Labor market inequality of opportunity: A comparative analysis of Brazil and India

Tista Kundu[1]*, Daniel Duque[2], Valeria Pero[3] and Christina Terra[4]

[1] Department of Economics, CHRIST (Deemed to be University),
Yeshwanthpur Campus, Bangalore 560073, India.

[2] Norwegian School of Economics,
Post box 3490 Ytre Sandviken, Bergen 5045, Norway

[3] Institute of Economics, Federal University of Rio de Janeiro,
Avenida Pasteur, 250, University Palace, Praia Vermelha Campus, Urca,
Rio de Janeiro, RJ - CEP 22.290-902, Brazil

[4] Department of Economics, ESSEC Business School Paris,
3 avenue Bernard Hirsch, CS 50105 Cergy,
95021 Cergy-Pontoise Cedex, France

*Email: tista.kundu@christuniversity.in, ORCD ID:0000-0003-3581-9203

Abstract

This paper examines the structure of inequality of opportunity (IOP) in the Brazilian and Indian labor markets and its evolution between the late 1990s and early 2010s. We apply the machine learning approach of regression tree to study IOP with respect to wage earnings, performing a comparative assessment of the role of caste in India and skin color in Brazil and their interactions with gender, region and parental background. The main results of the regression trees are the following: (i) family education is the circumstance that contributes the most to the IOP in both countries, and this result is stable over time; (ii) the upper castes in India and "whites" in Brazil have persistent higher labor earnings; (iii) the importance of gender in IOP increases over time, with a double disadvantage of women belonging to the less advantaged social groups of race and caste; (iv) the Brazilian opportunity tree becomes less stratified over time, whereas India evolves towards a more complex stratification. Regarding IOP, we find some improvement for India over time, and a virtually stagnant level for Brazil. The IOP evolution is in the opposite direction of the evolution of earnings inequality. India's IOP remains higher than that of Brazil, with at least 30% of the earnings inequality being due to circumstances. Taken together, our results call for special attention to be paid to public policies aimed at reducing the differences in earnings between social strata in the labor market.

Keywords: Inequality of opportunity, race, caste, gender, regression tree, country comparison

JEL: E24, D31, O57

Acknowledgment

We gratefully acknowledge helpful comments and suggestions from Ashwini Deshpande, Atul Sood, Partha Chatterjee and participants of the 63rd Annual Conference on Labor Economics, conducted by the Indian Society of Labor Economics. Tista Kundu also thankfully acknowledges the AXA Research Fund Post-Doctoral Fellowship.

1 Introduction

Proponents of social justice argue that a potential source of persistent inequality is the one associated with the innate characteristics of individuals that are beyond their control or responsibility (Arneson 1989, Cohen 1989), which was formalized as Inequality of Opportunity (IOP) by Roemer (1998). There is a consensus in the literature that IOP is worse from the perspective of social justice compared to inequality of outcome (World Development Report 2006). Moreover, the persistence of IOP, by systematically excluding some population groups from participation in economic activity, can lead to a loss of productive potential, compromising economic growth and institutional stability. Indeed, there is evidence suggesting an inverse relationship between growth and IOP (Marrero & Rodríguez 2013, Aiyar & Ebeke 2020). Thus, high and persistent IOP is a matter of grave concern that justify appropriate redistributive policies in order to "level the playing field".

Brazil and India are two countries marked by IOP at the very root of their development history. On the one hand, Brazil was the American country with the largest African slave migration, and the last one to enact the liberation of slaves. Skin-color or race discrimination in Brazil is an active field of research, with evidence of "white premium" in the labor market persisting in the twenty-first century (Gerard et al. 2021). On the other hand, the Indian caste system was a social division based on occupation, where people related to "purer" jobs of worshiping or teaching were at the top layer, while those related to the polluted jobs of manual scavenging or burning corps were at the bottom. In spite of being the first country to implement affirmative action policies, the historically marginalized caste groups in India are still found at the lowest rung in almost every economic spheres and are often subject to social stigma (Madheswaran & Singhari 2016). Therefore, skin color and caste are prone to be important circumstances explaining IOP in these countries.

We contribute to the economic literature on IOP by performing a comparative assessment of the role of castes and skin color in the IOP of India and Brazil, focusing on their interactions with other common sources of income inequality (gender, region, and the educational status of parents), and its evolution from the late 1990s to the early 2010s. We apply the state-of-the-art machine learning approach of regression trees by Brunori & Neidhöfer (2020), Brunori et al. (2023) to study IOP with respect to wage earnings in Brazil and India. While some scholars separately identify Brazil and India as countries with high measures of IOP (see for example Bourguignon et al. 2007, Ferreira & Gignoux 2011 for Brazil and Asadullah & Yalonetzky 2012, Singh 2012 for India, among others), none of them have applied the machine learning approach to identify the inter-linkage among the discriminating circumstances in either country, let alone in a comparative set-up, which is a novel contribution of this paper.

Inequality of opportunity is defined as the inequality between different 'types' of individuals, where 'types' are the permutation of several circumstances that divide the population under study into mutually exclusive layers. The between-type inequality, as a measure of IOP, is identified through the construction of counterfactual distributions by replacing each type with their respective mean outcomes (and thereby eliminating within-type inequality) (Bourguignon et al. 2007, Checchi & Peragine 2010, Ferreira & Gignoux 2011). Another robust approach is to compare the entire distributions of outcome differences between the types, instead of just their means (Lefranc et al. 2009, Andreoli et al. 2019). However, in either of these approaches, it is at the discretion of the researcher to decide which combination of circumstances (i.e. types) to consider for evaluating the IOP, which may lead to estimation biases (Brunori et al. 2019, 2023). Brunori et al. (2023) show that, by limiting the subjective choices made by researchers, the machine learning approach of regression trees offers more reliable estimations of IOP. This method relies on a well-defined algorithm with minimum assumptions regarding how circumstances affect the outcome variable. Dealing with these issues is particularly important when carrying on international comparative analysis, since each country is prone to have their own biases. We therefore use the machine learning approach to have comparable IOP estimates for Brazil and India.

Besides delivering more reliable IOP estimates, the machine learning approach offers at least two further advantages over the classical methods. First, it allows the identification, in a statistically significant way, of the subset of possible types that are relevant to generating IOP. Second, it provides a visually appealing opportunity tree that depicts the hierarchical order among the circumstances along with how they intertwine. These two features help us to form a more accurate and complete picture of the IOP determinants. In particular, when comparing Brazil and India, we can identify differences in which types are relevant to IOP in each country, as well as their relative importance.

On top of caste for India and skin color for Brazil, we use gender, region, and the educational status of parents as discriminating circumstances. Indeed Indian women from the lowest caste groups and black Brazilian women are often found to have the least employment opportunities (Deshpande 2007, Cacciamali & Rodgers 2015). Regarding regional inequalities, the North and North-East of Brazil are relatively more rural and poor compared to the South and South-East regions, and they also have the largest share of non-white population. In India, its West and Southern regions are relatively richer while the East is poorer. Finally, previous literature has identified parents' education as one important driver of IOP (Bourguignon et al. 2007, Brunori & Neidhöfer 2020, Roemer

& Trannoy 2016).

We use the National Household Sample Survey (PNAD) micro-database for Brazil, conducted by the Brazilian Institute of Geography and Statistics (IBGE), for 1996 and 2014, which are the two latest years containing information on family education. For India, we use the latest two rounds of Employment-Unemployment survey from the National Sample Survey (NSS) micro-database, 1999/2000 and 2011/2012. We are then able to estimate the evolution of IOP in these two countries over the two decades from the late 1990s and to the early 2010s. To make results comparable, we apply the same sample selection criteria for both datasets and we harmonize the number of categories for each circumstance.

The regression trees plots of Brazil and India depict four noteworthy patterns regarding the contribution of the different circumstances to the IOP. First, family education is the circumstance that contributes the most to the IOP in both countries, and this result is stable over time. This finding is in line with those of Roemer & Trannoy (2016), that identify parental background as crucial in determining IOP in several developing and developed countries. Second, we find a clear earnings advantage for the upper castes in India and for "whites" in Brazil, which corroborates the literature on labor market discrimination in these countries (Arcand & D'hombres 2004, Deshpande 2011). Third, the interconnections between gender and race are stronger in Brazil than those between gender and caste in India. However, despite the very different dynamics of female labor force participation in the two countries, we find two common features: (i) an increased importance of gender in IOP over time; and (ii) a double disadvantage of women belonging to the less advantaged social groups of race and caste, which persists throughout the time span considered. Finally, the Brazilian opportunity tree becomes less stratified in 2014 compared to 1996, reflected by the reduction of the number of types. India, however, evolves towards a more complex stratification to explain IOP over time.

Regarding IOP in the labor market, we find a "contrasting trajectory" between India and Brazil, on the opposite direction to what Barbosa et al. (2017) present in terms of income inequality. While the fall in earnings inequality was substantial in Brazil between 1996 and 2014, India presented almost unchanged inequality between 1999 and 2012. Our IOP estimates reveal the opposite picture: some IOP improvement for India over time, and virtually stagnant IOP level for Brazil. Still, India's IOP remains higher than that of Brazil, with at least 30% of the earnings inequality being due to circumstances. Taken together, our results call for special attention to be paid to policies to reduce the IOP in order to narrow the gap between social strata.

The rest of the paper is organized as follows. A brief account of the socio-political backgrounds of both countries is provided in section 2 to set the context of our study.

Section 3 explains the methodology followed by a detailed description of our dataset and sample summary in section 4. Section 5 presents our main results and section 6 concludes.

2 Socio-political backgrounds in India and Brazil

Brazil and India are two large emerging economies different in various social and economic dimensions, but similar in their high inequality of opportunities stemming from the history of their development. Marked by its history of slavery, Brazil was one the last countries in the world to enact the liberation of slaves. On average, Brazilian of African origin still have a lower level of schooling, lower income, and a higher poverty rate than whites . The Indian social structure, in turn, was built on a rigid caste system based on occupation status. Families in occupations related to "purer" jobs, such as worshiping or teaching, were at the top layer, while those in "dirty" jobs, such as manual scavenging, were at the bottom. Despite policies aimed at suppressing the caste system, inequalities related to it persist (Deshpande 2011).

Does the historical legacy of slavery in Brazil and the caste system in India still play a role in the inequality of opportunity in these two countries? Is there a persistence or a moving trend over time? To answer those questions, we apply a methodology that allows us to compare the inequality of opportunity in these two countries and its evolution between the late 1990s and the early 2010s. Before our empirical analysis, we present a brief description of the major socio-economic trends over that period in this session. Barbosa et al. (2017) provides evidence of an opposite path of growth and inequality in India and Brazil, particularly since the turn of the twenty-first century. We will look at the evolution of each country in turn.

Both countries launched a host of neo-liberal economic reform policies during the early nineties, with massive trade liberalization (Barbosa et al. 2017). In the 1980s and 1990s the Brazilian economy was marked by macroeconomic instability, with very high inflation rates and low economic growth. The introduction of the "Real plan" in 1994 was successful in fighting inflation, inaugurating an era of economic stability in the country (Averbug 2002). Our time frame for studying Brazil starts just after this turning point, in 1996.

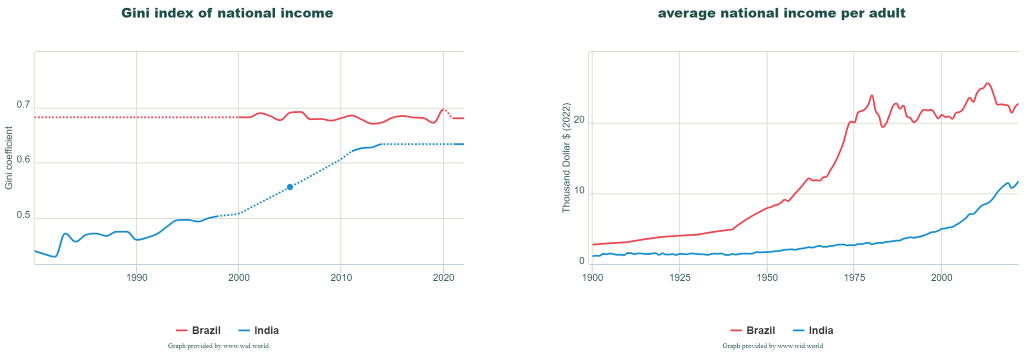
An important political turnover happened in Brazil in 2002 with the Workers' Party winning the presidential election and implementing new policies aiming at fighting poverty and reducing inequality. The Brazilian government expanded social programs, such as inclusion programs in higher education and the well known '*Bolsa Familia*' program, which is a conditional cash transfer scheme that provides financial aid to the poorest households. Regarding the labor market, the government pursued a policy of increasing the real minimum wage, and a mix of policies and economic performance led to an increase in the share of formal employment. The decrease in income inequality over the period has been attributed to the combination of these factors (Hall 2006, Firpo & Reis 2007).

India's economic performance was hardly impressive at that time, even though the country was barely perturbed by global financial crisis given its weak linkages to the international markets. The stable political period lead by the Indian National Congress (INC) during the 1980s no longer holds in the following decade until the National Democratic Alliance (NDA) came in power in the late 1990s, when we begin our analysis on India. Unlike the previous decade, NDA runs the Indian parliament for six consecutive years followed by a second stable phase of INC starting in 2004, with a rejuvenated political set-up and launching of large scale social policies. One of the most notable policy over this period is the National Rural Employment Generation Act (NREGA), which is the largest employment generation program. This program guarantees 100 days of paid work to every rural adult in India. Similar to the *Bolsa Familia* in Brazil, several studies find NREGA to have a significant impact on labor market inequality in India (Imbert & Papp 2015, Khurana & Mahajan 2020). Hence, our time frame corresponds to a period of economic stability and changing political priorities towards more inclusive policies in both countries.

The 2000s mark a favorable decade for Latin America countries, including Brazil, that experiences higher economic growth rates with a decrease in income inequality (Firpo & Portella 2019). Both indicators improved more rapidly since 2004, and began to show signs of stagnation from 2013. India, in turn, has witnessed rapid economic growth over the past two decades, but with a sharp rise in income inequality (Barbosa et al. 2017). As depicted in Figure 1, despite these contrasting trajectories between Brazil and India since 1990s, Brazil remains a richer country than India, with converging inequalities between them. In sum, while Brazil sees a fall in inequality afterwards along with a paltry growth rate, India witness a sharp rise in both (Rodgers 2016). It is in this socio-polical backdrop that we stage our comparative analysis of labor market IOP in India and Brazil.

3 Methodological framework

The canonical structure of IOP considers the outcome (y) to be a function of two mutually exclusive and exhaustive factors, circumstances (c) and effort (e), that is, $y = g(c, e)$. *Circumstances* consist of an array of variables that falls beyond individuals' control, while *effort* are variables that individuals typically can regulate. Given this set up, IOP is the inequality in outcome, y, generated by circumstances, c, only. The main methodological challenge is to capture the isolated effect of c inequality on y, leaving out the influence



(a) Gini index (b) National Income

Figure 1: Income inequality: Brazil and India[a]

[a]Source: World Inequality Database (WID). Naitonal income is in dollar PPP.

of e. That is a formidable task, primarily because of data availability, as the data on so-called *effort* are often not captured in most of the national scale surveys. Even with identifiable effort variables, one can not ensure the complete segregation between them and the circumstances. In fact, the choice of *effort* is often shaped and determined by the underlying *circumstances* in which people find themselves and it is virtually impossible to distill *pure effort* without any circumstantial effect on it. To solve this problem, *effort* is often assumed to be a function of c itself, that is, $e = \phi(c)$, leading to the reduced form outcome function as $y = f(c)$ (Ferreira & Gignoux 2011). Hence, a common first step of any IOP analysis is to capture as many circumstance variables c as possible so that, from them, we can obtain a representative set of types to generate a reliable estimate of "unfair inequality".

Consider a finite population set, $i \in \{1, ..., N\}$, characterized by $\{y_i, c_i\}$, standing respectively for outcome and circumstances for individual i. Furthermore, suppose that the vector c_i consists of J circumstances and each of them can take on x_j values or categories, so that the total sample space can be partitioned into a maximum of $K = \prod_{j=1}^{J} x_j$ "types". In this setting, IOP evaluation is reduced to the evaluation of inequality *between* these "types" and not *within* them.

One way to do that is to construct a counterfactual distribution, $\tilde{y}$, that represents each type by their respective mean outcome, such that $\tilde{y} = \{\mu_1, \ldots, \mu_K\}$, where μ_k denotes the mean of type k. This counterfactual distribution, unlike the actual one, contains only the variation between "types" by muting any variations within them. IOP in absolute terms can be measured directly by estimating the inequality of the counterfactual distribution, $\tilde{y}$.

While we can use any inequality index, the majority of the related literature resorts to the index of Mean Log Deviation (MLD) because of its path independent additive decomposability. By virtue of this property, the MLD of the actual distribution, y, can be written as a sum of inequalities within and between types, that is:

$$\text{Total inequality in MLD} = \text{MLD between types} + \text{MLD within types} \tag{1}$$

This allows an appealing representation of IOP, which is the relative measure of IOP as a share of total outcome inequality:

$$IOP = \frac{MLD(\tilde{y})}{MLD(\{y_i\})} \tag{2}$$

Brunori et al. (2023) points out two main problems with existing estimation approaches. The first one is related to the fact that, in these methods, researchers choose at their discretion a sub-set of all possible "types" for the IOP estimations. Failing to include relevant types leads to a downward bias in the estimation of IOP, while including too many of them generates upward-biased estimates. Second, in parametric approaches, researchers must specify the functional form that represents how different circumstances interact to generate the outcome of interest. On the one hand, too restrictive functional forms may limit the explaining power of circumstances on the outcome variable, thus generating a downward bias in the estimate of inequality of opportunity. On the other hand, estimating the model with too many interactions of circumstances may render the estimation infeasible. The authors show that the machine learning approach of regression trees offers more reliable estimations of IOP. They limit the subjective choices made by researchers, relying instead on a well-defined algorithm with minimum assumptions regarding how circumstances affect the outcome variable. Dealing with these issues is particularly important when carrying on international comparative analysis, since each country is prone to have their own biases.

The difference between non-parametric and tree-based IOP will be higher the larger the set of circumstances or the number of categories per circumstance variable. Moreover, in such cases, some "types" may contain very few data points, which would affect the resulting measure of non-parametric IOP. The advantage of the regression tree method is that the identification of "types" are based on the most relevant interactions of the circumstance variables and thereby minimizing the chances of spurious IOP in case of inadequate data availability per "type" (Brunori et al. 2019).

We therefore apply the regression tree methodology of Brunori & Neidhöfer (2020). This is a machine learning based approach where the researcher submits the full set of available circumstances, C_i, to the program and let the algorithm choose the relevant partitioning of the sample space into different 'types'. This method sorts the circumstance variables by how well they can predict the outcome, while letting these variables to interact with one another in a flexible manner to bring out their interactive effects. Hence, one advantage of this method is to provide a hierarchy of the types which it identifies as relevant. Another advantage of this approach is that the underlying algorithm generates a visually appealing opportunity tree depicting the hierarchy of the circumstances.

The algorithm of regression tree runs in two stages as follows:

- **Stage I**: Selecting the initial splitting circumstance
  - It starts with the simultaneous testing of the J partial hypothesis, H_0^{Cj} : $D(Y|C^j) = D(Y)$ for $j \in \{1, \ldots, J\}$. This is precisely the testing of the existence of IOP, to identify whether any of the circumstances have an effect on the outcome.
  - Adjusted p-values, p_{adj}^{Cj}, are then computed with the standard adjustment for multiple hypothesis testing[1], identifying the circumstance, C^*, with the highest degree of association, that is, the circumstance with the minimum p-value, $C^* = \{C^j : \text{argmin} \quad p_{adj}^{Cj}\}$.
  - The algorithm stops if the p-value associated to C^* is greater than some pre-specified significance level, α.[2] Hence, if $p_{adj}^{C^*} > \alpha$, the null hypothesis of equality of opportunity for the society cannot be rejected at $\alpha\%$ level of significance. Otherwise, the circumstance, C^*, is selected as the initial splitting variable.
- **Stage II**: Growing the opportunity tree
  - Once C^* is selected, it is split by the binary split criterion to grow the tree. For each possible binary partition, s, involving C^*, the entire sample can be split into two distinct parts as, $Y_s = \{Y_i : C_i^* < x_j\}$ and $Y_{-s} = \{Y_i : C_i^* \geq x_j\}$.
  - For each binary split, s, the goodness of split is tested by testing the discrepancy between Y_s and Y_{-s}. The split, s^*, with the maximum discrepancy, that is with the minimum p-value, is then selected as the optimum binary split point, based on which the sample is now partitioned into two sub-samples, constructing the initial two branch of the opportunity tree.

[1]The adjustment is the Bonferroni correction, $p_{adj}^{Cj} = 1 - (1 - p^{C^j})$.

[2]The level of significance bears the usual connotation as in any regression analysis. The length of opportunity tree can vary with different levels of α: a larger value of α may identify more types as being relevant, while the opposite is true for a more binding α value. The structure of the tree, however, does not depend on the value of α in the following sense: a larger value of α may add more layers to the bottom of the tree, but it does not change the structure of the first layers identified by a smaller value of α. Brunori et al. (2023) used a "K-fold cross validation technique" to tune the tree for optimal α. Their so-called "tuned" value of α is very close to the conventional value of 0.01 and they find that the IOP estimates do not vary by a large extent with other conventional α values (for example, 5% level of significance). We have checked that our IOP estimates and the top nodes of the opportunity tree do not change with $\alpha = 0.05$. Hence, we proceed with the conventional norm of setting $\alpha = 0.01$ for our analysis.

– The entire algorithm is then repeated for each branch separately, to construct the full opportunity tree.

Once the relevant types are revealed, IOP can be measured using equation (2), by calculating the between-type inequality. Therefore, "types" are no longer formed by a complete segmentation of the circumstances or by an arbitrary choice. Rather, the final IOP estimate takes into account the inequality between those "types" that the algorithm has identified as most relevant, and which are depicted on the opportunity trees.

4 Data, variables and sample selection

4.1 Data sources

India: We use two rounds of the employment-unemployment survey (EUS) database from the National Sample Survey (NSS): the 55th and 68th rounds, which correspond to survey years 1999-2000 and 2011-2012, respectively.[3] Each round surveys about 120,000 households on average, totalling over 400,000 individuals. These are large-scale surveys covering the entire country, with the exception of a few remote villages and conflict zones representing no more than 2% of the national population. We limit ourselves to households with inter-generational co-residence for extracting information on family backgrounds. As adult co-residence is the most widespread form of social norm in India, this sample restriction still allows us to obtain reliable estimates (Hnatkovska et al. 2013).

Brazil: For Brazil we use micro-data from the National Household Sample Survey (PNAD) conducted by the Brazilian Institute of Geography and Statisctics (IBGE). We take two rounds of the of PNAD, 1996 and 2014, which include a supplement with information on fathers' education. Each round covers over 100,000 households, corresponding to over 300,000 individuals. Surveys prior to 2003 do not include the rural areas of the North region, nor information on individuals who were not heads of household. We therefore exclude these data from the 2014 round to make it comparable with the 1996 round.

4.2 Variables of analysis

Given the available dataset, we focus on four circumstances for each country. Three of them are common to both countries: gender, region, and educational background of father. The fourth circumstance is country-specific: caste for India and race for Brazil.[4]

[3] NSS survey year corresponds to a typical agricultural year spanning from July to June of the corresponding years.

[4] We abstain from taking religion as another circumstance variable along with caste because of their strong correlation and possible overlapping. Casteism has its origin in ancient Hindu text. Many

Region: We have information on birth region for Brazil. For India, we take region of residence as a proxy of birth region. Given the low rate of inter as well as intra-state migration, this serves as a decent proxy for birth-region and has been used widely in related academic research (Singh 2012). For both countries we consider five broad regions according to the most prevalent form of categorization in their respective literature. We take North, East, Central, South and Western regional divisions for India. We exclude the special case of North-East India that consists of many remote villages. They are out of the survey coverage with tribal habitation and does not constitute more than 4% of the national population. For Brazil we consider North, North-East, South-East, South and Mid-West.

Family education: To allow the comparison between India and Brazil, we separate father's education into five levels: illiterate, incomplete primary, incomplete secondary, incomplete tertiary, and tertiary or above. Illiteracy denotes no exposure to formal schooling. Incomplete primary, in turn, denotes a beginning of schooling that ends even before finishing primary level school education. Similar is the case with incomplete secondary and tertiary.

Caste: Caste in India was an ancient social division based on hereditary occupation. The thousands of castes (technically sub-castes which are endogamous groups within the broad caste structure) are regrouped into four categories by Government of India in order to foster caste based affirmative policies (Deshpande 2011). The Scheduled Castes (SC) and the Scheduled Tribes (ST) are the earliest formed categories among them. Another group of socially and economically backward castes are put under the category of Other Backward Class (OBC) since mid-1980s. The rest of the Indian population who does not belong to any of the above, are automatically classified as the 'General category' with no entitlement of caste based benefits and whom we designate here as the higher castes. In this work we take three caste categorization as higher, middle (OBC) and lower (SC and ST taken together).

Race: Race in Brazil is determined by skin color and therefore bears more immediate visibility in public domain. In Brazilian surveys, individual respondents self-declare their race as an yardstick of social identity (Telles 2002). Caste in India on the other hand is determined objectively from the Government approved list, separately for each state. Therefore, race bears a less formal connotation than caste. IBGE provides five official racial grouping: white, multiracial, black, yellow, indigenous. We consider three race categories: white, brown and black. We take yellow and white together to form the

marginalized castes started converting their religion to escape from caste based oppression (Deshpande 2011). With more than 70% Hindu population, upper castes are always proportionately higher among Hindus, while a relatively larger share of marginalized castes can be found among Muslims, Christians and/or Buddhists.

'white' group, whereas 'brown' constitutes both indigenous and the multiracial.

Earnings: We consider the real wage as our outcome variable. Given the widespread informal character of the Indian labor market, wage is reported as total weekly earnings for India. Information on earnings is not available for the non-salaried workers who are reportedly self-employed. Wage for Brazil is the total monthly earnings, even for short-term contractual workers and self-employed workers. For keeping parity in our cross-country comparison, we limit ourselves to the salaried workers in both countries, that is, those who are not self-employed workers. We take the real value of wage as our outcome variable.[5]

4.3 Sample selection and descriptive statistics

We consider salaried workers aged between 18-45 years, to focus on the prime working-age population. As mentioned before, we also exclude self-employed workers, due to the lack of such date for India. That leaves us with nearly 60% and 75% of employed workers in India and Brazil, respectively.[6] Despite the exclusion of the self-employed, we still have a large sample with similar distribution of the major socio-economic characteristics. Moreover, the salaried labor market alone is of sufficient interest for both the countries, which is the focus of the present work.

There is a point of difference between the two survey rounds we use for Brazil. Family background in the 1996 PNAD supplementary survey is reported for all the household heads as well as their spouses, but for one random member from each household in 2014. Hence for comparability, we limit ourselves to those who are only heads or spouse of heads of the household for the latter round, which decreases the number of observations of our 2014 Brazilian sample significantly.[7]

For India, data on family background is not part of the direct questionnaire of any NSS survey and it is only available for co-resident households where individuals from both generations are enumerated simultaneously. For many countries this may cause severe selectivity bias in the working sample, but not for India where large joint-family structure is the prevalent social norm (Hnatkovska et al. 2013, Rosenzweig & Zhang 2014, Reddy 2015). In the database we use, over 70% of the surveyed households are co-resident

[5] For India, we use CPI for agricultural labor (CPI-AL) for 1999-2000 and 2011-2012 with 1986 as the base year, that is published by the Reserve Bank of India. Data on real wage for Brazil however, is directly published by IBGE which we have used here.

[6] Although earnings of self-employed is available for Brazil, we drop it for the sake of comparability. Besides it is difficult to get accurate information on earning for the self-employed, given their characteristics are quite different from that of the wage-earners, mainly in terms of income stability, autonomy and control over work.

[7] As a robustness check, we randomly drop data points from the larger sample set so that the sample size matches for both survey years. We find negligible difference in the regression tree estimates on all these smaller sub-sets of the larger sample set.

households where at least one of the parent is living with their adult child/children.[8] The distribution of caste categories, agricultural wage earners, levels of school completion, and rural habitats are nevertheless similar across the co-resident and the non-co-resident sample. The samples differ mainly in terms of age and gender composition. Co-resident salaried workers are younger. An obvious reason for this is that the older samples are likely to have deceased parents. This is why we focus on the relatively younger wage earners to maximize our information set on parental backgrounds. Also, the share of married female workers is relatively low in co-resident sample, since in India women usually migrate to the groom's home after marriage. To avoid having too few women workers in our sample, we have taken data of the groom's parent as a proxy for parental information of the married women in the house.

Table 1 reports the distribution of the circumstances in our samples for each country and the two survey rounds, separately. While majority of the Brazilian wage earners are "white", not even one-third of them in India belong to the higher castes. The share of "browns" in Brazil and OBC (middle castes) in India are increasing in the labor market. Table 1 further shows the low rate of female workforce in India, which decreases over time. This feature is consistent with the existing evidence (Mehrotra & Sinha 2017). In Brazil, the share of women and men in the labor market is equal by 2014 in Brazil.

Over half of the Brazilian sample are from the South or the South-East, two economically strong regions of the country (Bucciferro & Ferreira de Souza 2020). Compared to Brazil, regional heterogeneity is greater in India. The growing service sector of the Southern regions, not surprisingly, brings in a larger share of wage earners, whereas we see less of them from the relatively poor Eastern states. The Amazon basin in North Brazil and the Himalayas, as well as the conflict zones of Kashmir in North India, have the least share of wage laborers.

Apart from the gender ratio, the second most divergent circumstance between the two countries is family education. While half of the Indian sample is from illiterate families, Brazil has a larger share of those whose fathers with some primary education, as shown in Table 2. The distribution of family education in Brazil and India becomes more similar for those from families with a higher level of schooling. The share of college-educated families is higher in the latter period for both countries, and it is higher in Brazil compared to India.

[8]There is substantial evidence of elderly inter-generational co-residence in East, South and South-East Asia (Croll 2006) and the rate of adult inter-generational co-residence is found to be increasing in many developing countries including India (Ruggles & Heggeness 2008). In the world value survey of 2005, India stands highest in terms of co-residence. Ethiopia, Malaysia, Morocco, Indonesia, Iran, are among the few next countries with very high co-residence, while this is lowest for countries like New Zealand, Australia, Finland, Norway, Sweden, Switzerland. In 2010, India, is still the fourth highest country regarding co-residence for the inclusion of Algeria, Haiti and Pakistan, that have even higher co-residence.

Table 1: Distribution of working sample by circumstances

	Brazil		India	
Brazil / India	1996	2014	1999	2012
Race / Caste				
White / Higher caste	0.56	0.44	0.30	0.27
Brown / Middle caste	0.38	0.45	0.35	0.41
Black / Lower caste	0.06	0.11	0.35	0.32
Gender				
Male	0.58	0.50	0.81	0.84
Female	0.42	0.50	0.20	0.16
Region				
North / North	0.05	0.10	0.06	0.09
North-East / East	0.30	0.30	0.19	0.19
South-East / Central	0.35	0.31	0.20	0.22
South / South	0.22	0.21	0.35	0.30
Mid-West / West	0.08	0.08	0.20	0.20
Family education				
Illiterate	0.28	0.19	0.54	0.47
Incomplete primary	0.57	0.43	0.15	0.13
Incomplete secondary	0.06	0.13	0.20	0.22
Incomplete tertiary	0.06	0.17	0.08	0.12
Tertiary	0.04	0.08	0.03	0.06
Sample size	34,512	8,725	17,362	16,163

Sample: Non-self employed workers aged 18-45 years
Author's calculation from: PNAD (Brazil); NSS (India)

Table 2: Descriptive summary statistics of working sample

	Brazil		India	
	1996	2014	1999	2012
Mean household size	3.9	2.9	6.6	5.9
%Rural	0.14	0.08	0.71	0.63
%Agricultural workers	0.09	0.05	0.45	0.27
%Registered / Regular	0.70	0.76	0.66	0.54
%Illiterate	0.09	0.03	0.29	0.11
%Some college	0.12	0.25	0.09	0.19

Note: Working sample constitutes of non-self employed workers aged 18-45 years. Each row indicates the share of the corresponding criteria in the working sample. Registred workers in Brazil are those in the formal sector, whereas regular workers in India are those with a regular monthly or weekly salary, comprising of informal sector as well. 'Some college' indicates those who began college but may or may not have completed it. Source: Author's calculation from PNAD (Brazil); NSS (India)

Table 2 depicts some additional stylized facts in our working samples such as a larger household size, a higher share of agricultural workers as well as the predominantly rural character of India as compared to Brazil. The increasing formalization of labor market in Brazil is reflected by a higher share of registered workers in 2014, and the reverse is true for India. The descriptive statistics in Table 2 are in line with the labor market structures in Brazil and India (Narayanan 2015, Firpo & Pieri 2018).

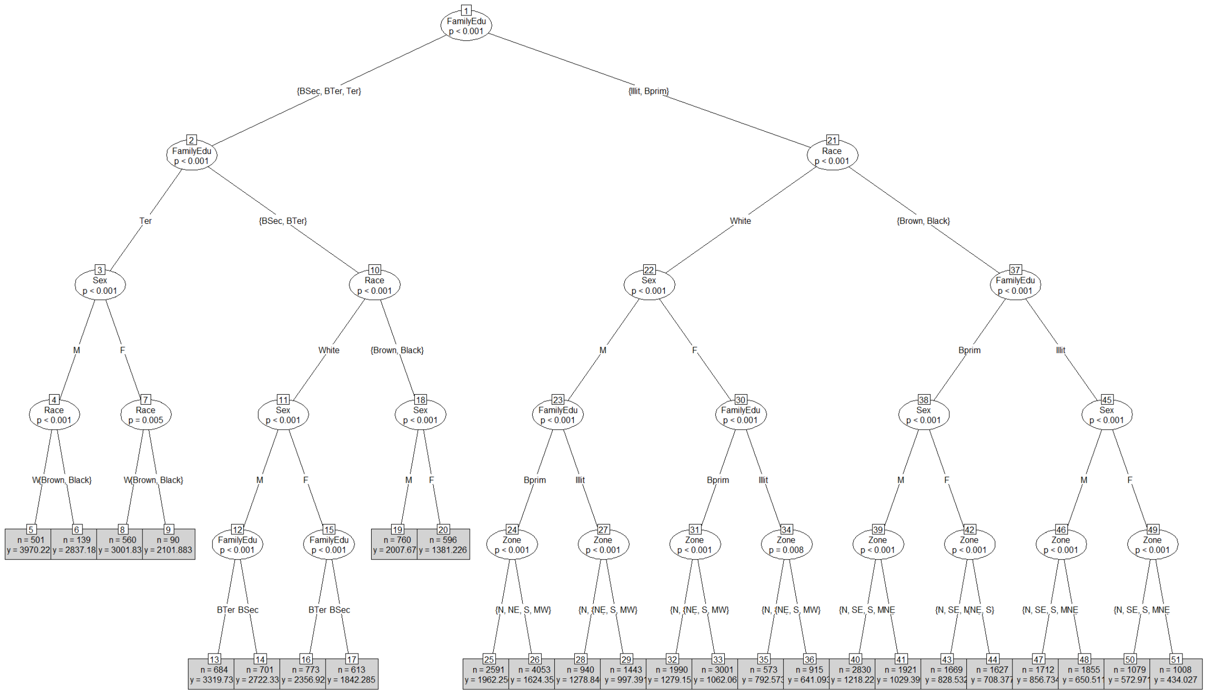
5 Results and discussion

5.1 Regression tree plots

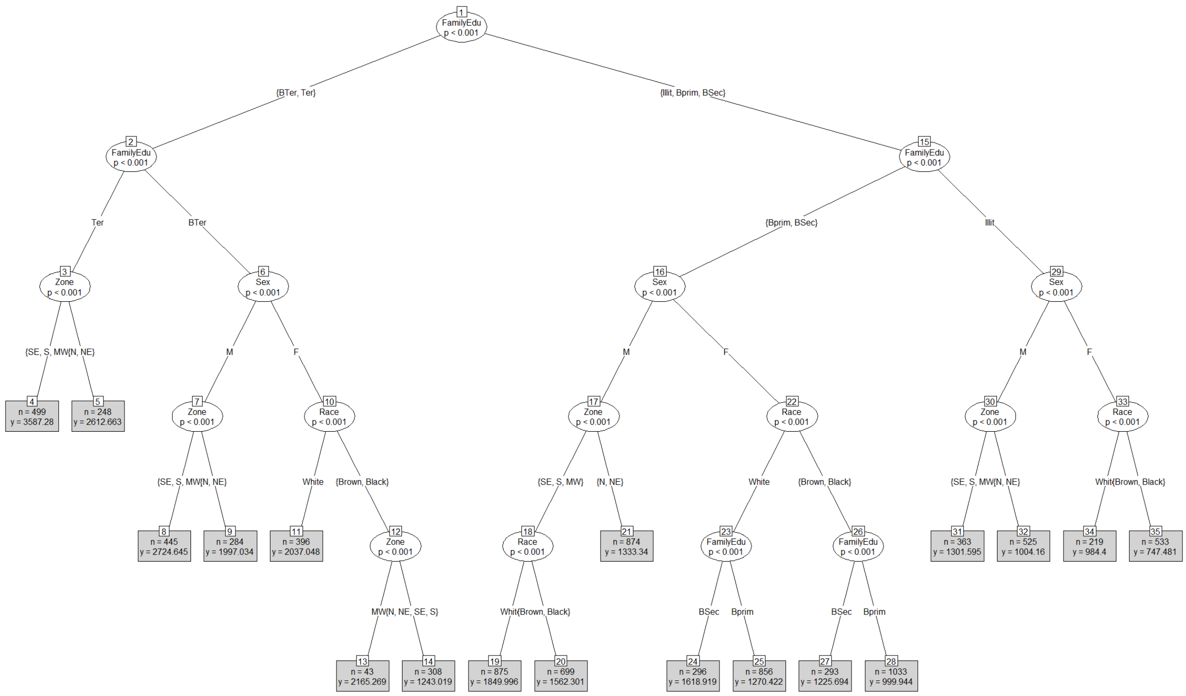
Figures 2 and 3 show the opportunity trees for Brazil and India for the two periods of time. The first node of each tree indicates the circumstance that contributes the most to the IOP. The tree grows in interaction with other circumstances, generating subsequent branches in lower order of hierarchy and it stops when the null of equal opportunity among the resulting nodes can no longer be rejected. Each terminal node of an opportunity tree denotes a "type" which is composed by the specific interaction of circumstances leading to that node. Hence, types are generated upon identifying the most statistically significant interactions of the circumstances following the algorithm of section 3. Point estimates of IOP is obtained by evaluating inequality in the distribution of mean earnings of the formed types.

Opportunity trees portray a complex labor market stratification in both countries, with intricate hierarchy among the considered circumstances. Nevertheless, for both Brazil and India, family background turns out to be the most important circumstance throughout the time period. On average, earnings are higher for workers whose fathers have college level education in both countries and periods. This result is in line with a sizable literature that finds parental background to be a crucial circumstance behind IOP in some developing as well as developed countries (Roemer & Trannoy 2016). In particular, Ferreira & Gignoux (2011) and Singh (2012) reach a similar result applying the classical parametric approach to Brazil and India, respectively. Moreover, the opportunity trees generate more ramifications of types for workers with lower levels of family education, indicating more intense labor market stratification for workers coming from illiterate or lesser educated families. This result, generated by the use of machine learning of opportunity trees, should be taken into account in the design of public policies to reduce the IOP.

The trees further suggest that skin color, in Brazil, and caste, in India, still matter in the twenty-first century. Our findings support a substantial body of literature on discrimination in the labor market between social groups in both countries (Arcand &



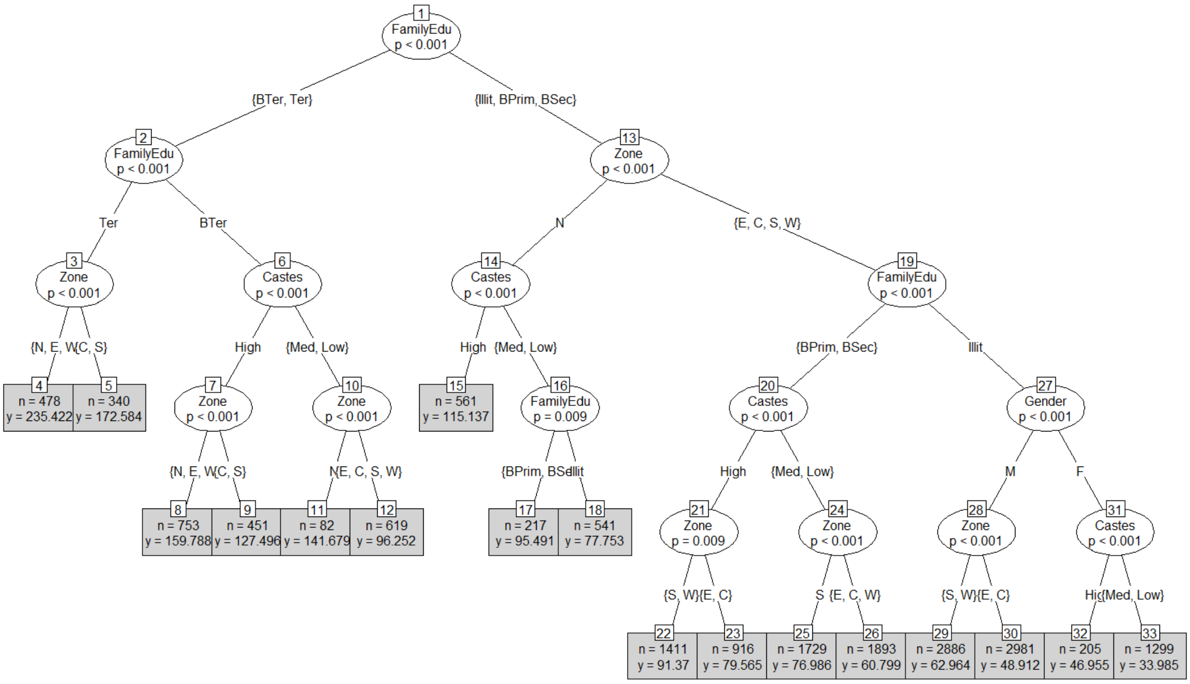
(a) 1996



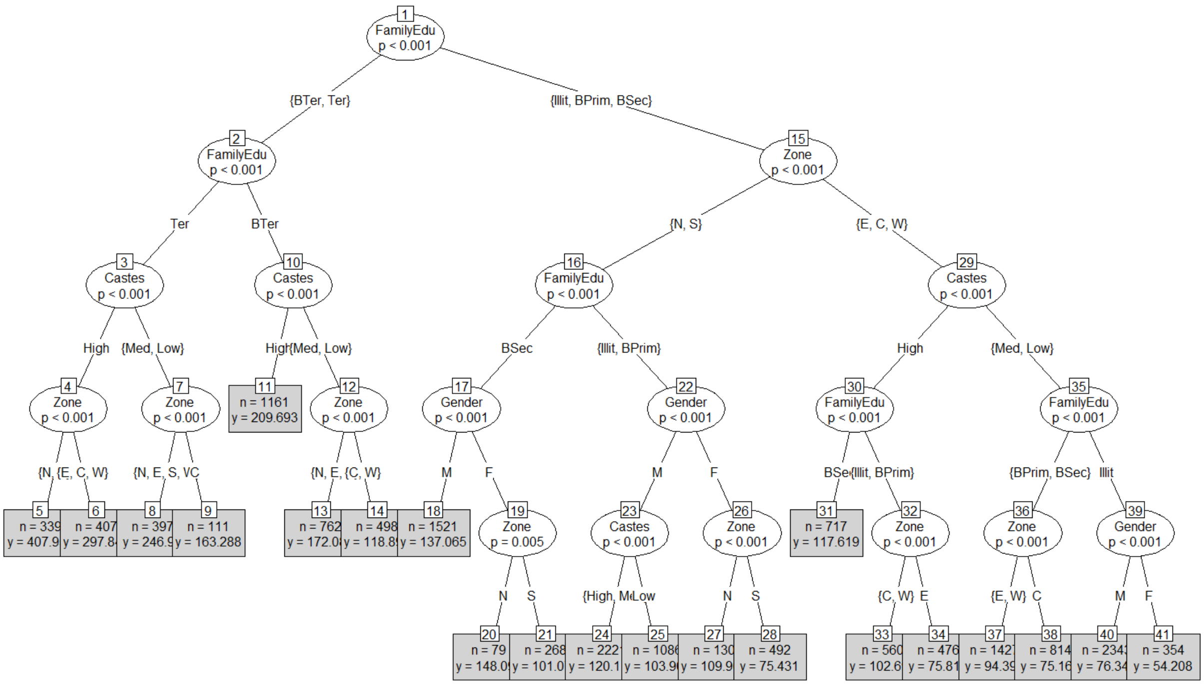
(b) 2014

Figure 2: Opportunity tree in Brazil (1996-2014)[a]

[a]'n' and 'y' denote number of individuals and earnings, respectively. Wage is reported in Brazilian Real (1BRL ≡ 0.19USD). Abbreviations: FamilyEdu - Family education, Ill - illiterate, BPrim/BSec/BTer - below primary, secondary, tertiary, respectively, Ter - tertiary, Zone - region, N - North, NE - North-East, S - South, SE - South-East, MW - Mid-West; Sex - Gender, M - Male, F - Female



(a) 1999



(b) 2012

Figure 3: Opportunity tree in India (1999-2012)[a]

[a]'n' and 'y' denote number of individuals and earnings, respectively. Wage is reported in Indian Rupee (1INR $\equiv$ 0.01USD). Abbreviations used: FamilyEdu - Family education, Ill - illiterate, BPrim/BSec/BTer - below primary, secondary, tertiary, respectively, Ter - tertiary, Zone - region, N - North, C - Central, E - East, S - South, W - West, M - Male, F - Female

D'hombres 2004, Deshpande 2011). We find that high castes in India and whites in Brazil continue to get a wage premium. Quite interestingly, in Brazil race and gender tend appear in successive nodes. In 1996, when race is chosen as a splitting variable, gender is typically the next circumstance variable to divide (Figure 2a). The only exception are individuals from the highest educational background, for whom gender appears to be more important than race in determining IOP. The situation is reversed in 2014, when gender becomes more important that race as a splitting variable (Figure 2b). For individuals from more educated backgrounds, those two variables are no longer significant circumstances to form types. Despite the changes over time, a common feature is the presence of black women in the least privileged type. These findings validate the recorded discrimination and struggle faced by Afro-Brazilian women (Lovell 2000).

The occurrence of caste and gender in subsequent nodes is far less prevalent in India. In 1999, caste is an important circumstance for all individuals, except for those from families with tertiary education. Gender appears only once as a splitting variable, for those from illiterate family backgrounds at the bottom of the wage distribution (Figure 3a). Interestingly, in 2012 gender appears as a relevant circumstance in all formed types, except for individuals from the most educated families (Figure 3b). Overall, there are two common features to the Brazilian and Indian labor markets: (i) the increased importance of gender in IOP over time; and (ii) the double disadvantage of women belonging to the less advantaged social groups of race and caste, which persists throughout the time span considered.

Indeed, we observe that female workers are at the bottom of the labor market hierarchy in both countries: women are at the right-most terminal nodes of all the trees depicted in Figures 2 and 3. Nevertheless, the participation rate of women in the labor market is very different in Brazil and India. In Brazil, they represented 42% of the labor market in 1996 and 50% in 2014 (see Table 1). This increase in participation does not change the fact that Brazilian brown and black women coming from illiterate families have the least earnings opportunities. A possible explanation could be the higher absorption of Brazilian women in non-registered employment or informal sector in general (Soares 2004). Gender is an important splitting variable in the Brazilian opportunity trees, which could be reflecting the existing scenario of pervasive involvement of women in the labor market. This is not the case for India (Klasen et al. 2021).

Regarding India, it presents one of the lowest share of female labor force in the world (Das et al. 2015). Despite the sporadic success of NREGA during this period, specially among the poor rural women (Khera & Nayak 2009), the 2012 opportunity tree does not indicate an improvement of the relative situation of Indian women with worst circumstances. The agricultural sector, which absorbs the highest share of unskilled female labor, is shrinking in India (Mehrotra & Parida 2017), and this decline is not accompanied by a proportional increase in non-farm employment for females (Ranjan 2009). At the same time, decreased poverty and improved household earnings, along with inadequate female-friendly infrastructure, lead to the withdrawal of skilled women from the labor market (Klasen & Pieters 2015).

Regional disparity in the opportunity trees is entangled with other circumstances in both Brazil and India. Differences in earning opportunity due to regions were visible in Brazil exclusively among those with unfavorable family backgrounds during 1996. This is not the case in 2014, when region is selected as a splitting variable even among the more privileged, and its importance decreases among the less privileged. This could spur from the enlargement of the *Bolsa Família*, a conditional cash transfer program to poor families, that is shown to have positive impact on some of the most underdeveloped regions of North and North-East Brazil (Hall 2006). Although no region is particularly conducive to better opportunities in India, the North seems to outperform others more frequently. It contains some of the richest states of India including the National Capital Territory of Delhi and embodies a higher share of regular salaried workers. Still, our results can not isolate any specific regional advantage, or lack thereof, for any of the countries and find it to be less conspicuous as compared to the other circumstances considered.

Finally, it is interesting to note that the Brazilian opportunity tree becomes less stratified in 2014 compared to 1996, as can be seen by the reduction of the number of types. India, however, evolves towards a more complex stratification to explain IOP, with more types being identified in 2012 compared to 1999.

5.2 IOP estimates

Table 3 presents (i) the total inequality of earnings, measured by MLD; (ii) the tree-based IOP estimates taking into account the inequality between "types" identified as the most relevant by the opportunity trees; and (iii) the IOP between an exhaustive partitioning of the circumstances into "types", resulting in a total of 150 "types".[9]

Inequality of earnings is considerably higher in Brazil compared to India in the late 1990's. We observe "contrasting trajectories" of inequality between the two countries similar to the findings of Barbosa et al. (2017): the total wage earnings inequality falls sharply in Brazil during this time span, while it remains virtually unchanged in India. In the 2010's, Brazilian earnings inequality falls behind that of India.

The picture is very different regarding IOP. First, India presents higher IOP compared to Brazil. Second, IOP in wage earnings shows some improvement in the Indian labor

[9]With 3 categories of caste/race, 2 categories of gender, and 5 categories of region and of father's education, we have a total of $3 \times 2 \times 5 \times 5 = 150$ "types".

market over time, with a slight increase for Brazil. Nevertheless, IOP remains higher in India than in Brazil in later period. This pattern also holds true for the non-parametric approach. Notice that the non-parametric method, as elaborated in section 3, operates by exhaustive partitioning of all the circumstances and thereby generating the maximum possible 'types'. Whereas the algorithm of regression tree identifies 'types' based on the most statistically significant combinations of the available set of circumstances. IOP in either approach is estimated as the measure of inequality between the 'types'. However, by construction, there will be more 'types' in the non-parametric method, since it fully spans the set of circumstances to form 'types'. Therefore the non-parametric IOP estimates are always higher than tree-based IOP estimates, as found by Ferreira & Gignoux (2011), Brunori et al. (2023) and as can be seen from our results too.

Table 3: Wage IOP measures

	Brazil		**India**	
	1996	2014	1999	2012
Inequality of earnings	0.4257	0.2924	0.3173	0.3168
IOP (tree-based method)	27.1%	27.7%	32.8%	30.5%
IOP (non-parametric method)	28.0%	31.1%	34.6%	32.3%

Notes: Non-self employed workers aged 18-45. Sources: PNAD, Brazil, and NSS, India. The non-parametric method estimates IOP considering all possible permutations to partition the sample into 'types', whereas the tree-based method partition the sample space according to the most important interactions. IOP is reported in relative terms, that shows the share of absolute IOP in total inequality.

Our results indicate that, in spite of the drastic fall in labor market inequality, Brazil is far from being inclusive, with an IOP estimate around 27%. Hence, despite efforts to promote social inclusion, the legacy of slavery and structural discrimination continue to perpetuate social hierarchy and marginalization in Brazil. India, on the other hand, always depicts a higher IOP than Brazil. Unalterable circumstances explain 32.8% of the wage inequality among the salaried workers in India in 1999, with a slight improvement to 30.5% in 2012. Despite the economic growth over this period, India continues to grapple with entrenched social inequalities rooted in caste and class divisions.

Our results unveil the persistence of IOP in Brazil and India, indicating a difficulty of fighting discrimination and segregation in the labor market. Both countries are still struggling to overcome the obstacles in the way of obtaining greater social mobility and equality. Policies that seek to lessen inequality generally fail to recognize the degree of persistence of the social hierarchy.

Robustness check It is important to note that there are large variations in sample size for the Brazilian cohorts due to differences in the questionnaire on family backgrounds in the two waves of PNAD. The regression tree methodology is susceptible to differences in sample size. Brunori & Neidhöfer (2020) shows that changes in the sample size could result in different opportunity trees, as well as differences in the point estimates of IOP. Our results are thus subject to this caveat. To check the robustness of our results, we estimate the regression tree in 1996, that originally has a sample size of 34,512, using a repeated random draw of 8,725 observations, which is the sample size of 2014 (see Table 1). We estimate the IOP over 100 such random draws in Brazil for 1996, and the average IOP of our estimations is 26.8%, which indeed is very close to the full-sample estimate of 27.1% as reported in Table 3. We perform an additional robustness check by randomly dropping 10% of the observations and we verify that the final IOP estimates do not change by more than 0.5%.[10]

6 Conclusion

Brazil and India have high inequalities of opportunity, which can be traced back to their historical heritage: slavery in Brazil and the caste system in India. The analysis of the structure and evolution of the inequality of opportunity is important to understand the patterns of inequality and development in these countries. There is a large body of literature studying income inequality, but few studies look at the issue of inequality of opportunity in the labor market. In this article, we contribute to this literature by investigating the inequality of opportunities in the labor market in these two countries using the regression tree methodology, which allows for comparisons between countries.

Our main results of the regression three analysis are the following. Similar to the extant literature on IOP for developed as well as developing countries, we find the family education to be the most important circumstance factor for both Brazil and India, and this result is stable over time. We also find a clear earnings advantage for the upper castes in India and for "whites" in Brazil, which corroborates the literature on labor market discrimination in these countries.

The more interesting feature of the opportunity tree, however, stems from its growth into branches and nodes, revealing the various linkages between family education and caste/race, gender, and region. We find the interconnections between gender and race to be stronger in Brazil compared to those between gender and caste in India, with two common features: (i) an increased importance of gender in IOP over time; and (ii) a double disadvantage of women belonging to the less advantaged social groups of race and caste, which persists throughout the time span considered. Another notable result

[10] The IOP estimates after randomly dropping 10% of the observations are the following: 27.4% in 1996 and 28.7% in 2014 for Brazil; and 32.9% in 1999 and 30.7% in 2012 for India

is that the Brazilian opportunity tree becomes less stratified in 2014 compared to 1996, as indicated by the reduction in the number of types. Conversely, India shows a trend towards increased stratification over time to explain IOP.

Based on the "types" identified by the opportunity trees, we compute the IOP in the labor market. We find a higher IOP in India compared to Brazil, with different trajectories of inequalities over time. Earnings inequality falls sharply in Brazil from 0.42 in the late 1990s to 0.29 in 2014, while in India it remains constant at 0.32. Regarding IOP, it improved marginally in India, falling from 33% to 31% during 1999-2012, while it remained around 27% in Brazil over the time span of 1996-2014. Thus, we find a "contrasting trajectory" in IOP between India and Brazil, on the opposite direction of the "contrasting trajectory" in terms of income inequality, albeit the contrast in IOP is not as sharp as it is in terms of income inequality.

Finally it is important to mention that the present study has no pretension of delivering a causal inference as to the effectiveness of the several social policies undertaken by both the countries during the study period. Our objective is rather modest, that to analyze labor market IOP in the course of this period and interpret the results with respect to the ongoing socio-political changes. Both countries implemented their own public policies to reduce inequalities and promote social inclusion, nevertheless our results have not captured a significant improvement of labor market opportunities across the social strata, in neither country. Our analysis overall suggests that the design of public policies aimed at decreasing inequality in both Brazil and India, should also persevere on reducing IOP in their respective labor markets.

References

Aiyar, S. & Ebeke, C. (2020), 'Inequality of opportunity, inequality of income and economic growth', *World Development* **136**, 105115.

Andreoli, F., Havnes, T. & Lefranc, A. (2019), 'Robust inequality of opportunity comparisons: Theory and application to early-childhood policy evaluation', *The Review of Economics and Statistics* **101**(2), 355–369.

Arcand, J.-L. & D'hombres, B. (2004), 'Racial discrimination in the brazilian labour market: Wage, employment and segregation effects', *Journal of International Development: The Journal of the Development Studies Association* **16**(8), 1053–1066.

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.

Asadullah, M. N. & Yalonetzky, G. (2012), 'Inequality of educational opportunity in India: Changes over time and across states', *World Development* **40**(6), 1151–1163.

Averbug, A. (2002), 'The brazilian economy in 1994–1999: from the real plan to inflation targets', *World Economy* **25**(7), 925–944.

Barbosa, A. d. F., Cacciamali, M. C. & Rodgers, G. (2017), *Growth and inequality: The contrasting trajectories of India and Brazil*, Cambridge University Press.

Bourguignon, F., Ferreira, F. H. & Menéndez, M. (2007), 'Inequality of opportunity in Brazil ', *Review of income and wealth* **53**(4), 585–618.

Brunori, P., Hufe, P. & Mahler, D. (2023), 'The roots of inequality: estimating inequality of opportunity from regression trees and forests', *Scandinavian Journal of Economics* **125**(4), 900–932.

Brunori, P. & Neidhöfer, G. (2020), 'The Evolution of Inequality of Opportunity in Germany: A Machine Learning Approach ', *Review of Income and Wealth* .

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

Bucciferro, J. R. & Ferreira de Souza, P. H. (2020), The evolution of regional income inequality in brazil, 1872–2015, *in* 'Time and Space', Springer, pp. 131–156.

Cacciamali, M. C. & Rodgers, G. (2015), 'Wage inequality in brazil and india: A quantitative comparative analysis', *Institute for Human Development Working Paper WP03/2015, New Delhi* .

Checchi, D. & Peragine, V. (2010), 'Inequality of opportunity in Italy ', *Journal of economic inequality* **8**, 429–450.

Cohen, G. A. (1989), 'On the currency of egalitarian justice', *Ethics* **99**(4), 906–944.

Croll, E. J. (2006), 'The intergenerational contract in the changing asian family', *Oxford Development Studies* **34**(4), 473–491.

Das, M. S., Jain-Chandra, M. S., Kochhar, M. K. & Kumar, N. (2015), *Women workers in India: why so few among so many?*, International Monetary Fund.

Deshpande, A. (2007), 'Overlapping identities under liberalization: Gender and caste in india', *Economic Development and Cultural Change* **55**(4), 735–760.

Deshpande, A. (2011), *The grammar of caste: Economic discrimination in contemporary India*, Oxford University Press.

Ferreira, F. H. & Gignoux, J. (2011), 'The measurement of inequality of opportunity: theory and an application to Latin America ', *The review of income and wealth* **57**(4).

Firpo, S. P. & Pieri, R. G. d. (2018), 'The labor market in brazil, 2001-2015', *IZA World of Labor* (441).

Firpo, S. & Portella, A. (2019), 'Decline in wage inequality in brazil: a survey', *World Bank Policy Research Working Paper* (9096).

Firpo, S. & Reis, M. C. (2007), 'O salário mínimo ea queda recente da desigualdade no brasil', *Desigualdade de renda no Brasil: uma análise da queda recente* **2**, 499–506.

Gerard, F., Lagos, L., Severnini, E. & Card, D. (2021), 'Assortative matching or exclusionary hiring? the impact of employment and pay policies on racial wage differences in brazil', *American Economic Review* **111**(10), 3418–57.

Hall, A. (2006), 'From fome zero to bolsa família: social policies and poverty alleviation under lula', *Journal of Latin American Studies* **38**(4), 689–709.

Hnatkovska, V., Lahiri, A. & Paul, S. B. (2013), 'Breaking the caste barrier intergenerational mobility in india', *Journal of Human Resources* **48**(2), 435–473.

Imbert, C. & Papp, J. (2015), 'Labor market effects of social programs: Evidence from india's employment guarantee', *American Economic Journal: Applied Economics* **7**(2), 233–63.

Khera, R. & Nayak, N. (2009), 'Women workers and perceptions of the national rural employment guarantee act', *Economic and political weekly* pp. 49–57.

Khurana, S. & Mahajan, K. (2020), Evolution of wage inequality in india (1983-2017): The role of occupational task content, Technical report, WIDER Working Paper.

Klasen, S., Le, T. T. N., Pieters, J. & Santos Silva, M. (2021), 'What drives female labour force participation? comparable micro-level evidence from eight developing and emerging economies', *The Journal of Development Studies* **57**(3), 417–442.

Klasen, S. & Pieters, J. (2015), 'What explains the stagnation of female labor force participation in urban india?', *The World Bank Economic Review* **29**(3), 449–478.

Lefranc, A., Pistolesi, N. & Trannoy, A. (2009), ‘Equality of opportunity and luck: definitions and testable conditions, with an application to income in France (1979-2000)’, *Journal of public economics* **93**, 1189–1207.

Lovell, P. A. (2000), ‘Gender, race, and the struggle for social justice in brazil’, *Latin American Perspectives* **27**(6), 85–102.

Madheswaran, S. & Singhari, S. (2016), ‘Social exclusion and caste discrimination in public and private sectors in india: A decomposition analysis’, *The Indian Journal of Labour Economics* **59**(2), 175–201.

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

Mehrotra, S. & Parida, J. K. (2017), ‘Why is the labour force participation of women declining in india?’, *World Development* **98**, 360–380.

Mehrotra, S. & Sinha, S. (2017), ‘Explaining falling female employment during a high growth period’, *Economic & Political Weekly* **52**(39), 54–62.

Narayanan, A. (2015), ‘Informal employment in india: Voluntary choice or a result of labor market segmentation?’, *Indian Journal of Labour Economics* **58**(1), 119–167.

Ranjan, S. (2009), ‘Growth of rural non-farm employment in uttar pradesh: Reflections from recent data’, *Economic and Political Weekly* pp. 63–70.

Reddy, A. B. (2015), ‘Changes in intergenerational occupational mobility in india: Evidence from national sample surveys, 1983–2012’, *World Development* **76**, 329–343.

Rodgers, G. (2016), ‘Inequality and the labour market: comparing india and brazil’, *The Indian Journal of Labour Economics* **59**(1), 39–55.

Roemer, J. (1998), *Equality of Opportunity*, Harvard University Press, Cambridge, MA.

Roemer, J. E. & Trannoy, A. (2016), ‘Equality of opportunity: Theory and measurement’, *Journal of Economic Literature* **54**(4), 1288–1332.

Rosenzweig, M. & Zhang, J. (2014), ‘Co-residence, life-cycle savings and inter-generational support in urban china’.

Ruggles, S. & Heggeness, M. (2008), ‘Intergenerational coresidence in developing countries’, *Population and development review* **34**(2), 253–281.

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.

Soares, F. V. (2004), *Some stylized facts of the informal sector in Brazil in the 1980's and 1990's*, number 1020, Ipea.

Telles, E. E. (2002), 'Racial ambiguity among the brazilian population', *Ethnic and racial studies* **25**(3), 415–441.

World Development Report, W. B. (2006), *World Development Report 2006: Equity and Development*, The World Bank publications.