to be relatively less sensitive to both extremes than MLD or Theil.

In the context of inequality of opportunity, this index-specific sensitivity is inherited by both the inequality of opportunity component and the inequality within types. These specific sensitivities may have supporters and detractors; therefore, relying on only one measure, such as MLD, may be problematic. For example, people more concerned with population groups left behind during the economic development process may like indices of inequality of opportunity that are more sensitive to the bottom of the distribution. MLD can be a good option (with GE_{-1} being a more radical alternative). However, people more concerned with the concentration of income and, therefore, political power in the hands of small but very affluent population groups may prefer an index more sensitive to the top of the income distribution and Theil, or a more extremist GE_2 would be a better option. Between these two views, the Gini index is also a reasonable intermediate solution without such a strong preference for any of both extremes. In this line, (Brunori, Palmisano, et al., 2019) focusing on the effect of each index on equalizing within-type incomes, favored the use of the Gini index, claiming that the MLD underestimates inequality of opportunity (i.e., removes too much inequality after equalizing within-type incomes). Others followed this trend (although the MLD or other indices may also be computed).[6]

The paradox of path-dependent measures

Here, we argue that directly interpreting the ratio of inequality between types, IO^b, as the contribution of inequality of opportunity to overall inequality (inequality of outcomes) is problematic when using path-dependent measures. Suppose one correctly estimates the importance of inequality within types (by removing income differences between types). In that case, the sum of inequality between and within types will exceed total inequality when using the Gini index. At the same time, it may fall short when using entropy measures other than MLD, as shown in the empirical analysis. This inconsistency creates a paradox. Based on the IO^b ratio of the Gini index, inequality of opportunity represented 56 percent of overall market income inequality in Chile in 2022. This may create the perception that this is also the amount of inequality

[6] For example, Brunori (2016); Brunori, Ferreira, & Neidhöfer (2023); Brunori, Ferreira, & Salas-Rojo (2023; Brunori, Palmisano, et al. (2019) and Cabrera et al. (2021), among many others. Also, the recently launched GEOM project (geom.ecineq.org; accessed in June 2024) compares estimates of absolute and relative measures of inequality of opportunity and intergenerational mobility for 72 countries and 196 household surveys using the Gini coefficient alongside the MLD.

that would be removed if a complete equalization of incomes across types that eliminates inequality of opportunity were possible (only within-type inequality remained). This is another reasonable measure of the relevance of inequality of opportunity that we can label as IO^w as it was obtained comparing existing inequality with inequality in the within-type distribution. However, had the government been able to eliminate all average differences between population types by compensating those below the country mean with the surplus of the other groups, it would observe that only 15 percent of the original inequality disappeared, far less than the claimed 56 percent. This is because inequality within types, so defined, represents an even more significant 85 percent of overall inequality using the Gini index. The sum of inequality between and within types adds up to 141 percent of overall inequality. This means that the original 56 percent was obtained out of 141 percent explained by both sources, not out of 100 percent (which only happens if the measure is path-independent). This extra 41 percentage points, which we need to subtract to reach 100 is the gap between both reasonable measures of the magnitude of inequality of opportunity and stems from the significant (negative) impact that the interaction of the between-type and within-type distributions produces on overall inequality. The first measure (IO^b) attributes the entire interaction to inequality within types, while the second measure (IO^w) attributes it entirely to inequality between types. That is, IO^b of the Gini index overestimates the importance of inequality of opportunity, while IO^w underestimates it. To a different extent, a similar problem affects other path-dependent measures. However, the sign of the interaction is positive in our empirical application, so IO^b underestimates how important inequality of opportunity is, while IO^w overestimates it.

Formally, let us consider again the between-type distribution y_b obtained after replacing each observed income in y with the average income of the type the person belongs to, keeping average income by type unchanged. This gives the level of inequality between types, $I(y_b)$, to compute the IO^b ratio defined above. It also produces the level of inequality that goes away after equalizing earnings within types. This is a measure of the residual inequality not directly explained by observed circumstances (that could be the result of effort, luck, or the unobserved circumstances):

$$IR^b = \frac{I(y) - I(y_b)}{I(y)}$$

Additionally, following the alternative path, we can obtain the within-type distribution y_w by eliminating between-type inequality in y by redistributing income across types, i.e., rescaling each person's incomes so that all types have the same average earnings while keeping within-type inequality $I(y_w)$ unchanged:[7]

$$y_w = \left\{y_1 \frac{\mu}{\mu_1} \dots, y_i \frac{\mu}{\mu_i} \dots\dots, y_n \frac{\mu}{\mu_n}\right\} \in \mathbb{R}_+^N$$

From this distribution, one could quickly obtain the proportion of overall inequality explained by the within-type distribution as:

$$IR^w = \frac{I(y_w)}{I(y)}$$

This also gives us the inequality that would be gone after equalizing incomes across types, for example, after redistribution, keeping inequality within types unchanged: $I(y) - I(y_w)$. This provides an alternative measure of inequality of opportunity, likely the most relevant from a policy perspective since it indicates how far the Government can go in fighting inequality by equalizing circumstances, as:[8]

$$IO^w = \frac{I(y) - I(y_w)}{I(y)}$$

In the MLD case, we have that $I(y) = I(y_b) + I(y_w)$, and therefore, there is a consistency between the remaining inequality between types, $I(y_b)$, and the inequality that is gone if we eliminate inequality of opportunity, $I(y) - I(y_w)$, with $IO^b = IO^w$. Similarly, there is a consistency between inequality within types, $I(y_w)$, and inequality that would be gone had all within-type differences been removed, $I(y) - I(y_b)$: $IR^b = IR^w$. Also, we get that $IO^b + IR^b = IO^w + IR^w = 1$.

However, for path-dependent measures, such as the other entropy members or the Gini index among others, they can be rewritten as:

$$I(y) = I(y_b) + I(y_w) + I_{bw}$$

where I_{bw} is an interaction term of the between-type and within-type distributions.

[7] Note that in the case of entropy measures, inequality of the within-type distribution, as defined here, equals a population-weighted sum of type inequality: $GE_\alpha(y) = \sum_{i=1}^{n} \frac{N_i}{N} GE_\alpha(y_i)$, in which weights are not influenced by the between-group distribution (unlike in Shorrocks 1984). In the case of the Gini index, it is something close, but mediated by an index of overlapping between each type and the entire distribution that will be discussed later in this section, measured on the rescaled distribution: $G(y_w) = \sum_{i=1}^{n} \frac{N_i}{N} O(y_{w_i}, y_w) G(y_i)$. Since $O(y_{w_i}, y_w)$ revolves around 1 in most cases given that all types are recentered at the country's mean earnings, $G(y_w) \approx \sum_{i=1}^{n} \frac{N_i}{N} G(y_i)$.

[8] The different inequality components are summarized in Table 2 in the discussion of the results.

For example, due to its additive decomposability, the interaction with entropy measures is a weighted sum of the average incomes by type (to the power of α), multiplied by their corresponding within-type inequality, $GE_\alpha(y_i)$:[9]

$$GE_{\alpha_{bw}} = \sum\nolimits_{i=1}^{n} \frac{N_i}{N}\left[\left(\frac{\mu_i}{\mu}\right)^{\alpha} - 1\right] GE_\alpha(y_i)$$

Only if $\alpha = 0$, i.e., the MLD case, this term is always zero. In the other cases, the interaction effect associated with a type is zero only if the group has the country's mean or no inequality. Otherwise, inequality will increase if more affluent groups are more unequal ($\alpha > 0$). Overall inequality will disproportionally increase with inequality of the poorest groups if $\alpha < 0$.

For the Gini index, I_{bw} is like Theil's case ($\alpha = 1$), inequality also tends to be higher if more affluent groups tend to be more unequal, but with the critical peculiarity that this effect is also affected by the degree by which each type overlaps with the entire distribution (the other types and itself) over the income space.[10] Formally:

$$G_{bw} = \sum\nolimits_{i=1}^{n} \frac{N_i}{N}\left[\frac{\mu_i}{\mu} O(y_i, y) - O\left(y_{w_i}, y_w\right)\right] G(y_i)$$

Where $O(y_i, y)$ is the overlapping index presented in Gradín (2000).[11] This overlapping index is a population-weighted sum of overlapping of each type with every other type in the distribution and itself. It takes a minimum value of N^i/N if the group does not overlap with other groups (it

---

[9] In the conventional Shorrocks (1984) additive decomposition, where $GE_\alpha(y) = GE_\alpha(y_b) + \sum_{i=1}^{n} \frac{N_i}{N}\left(\frac{\mu_i}{\mu}\right)^{\alpha} GE_\alpha(y_i)$, the second term is known as the within-group inequality component. However, it differs from $GE_\alpha(y_w)$ because it also depends on the group relative average incomes, and not only on population and group inequality, as in the case of the MLD ($\alpha = 0$). This is why there is an interaction term for all entropy measures with $\alpha \neq 0$.

[10] Given that $G(y) = G(y_b) + \sum_{i=1}^{n} \frac{N_i}{N}\frac{\mu_i}{\mu} O(y_i, y) G(y_i)$, with $O(y_i, y) = \sum_{j=1}^{n} \frac{N_i}{N} O(y_i, y_j) = \sum_{j=1}^{n} \frac{N_j}{N}\frac{d_{ij}}{d_{ii}}$, with $d_{ij} = \frac{1}{2}\frac{1}{N_i N_j}\sum_r \sum_s \left|y_i^r - y_j^s\right| - \frac{1}{2}\left|\mu_i - \mu_j\right|$ related to Dagum' (1980) economic distance (Gradín, 2020). Note that this interaction term differs from the residual term in the conventional decomposition by Bhattacharya & Mahalanobis (1967) or Pyatt (1976), which was interpreted graphically by Lambert & Aronson (1993) in the Lorenz setting. Such residual is equal to $G(y) - G(y_b) - \sum_{i=1}^{n} \frac{N_i}{N}\left(\frac{N_i}{N}\frac{\mu_i}{\mu}\right) G(y_i)$, where the last term, although usually called within-group inequality, is influenced by the between-type distribution. The interaction discussed here is $G(y) - G(y_b) - G(y_w)$, where $G(y_w) = \sum_{i=1}^{n} \frac{N_i}{N} O(y_{w_i}, y_w) G(y_i)$ is independent of y_b. In our example, the residual and interaction are of similar size but different sign: the interaction is -41 percent, and the residual is 42 percent. For an alternative decomposition of the Gini index in the context of inequality of opportunity, see Moramarco (2023).

[11] This overlapping index is close to the overlapping measure discussed by Yitzhaki (1994), the main difference is that the one presented here is obtained using the conventional between-group inequality term $I(y_b)$, while Yitzhaki uses an alternative definition as twice the covariance between the relative mean income of each group and its mean rank in the overall population (instead of the rank of the group average income that produces $I(y_b)$). Overlapping between two groups results from *transvariations* (Dagum, 1960; Gini, 1916, 1955), situations in which people from the relatively poorer group have incomes above those of people in the richer group.

only overlaps with itself), increases with overlapping, and reaches one if it perfectly overlaps with the entire distribution.[12] Intuitively, this interaction highlights that the impact on inequality as measured by the Gini index of a given income distance between two individuals of the same type, as well as between the average incomes of two types, increases with the number of people of other types with incomes in between them, because in the Gini index not only the distance in incomes matter, but also the distance in ranks.[13] The presence of overlapping in the interaction term indicates that using the Gini index, ceteris paribus, the more stratification (less overlapping) among types, the lower the overall inequality, a value judgment embedded in this measure of inequality that Yitzhaki (1982, 1994) related with Runciman's (1966) theory of relative deprivation, in which stratified societies can tolerate higher inequality.

In the empirical analysis, we show that the interaction is positive with entropy measures, with its size ranging during the 2009-22 period around 3-9 percent of overall inequality using the Theil index, 13-18 percent using GE_{-1}, and 21-48 percent with GE_2. The Gini index is also large but negative, as previously indicated, due to the effect of overlapping (between -40 and -42 percent).

Despite being related to both inequality sources (between and within), the IO^b ratio implicitly attributes the entire interaction term to within-type inequalities since it is not part of the contribution of circumstances. This is so because it is computed when there is no within-type inequality and, therefore, no interaction. This will underestimate the actual contribution of inequality of opportunity if the interaction is positive (like the entropy measures in our empirical analysis) or overestimate it if it is negative (like for the Gini index). The opposite is true when we measure how much inequality goes away after removing inequality of opportunity, IO^w, as we attribute the entire interaction to inequality between types (inequality of opportunity) and therefore this is underestimated (overestimated) whenever the interaction is negative (positive).

The paradox of path-dependent measures is thus explained by the between-type distribution having two effects on overall inequality: a direct effect that is measured by the IO^b ratio and another indirect effect that results from the interaction of the between and within-type distributions. Therefore, the impact of a given gap between rich and poor groups measured by the Gini index or Theil will be aggravated if more affluent groups tend to be more unequal. In the case of the Gini index, this effect is also strongly mitigated if these types do not overlap along the income distribution. Not accounting for this interaction effect leads to a misestimation of the

[12] It can be higher than 1 in the case in which a relatively poorer group has a substantial part of its population above the richest people from richer groups.

[13] If expressed as twice the covariance between relative incomes and ranks (Lerman & Yitzhaki, 1984), the Gini index can measure how much individual relative incomes increase with a higher rank in the overall distribution.

actual contribution of inequality of opportunity to overall inequality.

The Shapley decomposition

To overcome this problem while keeping the freedom to use any inequality measure, not only MLD, we propose using the Shapley decomposition of inequality measures into the contribution of two factors (Chantreuil & Trannoy, 2013; Shorrocks, 2013), namely here, inequality between and within types.[14] This implies obtaining the average between both estimates of inequality of opportunity, i.e., inequality between types $I(y_b)$, and the inequality gone after removing inequality between types, $I(y) - I(y_w)$.

$$I_{sb} = \frac{1}{2}[I(y_b) + I(y) - I(y_w)]$$

$$I_{sw} = \frac{1}{2}[I(y_w) + I(y) - I(y_b)]$$

The two relevant ratios become:

$$IO^s = \frac{I_{sb}}{I(y)} = \frac{1}{2}(IO^b + IO^w)$$

$$IR^s = \frac{I_{sw}}{I(y)} = \frac{1}{2}(IR^b + IR^w)$$

The rationale of the Shapley decomposition in our context is to equally split the interaction term between both sources of inequality involved (between and within types), such that:

$$IO^s = IO^b + \frac{1}{2}\frac{I_{bw}}{y}$$

$$IR^s = IR^w + \frac{1}{2}\frac{I_{bw}}{y}$$

With

$$IO^s + IR^s = 1$$

[14] Since inequality is the result of two sources, the contribution of each source can be obtained as the change in inequality after equalizing that source. However, there are two paths to remove all inequality. In one path we first remove between-type differences, so the between-type contribution $I(y_b)$ is obtained when the other source has already been removed. In the other one, we first remove between-type inequality when the other source is still present, $I(y) - I(y_w)$. The Shapley approach eliminates this path-dependency by averaging across both paths. For other applications see Davies & Shorrocks (2021) and Gradín (2024) for global inequality between and within countries or the country studies in Gradín et al. (2023) for estimating the contribution of occupations to earnings inequality.

In the MLD case, this does not make any difference, $IO^s = IO^b$. In our empirical example, we show that $IO^w < IO^s < IO^b$ for Gini, but $IO^w > IO^s > IO^b$ for entropy measures given the sign of the respective interaction effects.

For example, in the case of the Gini index, the Shapley inequality of opportunity ratio is 36 percent, the midpoint between 56 percent and 15 percent (i.e., after absorbing half of the -41 percentage-point interaction). At the same time, we will attribute the remaining 64 percent to inequality within types (the midpoint between 44 and 85 percent). By doing this, the Shapley contribution guarantees comparability across inequality measures, regardless of their original decomposability properties, since now the contributions of the distribution between types and within types add up to overall inequality for all measures and account for both direct and interaction effects of types. This is particularly useful if one wants to analyze how much the ratio varies with different sensitivities to inequality taking place at various parts of the distributions or other value judgments embedded in inequality measures, thus accommodating all legitimate views.

We believe that the percentage of IO^s measures more adequately than IO^b the contribution of inequality of opportunity to overall inequality because $I(y_b)$ does not account for the full effect of circumstances on overall inequality. The reason why IO^s differs from IO^w, the share of inequality that would go away after eliminating inequality of opportunity, becomes now more transparent. When the government eliminates inequality of opportunity making $I(y_b) = 0$, it also affects the contribution of the within-type component of inequality due to the interaction, even if the within-type distribution did not change at all (i.e., $I(y_w)$ does not change). This contribution of within-type inequality increases in the case of the Gini index because of the elimination of the mitigating effect that stratification across types had on overall inequality. The within-type distribution is reduced instead in the case of entropy measures because the government is also eliminating the aggravating effect that the correlation between inequality and average earnings across types had on overall inequality. The latter also influenced the Gini index, but this effect is small compared to the size of the effect of overlapping.

There is a well-known relationship between the Gini index and the empirical Lorenz curve $L(p)$, by which IO^b and IO^w can be easily obtained. The Lorenz curve is also path dependent, the distances between the diagonal and the Lorenz curves between and within types (L_b and L_w) at each population share p do not add up to the overall distance: $p - L(p) < (p - L_b(p)) + (p - L_w(p))$. For that reason, we can also compute the correspondent Shapley Lorenz curves $L_{sb}(p)$

and $L_{sw}(p)$, such that $p - L(p) = (p - L_{sb}(p)) + (p - L_{sw}(p))$.[15] The IO^s of the Gini index can be obtained as twice the area between the diagonal and L_{sb} or, alternatively, due to path independence, twice the area between L_{sw} and L.

We focused here on relative inequality measures (i.e., scale-invariant) like the Gini and entropy indices since this is the most common approach in the inequality literature. However, the Shapley decomposition can also accommodate other indices and inequality views, including absolute inequality measures (i.e., translation invariant) like the absolute Gini or the standard deviation, or intermediate ones, once the construction of y_b and y_w are adequately adjusted to ensure that respectively $I(y_b)$ and $I(y_w)$ remain unchanged.[16]

The Shapley approach is not the only possible way to split the interaction between the two sources. For example, consider splitting the interaction term proportionally to the relevance of inequality in each source. This would be equivalent to normalizing I_b by the sum of inequality in both sources, $I_b + I_w$:

$$IO^n = \frac{I(y_b)}{I(y_b) + I(y_w)}$$

$$IR^n = \frac{I(y_w)}{I(y_b) + I(y_w)}$$

Given that:

$$I(y_b) + \frac{I(y_b)}{I(y_b)+I(y_w)} I_{bw} = \frac{I(y_b)}{I(y_b)+I(y_w)} I$$

The main point here is that the interaction term should contribute to the importance of both sources. The equal split used by the Shapley decomposition aligns with how interaction terms are generally split.[17] In our context, it considers that a small component can have a disproportional effect on the interaction term because if it goes to zero, the whole term goes to zero as well, no matter the size of the other component. The Shapley approach can be easily applied when the counterfactual distribution y_b is estimated from explicit types, whether they are the mean income of types defined ad-hoc or they are estimated as the predicted incomes of types obtained by

[15] Note that the Shapley Lorenz curves are obtained as: $p - L_{sb}(p) = \frac{1}{2}\left[p - L_b(p) + p - L(p) - \left(p - L_w(p)\right)\right]$, and $p - L_{sw}(p) = \frac{1}{2}\left[p - L_w(p) + p - L(p) - \left(p - L_b(p)\right)\right]$.

[16] For example, for absolute inequality measures, this implies obtaining y_w by recentering incomes at the country's mean (by adding a common factor, $\mu - \mu_i$).

[17] It originated in cooperative game theory, where it is the normative prescription for individual payoffs based on their average marginal contribution to each coalition (Serrano, 2007).

statistical approaches (e.g., conditional inference trees), by simply computing and y_w with the same classification. In the case of conditional inference, random forests y_b is usually obtained as an average of the predicted incomes across many different tree classifications. Applying the Shapley approach requires, in this case, consistently obtaining y_w following an equivalent aggregation procedure (e.g., as the average of the within-distributions obtained in all trees in the forest).[18]

The contribution of types to inequality

Following the method proposed in Gradín (2020), the contribution of a group (type) to inequality $I(y)$ will be obtained using a RIF-regression, that is, an OLS regression of the Recentered Influence Function (RIF) of the inequality measure $I(y)$ over the set of type dummies: $A_i^j = 1$ if person j belongs to type i, 0 otherwise, with $j = 1, \dots N^i$; $i = 1, \dots, n$ (without an intercept):[19]

$$RIF\left(y_i^j\right) = \sum_{i=1}^{n} \beta_i A_i^j + u^j$$

The $RIF(y)$ of an inequality measure is just the expected change in the selected outcome (here overall inequality) after marginally increasing the population with income y (Firpo et al., 2009, 2018; Hampel, 1974).

The β_i coefficient indicates the per capita contribution of type i, with:

$$\beta_i = \frac{1}{N^i} \sum_{j=1}^{N^i} RIF\left(y_i^j\right)$$

The total contribution of this type can be obtained by multiplying β_i by its population share:

$$\frac{1}{N} \sum_{j=1}^{N^i} RIF(y_i^j) = \beta_i p_i$$

With

$$p_i = \frac{1}{N} \sum_{j=1}^{N} A_j^i = N_i / N$$

Overall inequality can then be written as the average per capita contribution:

$$I(y) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{N^i} RIF(y_i^j) = \sum_{i=1}^{n} \beta_i p_i$$

By doing the same with inequality in the between and within-type distributions, y_b and y_w, we get the corresponding parameters β_{b_i} and β_{w_i}. The per capita contribution β_{b_i} mainly depends on the

18 While the main results presented here will use the ad-hoc classification of types, the appendix shows the results after implementing the Shapley approach with conditional inference trees and random forests.

19 The influence function of an inequality measure at an income y is the expected change in inequality after marginally increasing the population with income y (Firpo et al., 2018; Hampel, 1974). The influence function is recentered so that its average is the observed level of inequality, thus obtaining the RIF.

distance between the average incomes of the type and the population with a U-shaped relationship reflecting the sensitivity of the index along the distribution. The per capita contribution β_{w_i} is given by inequality in the type, $GE_\alpha(y_i)$, in the case of entropy measures, and a value close to $G(y_i)$ in the case of the Gini index.[20] However, like the aggregate analysis above, both per capita effects have in common that they do not account for the impact of the type operating through the interaction effect. This is what the Shapley counterparts, β_{sb_i} and β_{sw_i}, do, by adding a half of the corresponding interaction effect, β_{bw_i}:

$$\beta_{sb_i} = \frac{1}{2}\left(\beta_{b_i} + \beta_i - \beta_{w_i}\right) = \beta_{b_i} + \frac{1}{2}\beta_{bw_i}$$

$$\beta_{sw_i} = \frac{1}{2}\left(\beta_{w_i} + \beta_i - \beta_{b_i}\right) = \beta_{w_i} + \frac{1}{2}\beta_{bw_i}$$

The aggregate Shapley between- and within-type contributions can then be rewritten as:

$$I_{sb} = \sum_{i=1}^{n} \beta_{sb_i} p_i$$

$$I_{sw} = \sum_{i=1}^{n} \beta_{sw_i} p_i$$

Thus, the Shapley per capita effect of a type to inequality of opportunity, β_{sb_i}, accounts for both the direct effect based on its relative income and for how this effect is mediated by the level of inequality in the group (as well as its overlapping with others in the case of the Gini index). This allows a complete decomposition of inequality as:

$$I(y) = I_{sb} + I_{sw} = \sum_{i=1}^{n} \beta_{sb_i} p_i + \sum_{i=1}^{n} \beta_{sw_i} p_i$$

Finally, a Blinder-Oaxaca type of decomposition will allow us to disentangle the contribution of every type to a change in inequality over time, ΔI, through a *distributive* change (change in the between- and within-type distributions with constant population) ΔD_{sb} and ΔD_{sw}, and their corresponding *composition* effects ΔC_b and ΔC_w (due to changes in the population shares with respectively constant between and within-type distribution).

$$\Delta I = I^1 - I^0 = \Delta D_{sb} + \Delta D_{sw} + (\Delta C_b + \Delta C_w)$$

[20] Note that in this context $\beta_{w_i} \geq 0$ for all measures, while β_{b_i} can be negative with $\alpha \neq 0$. A negative per capita contribution indicates that marginally increasing the population in that type (typically with average earnings around central values) reduces inequality between types using these indices instead of increasing it However, $\beta_{b_i} \geq 0$ with MLD (reaching 0 if the type has the country's average earnings), and $\beta_{b_i} > 0$ with Gini (with its minimum close to the country's average). See Figure 6a.

For example, evaluating the distributive effects with the original population in t=0 and the composition effects with the final distribution (t=1), these are:

$$\Delta D_{sb} = \sum_{i=1}^{n} p_i^0 \Delta\beta_{sb_i}$$

$$\Delta D_w = \sum_{i=1}^{n} p_i^0 \Delta\beta_{sw_i}$$

$$\Delta C_b = \sum_{i=1}^{n} \Delta p_i \beta_{bi}^1$$

$$\Delta C_w = \sum_{i=1}^{n} \Delta p_i \beta_{w_i}^1$$

3. Data

The data comes from the Chilean income survey *Encuesta de Caracterización Socioeconómica Nacional*, CASEN. The Ministry of Social Development and Family (Ministerio de Desarrollo Social y Familia) conducts it biannually. We use the 2009, 2011, 2013, 2015, 2017, and 2022 surveys. This cross-sectional survey collects information about households and household members and is representative at the national and regional levels.

While many studies in the field use household per capita income or consumption as the primary outcome variable, we follow the branch of the literature (e.g., Aaberge et al., 2011; Checchi & Peragine, 2010; Piraino, 2015) that focuses the analysis on individual labor income to study inequality of opportunity in the labor market, the primary source of household income. The main outcome variable will be individual monthly net market income from labor. Monetary amounts are expressed in 2022 constant Chilean pesos. The sample is restricted to household heads, men and women between 25 and 60 years old, who are active in the labor market, have positive income, and have available information on circumstances included in this study. The reason for including only household heads is that some critical observed circumstances are only associated with the household head in 2017 and 2022. Participation rates are lower for individuals under 25 and over 60, which justifies the age restriction.

Circumstances are exogenous factors affecting income earning. Some are observable, such as gender, place of birth, ethnicity, and family background. Others, such as a family's cultural and social capital and genetic traits, are unobserved. Some of them will directly affect income earning, such as genetic traits, gender, ethnicity, or family connections —if those help individuals get a better job (Aaberge et al., 2011; Becker & Tomes, 1986; Checchi & Peragine, 2010; Piraino, 2015). Others, like the place of birth, might do so indirectly by affecting preferences and attitudes toward effort or through access to a quality education. In practice, studies are limited by data constraints and usually use all available circumstances in the data set.

Among circumstances, the literature has stressed the role of parental background in determining a person's economic outcomes, and it is the most popular variable used to measure inequality of opportunity. Families influence a child's educational (and professional) success in many ways. Among the most obvious ones are inherited abilities, financial capital that allows investment in human capital, and parental education that supports the cognitive development of children (Becker et al., 2015; Björklund & Jäntti, 2011). Although information on parental background is available from 2006, it has only been since 2009 that it has been possible to differentiate between complete and incomplete levels of parental education. Therefore, the number of people with parents who had no formal education was much lower in 2006 than in 2009 and the following years, which is inconsistent with demographic changes in education. The CASEN survey asks about the level of education reached by each parent and the highest course the person achieved at this level. For example, if the primary level and the highest course is sixth grade, the person did not complete primary education, which in Chile has eight levels. From these variables, a new one was generated indicating the highest level obtained by either of the parents, as in Checchi & Peragine (2010), to not drop an observation in the case of only one parent reporting their level of education. The new variable measures four levels of parental education completed, i.e., parents with no formal education, primary, secondary, and post-secondary (higher) education. Besides parental education, we also use family composition as a circumstance variable, indicating if the person grew up with both parents until age 15. The literature has shown that children who grow up without a biological parent do worse, on average, than other children (Lang & Zagorsky, 2001; McLanahan & Sandefur, 2009). Other circumstances used in this study are gender, Indigenous background (self-reported), and region of birth grouped into four categories: North, Centre, South, and Metropolitan Region. We do not include age because income differences tend to level out over the life cycle. Table 1 shows the distributions of circumstances in the target population.

We need to restrict the sample to individuals with available information on circumstances included in the analysis. This initially affects 2 percent or less of household heads with missing information on the region of birth and between 17 and 24 percent with missing parental education (Table A1 in the appendix).[21]

[21] Other circumstances, such as Indigenous background and family composition, were constructed as dummy variables with value 1 for people who declared in the survey living with both parents until age 15 and who declared having an Indigenous background, and zero otherwise.

Since about 80 percent of the sample lives in the same region of birth, we use the region of residence as a proxy when the former is missing. Observations with missing information on parental education are excluded from the analysis. The excluded observations are unlikely to be random. They tend to be poorer, with less attained education, and more women and people live outside the metropolitan area. There are some statistically significant differences in mean earnings in several male age groups, especially in the central region (Table A2). Their exclusion increases earnings inequality by around 0.9-1.3 Gini points (except in the last survey, where the difference is smaller). To reduce this bias, we reweight observations with non-missing parental education so that we recover the distribution of observable characteristics in the population, roughly eliminating the difference with the original sample in attained education or region of residence (Table A3), as well as in earnings inequality before 2022 (that goes from around 1.1 Gini point on average to only 0.2, Table A4).[22] This is equivalent to imputing parental education based on observable characteristics.

We will present the results using the most disaggregated definition of types. However, for the sake of robustness, the appendix will show the main results when circumstances are introduced in a sequence. It will show that most of the inequality of opportunity with its main features is already observed if the two core circumstances (parental education and sex) are considered, with a low risk of overfitting. We will also compute conditional inference trees and random forests in the appendix, leading to the same conclusion.[23]

[22] This is done using the inverse probability of being in the final sample, estimated with a logit regression, conditional on individual characteristics and their interactions.

[23] For example, using the classification of types with parental education and sex, we have already reached 94 percent of our estimate for (Gini) inequality of opportunity in 2022. Similarly, we reach 96 percent with the conditional inference tree and at least 94 percent with the random forest (if parental education and sex are used in all trees). The proportion is smaller (87 percent) with the random forest that sets the usual number of preselected variables to the square root of the number of input variables. We argue that this is driven by a certain number of trees in which parental education or sex are relegated (implying an artificially low inequality of opportunity).

Table 1. Distribution of circumstances

	2009	2011	2013	2015	2017	2022
	(Percentage %)					
Gender						
Male	76.6	70.6	70.0	67.7	63.8	57.9
Female	23.4	29.4	30.0	32.3	36.2	42.1
Parental education						
No education or primary incomplete	41.4	40.5	34.6	32.6	31.1	27.9
Primary complete	30.2	29.8	31.3	32.5	30.2	22.3
Secondary complete	21.8	22.9	25.3	24.6	26.3	31.4
Higher Education	6.7	6.8	8.8	10.3	12.5	18.5
Place of birth						
North	10.7	10.6	11.3	11.1	12.0	12.3
Centre	38.6	38.7	37.7	38.2	37.4	34.9
South	16.9	16.8	16.9	17.2	17.3	14.9
Metropolitan	33.8	33.9	34.2	33.5	33.3	37.8
Grew up with both parents	73.2	75.3	76.5	76.9	74.9	81.1
Indigenous background	6.8	7.7	8.6	8.5	9.3	10.8
Total sample (n. observations)	27,923	25,452	28,401	36,455	29,064	24,265

Note: The studied population is all 20–60-year-old household heads. All observations with information on parental education are reweighted to represent the total population of household heads.
Source: Author's estimations based on [Dataset] CASEN (2009, 2011, 2013, 2015, 2017, 2022).

4. RESULTS

The context

Chile is an interesting case study. Since its return to democracy in 1990, Chile has shown strong economic growth, with GDP per capita multiplied by 2.6 in constant 2017 USD (from $9,702 to $25,886 in 2022 (World Bank, 2024a). This has positively increased living standards and drastically reduced absolute poverty, with the median income being three times higher in 2020 and extreme poverty declining from 11 percent to less than 1 percent based on the international poverty line (World Bank, 2024b). However, the country has been much less successful in reducing income inequality in the long term (from 0.527 to 0.504 using the Gini index between 1990 and 2020 (UNU-WIDER, 2023), remaining alongside Costa Rica as the most unequal OECD country (OECD, 2024).[24] In this context, there is a reasonable concern about whether this growth is reaching all social groups or leaving 'winners' and 'losers' in the distribution of opportunities (Klein & Tokman, 2000).

[24] Income inequality declined substantially in Chile between 2000 and 2015 (7.3 Gini points), in line with other Latin American countries, but inequality seems to have bounced back right afterward (2.8 Gini points in 2020).

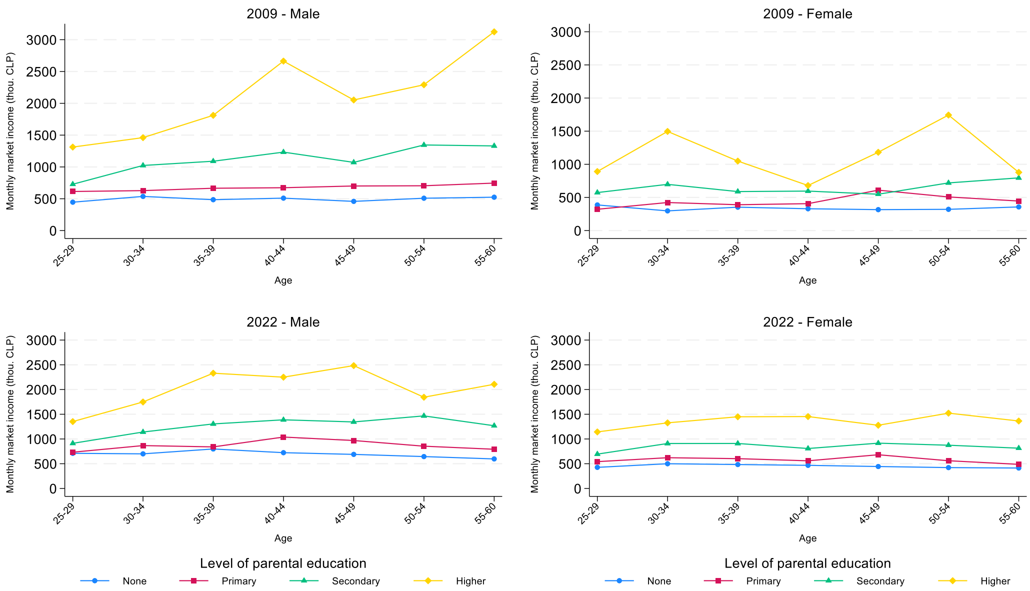
When trying to understand how social and economic factors have affected the distribution of opportunities in Chile, the literature has stressed the effect of the structural transformations the country has experienced over the opportunity structure (Torche & Wormald, 2004). Chile went from being a closed economy with an active, productive state in the early 1970s to one of the world's most open and least regulated economies in the 1980s (during the military dictatorship that was in power from 1973 to 1989). At this time, most of the public companies were privatized, and a significant reform was made to the educational system, allowing the private sector to provide education, increasing coverage but creating intense polarization within society, where the quality of education is directly correlated with the wealth of the family (Villalobos & Valenzuela, 2012).

In this scenario, the social structure suffered significant changes, particularly at the extremes of distribution, which was in line with the transformation of the country's productive structure. The expansion of educational opportunities benefited different social groups unevenly. The most to benefit was those with access to higher education, allowing them to remain in or enter the upper social class. In a significant proportion, this educational achievement was related to the educational heritage of their parents. This may be one of the reasons behind the observed inequality of schooling levels among different social classes. Particularly the significant difference between the level of the upper class and the rest of the social structure (Torche & Wormald, 2004). As a result, according to the World Inequality Lab's estimates for 2022, the richest one percent of Chile's population accumulated 24 percent of national income before taxes and owned 50 percent of total net personal wealth, among the highest concentration worldwide (WID, 2024).

Therefore, investigating the inequality of opportunity in Chile might help us to understand the mechanisms that produce these extreme inequalities in welfare, education, and other individual achievements (Fleurbaey & Peragine, 2013). Different sources of inequality might have different —even opposite— effects on growth and development (Ferreira et al., 2018; Marrero & Rodríguez, 2013) as they may affect different economic incentives (Bourguignon, 2018) and because they shape attitudes towards redistribution (Ferreira & Gignoux, 2011). People who believe their societies offer equal opportunities are more averse to redistribution (Alesina & La Ferrara, 2005).

The relevance of inequality of opportunity in Chile can be appreciated by the substantial average income differences by level of parental education and gender at different age intervals (Figure 1). The graph shows that higher averages of net market income are observed at higher parental education levels for all cohorts. The highest premiums are awarded to individuals from highly educated parents. The income gaps between types widen as we move toward higher parental educational achievements. Wider gaps are observed among males. Female wages are substantially lower than men's at all skill levels (Gaentzsch & Zapata-Román, 2020).

Figure 1. Average monthly earnings by level of parental education, age, and gender, 2009 and 2022



Note: Real earnings (in 2022 CL Pesos).

Source: Author's estimations based on [Dataset] CASEN(2009, 2011, 2013, 2015, 2017, 2022).

Inequality of opportunity estimates

a) Smoothing within-type inequality: IO^b

In the most common approach, inequality of opportunity is obtained as the remaining inequality between types, $I(y_b)$, after equalizing within-type distributions. In this decomposition, the rest of inequality, $I(y) - I(y_b) = I(y_w) + I_{bw}$, is the level of inequality gone after the equalization of incomes within types. However, due to the presence of the interaction term I_{bw}, its interpretation varies across indices, and it might be largely influenced by the between-type distribution as well. It can be substantially different from $I(y_w)$, which by construction is free of the between-type distribution. The $I(y) - I(y_b)$ term is the 'within-type' component of the additively decomposable entropy measures as defined by Shorrocks (1984), i.e., the sum of inequality in each type weighted by a function of their population and relative incomes (to the power of α). For the Gini index, it is the sum of the within-type inequality (weighted by the product of their population and income shares) plus a residual term in the classical decomposition (Bhattacharya & Mahalanobis, 1967; Pyatt, 1976).

The results of this decomposition for the Gini index and four members of the entropy family are reported in Table 2 for 2022. Figure 2a displays the corresponding share of overall inequality, IO^b, for every year. There is a substantial variability in the size of IO^b across indices. This variability can result from the different sensitivity of these indices to various parts of the distribution (captured by α among the entropy measures) or from other properties embedded in the indices, particularly the nature of their interaction term (i.e., their decomposability properties). In every year, there is an inverted-U shaped pattern among the entropy measures sorted by α, from more sensitive to the very bottom ($\alpha = -1$) to more sensitive to the very top ($\alpha = 2$), with the highest level in 2022, around 27 percent of inequality, attained with the Theil index ($\alpha = 1$) as well as with the most popular measure used in the early literature, and the only one that is path independent, the MLD ($\alpha = 0$).[25] The lowest level in 2022 is attained with $\alpha = -1$, 11 percent. Furthermore, the Gini index stands out for its much higher level of IO^b, 56 percent, twice the level indicated by the MLD, in line with other results in the literature that chose this index.[26]

b) Smoothing between-type inequality: IO^w

It is worth noting that, except for the MLD, the $I(y_b)$ term and IO^b do not respond to the key policy question of how much inequality would be reduced if the government managed to eliminate all existing income differences across types, thus eliminating inequality of opportunity. The answer to this question can be obtained by computing $I(y) - I(y_w) = I(y_b) + I_{bw}$ in an alternative decomposition, where $I(y_w)$ is the inequality in the within-type distribution remaining after removing any between-type difference. A distribution that does not depend on average incomes by type (unlike the 'within-term' in additively decomposable measures as defined by Shorrocks (1984). In the case of all entropy measures, $I(y_w)$ is just the population-weighted sum of inequality within types. In the case of the Gini index, it is approximately similar, as previously discussed.[27]

[25] This share for Theil is larger than that of Contreras et al. (2014), 16 percent for 2004-09.

[26] Most of the remaining inequality, $I(y) - I(y_b)$, in the case of the Gini index is explained by its residual term (42 percent of overall inequality), with only a very small proportion explained by the weighted within-type term (2 percent) in this decomposition (Table 2). This is the result of its peculiar weighting scheme, the product of income and population shares, which add up to a small amount (e.g., 0.027 in 2022, very far from the standard 1 using MLD or Theil, the two entropy measures where the sum of weights in this decomposition is fixed and, contrary to $\alpha = -1$ or $\alpha = 2$, does not depend on the distribution). Our estimates are close to those reported by Brunori, Ferreira, & Neidhöfer (2023) for Chile 2006-15, 50-56 percent, using 27-32 types with conditional inference trees.

[27] For example, the population-weighted Gini was 0.392 in 2022, while the Gini index of the within-type distribution y_w was 0.396. These two values would be identical if all the $O(y_{w_i}, y_w)$ were equal to 1, but in this case its population-weighted average was 1.02.

The decomposition following this alternative path is done in the second part of Table 2 for 2022. Figure 2b displays the corresponding shares of inequality explained by circumstances in every year, IO^w. The main issue here is that inequality of opportunity IO^b is 27 percent of overall inequality with MLD and, because of its path independence, removing inequality of opportunity would also eliminate 27 percent of the observed inequality, IO^w. However, this does not happen with the path-dependent indices. The inequality removed by eliminating inequality of opportunity is more extensive with the other entropy measures: with Theil ($\alpha = 1$), IO^w is 30 percent (while IO^b was 27 percent), with GE_{-1} is 15 percent (as opposed to 11 percent), and with GE_2 is 38 percent (as opposed to 17 percent). The latter index is very sensitive to the presence of top incomes, and the gap was even larger in 2017 (62 as opposed to 17 percent). In the case of entropy measures, it looks like now there is an increasing share of inequality that is gone when smoothing between-type differences for higher α in 2022, while there was a U-shaped relationship in previous years since the percentage with $\alpha = -1$ was larger than with $\alpha = 0$. The same type of paradox but in the opposite direction is observed with the Gini index, with inequality of opportunity representing as much as 56 percent. However, its removal only eliminated 15 percent of observed inequality. If the Gini index reduces inequality less than other measures when smoothing within-type differences (44 percent), it appears to reduce inequality even less when smoothing between-type differences (15 percent).[28]

c) The interaction term: I_{bw}

The entire gap between both estimates of the relevance of inequality of opportunity, IO^w and IO^b, results from the existence of an interaction term I_{bw} in the decomposition of path-dependent measures that is entirely attributed to the distribution being smoothed (i.e., the within-type distribution with IO^b, and the between-type distribution with IO^w), that in our application can be either positive (entropy measures) or negative (Gini), and can be significant (GE_2 and Gini) (see Figure 2c).

The significant discrepancy found between both measures of inequality of opportunity using the Gini index, as compared to the Theil index, is primarily the effect of overlapping. Indeed, there is substantial stratification among types; the average level of overlapping, weighted by population, revolves around 0.57-0.61 in Chile in the analyzed period, while 1 would mean perfect overlapping. The minimum with the observed population shares by type would be 0.025, achieved with the

[28] This effect of equalizing income across circumstances in Chile is smaller than the one reported by Núñez & Tartakowsky (2011), around 22-25 percent in 2006.

between-type distribution (see Figure 2e).[29] Types at both ends of the earnings distribution and those with the highest inequality tend to be the most stratified (i.e., less overlapping with the country's distribution; see Figure A1). This stratification has a strong mitigating effect on overall inequality.

If we remove the effect of overlapping on the Gini index by setting overlapping indices equal to 1, the resulting interaction term would be positive and similar in magnitude to Theil's: 0.010, about 2 percent of overall inequality.[30] This positive interaction term in the Gini index, after removing the effect of overlapping, reflects a certain tendency of inequality to be more prominent in richer types, like in the case of Theil (see Figure A2). In this case, the IO^b term for the Gini index would be 39 instead of 56 percent. This points out that 17 percentage points of the excess of the IO^b of Gini, compared to Theil, are driven by the effect of overlapping, and the other 13 percentage points by its differential metrics (likely, the lower sensitivity to the extremes).

d) The Shapley inequality of opportunity: IO^s

Both IO^b and IO^w misestimate the contribution of inequality of opportunity to overall inequality when using path-dependent measures. They respectively ignore the impact of the between- and within-type distributions through the interaction term. There are two possible solutions to this problem.

One solution, adopted in the earlier literature, is to stick to the use of only the MLD, guaranteeing that $IO^b = IO^w$. However, this is very restrictive as it imposes a specific sensitivity to different parts of the distribution and, as the previous analysis suggests, the share of inequality of opportunity may vary with different sensitivities. The other solution is to split the interaction term between the two sources of inequality (between and within types). This is what the Shapley approach does by averaging between both paths to obtain the importance of each source, or in other words, by evenly splitting the interaction term between both sources of inequality that take

---

[29] The minimum average overlapping is achieved when types are perfectly stratified and each type only overlaps with itself, $O(y_i, y) = \frac{N_i}{N}$. The aggregate measure of overlapping is then equal to the Herfindahl-Hirschman index of population concentration across types, $\sum_{i=1}^{n} \left(\frac{N_i}{N}\right)^2$. This is the situation found in the between-type distribution, for example.

[30] An overall Gini index constructed with the same additive structure as Theil, i.e., after removing the effect of overlapping, as $G^*(y) = G(y_b) + \sum_{i=1}^{n} \frac{N_i}{N} G(y_i) + \sum_{i=1}^{n} \frac{N_i}{N} \left[\frac{\mu_i}{\mu} - 1\right] G(y_i)$. The between-type term would represent 39 percent of inequality in 2022, the within-type term another 59 percent, and the interaction about 2 percent. $G^*(y)$ is 0.664, while the observed $G(y)$ is 0.467, which highlights the large mitigating effect of stratification in this index (0.197 or 30 percent of G^*).

part in it. This is done in the last part of Table 2 for 2022 (the share for all years is displayed in Figure 2d).

The result is that the contribution of inequality of opportunity to overall inequality, IO^s, ranges between 17 and 28 percent using the four entropy measures in 2022, with its importance following an inverted-U shaped pattern with the sensitivity to higher incomes. In most years, however, the choice of this sensitivity does not make much of a difference (only 1 or 2 percentage points, i.e., around 27-28 percent), unlike we want to give the very bottom a large protagonism, e.g., using $\alpha = -1$, in which case, it is substantially smaller (17 percent). There are exceptions or higher sensitivity to the top, like in 2017, when switching from $\alpha = 0$ to 2 raises the inequality of opportunity share by 7 percentage points. In general, it seems that in Chile, giving more relevance to what occurs at both ends of the distribution tends to reduce the importance of inequality of opportunity in favor of within-type inequality unless there is a very rich group, like in 2017, in which case a higher sensitivity to the top of the distribution ($\alpha = 2$) raises the relevance of inequality of opportunity.[31]

With the Shapley approach, the Gini index shows a percentage of inequality of opportunity in 2022, 36 percent, still above the upper end of the entropy range. However, it is no longer the big outlier that was with IO^b. The higher IO^s using the Gini index, is likely driven by its metrics, mainly because it is less sensitive to both extremes. The other distinguishing fact of this index, the large mitigating effect of stratification on inequality, is now divided into the effects of between-type and within-type distributions and is therefore not likely to be playing a substantial role in explaining the higher IO^s of the Gini index compared to Entropy measures. The IO^s of the Gini index (36 percent) is not very different from the IO^b estimated earlier after removing the effect of stratification on this index, 39 percent.

Inequality of opportunity based on the limited circumstances analyzed here plays an important role. However, still IO^s falls below 50 percent with all measures in 2022, including the Gini index. This is consistent with the Lorenz dominance of the between-type over the within-type distribution, as displayed in Figure 3a (graph on the right), pointing at much more inequality in the within-type distribution.[32] This fact is at odds with the between-type component representing

[31] Non-Indigenous men born in the northern region who grew up with both parents, at least one of them with higher education.

[32] Note that Figure 3a represents the Lorenz curves of y_b and y_w. The double of the area defined by the latter and the diagonal, as earlier discussed, differs from the weighted sum of within-group inequality from Pyatt's (1976) decomposition represented by Lambert & Aronson (1993), is independent of the between-type distribution and is approximately equal to the population-weighted average instead.

more than half of overall Gini inequality according to IO^b. This explains that the Gini index in the distribution within types (85 percent of overall inequality) is larger than the between-type distribution (56 percent), noting that this is only possible because the sum of both is 141, not 100 percent, the result of the large negative interaction. If instead of using the Shapley approach, the interaction is alternatively attributed based on the relevance of each source, this would be 40 percent (56 divided by 141), not far from the Shapley estimate of 36 percent.

Figure 3b (graph on the left) shows the corresponding Shapley Lorenz curves, $L_{sb}(p)$ and $L_{sw}(p)$, from which IO^s can be retrieved for the Gini index, as twice the area between the diagonal and L_{sb} or, alternatively, twice the area between L_{sw} and L.

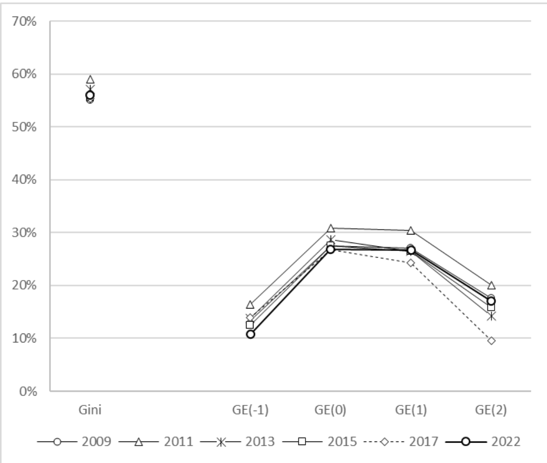
Table 2. Decomposition of overall inequality by type in 2022

			$Gini$	GE_{-1}	GE_0 (MLD)	GE_1 ($Theil$)	GE_2 ($1/2\ CV^2$)
	Overall	$I(y)$	0.467	1.063	0.401	0.409	0.707
Smoothing within-type inequality	Between	$I(y_b)$	0.262	0.115	0.108	0.109	0.121
		IO^b%	56.0	10.8	26.9	26.7	17.1
	Within	$I(y) - I(y_b) = I(y_w) + I_{bw}$	0.205	0.947	0.293	0.300	0.586
		IR^b%	44.0	89.2	73.1	73.3	82.9
	of which	'within' weighted sum	0.011				
		%	2%				
		residual	0.195				
		%	42%				
Smoothing between-type inequality	Between	$I(y) - I(y_w) = I(y_b) + I_{bw}$	0.071	0.252	0.108	0.122	0.267
		IO^w%	15.2	23.7	26.9	29.9	37.7
	Within	$I(y_w)$	0.396	0.810	0.293	0.287	0.440
		IR^w%	84.8	76.3	73.1	70.1	62.3
Interaction		$I_{bw} = I(y) - I(y_b) - I(y_w)$	-0.191	0.137	0.000	0.013	0.146
		$I_{bw}/I(y)$%	-40.9	12.9	0.0	3.2	20.6
Shapley (average)	Between	I_{sb}	0.166	0.184	0.108	0.116	0.194
		IO^s%	35.6	17.3	26.9	28.3	27.4
	Within	I_{sw}	0.301	0.879	0.293	0.293	0.513
		IR^s%	64.4	82.7	73.1	71.7	72.6

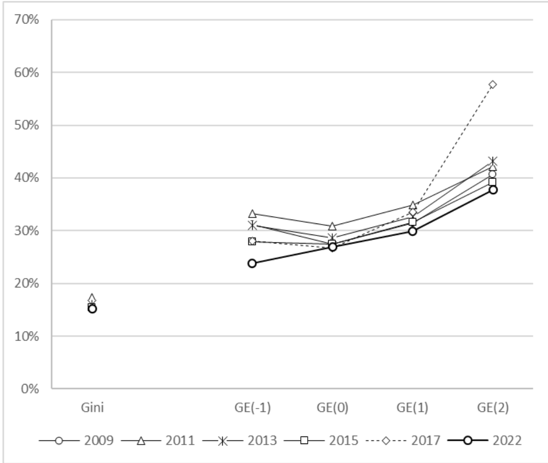
Source: Author's estimations based on ([Dataset] CASEN, 2009, 2011, 2013, 2015, 2017, 2022).

Figure 2. The contribution of inequality of opportunity as a percentage of overall inequality

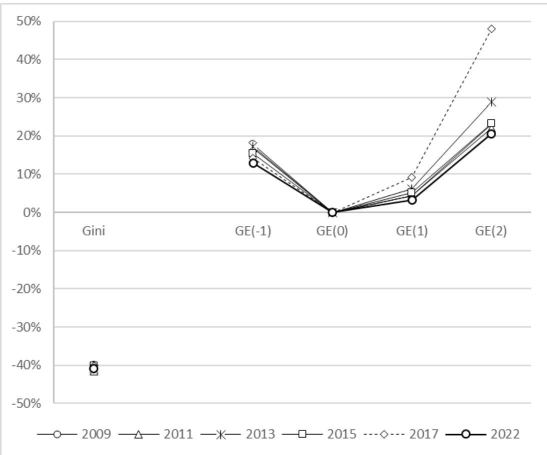
a) Smoothing 'within', IO^b



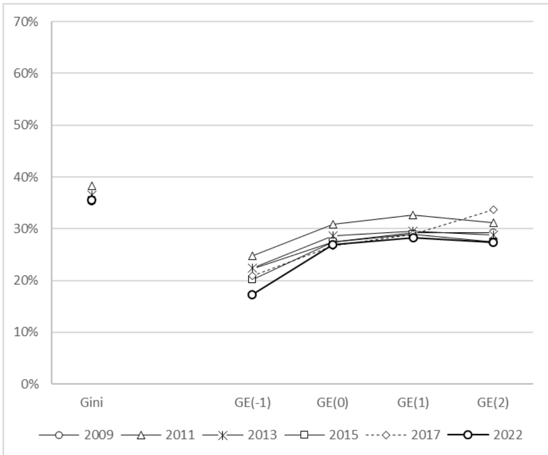
b) Smoothing 'between', IO^w



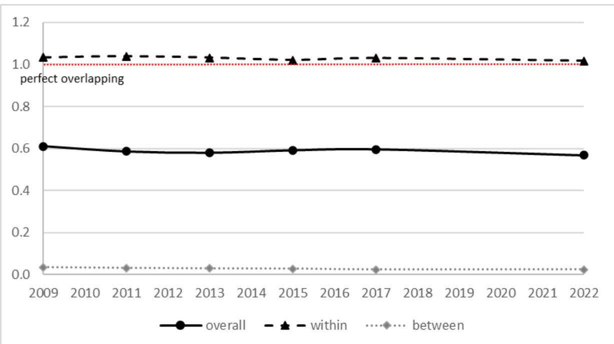
c) Interaction term, I_{bw}



d) Shapley, IO^s



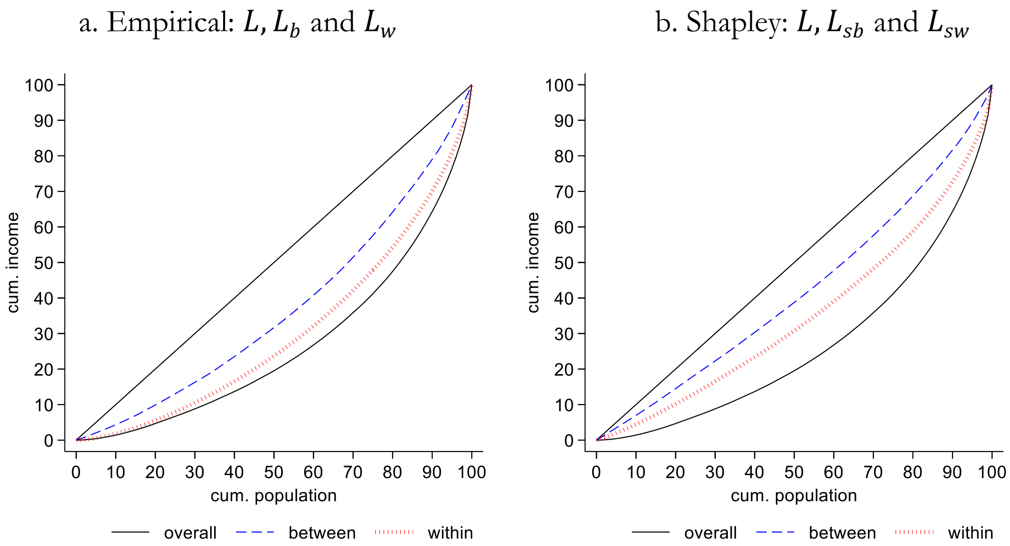
e) Overlapping



Notes: a) Smoothing within-type inequality (interaction assigned to inequality within types); b) Smoothing between-type inequality (interaction assigned to inequality between types); d) Shapley, interaction is split into between- and within-type inequality; e) Aggregate measure of overlapping is the population-weighted sum of overlapping of all types, the overlapping of each type is the weighted sum of overlapping with every type (Gradín, 2000). The between-type distribution shows the minimum level of overlapping (equal to the Herfindahl-Hirschman concentration index), 1 indicates perfect overlapping, and values above 1 indicate transvariations.

Source: Author's estimations based on [Dataset] CASEN (2009, 2011, 2013, 2015, 2017, 2022).

Figure 3. Lorenz curve for overall (y), between-type (y_b), and within-type (y_w) distributions in 2022



Note: a) see footnote 32.
Source: Author's estimations based on [Dataset] CASEN (2009, 2011, 2013, 2015, 2017, 2022).

Changes in inequality over time

One advantage of the Shapley decomposition is that it facilitates estimating changes over time for all measures. Thus, the sum of changes in overall inequality can be attributed to either inequality of opportunity or within-type inequality, with both components adding up to the total change.

The trend in overall earnings inequality in Chile over the 2009-22 period depends to some extent on the measure used, mainly whether it is extremely sensitive to either the bottom or the top of the distribution (Figure 4).

With some caveats, the Gini index and the MLD tell us a similar story about the recent trends in the level and the nature of inequality in Chile. Both point to two distinguished phases. Inequality first slightly increased by 1 and 3 percent between 2009 and 2013. This happened during the 2010 financial crisis, which in Chile was aggravated by the considerable damage caused by a magnitude 8.8 earthquake on February 27, 2010, followed by a tsunami. Inequality declined afterward as the country recovered (by 7 and 11 percent during 2013-22). The role of inequality of opportunity was vital for these dynamics during and right after the financial crisis, with a more active role in the case of the Gini index.

Indeed, the increase in the Gini index during the financial crisis and two-thirds of the rise in the MLD were driven by higher inequality of opportunity. At the same time, inequality within types declined with the Gini index and explained the other third of the increase with MLD. Later, lower inequality of opportunity explained half of the total decline in earnings inequality, followed by the

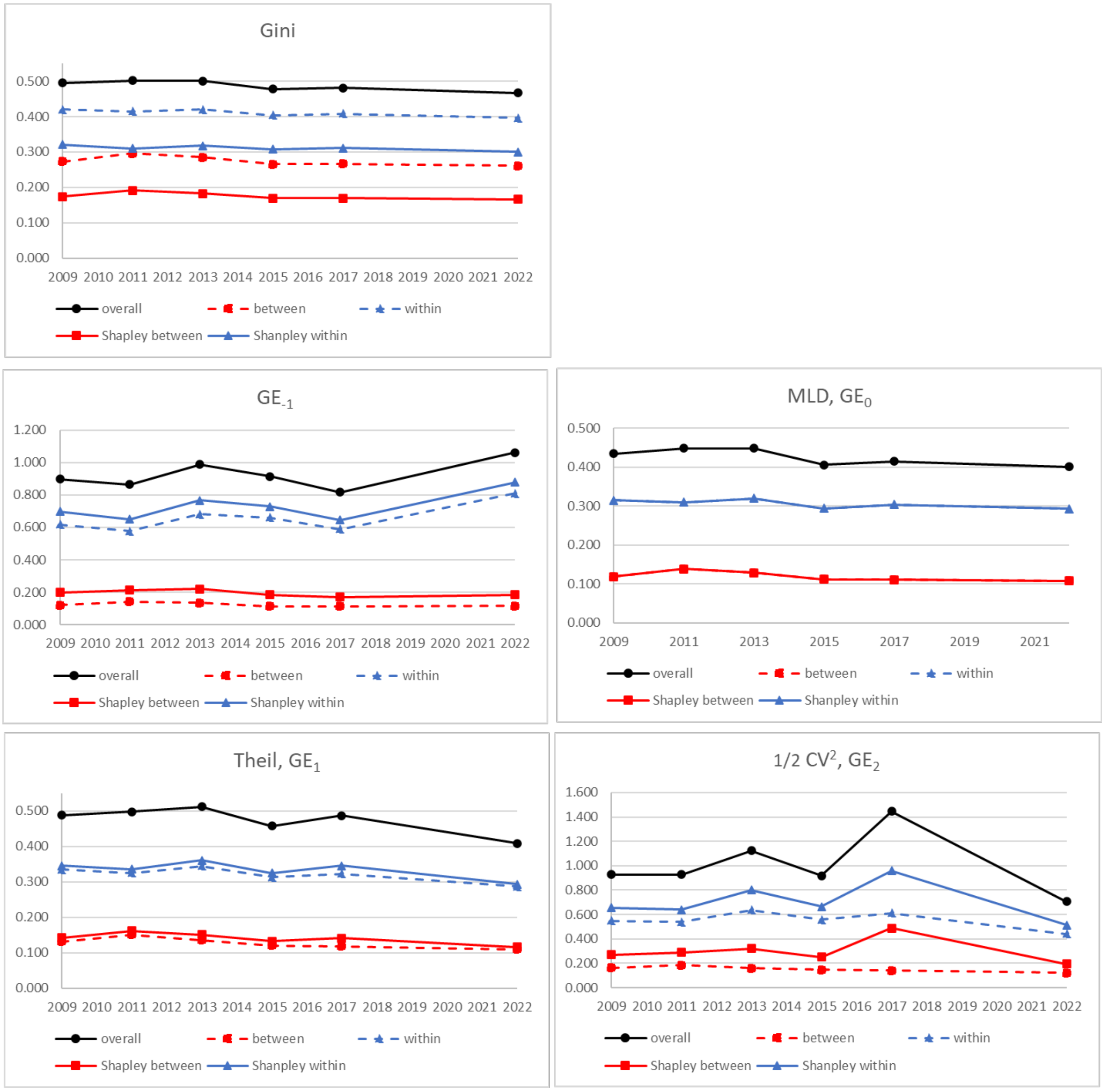
Gini index and 44 percent with the MLD. However, this active role of inequality of opportunity was concentrated in 2013-15, when it explained 50 percent of the decline with Gini and 40 percent with MLD. The slight rebound of inequality between 2015 and 2017 and the three-quarters decline between 2017 and 2022 were driven by inequality within types.

However, with $\alpha = -1$ and $\alpha = 2$ overall inequality is primarily driven by the income share of the bottom and top respectively (which are displayed in Figure A3). Therefore, they show opposite trends in 2015-17, when the bottom and top 1 percent income shares improved, and in 2017-22, when both income shares largely declined. Inequality of opportunity played a substantial role in the case of $\alpha = 2$ (44 percent in the first period, 40 percent in the second one).

As a result, the relevance of inequality of opportunity increased unanimously across all measures during 2009-11 and declined afterward (except for 2015-17 with $\alpha = -1$ and $\alpha = 2$, as reflected in Figure 5).

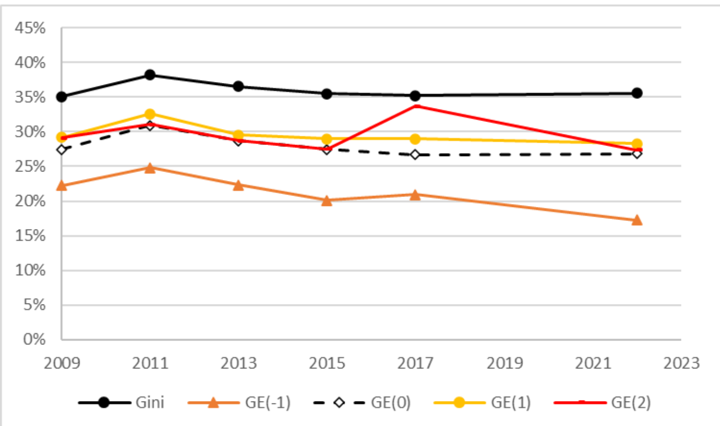
In summary, inequality of opportunity rather than being immutable largely drove the overall trend in earnings inequality in Chile during and right after the financial crisis, even if it shows a high level of persistence. However, assessing its evolution over time is subject to the same level of ambiguity that overall inequality exhibits when the distributions between or within types involve a combination of equalizing and inequality-enhancing distributional changes based on the principle of transfers. In these cases, the evolution will inevitably depend on how much weight we want to put on what happens in different parts of the distribution. In this context, imposing one single index to measure inequality of opportunity with a specific sensitivity does not seem very reasonable.

Figure 4. The trend in inequality in individual market income in Chile, 2009-22 using various indices: overall, between- and within-type distributions, as well as Shapley values



Source: Author's estimations based on [Dataset] CASEN (2009, 2011, 2013, 2015, 2017, 2022).

Figure 5. The trend in inequality of opportunity in Chile, 2009-22 using various indices: (Shapley) share of overall inequality, IO^s



Source: Author's estimations based on [Dataset] CASEN (2009, 2011, 2013, 2015, 2017, 2022).

Contributions to changes in inequality of opportunity by type

The per capita relative contribution to inequality of opportunity $I(y_b)$ of a type, those sharing the same set of circumstances, primarily depends on the income distance to the mean evaluated by the corresponding inequality measure. Figure 6a displays the relative contributions in 2022, with types sorted by their average earnings, unraveling a clear U-shaped relationship of this direct effect on inequality of opportunity. This shape is the distinctive feature of any inequality measure shared by all. Although the intensity of the relative contribution of rich and poorest groups varies across indices in the expected direction, the Gini index being the least sensitive to both ends among the selected measures, and GE_{-1} and GE_2 being the most sensitive to the bottom and top, respectively. These discrepancies among indices perfectly highlight the critical value judgments involved in choosing a measure of inequality of opportunity. However, as previously discussed, this is not all the influence of types on inequality of opportunity since this is modulated by the interaction with the within-type distribution (and with overlapping in the Gini index), making that two types with the same mean may affect differently overall inequality based on how unequal they are (or how stratified). The Shapley contribution is displayed in Figure 6b, which shows that a general U-shape pattern remains. Similarly, the per capita contribution of a type to inequality within types primarily depends on its level of inequality (this is only approximately so in the case of the Gini index). The Shapley contribution also must add the effect operating through the interaction effect.

How much inequality of opportunity and inequality within types is explained by a type is the result of multiplying the per capita contribution by its population size. Figure 7 presents the RIF contribution of every type to inequality between and within types in 2022 using the Shapley decomposition so that the sum of all contributions is overall inequality as measured by the Gini index. Types are sorted by their contribution to inequality of opportunity, and it is straightforward to observe how a few types have a disproportional effect on inequality of opportunity while other types contribute more to inequality within types. This disproportionality in the contribution by a few groups is even larger with the entropy measures, as could be expected (Figure A4 in the appendix).

Table 3 summarizes the contribution of the seven types that are part of the top 5 of any measure. The type with the largest contribution to inequality of opportunity with all measures (second in the case of GE_2) is made up of non-Indigenous men born in the metropolitan area who grew up with both their parents, and at least one of them completed higher education (Id=80). This is one of the most affluent types (2.4 times the mean income), and their contribution ranges between around 13 percent with Gini and GE2 and 20 percent with GE_{-1}, while they represent less than 4 percent of the population. The second largest contribution with the Gini, Theil, and MLD indices comes from one of the poorest types (less than half the average earnings) made up of non-Indigenous women born in the country's center who grew up with uneducated parents (Id=82). Their contribution ranges around 6-7 percent to these measures, while they represent 3 percent of the population.

In the case of indices that are more sensitive to either end of the distribution, small types with 1 percent or less of the population are found among the most significant contributors. In the case of GE_2, the most affluent type (2.7 times the average), made up of non-Indigenous men who grew up with highly educated parents in the northern region (Id=77), contributes the most to inequality of opportunity (14 percent), even if they only represent 0.7 percent of the household heads, and contribute with about 3-6 percent to the other measures.[33] Similarly, the type that is similar to Id=82 above but who did not grow up with both parents (Id=114), which is relatively poorer than the others (38 percent of the mean), has the second most significant contribution to GE_{-1} (9.1 percent, compared to 2-3 percent with the other indices) with only 1 percent of the population.

Types may also differentially affect changes in inequality of opportunity over time. Figure 8 decomposes the total change in overall inequality into the contribution of inequality of opportunity and inequality within types (pure distributive contributions estimated under constant population shares by type) and their corresponding composition effects (changes in population shares by type). The figure summarizes the results of all types sharing each circumstance (e.g., all male types, all female types, etc.).

Inequality increased in Chile between 2009 and 2011 due to the increase in inequality of opportunity, partially mitigated by the decline in inequality within types, with minimal net effect due to changes in composition. The rise in inequality of opportunity was driven by all parental educational levels, both genders, mainly in the majoritarian groups (non-Indigenous heads who grew up with both parents in the center and metropolitan regions). The increase in inequality within types was driven primarily by non-Indigenous male types, especially those with no parental education, born in the center of the country (with both parents).

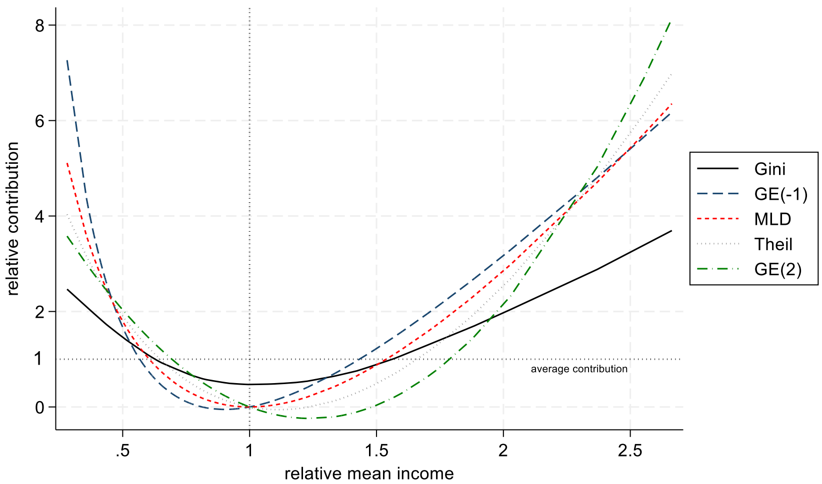
Inequality declined after 2011, almost entirely driven by a significant decline in inequality of opportunity (with constant population), partially mitigated by its associated composition effect. The improvement in the distribution of opportunities was mainly driven by a smaller advantage of non-Indigenous men born in the metropolitan area with higher parental education, growing up

[33] The contribution of this group to all measures in 2017 was even larger, 28 percent with GE_2 when the average earnings were 3.7 times the average (13 percent with Gini, 22 percent with Theil).

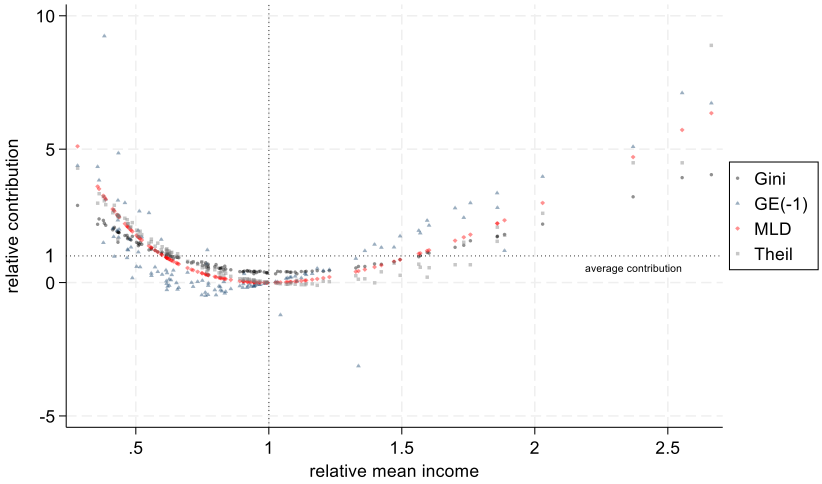
with both parents. The compositional effect that mitigated the reduction in inequality was related to the large increase of heads in the labor market with higher parental education (from 7 to 19 percent, Table 1), growing up with both parents (from 75 to 81 percent), born in the metropolitan region (34 to 38 percent), and the decline of those with a non-Indigenous background (from 92 to 89 percent).

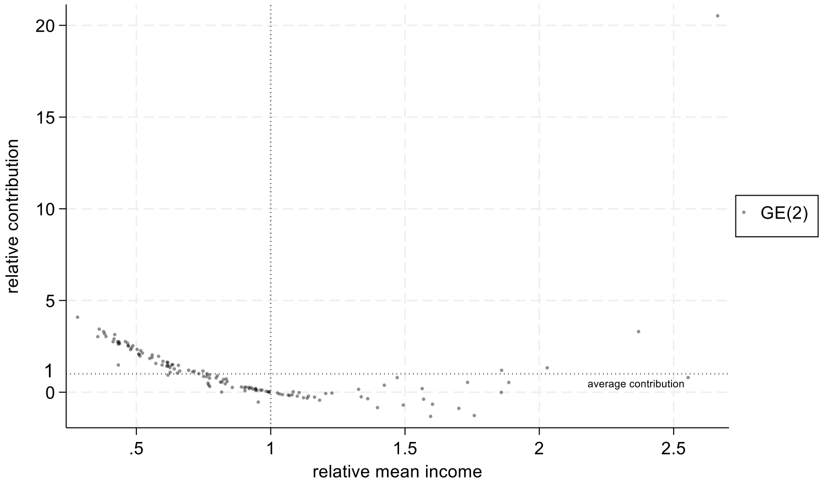
Figure 6. Mean income and relative per capita contribution to inequality of opportunity (between types) by type in 2022

a. Direct contribution to $I(y_b)$



b. Total contribution to inequality through inequality of opportunity, Shapley: $I(y_b) + \frac{I_{bw}}{2}$

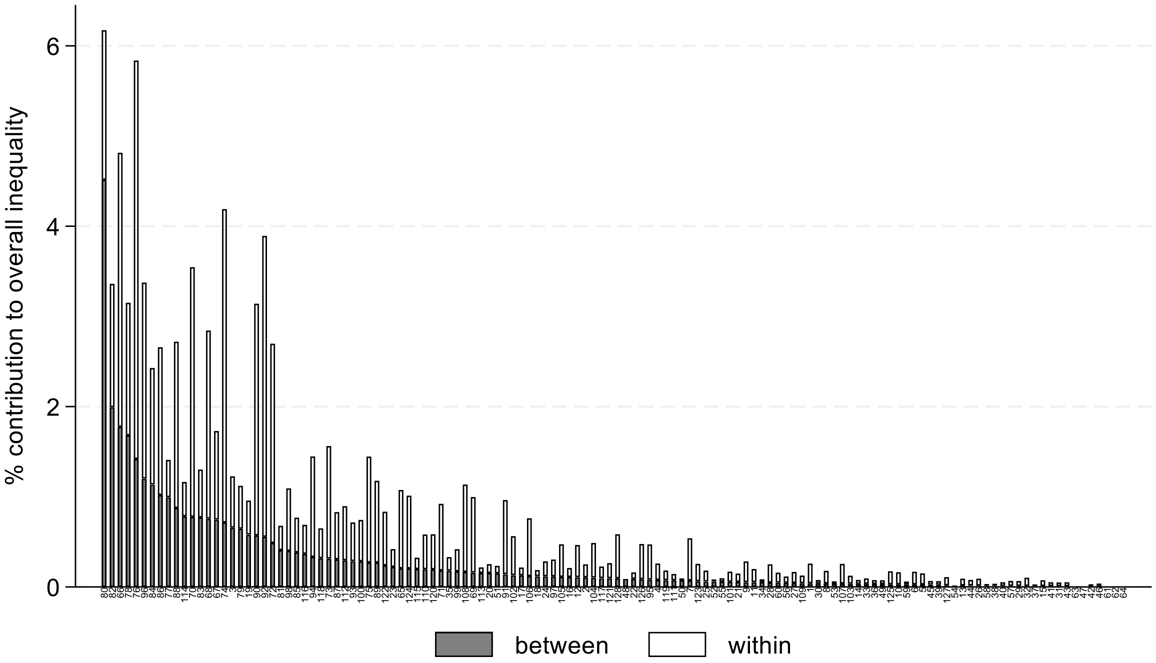




Note: The relative per capita contribution mapped here is the per capita contribution of individuals of each type, divided by the population's average contribution (i.e., inequality of opportunity). A value greater than 1 indicates a per capita contribution above the average. The total contribution of a type to inequality of opportunity is the product of its per capita contribution and population share.

Source: Author's estimations based on [Dataset] CASEN (2009, 2011, 2013, 2015, 2017, 2022).

Figure 7. (Shapley) Contribution of all types to inequality (Gini) between and within types (as a percentage of overall inequality), 2022



Note: Types are sorted in descending order by their contribution to inequality of opportunity ('between'). These are Shapley's contributions. The sum of all contributions adds up to overall inequality. Tables 3 (7 types) and A5 (all types, with relative income, population share, and inequality) report the circumstances associated with each type.

Source: Author's estimations based on [Dataset] CASEN (2009, 2011, 2013, 2015, 2017, 2022).

Table 3. Types that belong to the top 5 contributors to any of the five indices, 2022

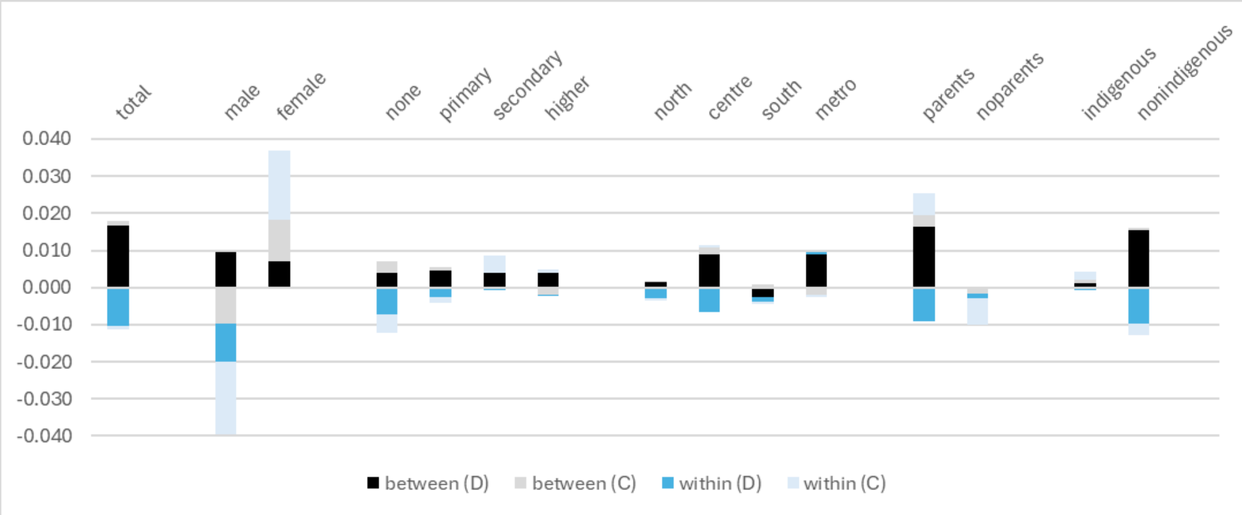
		Type				Population	Mean income		Shapley contribution to IO (%)				
Id	S	E	R	P	I	%	Pesos	% Mean	$Gini$	GE_{-1}	GE_0	GE_1	GE_2
80	M	H	Metro	yes	no	3.9	2,355,658	237	**12.7**	**20.1**	**18.6**	**17.7**	13.0
82	F	N	Center	yes	no	3.0	434,912	44	5.6	6.0	7.3	7.4	7.9
66	M	N	Center	yes	no	5.2	649,008	65	5.0	-0.1	3.9	5.2	5.5
78	M	H	Center	yes	no	2.6	1,875,211	189	4.7	3.1	6.1	4.6	1.4
76	M	S	Metro	yes	no	5.8	1,414,813	142	4.0	7.7	3.8	2.3	2.2
96	F	H	Metro	yes	no	3.3	1,560,141	157	3.4	6.2	3.7	1.9	-1.3
84	F	N	Metro	yes	no	2.2	510,420	51	3.2	6.0	3.7	3.9	4.4
77	M	H	North	yes	no	0.7	2,648,230	266	2.8	4.7	4.4	6.1	**14.2**
114	F	N	Center	no	no	1.0	379,601	38	2.2	9.1	3.2	3.0	3.2
			Sum			27.9			43.5	62.8	54.7	52.1	50.5

Note: In bold, the largest contribution to each index. S=Sex (Male, Female), E= Parental education (N=none or less than primary; S= Secondary, H= Higher); R= Region of birth, P=Grew with both parents, I=Indigenous.

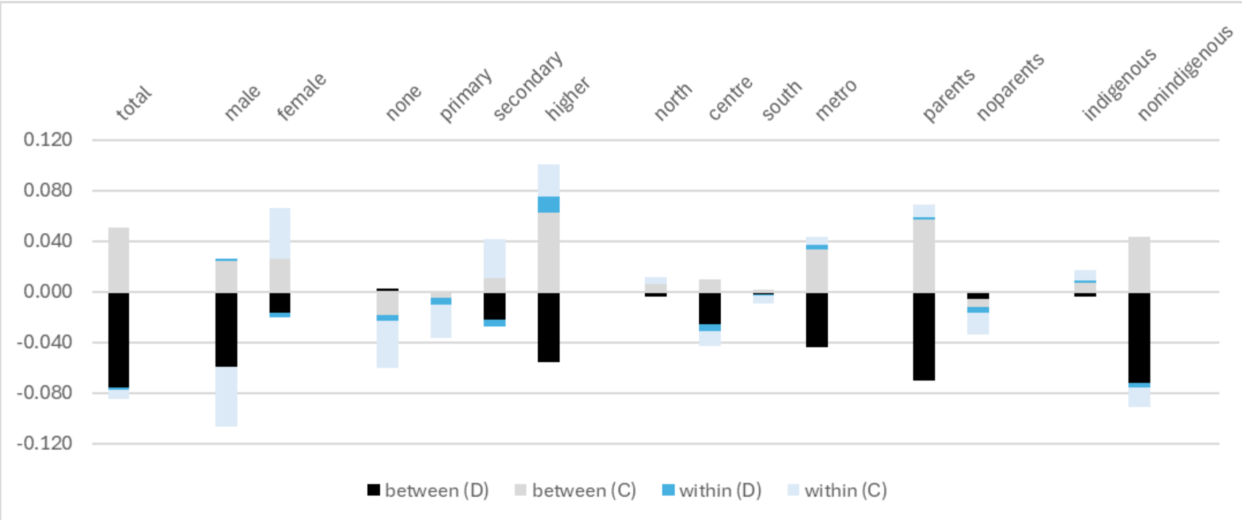
Source: Author's estimations based on [Dataset] CASEN (2009, 2011, 2013, 2015, 2017, 2022).

Figure 8. Decomposition of the contribution of types (aggregated by circumstance) to the change in overall inequality

a. 2009-11



b. 2011-22



Notes: Each category shows the contribution from all types sharing the same circumstance (e.g., the male contribution is the combined contribution of all male types). The components add to the overall inequality change for each characteristic (i.e., sex, parental education, region of birth, growing up with parents, and ethnicity). D=distributive effect (due to changes in the distribution between or within types with constant population); C=composition effect (due to changes in the population shares of types with constant distribution).

5. Source: Author's estimations based on [Dataset] CASEN (2009, 2011, 2013, 2015, 2017, 2022).

6. Concluding remarks

This paper has measured inequality of opportunity in Chile as inequality in market income among household heads due to the family of origin and other birth circumstances. We have argued that interpreting the between to overall inequality IO^b ratio using path-dependent measures (all except the MLD) can be problematic because this ratio does not factor in that the effect of circumstances also depends on the interaction with the within-type distributions. The resulting misestimation can be substantial. To overcome this limitation, we adopt the Shapley approach that allows us to expand the set of inequality measures while making more meaningful and consistent comparisons, in which half of the interaction terms are included as part of the effect of circumstances in our proposed IO^s ratio.

After constructing 128 types of people sharing similar circumstances based on limited available information on parental background in CASEN surveys, we estimate that these circumstances account for about 17-36 percent of market income inequality among household heads. This range reflects that people concerned with inequality of opportunity may have different views about inequality, and therefore, the relevance of circumstances depends on the measure used.

Entropy measures have shown little variability in the relevance of inequality of opportunity in Chile based on the sensitivity to distributional changes affecting different parts of the distribution, ranging between 27-28 percent in 2022 with three of the four measures. The main exception occurs in the extreme case where inequality becomes very sensitive to the very bottom of the distribution ($\alpha = -1$), where within-type inequality seems more relevant. Circumstances explain about 11 percent of inequality. Although this is a measure rarely used in empirical analysis of inequality and is subject to the influence of measurement error that takes place at the bottom, understanding the nature of inequality that gives more importance to the very poorest people may be of interest, in line with the *leave no one behind* approach adopted by the United Nations 2030 Agenda for Sustainable Development and the Sustainable Development Goals. In some cases, like in 2017, the presence of a very affluent small group, with only 0.6 percent of the population, explains 40 percent of inequality of opportunity, raising the relevance of circumstances to 34 percent with a high sensitivity to the very top ($\alpha = 2$, a monotonic transformation of the coefficient of variation, a variance-based popular measure). This is particularly relevant given the growing sensitivity to the concentration of income at the top of the distribution that has manifested both in academic and political arenas, combined with the known fact of the underestimation of the actual level of such concentration in household surveys.

The Gini index deviates from entropy measures not only for being generally less sensitive to both ends of the distribution (both between and within types) but also for being sensitive to the extent to which the types overlap along the income space (i.e., stratification by types has a mitigating effect). With this index, we obtain the largest contribution of circumstances to overall inequality, 36 percent in 2022, a figure above the upper bound obtained using the four entropy measures in the same year (28 percent). Remarkably, the Gini index is a particular case in which IO^b substantially overestimates the contribution of circumstances; at the same time, with entropy measures other than MLD and especially with $\alpha = 2$ and $\alpha = -1$, it is underestimated. The main problem was that while IO^b is the largest with Gini, 56 percent, its within-type equivalent, IE^w, is even larger, 85 percent. Both cannot be true simultaneously because they add up to 141 percent of overall inequality, the excess resulting from the abovementioned interaction. We get this paradox by assigning this -41 percent interaction, primarily the result of the mitigating effect of stratification across types of inequality, to the within and between distributions in each case, respectively. The Shapley approach splits this interaction effect into the two sources of inequality. This implies admitting the complex nature of circumstances' effect on path-dependent inequality measures. The impact of circumstances depends on the within-type distribution. Similarly, removing between-type inequality, the ultimate policy goal in this context, has an impact on the effect of the within-type inequality on overall inequality as well, explaining why inequality between types represents 56 percent of inequality in Chile in 2022, according to the Gini index, but the government can only reduce inequality 15 percent by smoothing average incomes across types. To go further, the government must reduce inequality within types. Our analysis has shown that these estimates are highly robust to the risk of upward bias due to small samples because they are driven by the core circumstances (parental education and sex), while expanding how circumstances are fully accounted for may still increase their importance substantially.

Our results also show that the inequality trend in Chile is not exempt from important nuances based on the sensitivity to both ends of the distribution. However, in general, the recent trend in inequality in Chile, especially the increase during the 2010 financial crisis aggravated by the earthquake and the decline that followed, has been largely driven by changes in inequality of opportunity as measured here. The relevance of circumstances to explain overall inequality has generally fallen but shows a high level of persistence. This indicates that the analysis of types is strongly relevant to understanding the dynamic of inequality more broadly, even if types are defined with limited information, particularly in a highly unequal country like Chile, where there are reasons to believe that observed inequalities are largely inherited. Although the channels of this intergenerational transmission of advantages are yet to be investigated, we know that coming from a wealthier family gives access to differentiated educational and employment opportunities later reflected in people's ability to generate income. This feature becomes extraordinarily relevant since reducing inequality of opportunity can generate more political consensus than other forms of inequality to adopt the necessary structural reforms to remove them, especially in a country that has also shown a solid political divide in the last decades.

7. References

Aaberge, R., Mogstad, M., & Peragine, V. (2011). Measuring long-term inequality of opportunity. *Journal of Public Economics*, *95*(3–4), 193–204.

Alesina, A., & La Ferrara, E. (2005). Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, *89*(5), 897–931.

Arneson, R. J. (1989). Equality and Equal Opportunity for Welfare. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *56*(1), 77–93.

Atkinson, A. (1970). On the measurement of inequality. *Journal of Economic Theory*, *2*(3), 244–263.

Atkinson, A. (2005). Comparing the Distribution of Top Incomes across Countries. *Journal of the European Economic Association*, *3*(2/3), 393–401.

Atkinson, A. (2015). *Inequality: What can be done?* Harvard University Press.

Becker, G. S., Kominers, S., Murphey, K., & Spenkuch, J. (2015). A Theory of Intergenerational Mobility. *Munich Personal RePEc Archive Paper No. 66334*.

Becker, G. S., & Tomes, N. (1986). Human Capital and the Rise and Fall of Families. *Journal of Labor Economics*, *4*(3), 1–39.

Bhattacharya, N., & Mahalanobis, B. (1967). Regional Disparities in Household Consumption in India. *Journal of the American Statistical Association*, *62*(317), 143–161.

Björklund, A., & Jäntti, M. (2011). Intergenerational Income Mobility and the Role of Family Background. In B. Nolan, W. Salverda, & T. M. Smeeding (Eds.), *The Oxford Handbook of Economic Inequality*. Oxford University Press.

Bourguignon, F. (2018). Measuring the inequality of opportunities. In *For Good Measure Advancing Research on Well-being Metrics Beyond GDP: Advancing Research on Well-being Metrics Beyond GDP*. OECD Publishing.

Bourguignon, F., Ferreira, F. H. G., & Menéndez, M. (2007). Inequality of Opportunity in Brazil. *Review of Income and Wealth*, *53*(4), 585–618.

Bourguignon, F., & Morrisson, C. (2002). Inequality among World Citizens: 1820-1992. *The American Economic Review*, *92*(4), 727–744.

Brunori, P. (2016). How to Measure Inequality of Opportunity: A Hands-On Guide. *Life Course Centre Working Papers Series*.

Brunori, P., Ferreira, F. H. G., & Neidhöfer, G. (2023). Inequality of Opportunity and Intergenerational Persistence in Latin America. *IDB Publications*.

Brunori, P., Ferreira, F. H. G., & Salas-Rojo, P. (2023). Inherited inequality: A general framework and an application to South Africa. *Working Papers ECINEQ, Society for the Study of Economic Inequality*, *WP 658*, Article 658.

Brunori, P., Palmisano, F., & Peragine, V. (2019). Inequality of opportunity in sub-Saharan Africa. *Applied Economics*, *51*(60), 6428–6458.

Brunori, P., Peragine, V., & Serlenga, L. (2019). Upward and downward bias when measuring inequality of opportunity. *Social Choice and Welfare*, *52*(4), 635–661.

Cabrera, L., Marrero, G. A., Rodríguez, J. G., & Salas-Rojo, P. (2021). Inequality of Opportunity in Spain: New Insights from New Data. *Hacienda Publica Espanola*, *237*, 153–185.

Chakravarty, S. R. (2009). *Inequality, Polarization and Poverty: Advances in Distributional Analysis* (1. Aufl., Vol. 6). Springer-Verlag.

Chantreuil, F., & Trannoy, A. (2013). Inequality decomposition values: The trade-off between marginality and efficiency. *Journal of Economic Inequality*, *11*(1), 83–98.

Checchi, D., & Peragine, V. (2010). Inequality of opportunity in Italy. *The Journal of Economic Inequality*, *8*(4), 429–450.

Checchi, D., Peragine, V., & Serlenga, L. (2010). Fair and Unfair Income Inequalities in Europe. *IZA Discussion Paper*, *No. 5025*.

Cohen, G. A. (1989). On the Currency of Egalitarian Justice. *Ethics*, *99*(4), 906–944.

Contreras, D., Larrañaga, O., Puentes, E., & Rau, T. (2014). Improving the Measurement of the Relationship between Opportunities and Income: Evidence from Longitudinal Data from Chile. *Development Policy Review*, *32*(2), 219–237.

Dagum, C. (1960). *Teoria de la transvariación: Sus aplicaciones a la economia*. Istituto di Statistica, Univ. degli Studi di Roma.

Dagum, C. (1980). Inequality Measures between Income Distributions with Applications. *Econometrica*, *48*(7), 1791–1803.

[Dataset] CASEN. (2009). *Base de datos CASEN 2009*. Ministerio de Desarrollo Social y Familia. https://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen-2009

[Dataset] CASEN. (2011). *Base de datos CASEN 2011*. Ministerio de Desarrollo Social y Familia. https://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen-2011

[Dataset] CASEN. (2013). *Base de datos CASEN 2013*. Ministerio de Desarrollo Social y Familia. https://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen-2013

[Dataset] CASEN. (2015). *Base de datos CASEN 2015*. Ministerio de Desarrollo Social y Familia. https://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen-2015

[Dataset] CASEN. (2017). *Base de datos CASEN 2017*. Ministerio de Desarrollo Social y Familia. http://observator17.ministeriodesarrollosocial.gob.cl/encuesta-casen-2017

[Dataset] CASEN. (2022). *Base de datos CASEN 2022*. Ministerio de Desarrollo Social y Familia. http://observator17.ministeriodesarrollosocial.gob.cl/encuesta-casen-2020

Davies, J., & Shorrocks, A. (2021). Comparing Global Inequality of Income and Wealth. In C. Gradín, M. Leibbrandt, & F. Tarp (Eds.), *Inequality in the Developing World* (p. 0). Oxford University Press.

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

Ferreira, F. H. G., Lakner, C., Lugo, M. A., & Özler, B. (2018). Inequality of Opportunity and Economic Growth: How Much Can Cross‐Country Regressions Really Tell Us? *The Review of Income and Wealth*, *64*(4), 800–827.

Ferreira, F. H. G., & Peragine, V. (2015). *Equality of Opportunity: Theory and Evidence*. The World Bank.

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional Quantile Regressions. *Econometrica*, *77*(3),

Firpo, S., Fortin, N. M., & Lemieux, T. (2018). Decomposing Wage Distributions Using Recentered Influence Function Regressions. *Econometrics*, *6*(2), 1–40.

Fleurbaey, M., & Peragine, V. (2013). Ex Ante Versus Ex Post Equality of Opportunity. *Economica*, *80*(317), 118–130.

Foster, J. E., & Shneyerov, A. A. (2000). Path Independent Inequality Measures. *Journal of Economic Theory*, *91*(2), 199–222.

Gaentzsch, A., & Zapata-Román, G. (2020). Climbing the Ladder: Determinants of Access to and Returns from Higher Education in Chile and Peru. *UNRISD Working Papers*, *2020*(2). http://www.unrisd.org/unrisd/website/document.nsf/(httpPublications)/904B2706D3EEBF7D8025854A00556683?OpenDocument

Gini, C. (1916). Il Concetto di Transvariazione e le sue Prime Aplicazioni. *Studi Di Economia, Finanza e Statistica, Editi Del Giornali Degli Economisti e Rivista Statistica*, *Reprinted in Gini (1955)*.

Gini, C. (1955). *Memorie di metodologia statistica*. Libreria Goliardica.

Gradín, C. (2000). Polarization by sub-populations in Spain: 1973-91. *The Review of Income and Wealth*, *46*(4), 457–474.

Gradín, C. (2020). Quantifying the contribution of a subpopulation to inequality an application to Mozambique. *Journal of Economic Inequality*, *18*(3), 391–419.

Gradín, C. (2024). Revisiting the trends in global inequality. *World Development*, *179*, 106607-.

Gradín, C., Lewandowski, P., Schotte, S., & Sen, K. (Eds.). (2023). *Tasks, Skills, and Institutions: The Changing Nature of Work and Inequality*. Oxford University Press.

Hampel, F. R. (1974). The Influence Curve and Its Role in Robust Estimation. *Journal of the American Statistical Association*, *69*(345), 383–393.

Klein, E., & Tokman, V. E. (2000). *Social stratification under tension in a globalized era*.

Lambert, P. J., & Aronson, J. R. (1993). Inequality Decomposition Analysis and the Gini Coefficient Revisited. *The Economic Journal (London)*, *103*(420), 1221–1227.

Lang, K., & Zagorsky, J. L. (2001). Does Growing up with a Parent Absent Really Hurt? *The Journal of Human Resources*, *36*(2), 253–273.

Lerman, R. I., & Yitzhaki, S. (1984). A note on the calculation and interpretation of the Gini index. *Economics Letters*, *15*(3), 363–368.

Marrero, G. A., & Rodríguez, J. G. (2013). Inequality of opportunity and growth. *Journal of Development Economics*, *104*, 107–122.

McLanahan, S., & Sandefur, G. D. (2009). *Growing Up with a Single Parent: What Hurts, What Helps*. Harvard University Press.

Moramarco, D. (2023). Fairness and Gini decomposition. *Economics Letters*, *233*, 111409-.

Núñez, J., & Tartakowsky, A. (2011). The relationship between income inequality and inequality of opportunities in a high-inequality country: The case of Chile. *Applied Economics Letters*, *18*(4), 359–369.

OECD. (2018). *Income distribution and Poverty: Overview*. http://www.oecd.org/social/inequality.htm#income

OECD. (2024). *Income Inequality*. OECD Data. https://data.oecd.org/inequality/income-inequality.htm

Piraino, P. (2015). Intergenerational Earnings Mobility and Equality of Opportunity in South Africa. *World Development*, *67*, 396–405.

Pyatt, G. (1976). On the Interpretation and Disaggregation of Gini Coefficients. *The Economic Journal (London)*, *86*(342), 243–255.

Roemer, J. E. (1993). A Pragmatic Theory of Responsibility for the Egalitarian Planner. *Philosophy and Public Affairs*, *22*(2), 146–166.

Roemer, J. E. (1998). *Equality of opportunity*. Harvard University Press.

Runciman, W. G. (1966). *Relative deprivation and social justice: A study of attitudes to social inequality in twentieth-century England*. Routledge & Kegan Paul.

Serrano, R. (2007). Cooperative games: Core and shapley value. *Working Papers, Department of Economics, Brown University*, *Working Paper No. 2007-11*. https://www.econstor.eu/handle/10419/80201

Shorrocks, A. F. (1984). Inequality Decomposition by Population Subgroups. *Econometrica*, *52*(6), 1369–1385.

Shorrocks, A. F. (2013). Decomposition procedures for distributional analysis: A unified framework based on the Shapley value. *The Journal of Economic Inequality*, *11*(1), 99–126.

Singh, A. (2012). Inequality of opportunity in earnings and consumption expenditure: The case of Indian Men. *The Review of Income and Wealth*, *58*(1), 79–106.

Torche, F., & Wormald, G. (2004). *Estratificación y movilidad social en Chile: Entre la adscripción y el logro*. CEPAL. http://repositorio.cepal.org//handle/11362/6089

UNU-WIDER. (2023). *World Income Inequality Database (WIID)* (28 November 2023) [Companion dataset (wiidcountry)]. https://www.wider.unu.edu/project/wiid-%E2%80%93-world-income-inequality-database

Van de Gaer, D. (1993). *Equality of Opportunity and Investment in Human Capital* [Dr. Econ]. Katholieke Universiteit Leuven.

Villalobos, C., & Valenzuela, J. P. (2012). Polarización y Cohesion del Sistema Escolar Chileno. *Revista de análisis económico*, *27*(2), 145–172.

WID. (2024). *Data Chile*. WID - World Inequality Database. https://wid.world/country/chile/

World Bank. (2024a). *GDP per capita, PPP (constant 2017 international $)—Chile*. Data Bank Microdata Catalog. https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD?locations=CL

World Bank. (2024b). *Poverty headcount ratio at $2.15 a day (2017 PPP) (% of population)—Chile*. Data. https://data.worldbank.org/indicator/SI.POV.DDAY?locations=CL&view=chart

Yitzhaki, S. (1982). Relative deprivation and economic welfare. *European Economic Review*, *17*(1), 99–113.

Yitzhaki, S. (1994). Economic distance and overlapping of distributions: The econometrics of labor market segregation and discrimination. *Journal of Econometrics*, *61*(1), 147–159.

ANNEX

Table A1. Distribution of missing type by year and missing circumstance

	2009	2011	2013	2015	2017	2022
Parental education	24.1	21.6	21.2	18.2	17.3	21.3
Region of birth	2.0	1.6	1.8	1.0	1.8	1.2

Note. We imputed the region of birth by the region of residence since between 73 and 87 percent of the sample live in their region of birth. All observations with information on parental education are reweighted to be representative of the total population.

Source: Author's estimations based on [Dataset] CASEN (2009, 2011, 2013, 2015, 2017, 2022).

Table A2. T-statistic for the difference in mean earnings (samples with missing and non-missing parental education)

		Men				Women			
	Region	North	Center	South	Metro	North	Center	South	Metro
	Age								
	25-39	1.24	1.32	1.58	1.69	0.53	0.38	0.50	0.45
2009	40-49	0.96	1.41	1.08	1.78	0.34	0.84	0.65	0.30
	50-60	0.89	1.84	0.87	1.55	0.24	0.46	0.21	0.66
	25-39	1.01	2.91	1.11	0.93	0.21	0.69	0.12	0.27
2011	40-49	2.01	2.93	0.94	2.13	0.61	0.68	0.37	0.46
	50-60	1.80	2.46	1.16	1.97	0.46	0.71	0.23	1.08
	25-39	1.30	2.93	1.14	1.19	0.43	0.46	0.31	0.07
2013	40-49	1.46	2.29	1.47	1.66	1.00	0.44	0.58	0.77
	50-60	1.06	2.00	1.12	1.89	0.57	1.39	0.77	0.79
	25-39	1.33	2.72	1.57	1.56	0.12	0.80	0.44	0.52
2015	40-49	1.21	2.32	1.83	1.63	0.44	1.29	0.73	1.01
	50-60	1.25	3.16	0.93	2.05	0.74	1.04	0.65	0.98
	25-39	1.17	1.73	1.29	1.23	0.51	0.44	0.70	0.85
2017	40-49	1.00	1.57	0.81	1.72	0.69	0.89	0.55	0.67
	50-60	0.64	1.75	1.28	1.58	0.96	0.96	0.95	0.96
	25-39	0.98	1.66	1.30	1.66	1.24	1.35	0.93	1.52
2022	40-49	1.75	2.42	1.37	1.98	1.50	1.46	1.28	1.41
	50-60	0.85	2.28	1.62	2.04	0.90	0.96	0.83	1.59

Note: values correspond to t-test H_0: $u_0 = u_1$ (mean income equality). Darker areas are groups for which the null hypothesis of equal means is rejected, t-value $>|1.96|$.
Source: Author's estimations based on [Dataset] CASEN (2009, 2011, 2013, 2015, 2017, 2022).

Table A3. Distribution of attained education and metropolitan region in the original and final samples (with and without reweighting)

		2009	2011	2013	2015	2017	2022
Education	*Sample*	(Percentage, % household heads)					
None	Original	14.6	13.1	11.5	10.3	9.7	5.4
	Not reweighted	13.2	11.3	9.7	9.0	8.3	4.2
	Reweighted	14.6	13.1	11.5	10.3	9.7	5.4
Primary	Original	26.9	28.0	25.5	23.6	21.6	14.9
	Not reweighted	24.5	25.5	23.2	21.0	19.5	12.7
	Reweighted	26.9	27.9	25.5	23.6	21.6	14.9
Secondary	Original	38.7	37.9	38.7	40.8	40.6	39.6
	Not reweighted	39.6	38.7	39.2	41.2	40.8	38.6
	Reweighted	38.8	38.0	38.7	40.8	40.6	39.6
Higher	Original	19.8	21.0	23.9	25.2	27.5	39.7
	Not reweighted	22.8	24.4	27.6	28.7	30.8	44.2
	Reweighted	19.8	21.1	24.0	25.2	27.5	39.7
Region	*Sample*						
Metropolitan region	Original	40.7	41.9	41.6	40.5	40.7	41.3
	Not reweighted	42.1	43.1	43.3	41.3	42.6	42.9
	Reweighted	40.7	41.9	41.5	40.5	40.6	41.3

Note: The original sample used sample weights. The final sample used sample weights without and with reweighting (the inverse probability of being part of the final sample using a logit on observable characteristics and interactions).
Source: Author's estimations based on [Dataset] CASEN (2009, 2011, 2013, 2015, 2017, 2022).

Table A4. Earnings (Gini) inequality: original and sample with information on parental education (with and without reweighting)

	2009	2011	2013	2015	2017	2022
Original	0.493	0.498	0.498	0.476	0.480	0.465
With non-missing parental education						
Without reweighting	0.506	0.511	0.510	0.486	0.489	0.468
With reweighting	0.495	0.502	0.500	0.477	0.481	0.467

Source: Author's estimations based on [Dataset] CASEN (2009, 2011, 2013, 2015, 2017, 2022).

Table A5. Circumstances, population share, relative income, and group inequality in 2022 of each type

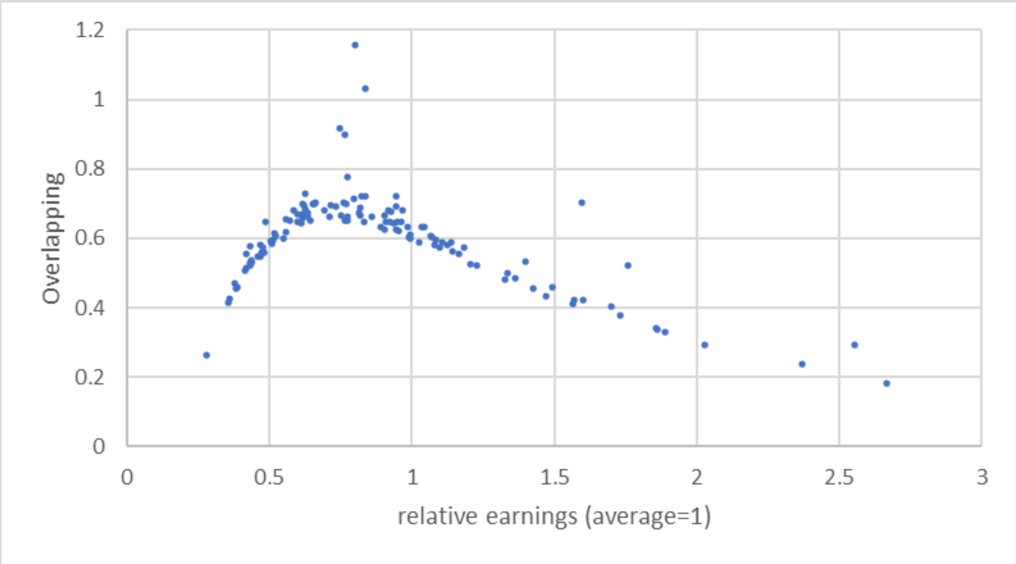
Id	Gender	Parental Education	Region of birth	Both parents	Indigenous	Population %	Relative Income (mean =1)	Inequality $Gini(y_i)$
1	Male	No education	North	yes	yes	0.28	0.89	0.431
2	Male	No education	Centre	yes	yes	0.28	0.63	0.316
3	Male	No education	South	yes	yes	1.23	0.52	0.326
4	Male	No education	Metropolitan	yes	yes	0.30	0.73	0.354
5	Male	Primary	North	yes	yes	0.18	0.95	0.378
6	Male	Primary	Centre	yes	yes	0.20	0.93	0.388
7	Male	Primary	South	yes	yes	0.53	0.95	0.459
8	Male	Primary	Metropolitan	yes	yes	0.25	0.97	0.309
9	Male	Secondary complete	North	yes	yes	0.36	0.96	0.354
10	Male	Secondary complete	Centre	yes	yes	0.20	1.07	0.369
11	Male	Secondary complete	South	yes	yes	0.26	0.82	0.310
12	Male	Secondary complete	Metropolitan	yes	yes	0.51	1.36	0.401
13	Male	Higher Education	North	yes	yes	0.11	1.03	0.368
14	Male	Higher Education	Centre	yes	yes	0.06	1.86	0.477
15	Male	Higher Education	South	yes	yes	0.08	0.99	0.425
16	Male	Higher Education	Metropolitan	yes	yes	0.18	1.86	0.393
17	Female	No education	North	yes	yes	0.18	0.41	0.371
18	Female	No education	Centre	yes	yes	0.17	0.42	0.318
19	Female	No education	South	yes	yes	0.86	0.43	0.338
20	Female	No education	Metropolitan	yes	yes	0.24	0.47	0.298
21	Female	Primary	North	yes	yes	0.15	0.62	0.385
22	Female	Primary	Centre	yes	yes	0.15	0.47	0.351
23	Female	Primary	South	yes	yes	0.39	0.48	0.371
24	Female	Primary	Metropolitan	yes	yes	0.28	0.57	0.384
25	Female	Secondary complete	North	yes	yes	0.21	0.72	0.339
26	Female	Secondary complete	Centre	yes	yes	0.10	1.08	0.406
27	Female	Secondary complete	South	yes	yes	0.21	0.80	0.330
28	Female	Secondary complete	Metropolitan	yes	yes	0.31	0.93	0.350
29	Female	Higher Education	North	yes	yes	0.08	1.14	0.368
30	Female	Higher Education	Centre	yes	yes	0.07	0.50	0.358
31	Female	Higher Education	South	yes	yes	0.06	1.04	0.378
32	Female	Higher Education	Metropolitan	yes	yes	0.09	1.03	0.493
33	Male	No education	North	no	yes	0.09	0.63	0.392
34	Male	No education	Centre	no	yes	0.08	0.49	0.273
35	Male	No education	South	no	yes	0.31	0.47	0.358
36	Male	No education	Metropolitan	no	yes	0.08	0.58	0.320
37	Male	Primary	North	no	yes	0.04	0.76	0.221
38	Male	Primary	Centre	no	yes	0.06	0.83	0.181
39	Male	Primary	South	no	yes	0.07	0.61	0.372
40	Male	Primary	Metropolitan	no	yes	0.07	0.84	0.329
41	Male	Secondary complete	North	no	yes	0.06	0.98	0.362
42	Male	Secondary complete	Centre	no	yes	0.03	0.94	0.461
43	Male	Secondary complete	South	no	yes	0.06	0.90	0.395
44	Male	Secondary complete	Metropolitan	no	yes	0.10	0.94	0.318
45	Male	Higher Education	North	no	yes	0.05	1.73	0.481
46	Male	Higher Education	Centre	no	yes	0.03	1.11	0.571
47	Male	Higher Education	South	no	yes	0.01	1.76	0.223
48	Male	Higher Education	Metropolitan	no	yes	0.07	2.55	0.223
49	Female	No education	North	no	yes	0.07	0.51	0.422
50	Female	No education	Centre	no	yes	0.07	0.28	0.294
51	Female	No education	South	no	yes	0.20	0.39	0.340
52	Female	No education	Metropolitan	no	yes	0.07	0.36	0.287
53	Female	Primary	North	no	yes	0.06	0.46	0.311
54	Female	Primary	Centre	no	yes	0.02	0.42	0.238
55	Female	Primary	South	no	yes	0.08	0.36	0.418
56	Female	Primary	Metropolitan	no	yes	0.08	0.43	0.554
57	Female	Secondary complete	North	no	yes	0.06	0.77	0.496
58	Female	Secondary complete	Centre	no	yes	0.03	0.62	0.337
59	Female	Secondary complete	South	no	yes	0.06	0.56	0.322
60	Female	Secondary complete	Metropolitan	no	yes	0.17	0.69	0.376
61	Female	Higher Education	North	no	yes	0.01	0.75	0.215
62	Female	Higher Education	Centre	no	yes	0.02	0.94	0.309
63	Female	Higher Education	South	no	yes	0.02	1.60	0.185
64	Female	Higher Education	Metropolitan	no	yes	0.01	0.80	0.157
65	Male	No education	North	yes	no	1.12	0.82	0.427
66	Male	No education	Centre	yes	no	5.23	0.65	0.351
67	Male	No education	South	yes	no	1.84	0.60	0.343
68	Male	No education	Metropolitan	yes	no	3.29	0.77	0.360
69	Male	Primary	North	yes	no	1.15	1.08	0.399
70	Male	Primary	Centre	yes	no	4.07	0.81	0.383
71	Male	Primary	South	yes	no	1.06	0.86	0.386

72	Male	Primary	Metropolitan	yes	no	3.20	0.92	0.379
73	Male	Secondary complete	North	yes	no	1.98	1.18	0.353
74	Male	Secondary complete	Centre	yes	no	4.49	1.23	0.426
75	Male	Secondary complete	South	yes	no	1.39	1.33	0.465
76	Male	Secondary complete	Metropolitan	yes	no	5.81	1.42	0.435
77	Male	Higher Education	North	yes	no	0.69	2.66	0.547
78	Male	Higher Education	Centre	yes	no	2.62	1.89	0.407
79	Male	Higher Education	South	yes	no	0.83	2.03	0.434
80	Male	Higher Education	Metropolitan	yes	no	3.94	2.37	0.414
81	Female	No education	North	yes	no	0.61	0.44	0.342
82	Female	No education	Centre	yes	no	2.99	0.44	0.348
83	Female	No education	South	yes	no	1.15	0.43	0.354
84	Female	No education	Metropolitan	yes	no	2.24	0.51	0.399
85	Female	Primary	North	yes	no	0.75	0.52	0.357
86	Female	Primary	Centre	yes	no	2.62	0.60	0.399
87	Female	Primary	South	yes	no	0.83	0.62	0.391
88	Female	Primary	Metropolitan	yes	no	2.72	0.64	0.417
89	Female	Secondary complete	North	yes	no	1.29	0.77	0.403
90	Female	Secondary complete	Centre	yes	no	3.35	0.83	0.427
91	Female	Secondary complete	South	yes	no	0.98	0.90	0.451
92	Female	Secondary complete	Metropolitan	yes	no	4.21	1.00	0.430
93	Female	Higher Education	North	yes	no	0.73	1.60	0.373
94	Female	Higher Education	Centre	yes	no	1.54	1.34	0.406
95	Female	Higher Education	South	yes	no	0.53	1.16	0.407
96	Female	Higher Education	Metropolitan	yes	no	3.34	1.57	0.399
97	Male	No education	North	no	no	0.35	0.66	0.326
98	Male	No education	Centre	no	no	1.21	0.66	0.346
99	Male	No education	South	no	no	0.45	0.62	0.341
100	Male	No education	Metropolitan	no	no	0.81	0.64	0.356
101	Male	Primary	North	no	no	0.23	0.77	0.272
102	Male	Primary	Centre	no	no	0.67	0.82	0.359
103	Male	Primary	South	no	no	0.15	0.76	0.344
104	Male	Primary	Metropolitan	no	no	0.65	0.92	0.328
105	Male	Secondary complete	North	no	no	0.40	1.47	0.500
106	Male	Secondary complete	Centre	no	no	0.87	1.10	0.405
107	Male	Secondary complete	South	no	no	0.28	0.99	0.420
108	Male	Secondary complete	Metropolitan	no	no	1.11	1.20	0.463
109	Male	Higher Education	North	no	no	0.17	1.40	0.304
110	Male	Higher Education	Centre	no	no	0.64	1.49	0.365
111	Male	Higher Education	South	no	no	0.16	1.70	0.318
112	Male	Higher Education	Metropolitan	no	no	0.86	1.56	0.424
113	Female	No education	North	no	no	0.19	0.38	0.277
114	Female	No education	Centre	no	no	0.99	0.38	0.315
115	Female	No education	South	no	no	0.29	0.43	0.325
116	Female	No education	Metropolitan	no	no	0.64	0.48	0.361
117	Female	Primary	North	no	no	0.22	0.56	0.384
118	Female	Primary	Centre	no	no	0.61	0.51	0.382
119	Female	Primary	South	no	no	0.17	0.55	0.421
120	Female	Primary	Metropolitan	no	no	0.56	0.61	0.431
121	Female	Secondary complete	North	no	no	0.28	0.63	0.374
122	Female	Secondary complete	Centre	no	no	0.91	0.71	0.390
123	Female	Secondary complete	South	no	no	0.28	0.75	0.398
124	Female	Secondary complete	Metropolitan	no	no	1.00	0.77	0.454
125	Female	Higher Education	North	no	no	0.20	1.12	0.394
126	Female	Higher Education	Centre	no	no	0.56	0.91	0.386
127	Female	Higher Education	South	no	no	0.13	1.07	0.380
128	Female	Higher Education	Metropolitan	no	no	0.64	1.14	0.414

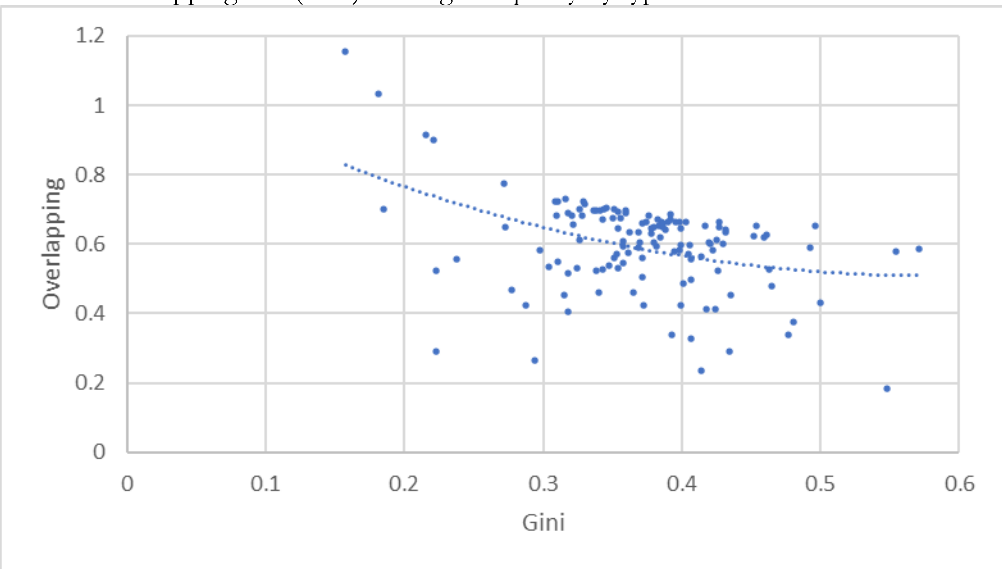
Source: Author’s estimations based on [Dataset] CASEN (2022).

Figure A1. Overlapping between types in Chile, 2022

a. Overlapping and average earnings by type

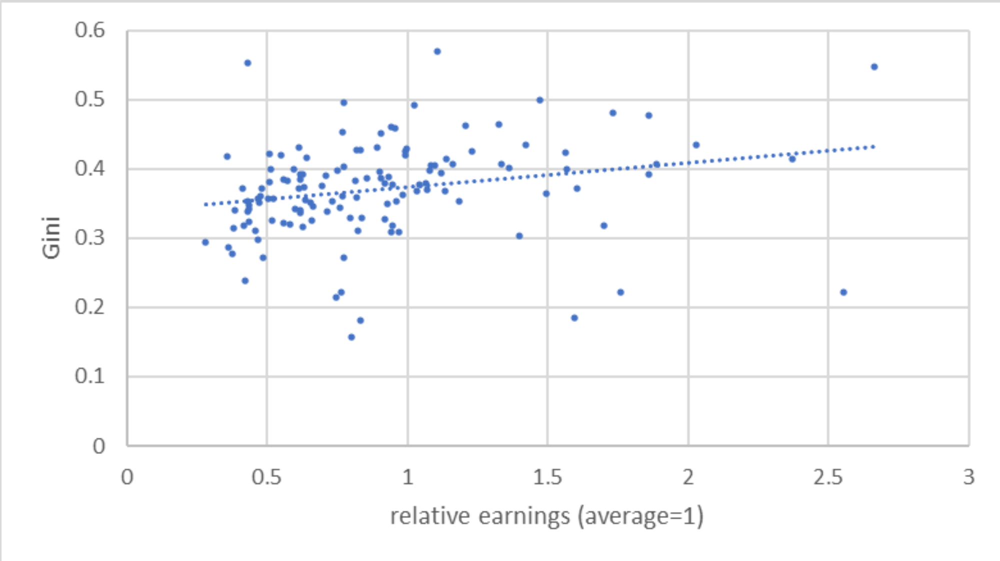


b. Overlapping and (Gini) earnings inequality by type



Notes: a) Overlapping of each type versus its level of inequality measured by the Gini index. b) Overlapping of each type versus its relative average earnings (average earnings divided by the country's mean).
Source: Author's estimations based on [Dataset] CASEN (2022).

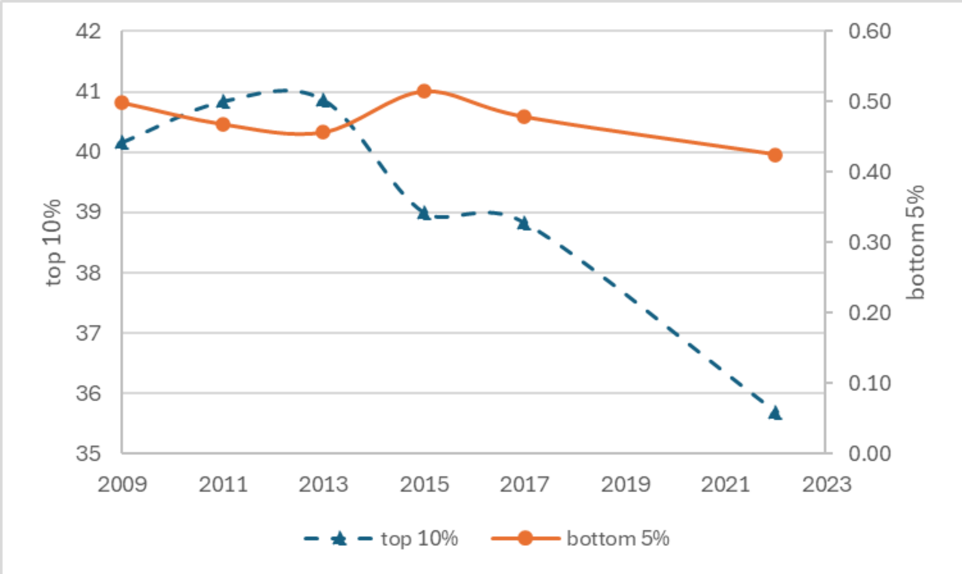
Figure A2. Inequality by type and relative average earnings in Chile 2022: Gini index



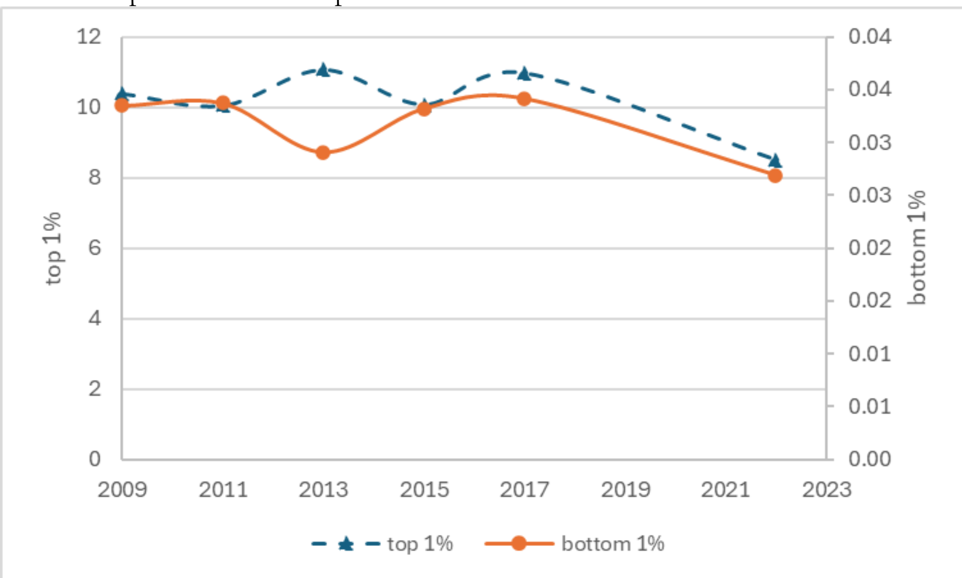
Note: The dashed line indicates the linear prediction.
Source: Author's estimations based on [Dataset] CASEN (2022).

Figure A3. The share of the bottom and top earnings groups in Chile, 2022

a. Top 10 and bottom 5 percent



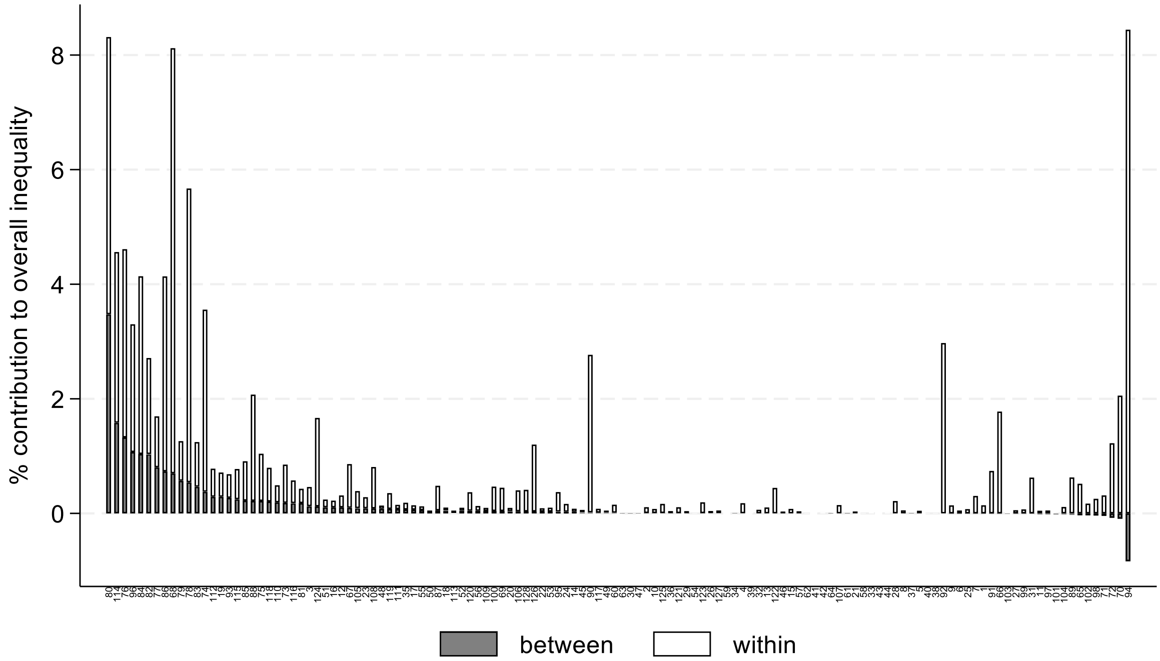
b. Top 1 and bottom 1 percent



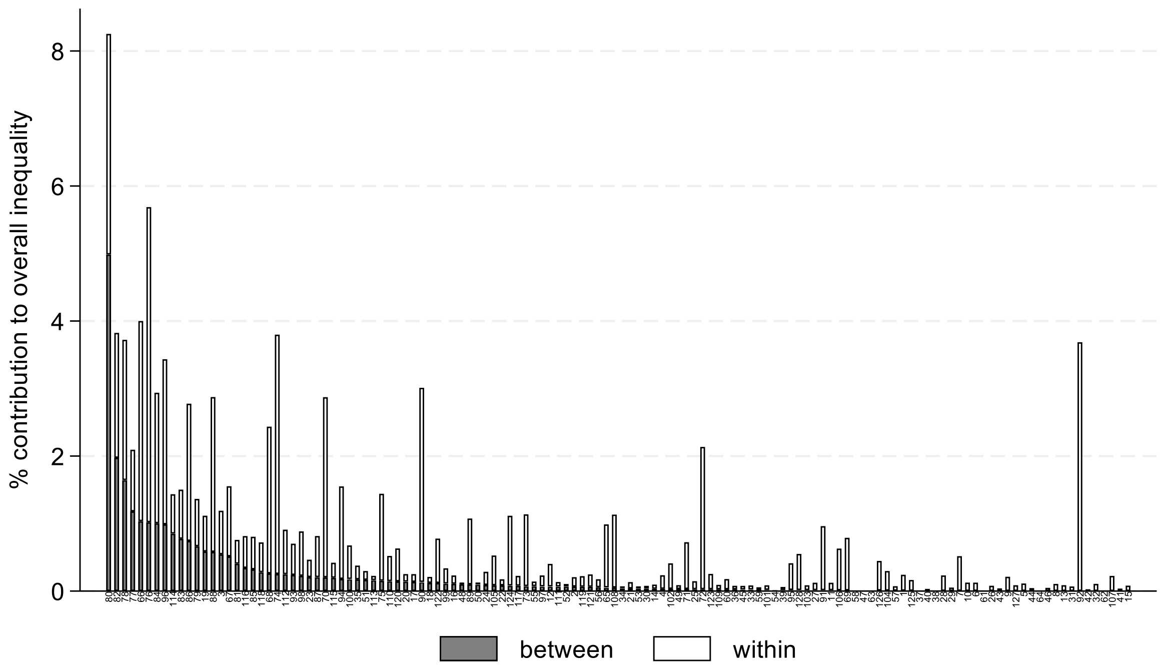
Source: Author's estimations based on [Dataset] CASEN (2009, 2011, 2013, 2015, 2017, 2022).

Figure A4. Mean market income and relative per capita contribution to inequality of opportunity (between types) by type in 2022: Entropy measures

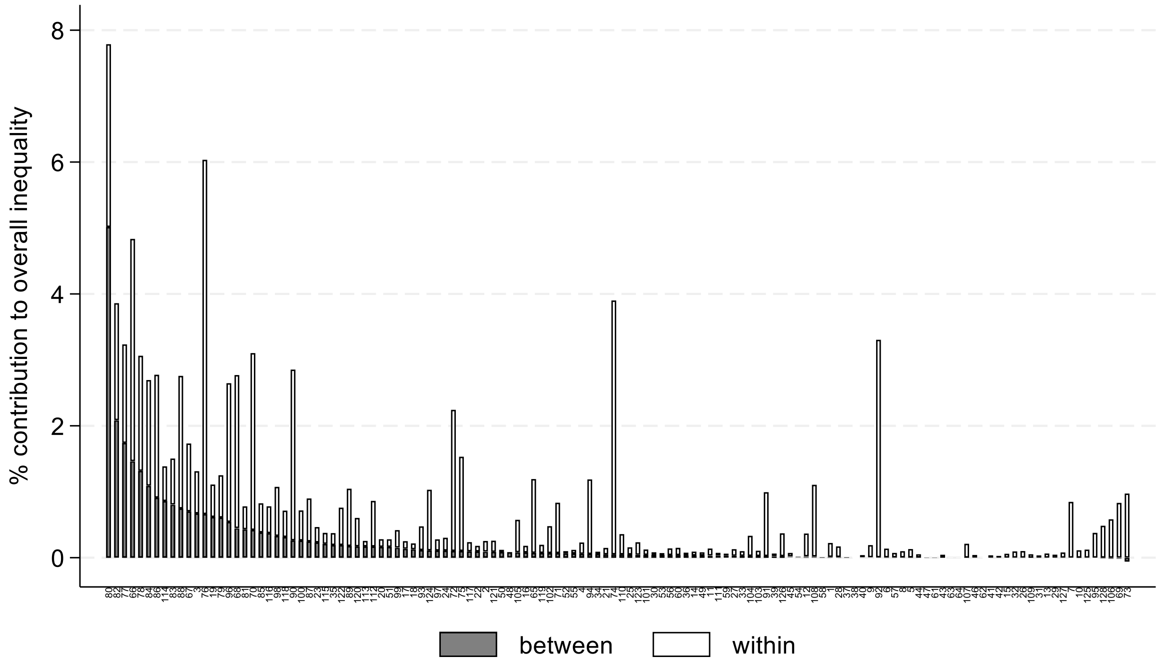
GE_{-1}



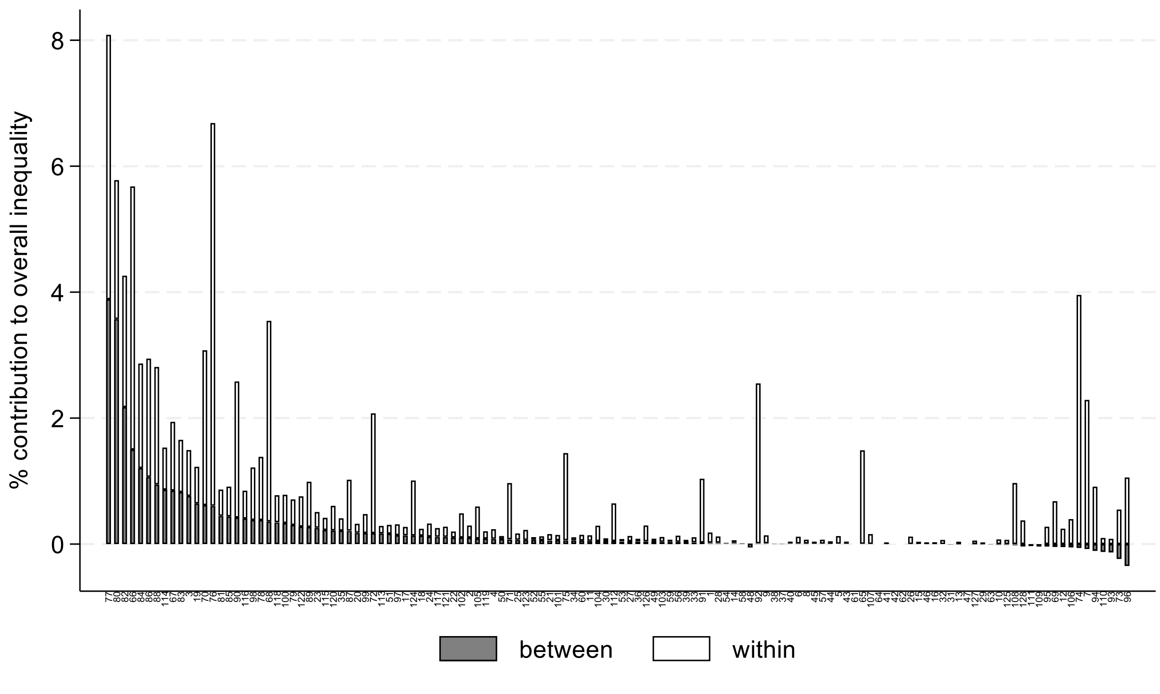
MLD, GE_0



Theil, GE_1



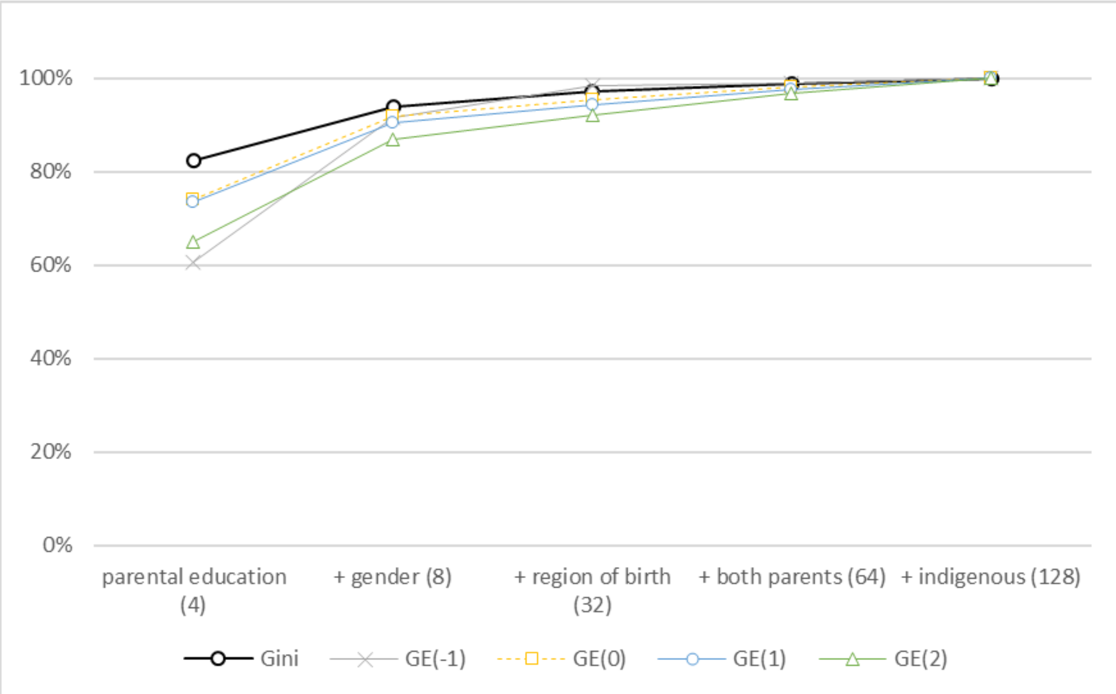
½ Squared Coefficient of Variation, GE_2



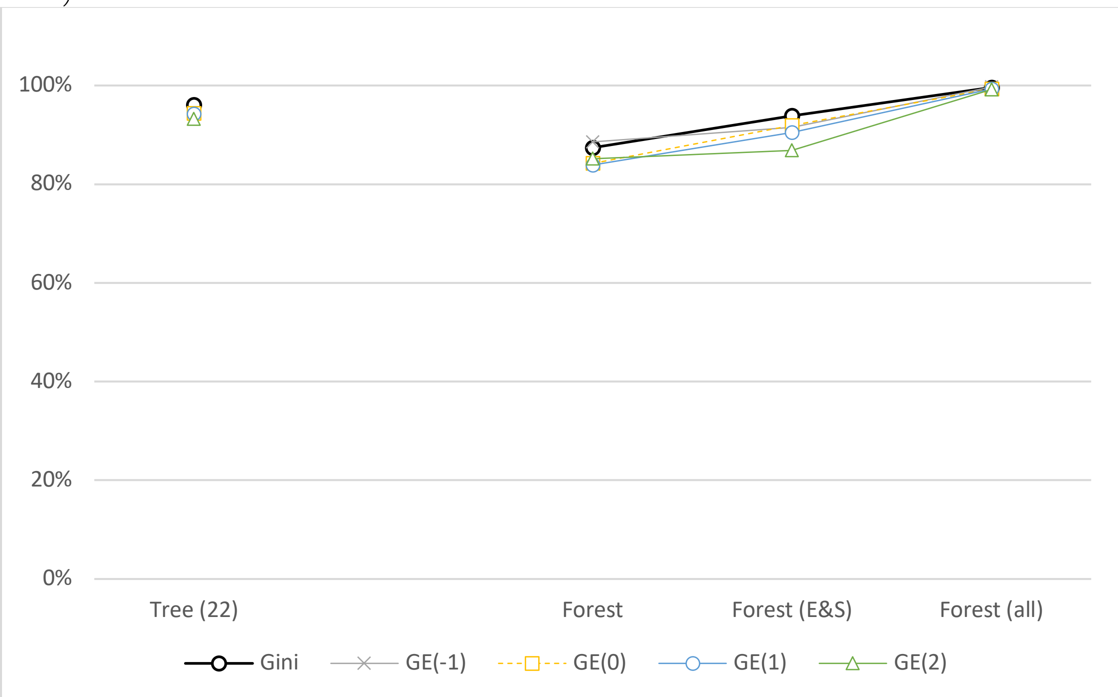
Source: Author's estimations based on [Dataset] CASEN (2022).

Figure A5. (Shapley) Inequality of opportunity in Chile in 2022, with various indices and type classifications: percentage of IO compared to IO using the most detailed classification of types used in this study

a) Variables added in a sequence



b) Tree and forests



Notes:

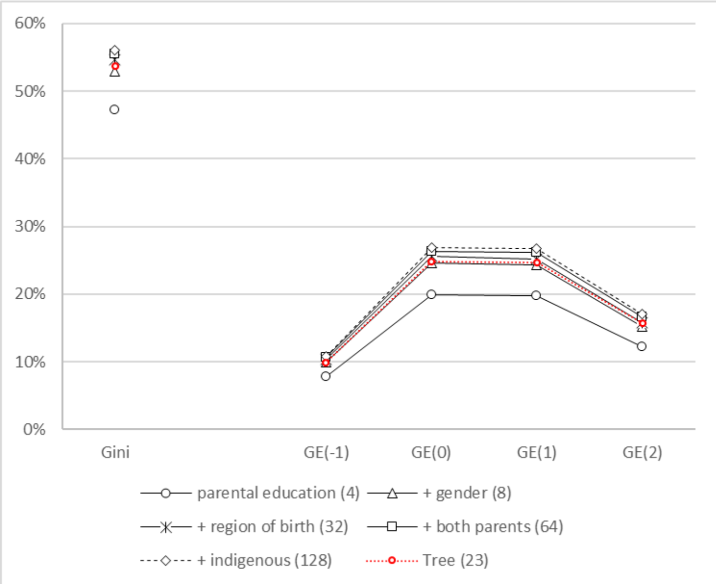
a) Circumstances are introduced sequentially. In parenthesis the resulting number of types.

b) 'Tree' is a conditional inference tree estimated with *ctree* (*Partykit* R package) with options similar to those used in the GEOM project (the proportion of observations needed to establish a terminal node as well as the significance level for variable selection is 1 percent). The tree with 22 final nodes (types) is reported in Figure A9. 'Forest' is a conditional inference forest estimated with *cforest* (*Partykit* R package) with options similar to those used in the GEOM project (1-alpha=0; the number of preselected random variables is the square root of the number of input variables). This forest computes inequality in the average predicted between-type distribution across 500 trees. Several of them do not use the information from the critical circumstances (parental education and gender), leading to a very low level of inequality between the different types in those cases (see Figure A10). Therefore, the graph also shows two alternative forests that always use these key variables: 'Forest' (E&S) is a forest constructed with parental education and sex, and 'Forest' (all) is a forest always using all circumstances. These show that as long as the two critical circumstances are used, the level of IO is very similar to the one reported in our study (e.g., 94 and 100 percent, respectively, for the Gini index; 87-92 and 99-100 percent for the entropy measures).

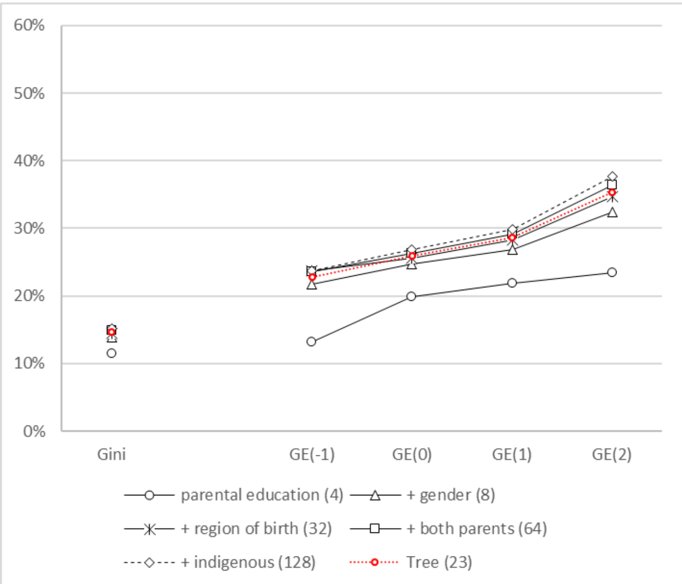
Each graph represents the percentage of IO reached with each set of circumstances as a proportion of the maximum. For example, the IO using parental education and gender is 94 percent of IO with all circumstances with the Gini index (87-92 percent with Entropy measures). With the tree, these percentages are 96 percent and 93-94 percent, respectively.
Source: Author's estimations based on [Dataset] CASEN (2022).

Figure A6. The contribution of inequality of opportunity in Chile in 2022 as a percentage of overall inequality with various indices and type classifications

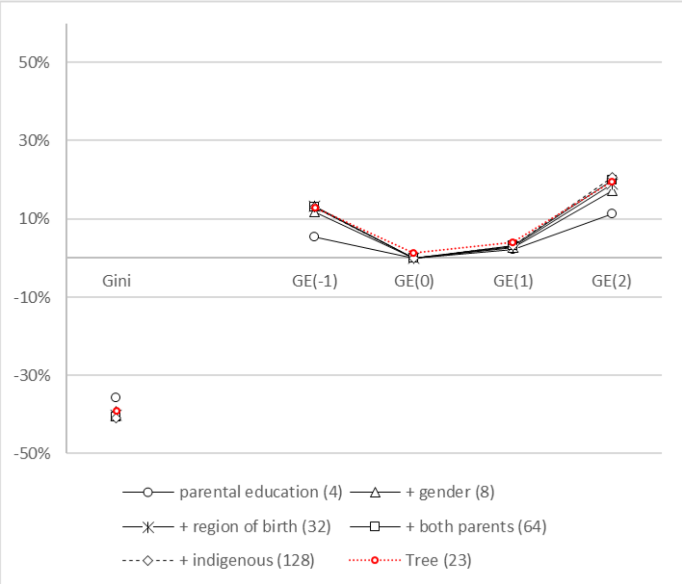
a. Smoothing within-type inequality, IO^b (the interaction term is assigned to inequality within types)



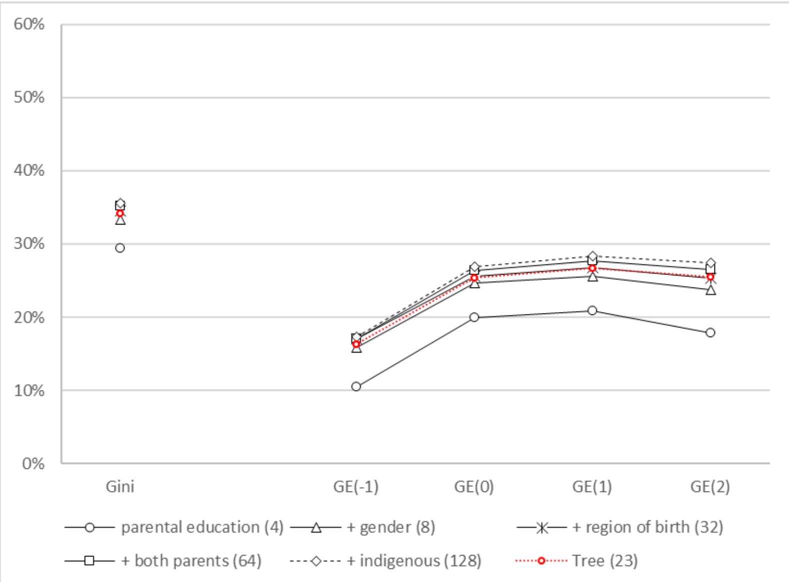
b. Smoothing between-type inequality, IO^w (the interaction term is assigned to inequality between types)



c. Interaction term, I_{bw}



d. Shapley, IO^s (average between both paths, assigning half of the interaction to inequality between types, the other half to inequality within types)



Notes. Circumstances are introduced sequentially. 'Tree' is a conditional inference tree estimated with *cforest* (*Partykit* R package) with similar options to those used in the GEOM project (the proportion of observations needed to establish a terminal node and the significance level for variable selection is set to 1 percent). In parenthesis the resulting number of types for each classification.

Source: Author's estimations based on [Dataset] CASEN (2022).

Figure A7. The trend in inequality in individual market income in Chile, 2009-22 using various indices: overall, between- and within-type distributions, as well as Shapley values

Notes. Circumstances are introduced sequentially. Tree is a conditional inference tree estimated with *cforest* (*Partykit* R package) with options similar to those used in the GEOM project (the proportion of observations needed to establish a terminal node and the significance level for variable selection is 1 percent). In parenthesis the resulting number of types.

Source: Author's estimations based on [Dataset] CASEN (2009, 2022).

Figure A8. The trend in inequality of opportunity in Chile, 2009-22 using various indices and type classifications: (Shapley) share of overall inequality, IO^s

Notes. Circumstances are introduced sequentially. 'Tree' is a conditional inference tree estimated with *cforest* (*Partykit* R package) with options similar to those used in the GEOM project (the proportion of observations needed to establish a terminal node and the significance level for variable selection is 1 percent). In parenthesis the resulting number of types.

Source: Author's estimations based on [Dataset] CASEN (2009, 2022).

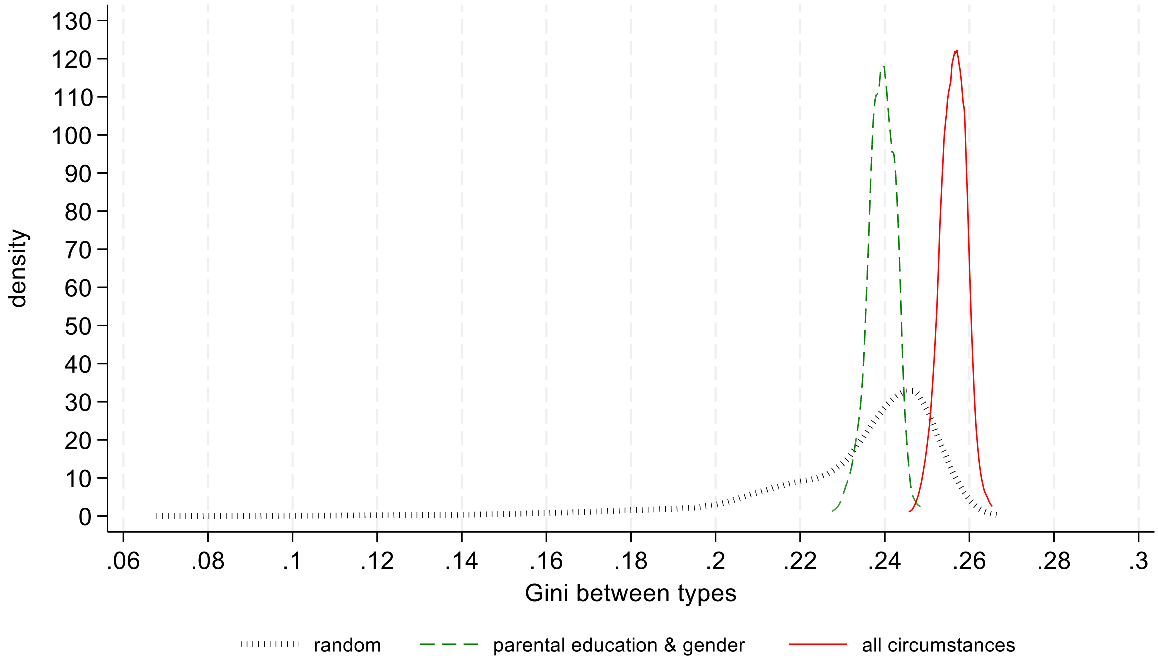
Figure A9. Conditional Inference Tree, Chile 2022

```
[1] Root
|   [2] Less than High Education
|   |   [3] None/Primary Education
|   |   |   [4] Male
|   |   |   |   [5] No Education
|   |   |   |   |   [6] North & Metropolitan: 772,324 (n = 1,228, err = 1.095316e+15)
|   |   |   |   |   [7] Centre & South
|   |   |   |   |   |   [8] Non-Indigenous: 637,689 (n = 2,567, err = 9.287274e+14)
|   |   |   |   |   |   [9] Indigenous: 548,084 (n = 624, err = 1.146159e+14)
|   |   |   |   [10] Primary Education: 874,767 (n = 2,942, err = 3.137395e+15)
|   |   |   [11] Female
|   |   |   |   [12] No Education
|   |   |   |   |   [13] Non-metropolitan: 424,369 (n = 2,933, err = 2.978126e+14)
|   |   |   |   |   [14] Metropolitan: 491,901 (n = 539, err = 9.791798e+13)
|   |   |   |   [15] Primary Education
|   |   |   |   |   [16] North & Centre: 525,245 (n = 1,595, err = 3.008945e+14)
|   |   |   |   |   [17] South & Metropolitan: 590,213 (n = 1,187, err = 3.533963e+14)
|   |   [18] Secondary Education
|   |   |   [19] Male
|   |   |   |   [20] Non-metropolitan
|   |   |   |   |   [21] Non-Indigenous: 1,153,029 (n = 2,364, err = 2.73138e+15)
|   |   |   |   |   [22] Indigenous: 959,242 (n = 315, err = 1.433552e+14)
|   |   |   |   [23] Metropolitan: 1,355,924 (n = 980, err = 2.173275e+15)
|   |   |   [24] Female
|   |   |   |   [25] Non-metropolitan
|   |   |   |   |   [26] Did not grow up with both parents: 650,563 (n = 519, err = 1.358989e+14)
|   |   |   |   |   [27] Grew up with both parents: 773,845 (n = 1,985, err = 1.036528e+15)
|   |   |   |   [28] Metropolitan
|   |   |   |   |   [29] Did not grow up with both parents: 708,729 (n = 187, err = 1.097295e+14)
|   |   |   |   |   [30] Grew up with both parents: 959,743 (n = 713, err = 6.707809e+14)
|   [31] Higher Education
|   |   [32] Male
|   |   |   [33] Did not grow up with both parents: 1,466,428 (n = 342, err = 5.363076e+14)
|   |   |   [34] Grew up with both parents
|   |   |   |   [35] North & Metropolitan: 2,293,867 (n = 750, err = 6.063652e+15)
|   |   |   |   [36] Centre & South: 174,0537 (n = 721, err = 1.633349e+15)
|   |   [37] Female
|   |   |   [38] Non-metropolitan
|   |   |   |   [39] Did not grow up with both parents: 904,762 (n = 261, err = 1.55536e+14)
|   |   |   |   [40] Grew up with both parents
|   |   |   |   |   [41] Non-Indigenous: 1,243,006 (n = 790, err = 9.733985e+14)
|   |   |   |   |   [42] Indigenous: 839,052 (n = 71, err = 3.096672e+13)
|   |   |   [43] Metropolitan: 1,486,433 (n = 652, err = 1.333122e+15)
```

Number of inner nodes: 21
Number of terminal nodes: 22

Source: Author's estimations based on [Dataset] CASEN (2022).

Figure A10. Distribution of Gini between types in all conditional inference trees in the random forests, Chile in 2022



Notes: 'random' refers to the standard case in which the squared root of the number of input variables are preselected; the other refer to the cases in which parental education and sex or all circumstances are preselected.
Source: Author's estimations based on [Dataset] CASEN (2022).