IHD WORKING PAPER SERIES

WP 07/2022

Decomposition of Inequality of Opportunity in India: An Application of Data-Driven ML Approach

Balwant Singh Mehta
Siddharth Dhote
Ravi Srivastava

INSTITUTE FOR HUMAN DEVELOPMENT
www.ihdindia.org

INEQUALITREES

WP 07/2022

Decomposition of Inequality of Opportunity in India: An Application of Data-Driven ML Approach

Balwant Singh Mehta, Siddharth Dhote, and Ravi Srivastava

2022

Suggested citation

Mehta B.S,. Dhote S,. & Srivastava, R. (2022). Decomposition of Inequality of Opportunity in India: An Application of Data-Driven ML Approach. Working Paper Series, N.X.

Published by:

INSTITUTE FOR HUMAN DEVELOPMENT
256, 2nd Floor, Okhla Industrial Estate, Phase-III
New Delhi - 110020
Tel: +91 11 41064676, +91 9871177540
E-mail: mail@ihdindia.org
Website: www.ihdindia.org

ISBN: 978-81-88315-83-3

https://inequalitrees.eu/

A Novel Look at Socio-Economic Inequalities Using Machine Learning Techniques and Integrated Data Sources

PROJECT'S DESCRIPTION

The INEQUALITREES project aims to investigate the levels and main drivers of two key manifestations of socio-economic inequality across the globe: poverty and inequality of opportunity (IOp). This project adopts a multidimensional, interdisciplinary and cross-national approach, by analysing IOp and poverty in three key individual outcomes (education, income and health) in four countries (Bolivia, Germany, India, Italy), and integrating contributions from economics, sociology, geography and computer science. A key innovative feature of this project is the application of cutting-edge machine learning techniques to integrate and analyze large scale datasets from various sources, including national and international surveys, administrative and register data, as well as innovative data extracted from satellite images.

The INEQUALITREES project aims to investigate the levels and main drivers of two key manifestations of socio-economic inequality across the globe: poverty and inequality of opportunity (IOp). This project adopts a multidimensional, interdisciplinary and cross-national approach, by analysing IOp and poverty in three key individual outcomes (education, income and health) in four countries (Bolivia, Germany, India, Italy), and integrating contributions from economics, sociology, geography and computer science. A key innovative feature of this project is the application of cutting-edge machine learning techniques to integrate and analyze large scale datasets from various sources, including national and international surveys, administrative and register data, as well as innovative data extracted from satellite images.

For details see https://inequalitrees.eu/

ACKNOWLEDGEMENTS

The authors are grateful to Prof. S.Madeshwaran, Prof. Arup Maitra, Dr. Pedro Salas Rojo, and the conference audience at the 63rd ISLE conference for their valuable comments and suggestions on the draft of this paper.

ABOUT THE AUTHORS

BALWANT SINGH MEHTA is 'Senior Fellow' at the Institute for Human Development (IHD). He has authored 10 books and over 50 articles in reviewed national and international journals on employment, inequality, poverty, education, child welfare and information and communications technology for development- and human development-related issues.

SIDDHARTH DHOTE is Senior Research Associate at the Institute for Human Development and has completed his master in Development Studies from the International Institute for Social Studies in the Hague. His main research interests focus on social justice, poverty, inequality, and urbanization.

RAVI SRIVASTAVA is Director, Centre for Employment Studies, Institute for Human Development, Delhi. He was earlier Professor of Economics and Chairperson, Centre for the Study of Regional Development, Jawaharlal Nehru University, and full-time Member (2006-09) of the National Commission for Enterprises in the Unorganised Sector (NCEUS), in the rank of Secretary, Government of India. He has published 6 books, 5 monographs, and more than 110 papers in the areas of labour economics, agriculture, rural development, and poverty.

CONTENTS

Abstract

This paper introduces a novel measure of inequality of opportunity (IOp) in India by comparing both ex-ante and ex-post results, which aligns with Roemer's (1998) theory. The study utilizes data-driven machine learning (ML) algorithms, namely conditional inference tree and conditional inference forest, to measure ex-ante IOp, and a transformation tree to estimate ex-post IOp. The findings indicate that, according to the ex-ante approach, approximately 58-61 per cent of overall income inequality can be attributed to variations in circumstances, while around 46 per cent of the overall income inequality is explained by differences in the degree of effort. The results from the tree-based analysis reveal that parents' occupation, sector (rural or urban), and geographical region, are the primary circumstances contributing to IOp, which is further confirmed by the Shapley decomposition exercise. Specifically, individuals residing in rural areas in the eastern and central parts of the country, whose parents are employed in low-skilled and unskilled occupations, whose parents have below secondary level education or no formal education, and those who belong to marginalized social groups, exhibit significantly lower average income. Consequently, it is crucial to implement regional-level development policies that specifically target marginalized groups in order to foster a more equitable society and mitigate overall income inequality.

Keywords: Inequality of opportunity; Machine learning algorithm; Conditional inference tree; conditional inference forest; transformation tree

JEL CODES: D63, D30, D31, G51, C60

Decomposition of Inequality of Opportunity in India: An Application of Data-Driven ML Approach

Balwant Singh Mehta, Siddharth Dhote and Ravi Srivastava[1]

I. INTRODUCTION

The relationship between economic growth and income distribution has been a prominent topic of research over the past several decades. Kuznets (1955) conducted a seminal study that examined the historical patterns of income distribution and economic growth. He proposed the idea that income inequality widens as countries experience high economic growth in the initial stages of development. However, as countries develop further and reach higher levels of income, the benefits of growth become more widespread, leading to a narrowing of income inequality. This creates an inverted U-shaped relationship between growth and inequality. Kuznets's inverted-U hypothesis has sparked considerable debate in recent years. Numerous empirical studies, conducted across different countries and regions, have both supported and challenged this hypothesis. The latter studies emphasize that economic growth alone is not sufficient to reduce poverty and inequality. Rather, they emphasize the importance of distinguishing between the 'growth effect' and the 'inequality effect' in an economy (Ravallion & Chen, 2003; Bourguignon, 2004). In the current global scenario, many countries are facing this dual challenge of achieving poverty reduction and a simultaneous decline in income inequality within their economic growth processes.

According to the recently published World Inequality Report (WIR, 2022), income and wealth inequalities have experienced a widespread increase since the 1980s. The increase in inequality is attributed to various deregulation and liberalization initiatives implemented in different countries. This trend of rising inequality highlights a growing gap between the rich and the poor, with the wealthiest

1. Senior Fellow, Senior Research Associate and Director, Centre for Employment Studies, Institute for Human Development, Delhi respectively.

10 per cent of the global population capturing 52 per cent of global income, while the poorest 50 per cent of the world's population own only 8.5 per cent of it. Wealth inequalities are even more pronounced, with the poorest 50 per cent of the global population possessing a mere 2 per cent of total wealth, while the richest 10 per cent own a staggering 76 per cent. Although income inequalities between countries have decreased since the 1990s, marked by a decline in the gap between the average income of the richest 10 per cent of countries and the average incomes of the poorest 50 per cent of countries by nearly 50 times, the ratio currently stands at slightly less than 40 times. This convergence of inequalities between countries is notable; however, it is important to note that within-country inequalities have significantly increased. The gap in average incomes between the richest 10 per cent and the poorest 50 per cent of individuals within countries has nearly doubled, rising from 8.5 times to 15 times. This rise in income inequalities within countries is particularly prominent in developing nations such as India and China, and underscores the challenges faced by these countries in reducing inequality and poverty simultaneously, through their economic growth processes.

In particular, India has experienced significant economic growth over the past two decades, with an annual average of 7 per cent (Anand & Thampi, 2016; Chancel & Piketty, 2017). However, this period of growth has been accompanied by an increase in income inequality. Despite overall economic progress, the benefits of economic growth have not reached the poorest individuals, leading to India being described as one of the "most unequal countries in the world" (WIR, 2022). Research conducted by Chancel and Piketty (2019) highlights the historically high levels of inequality in India, with the richest 10 per cent of the population holding 57 per cent of the national income, compared to only 13 per cent held by the poorest 50 per cent. Household surveys also reveal a significant income disparity, with a Gini coefficient of 0.543 in 2012 (Anand & Thampi, 2016). The situation assessment survey further indicates a high Gini coefficient of 0.587 for per capita income among agricultural households in 2013.

The state of inequality report prepared by the Institute for Competitiveness reveals a substantial divergence between the earnings of the top and bottom percentiles of workers, as well as in the average earnings of workers. The cumulative annual earnings of the richest 1 per cent are nearly three times higher than those of the poorest 10 per cent. Between 2017-18 and 2019-20, the average annual earnings share of the richest 1 per cent earners increased from 6.1 per cent to 6.8 per cent, while the share of the poorest 50 per cent remained stagnant at around

22 per cent. During the same period, the income of the richest 1 per cent grew by 15 per cent, and the income of the richest 10 per cent rose by 8.1 per cent. In contrast, the income of the poorest 10 per cent declined by 1 per cent (EPW, 2022). These recent trends in income inequality reaffirm that the concepts of trickle-down theories do not hold true in India. Other studies have also highlighted the dangers associated with the uneven growth process in India. While Anand and Thampi (2016) discuss the high and rising levels of wealth inequality in neoliberal times, Chancel and Piketty (2019) foreground the unprecedented rise in the income share of the richest 1 per cent of the population.

It is argued that if income is concentrated at the top tiers of the population without a simultaneous increase in average income levels at the bottom, it leads to a rise in income inequality (Deaton & Stone, 2013). However, in the case of India, the rise in income inequality is not solely due to the skewed nature of the income distribution, but is also driven by persistent social disparities and hierarchies. The literature extensively documents the persistence of resource inequalities in terms of land ownership, income, and wealth, and the continued practice of caste-based discrimination in the Indian labor market (Tagade et al., 2018). Gender inequalities are also evident in the limited participation of women in the labor market and the disproportionate burden of unpaid work on women in India (Ghose, 2019). Further, rural-urban wage gaps and gender-based wage disparities also highlight the presence of gender discrimination in the labor market (Deshpande et al., 2018; IHD, 2014). Given these circumstances, it is argued that economic equality is necessary for the creation of a society that treats individuals as fundamentally equal, bridges the gaps between identity groups, provides greater representation for historically marginalized populations, and ensures equality of opportunity for all (Weisskopf, 2011).

In the context of fast-growing countries like India, the persistent trend of rising income inequality has led to increased research in this field in recent years, with a special focus on identifying its main causes. There is a general consensus that it is essential to provide people with equal chances for success and the pursuit of interests through fair and equal opportunities. Roemer (1998) argues that inequality of opportunity (IOp) arises from the interplay between ‘circumstances’ and the degree of ‘effort’ exerted by individuals (discussed in detail later). The study of IOp has gained increasing attention in empirical studies in recent years, especially in exploring the unfair aspects of societal inequality (Fleurbaey, 2009; Checchi & Peragine, 2010; Ferreira & Gignoux, 2011; Ferreira & Peragine, 2015; Roemer & Trannoy, 2016; Brunori et al., 2019a, 2019b; Brunori & Neidhöfer, 2020; Hothorn, T., & Zeileis, A., 2021; Salas-Rojo & Rodríguez, 2022).

However, a majority of these studies on IOp have predominantly focused on developed nations, with limited research focusing on developing countries such as India. In order to address the problem of rising inequality, India requires significant redistribution of income and wealth. The initial conditions or circumstances, such as the social group an individual belongs to, their religion, location (rural or urban), and geographical region, have divided Indian society, resulting in differential privileges for different groups (Singh, 2012; Das & Biswas, 2022). Moreover, at the individual level , India also faces a significant gender divide. Several studies have estimated IOp in India by examining consumption and earning levels as outcome variables and considering the impact of circumstances such as social group, gender, place of birth, rural-urban location, parental education, and parental occupation (Motiram, 2018; Asadullah & Yalonetzky, 2012; Chaudhary et al., 2019; Lefranc & Kundu, 2020; Das & Biswas, 2022). These studies emphasize that a significant portion of income or consumption inequality can be attributed to unequal circumstances, with parental education as an important determinant. However, many of these studies rely on statistical assumptions and model selection approaches that are biased or arbitrary. In particular, the empirical studies conducted so far have estimated IOp without explicitly considering the role of effort, which is a key aspect of Roemer's theory (Ramos & Van de gaer, 2021). Incorporating the notion of effort in the estimation of IOp in India would provide a more comprehensive understanding of the factors influencing IOp and inform policy interventions aimed at reducing inequality.

The rest of the paper is structured as follows: Section 2 presents the conceptual framework, followed by Section 3 which focuses on the measurement approaches of IOp. Section 4 describes the details of the data sources and variables used for the analysis. Section 5 discusses the descriptive statistics of the sample, presents the results, and examines the important variables that generate IOp by utilizing a conditional inference regression tree and transformation tree. Finally, Section 6 concludes the paper by highlighting key findings and providing policy remarks.

II. CONCEPTUAL FRAMEWORK

Rawls (1958, 1971) proposed that justice in an egalitarian society can be achieved through equality of opportunity, often depicted through metaphors such as 'levelling the playing field' or 'equality at the starting gate'. In Rawls' vision of a just society, individuals are provided with fair and equal chances to pursue their interests, with a particular emphasis on 'primary goods'. This ethical justification for equality of opportunity sparked a discussion that reshaped the understanding of equality and contributed to several philosophical debates (Arneson, 1989; Cohen, 1989; Dworkin, 1981a, b; Sen, 1980). Following Rawls, Roemer (1993, 1998) and Fleurbaey (1995, 2008) developed a systematic approach to measure inequality of opportunity (IOp). They identified two distinct sets of factors influencing individuals' outcomes, namely effort and circumstances. Effort refers to factors within individuals' control, such as the number of hours dedicated to work or study, the quality of work provided, and occupational choices. Circumstances, on the other hand, encompass factors beyond individuals' control, such as family background, socioeconomic status, ethnicity, gender, and age (Roemer & Trannoy, 2016). Mathematically, this can be represented as follows: In a population of individuals from 1…..N, individual i achieves an outcome of interest, denoted as , as a result of the interplay between circumstances (and effort (.

$$y_i = \mathrm{g}(C_i, e_i), \forall i = 1, \ldots N \tag{1}$$

In equation (1), the function g captures how circumstances and effort jointly determine the outcome for each individual.

Roemer proposes a framework for achieving equality of opportunity by partitioning the population into categories based on circumstances (types) and groups with similar effort levels (tranches). The types are constructed in a way that individuals belonging to a specific category share the same circumstances. Similarly, individuals within a tranche have the same effort level. When individuals exert effort, those belonging to the same type have equal ability to convert resources into outcomes. Therefore, an equal opportunity policy requires disregarding within-type variability in outcomes, which is attributable to individual effort, and addressing any between-type inequality (Roemer, 2002).

In the literature on inequality of opportunity (IOp), two ethical principles are commonly discussed: the 'compensation' and the 'reward'. The compensation principle suggests that inequalities resulting from circumstances should be compensated, while the reward principle incorporates the notion of individual responsibility, advocating for higher outcomes for additional effort (Plassot et al., 2022). The reward principle does not oppose inequality between individuals with different effort levels, whereas the compensation principle focuses on unfair inequalities and can be approached through either the ex-ante or ex-post method (Plassot et al., 2022; Fleurbaey & Peragine, 2013). The ex-ante approach focuses on inequalities between individuals with different circumstances or types, whereas the ex-post approach examines individuals with the same effort levels or tranches. The difference in approaches stems from divergent views on the nature of the effort variable (Fleurbaey & Peragine, 2013; Ramos and Van de gaer, 2016).

In the ex-post approach, equality of opportunity is achieved when individuals exerting the same effort obtain the same outcome regardless of their type. Measuring the extent of violation of this principle involves comparing the outcomes of individuals belonging to different types but exerting the same effort (Brunori & Neidhofer, 2021). Since effort is often unobservable, Roemer (2002) proposed a method based on twoassumptions: firstly, individuals are assigned to types, and secondly, the outcome must monotonically increase with effort. In simple terms, greater effort within each type should result in a higher outcome, expressed mathematically as:

$$y^k(e_i) >= y^k((e_j) <=> e_i^k >= e_j^k,\ \forall\ k=1\ldots..K;\ \forall\ e_i, e_j \in R \quad (2)$$

Where $(y^k(e_i))$ is the outcome of an individual in type k with degree of effort i, represented by e_i, and $y^k(e_j)$ is the outcome of an individual in type k with degree of effort j, represented by e_j and K is the total number of types.

It is assumed that the distribution of effort is a characteristic of the type, meaning that when comparing effort levels across individuals in different types, adjustments should be made to account for the fact that these effort levels are drawn from different distributions, for which individuals should not be held responsible. Roemer distinguishes between the 'level of effort' and the 'degree of effort' exerted by an individual. The 'degree of effort' is a morally relevant variable of effort and is identified as the quantile of the effort distribution for the specific type to which an individual belongs. The underlying assumption is that all circumstances have been identified and are exogenous to each individual. If individuals belonging to different types face different incentives and constraints in exerting effort, this is considered a characteristic of the type and thus falls under circumstances beyond individual control (Brunori & Neidhofer, 2021). For instance, a student with highly educated parents may have greater ease in dedicating long hours to studying, while a student with less educated parents may face more challenges in studying.

The distribution of effort within type k and quantiles $\pi \in [0,1]$ is denoted as $G^K(e)$. In cases where effort is unobservable but the outcome monotonically increases with effort, Roemer suggests identifying the 'degree of effort' exerted by an individual with their quantile position in the type-specific outcome distribution (y), represented as $y^k G^k(e)=y^k(\pi)$. This definition of effort accounts for differences in the absolute level of effort exerted, which Roemer considers to be influenced by circumstances beyond individuals' control. It also allows for the comparison of effort levels among individuals in different types. The requirement of the same outcome (y) for individuals exerting the same effort in terms of type-specific outcome distributions is represented mathematically as:

$$y^k(\pi)= y^l(\pi) \Leftrightarrow F^k(\text{y})= F^l(\text{y}), \forall \text{ k, l=1,...K, and } \pi \in [0,1] \quad (3)$$

Where $F^k(y)$ represents the type-specific cumulative distribution of outcome in type k.

A measure of IOp quantifies the extent to which the principle of equal opportunity is violated. Checchi and Peragine (2010), and Ferreira and Gignoux (2011) propose an ex-post measure of IOp that evaluates inequality in a standardized distribution. This measure takes into account the variability of the outcome distribution among individuals exerting the same effort. When equation (3) is satisfied, indicating that individuals with the same effort achieve the same outcome, the measure takes a value of zero. As the difference in outcomes among individuals with similar degrees of effort increases, the measure of IOp increases accordingly. The standardized distribution, denoted as $\bar{Y}_{EP}$, is obtained by replacing individual outcomes with standardized values, denoted as:

$$\acute{y}_i^k(\pi)= y_i^k(\pi)\frac{\mu}{\mu^{\pi}}, \forall \text{ i=1...N; k=1....K; } \forall \pi \in [0,1] \quad (4)$$

Where $y_i^k(\pi)$ is the outcome of individual i, belonging to type k, at quantile π, of the type-specific effort distribution; where denotes the average outcome of individuals at quantile π across all types, and is the population mean outcome. In the standardized distribution, the average value for individuals in all the quantiles is the same, which eliminates the between-quantile inequality, while preserving within-quantile relative distance in outcomes. Hence, the ex-post measure of IOp, denoted as IOp_{EP}, quantifies inequality in the standardized distribution, denoted as:

$$\text{Ex-post IOp}_{EP} = I(\bar{Y}_{EP}) \quad (5)$$

Where I is an inequality measure that satisfies standard properties, including scale invariance. It should be noted that ex-post measures of IOp are not commonly used in empirical analyzes, with most studies focusing on ex-ante measures of IOp. The ex-ante IOp measure proposed by Van de gaer (1993) is also known as the 'weak equality of opportunity' criterion. It allows for some inequality within groups of individuals exerting the same effort but requires that mean advantage levels are the same across types (Ferreira & Gignoux, 2011). This approach considers the type-specific outcome distribution as the opportunity set for individuals in each type. The value of the opportunity set for each type is determined by the mean outcome of that type. Therefore, in this case, IOp is simply the inequality between types, and the counterfactual distribution, denoted as $\bar{Y}_{EA}$, is obtained by replacing individual outcomes with the mean outcome, denoted as:

$$\acute{y}_i^k(\pi) = \mu^k,\ \forall\ i=1\ldots.N;\ \forall\ k=1\ldots..K;\ \ \forall\ i,\ \pi \in [0,1] \quad (6)$$

Where μ^k is the mean outcome of type k.

$$\text{Ex-ante IOp}_{EA} = I(\acute{Y}_{EA}) \quad (7)$$

These measures provide different perspectives on IOp, with the ex-post measure focusing on within-group inequality among individuals exerting the same effort, and the ex-ante measure considering inequality between types based on mean advantage levels.

III. MEASUREMENT APPROACHES FOR ESTIMATING IOP: DATA-DRIVEN MACHINE LEARNING TECHNIQUES

The conventional approaches of measuring IOp suffer from many limitations such as researcher's discretion in selecting circumstances or effort variables. This arbitrary selection may result in the exclusion of relevant variables or the inclusion of too many variables. Omitting important variables restricts the explanatory power of the model and leads to downward biased estimates, while including excessive variables can yield upward biased estimates (Brunori et al., 2019a; Hufe et al., 2017; Ferreira & Gignoux, 2011). To overcome these limitations, machine learning (ML) algorithms offer a promising solution for measuring IOp. These algorithms follow the principle of 'let the data talk', aiming to make data-driven decisions, and minimize the risk of arbitrary and ad-hoc selections. Additionally, this approach adopts the standardized approach to balance upward and downward biases (Salas-Rojo & Rodríguez, 2022; Hothorn et al., 2006; Hothorn & Zeileis, 2021; Brunori et al., 2019a; Brunori & Neidhöfer, 2020).

In this paper, the ML algorithm called conditional inference regression tree and conditional inference forest have been used to estimate the ex-ante IOp, while the ML algorithm known as transformation tree has been used to estimate the ex-post IOp. These ML techniques are used to identify the types and tranches (or degrees of effort) to calculate IOp by using the ex-ante and ex-post approaches. It is important to note that both the ex-ante and the ex-post approach have the same first step, i.e., the identification of types by dividing the sample into subgroups that share identical circumstances (Brunori et al., 2023, p.8).

3.1 Identification of Types: Conditional Inference Tree and Conditional Inference Forest

The identification of types based on individual circumstances is an important component of the empirical analyzes of IOp (Ferreira & Gignoux, 2011; Brunori et al., 2019). For this, data-driven ML algorithms called conditional inference trees and conditional inference forests are applied to aid in the identification of types (Brunori & Neidhöfer, 2021). These techniques have been widely used in some recent empirical studies (Brunori et al., 2018, 2019a, 2019b; Brunori & Neidhöfer, 2020; Lefranc & Kundu, 2020).

Conditional inference trees offer a visually intuitive depiction of the structure of opportunities by recursively splitting the complete range of circumstances, and allowing for the identification of subgroups with similar circumstances. On the other hand, conditional inference forests, which are a variant of conditional inference trees, created through bootstrapping, incorporate data-specific characteristics by aggregating the trees into a forest, thereby enhancing the reliability of IOp estimates. Further, a notable feature of conditional inference forests is their ability to determine the relative importance of factors beyond the structure of the trees.

The algorithm for conditional inference trees involves two stages: (i) selection of the initial splitting circumstance, and (ii) growth of the opportunity tree. In the first stage, a hypothesis test, typically a t-test, is performed before each split to assess whether equal opportunities exist within a given sample or subsample. If the algorithm determines that a split is not warranted, it fails to reject the null hypothesis of equal opportunity. This occurs when the p-value associated with the circumstance being considered (C^*) is greater than a pre-determined significance level (α). Conversely, if a split is justified, the selected circumstance C^* becomes the splitting variable, and the algorithm proceeds to grow the opportunity tree. This iterative process generates a hierarchical arrangement of circumstances, reflecting the significant associations with the outcome. Only those circumstances which demonstrate statistically meaningful relationships with the outcome are considered, particularly when a large set of circumstances is present. The tree identifies the interactions that contribute to the variance in the outcome. The terminal nodes, located at the bottom of the tree, represent the average predicted outcome for individuals assigned to each specific type or group. The points along the tree where the predictor space is divided are referred to as internal nodes, and each split generates new branches within the tree. The final prediction made by this algorithm is the average outcome of each identified group or type.

The conditional inference forest algorithm generates a set number of conditional inference trees and then combines their results by taking an average. The repetitive extraction of subsamples guarantees the independence of each tree, resulting in different estimates for each subsample. Each tree within the conditional inference forest follows the same two-step structure described earlier for the conditional inference tree.

3.2 Identification of Tranches (Degrees of Effort): Transformation Trees

The conditional inference trees and conditional forests primarily focus on estimating the mean differences between types to compute IOp. However, these approaches overlook higher moments of the within-type distribution and the importance of effort ranks in generating types. To address these limitations, a model based on the type-specific outcome distribution, known as the Transformation Tree (TrT) has been used in this paper. This approach is an ex-post method and employs an algorithm that estimates the outcome distribution within each type using coefficients of Bernstein polynomials[2]. The TrT model predicts the shape of the outcome distribution by partitioning the regressors' space. It identifies the heterogeneity among the distributions defining each type. This process involves estimating the unconditional distribution and searching for binary splitting variables. Splitting is allowed if the resulting conditional distributions exhibit sufficient dissimilarity in shape. The shape of the distributions is an approximation with a linear combination of Bernstein basis polynomials that, for an order m and some positive continuous variable y € [a, z], is defined as a set of polynomials, denoted as:

$$B_m(\text{y, a, z}) = \sum_{i=0}^{m} \beta_i \, b_{j,m}(\text{y a, z}) \tag{8}$$

For a distribution approximated with a Bernstein polynomial of order m, we get m+1 parameters defining the shape of the objective distribution. For each type, the TrT algorithm follows these steps: First, a confidence level (α) and a polynomial

2. It is widely used in computer graphics to model smooth curves (Farouki, 2012). It outperforms competitors such as kernel estimators in approximating distribution function (Lablanc, 2012).

order (m) are set. Second, the unconditional distribution is estimated using the Bernstein polynomial approximation. The algorithm tests the null hypothesis of polynomial parameter instability for all possible partitions based on the regressors and stores the corresponding p-values. This stability test ensures the reliability of the predicted distributions. If the parameters are stable, the conditional distributions fall into the same terminal node of the TrT. In other words, if for all possible partitions/splits, the Bonferroni-adjusted p-value > α, it means that the underlying distributions are the same and the algorithm stops. On the other hand, if the parameters are unstable, indicating different conditional distributions, the algorithm performs a binary split. In other words, if the p-value <α, then the parameters are unstable and the algorithm makes a binary split and repeats until the stability condition is met. As a result, once the TrT generates two groups or types, the Bernstein polynomial is used to interpolate the shape of the distributions. To estimate IOp, each individual's average outcome value ($\mu^{(p,t)}$) for type t and tranch p is divided by the population mean (μ^{p}) within the respective quantile p to obtain the adjusted value $\acute{y}_i^{p,t}$. IOp is then measured using any inequality measure applied to the adjusted values. The adjusted values is denoted as:

$$\text{IOP} = \text{f}\,(X_{11}\ldots\ldots X_{N_1 1}, X_{12}\ldots\ldots X_{N_1 2}, X_{13}\ldots\ldots X_{N_1 3}) \tag{9}$$

IOp is estimated with any inequality measure applied over $\acute{y}_i^{p,t}$.

3.3 Decomposition of IOp Measure

The Shapley decomposition method is based on the well-known concept of Shapley value in cooperative game theory (Shapley 1953). It is used to estimate the relative contribution of various factors or circumstances in total income IOp. Shapely values are order independent and the main idea behind them is to compute the value of a function considering all the possible combinations of circumstances. The functional form of the index is represented as:

$$\text{IOP} = \text{f}\,(X_{11}\ldots\ldots X_{N_1 1}, X_{12}\ldots\ldots X_{N_1 2}, X_{13}\ldots\ldots X_{N_1 3})$$

Where X_{ij} denotes the income of i[th] individual (i=1,……..N_j), within the subgroup j (j=1,2,3).

Additive decomposition is achieved by considering the impact of inequality within subgroups, inequality between subgroups, and the ranking and relative size within each subgroup. By the application of Shapely decomposition, one can derive the marginal impact of each circumstance by measuring the difference in the value of the inequality index between the observed situation and a reference scenario, where the income does not change with the circumstance (Das & Biswas, 2022).

IV. DATA SOURCES AND VARIABLES

In this paper, the annual Periodic Labor Force Survey (PLFS) data for the year 2018-19 has been used to calculate income IOp. Being a pre-pandemic year, it provides more realistic data compared to the pandemic years of 2019-20 and 2020-21. The PLFS is conducted by the National Sample Survey Office (NSSO) and is a cross-sectional survey representative of both national and state level data. The outcome variable used in the analysis is Household Monthly Per Capita Income (MPCI), which has been calculated by aggregating the income of regular, self-employed, and casual wage workers within a household, and then dividing it by the household size. The PLFS provides only weekly income data for casual wage workers, which has been converted into monthly income. On the other hand, the PLFS directly provides monthly income data for regular salaried and self-employed individuals. The circumstances variables available in the PLFS data and included in the analysis are: parents' level of education, categorized as no education, education up to primary level, secondary and higher secondary education, or graduate and above; parent's occupation, classified based on skill levels, including non-routine cognitive or high-skilled, routine cognitive or medium-skilled, non-routine manual or low-skilled, or routine manual or unskilled; social group variables, including scheduled castes (SC), scheduled tribes (ST), other backward classes (OBC), and general caste (GC); gender, categorized as male and female; place of birth or region, classified as north, east, central, north-east, south, and west; and location, categorized as rural or urban.

A sample of 105,492 individuals, out of the total 420,757 covered in the PLFS 2018-19, has been selected for the analysis. This includes only those individuals for whom parental background information is available. Furthermore, the analysis is limited to working-age individuals between the ages of 15 and 64 years only. More detailed information regarding the sample selection and variable selection procedures is given in Appendix 1.

V. RESULTS AND DISCUSSION

5.1 Profile of the Sample

Table 1 provides an overview of the sampled individuals in terms of demography, region, gender, social groups, and employment characteristics. The bulk of the sampled individuals reside in rural areas. Geographically, one-fourth of the sample belongs to the central region, while one-fifth each belongs to the eastern and southern regions. In contrast, the northern and north-eastern regions have a smaller representation compared to other regions, with only around 14.5 per cent belonging to the former and 3.7 per cent to the latter. Males make up 70 per cent of the sample, while females constitute 30 per cent. Around 44 per cent of individuals belong to the OBC category, followed by GC (28%), SC (20%), and ST (9%). Nearly half of the sampled individuals are involved in self-employment activities, followed by around one-third in regular salaried jobs, and one-fourth in casual wage work.

Table 1

Characteristics of Sample (in %)

		%
Sector	Rural	68.2
	Urban	31.8
Region	North	14.5
	East	20.3
	Central	25.8
	North East	3.7
	South	20.2
	West	15.5
Social Group	ST	8.5
	SC	19.6
	OBC	44.3
	GC	27.6
Gender	Male	71.4
	Female	28.6
Status of Employment	Self-employment	46.8
	Regular salaried	31.0
	Casual labor	22.2
Total		100.0

Source: Periodic Labor Force Survey, 2018-19

Table 2 provides insights into the educational qualifications and occupations of the sampled individuals' parents. Among the sampled individuals, a larger proportion of parents have secondary and higher secondary education or education below secondary level, while a smaller percentage of parents have a graduate level qualification and above or are illiterate. In terms of parental occupation, nearly half of the sampled individuals have parents engaged in non-routine manual or low-skilled jobs, and around one-fourth have parents in routine manual or unskilled jobs. Parents involved in non-routine cognitive or high-skilled jobs (15.8%) and routine cognitive or medium-skilled jobs (11.5%) represent a smaller percentage.

Table 2

Characteristics of Parents of Sampled Individuals (in %)

		%
Education Levels	No Education	5.7
	Below Secondary	35.7
	Secondary/Higher Secondary	43.0
	Graduate and above	15.5
Occupation by Skill Levels	Non-routine cognitive (high-skilled)	15.8
	Routine cognitive (medium-skilled)	11.5
	Non-routine manual (low-skilled)	46.7
	Routine manual (unskilled)	26.0
Total		100.0

Source: Periodic Labor Force Survey, 2018-19

Table 3 provides insights into the average monthly per capita household income (MPCI), highlighting notable differences between urban and rural areas. The data reveals that urban areas exhibit significantly higher average MPCI compared to rural areas. This suggests that households residing in urban regions generally have higher levels of income. Furthermore, the analysis shows that average MPCI is relatively higher among male-headed households. This indicates a gender disparity in income, with male earners tending to have higher earnings levels compared to female earners. Additionally, the data highlights variations in average MPCI across social groups. The households belonging to the GC group demonstrate the highest average MPCI, followed by OBC, SC, and ST. These findings underscore the existence of income disparities across social groups.

Table 3

Average, Median, and Standard Deviation of Household Monthly Per Capita Income (in Rs.) by Sector, Region, and Social Group

		Mean	*Median*	*SD*
Sector	Rural	2941	2400	2253
	Urban	5531	4000	7688
Gender	Male	3838	2810	5213
	Female	3572	2500	3822
Social Group	ST	2869	2125	2603
	SC	3282	2563	2688
	OBC	3414	2571	3081
	GC	4949	3375	7844
Total		3764	2750	4869

Source: Periodic Labor Force Survey, 2018-19

Table 4 presents insightful data regarding the association between parental education, parental occupation, geographical location, and MPCI. The findings reveal noteworthy disparities in average MPCI based on these factors. Firstly, the data demonstrates a substantial difference in average MPCI based on parental education levels. Individuals whose parents are graduates enjoy nearly two and a half times higher average MPCI compared to individuals whose parents have no education. The average MPCI is twice as high for individuals with graduate parents compared to those with below secondary level education, and it is significantly higher than those with secondary and higher secondary level education. Similarly, the analysis indicates a strong association between parental occupation and MPCI. Individuals whose parents are engaged in high and medium-skilled jobs exhibit significantly higher average MPCI compared to those whose parents are involved in low-skilled and unskilled manual jobs. This highlights the impact of parental occupation on the economic well-being of individuals. Furthermore, the data also suggests variations in average MPCI based on geographical location. Individuals born in the southern and western regions of India tend to have relatively higher average MPCI compared to those born in other regions. This regional disparity indicates the influence of geographical factors on income levels.

Table 4

Average, Median, and Standard Deviation of Household Monthly Per Capita Income (MPCI) (in Rs.) by Level of Parental Education and Parental Skill Level

		Mean	*Median*	*SD*
Educational Levels	No Education	2499	2000	1676
	Below Secondary	2874	2400	2012
	Secondary/Higher Secondary	3734	2900	3200
	Graduate and above	6388	4300	10259
Occupation by Skill Levels	Non-routine cognitive (high-skilled)	6580	4667	6893
	Routine cognitive (medium-skilled)	4538	3444	3807
	Non-routine manual (low-skilled)	2854	2286	2236
	Routine manual (unskilled)	3432	2900	2300
Region	North	4562	3200	4931
	East	2848	2167	2634
	Central	2645	2000	2475
	North East	3925	3000	2990
	South	4948	3929	4303
	West	4442	3100	8771
Total		3764	2750	4869

Source: Periodic Labor Force Survey, 2018-19

The sample characteristics observed provide valuable insights into the factors that contribute to variations in household MPCI. Key circumstances that play a significant role in shaping these differences include caste or social group, sector (rural or urban), parental education, parental occupation, and geographical location. On analysing the data, it becomes evident that certain categories associated with higher average MPCI values exhibit notable characteristics in terms of median value and standard deviation. These categories, such as urban areas, southern regions, highly educated parents, and those belonging to the GC and OBC social groups, display significantly higher levels of variability. The higher median values indicate that these categories tend to have households with relatively higher MPCI. Additionally, the larger standard deviations reflect a greater degree of variation within these categories. This implies that there is a wider range of MPCI values among households and individuals in these circumstances, suggesting diverse economic conditions and opportunities. These findings underscore the importance of considering various circumstances, including social group, sector, geographical location, parental education, and parental occupation, when analyzing inequality in an income variable such as MPCI.

5.2 Ex-Ante Inequality of Opportunity

In this section, a comparative analysis of ex-ante IOp results for MPCI (income, henceforth) has been conducted by employing three distinct approaches. These include the parametric approach, the conditional inference tree approach, and the conditional forest approach. The parametric approach uses the ordinary least squares (OLS) regression to estimate the IOp measures for income. This method allows us to model the relationship between the outcome variables and various circumstances, while taking into account the potential confounding factors. The regression analysis enables us to identify the factors that significantly contribute to IOp in terms of income.

5.2.1 Parametric Approach

The parametric approach is based on the methodology developed by Ferreira and Gignoux (2014), and Wendelspiess and Soloaga (2014). This is based on an OLS regression model, where MPCI serves as the dependent variable or outcome variable, while sector, gender, caste, parents' occupation, parents' education, and region are considered as explanatory or circumstance variables. By using the estimated coefficients obtained from the regression, a counterfactual distribution is derived for the sample data. This enables the decomposition of MPCI inequality within the sample population, part of which is attributed to unequal circumstances (IOp). The Gini coefficient of the predicted income values obtained from the regression provides an absolute measure of IOp. A relative measure of IOp is obtained by dividing the absolute Gini measure of IOp by the Gini measure of overall inequality.

Table 5.1 presents the results of the parametric approach, showing the number of types, and the IOp estimates in absolute and relative terms using the Gini coefficient. This approach utilizes 1536 types, which are groups of the sample population having similar circumstances. The overall income inequality is estimated to be 0.408, indicating a moderate level of inequality among the sampled population. The opportunity Gini coefficient derived from the parametric approach is 0.279. This suggests that the difference in average income between the 1536 sub-groups of the sample population is lower than the overall Gini coefficient which indicates a relatively smaller inequality within these subgroups. The relative IOp is estimated to be 0.682, meaning that around 68 per cent of the overall income inequality can be attributed to circumstances such as sector, gender, caste, parents' occupation, parents' education, and region.

Table 5.1

Parametric: Ex-Ante Income IOp

Type	*Overall Gini*	*Absolute Gini*	*Relative IOp*
1536	0.408	0.279	0.682

Source: Author's Calculation from Periodic Labour Force Survey 2018-19

5.2.2 Conditional Inference Trees

As mentioned previously, tree algorithms divide a dataset into mutually exclusive groups of observations based on sequential and hierarchical criteria. Once all the partitions are made, the algorithm assigns the average value of the dependent variable to each observation (Salas-Rojo & Rodriguez, 2022). However, one of the main drawbacks of tree-based algorithms is their strong reliance on various factors, including the chosen alpha level, which determines the threshold for accepting or rejecting the null hypothesis. To address this issue, the Grid Search Cross Validation method has been utilized to obtain an endogenously tuned alpha level (Appendix 2). The alpha level with the lowest root mean squared error (RMSE) is 0.07. The results at the alpha level of 0.07 are also compared with standard measures of alpha at 1 per cent and 5 per cent, as explained in detail in Appendix 2.1.

Table 5.2
Conditional Inference Tree: Ex-Ante Income IOp

Type	*Overall Gini*	*Absolute Gini*	*Relative IOp*
14	0.408	0.239	0.584

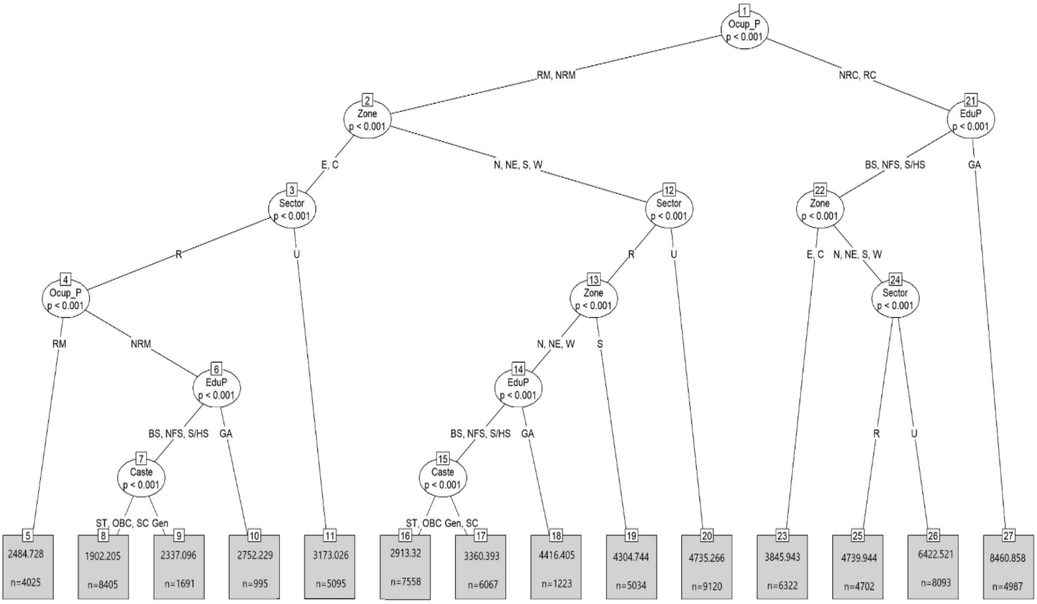
Source: Author's Calculation from Periodic Labour Force Survey 2018-19

Table 5.2 displays the results based on conditional inference tree using an endogenously chosen alpha level. The opportunity Gini coefficient for IOp is calculated to be 0.239, indicating that differences in average income among the 14 subgroups of the sampled population are significantly less than the overall income inequality. The relative IOp using the conditional inference tree approach is estimated to be 0.584. This suggests that around 58 per cent of the overall income inequality is attributable to various circumstances such as sector, gender, caste, parents' occupation, parents' education, and region.However, the relative IOp estimates obtained from the conditional inference tree method are comparatively lower than those obtained from the parametric method. This difference can be attributed to the machine learning (ML) algorithm used in the conditional inference tree method, which automatically generates a smaller number of types compared to the parametric method. These types correspond to distinct circumstances contributing to inequality and consequently, provide a more robust measure of relative IOp compared to the parametric approach.

Additionally, the conditional inference tree graphically illustrates the key circumstances that influence income IOp, which are shown in Figure 1. Among these circumstances, parents' occupation emerges as the most important factor, demonstrating statistically significant differences in average income between two groups. Specifically, there are statistically significant variations observed between individuals whose parents are engaged in routine manual and non-routine manual jobs (or low-skilled and unskilled jobs) as compared to those whose parents are employed in non-routine cognitive and cognitive jobs (or medium and high-skilled jobs). The first group of individuals with parents involved in low-skilled and unskilled jobs is further divided into two broad region (zone) types: North, North-East, South, West (NNESW) and East, Central (EC). Within the EC region, there is a subsequent split by rural and urban areas. In rural areas of the EC region, there is an additional split based on the occupation of parents, distinguishing between routine manual (unskilled) and non-routine manual (low-skilled) occupations. Within the low-skilled category, further divisions are made based on the educational levels of parents. One group includes parents whose educational qualification is that of graduate and above, while the other group comprises those with secondary and higher secondary education, below secondary education, and no formal education. Within the second category, a subsequent split is made based on social groups, with one group belonging to GC and the other consisting of SC, ST, and OBC. Similarly, the split in NNESW regions is also by rural and urban areas. The rural type is further subdivided into regions with North, North East, and West (NNEW), which further splits based on the education levels of parents and social groups. The final nodes indicate that individuals belonging to SC, ST, and OBC categories, whose parents are employed in low-skilled and unskilled jobs, and have an education below secondary level or no formal education, exhibit the lowest average income. More specifically, those residing in rural areas of the EC region demonstrate the lowest average income.

Figure 1

Conditional Inference Trees for MPCI



Source: *Source:* Author's Calculation from Periodic Labour Force Survey 2018-19

Note: R: Rural; U: Urban; N: North; NE: North East; S: South; W: West; E: East; C: Central; Sec/HS: Secondary/Higher Secondary; GradAbv: Graduate and Above; NoEdu: Illiterate or No Formal Schooling; BS: Below Secondary; NR: Non Routine Cognitive; RC: Routine Cognitive; NRM: Non Routine Manual; RM: Routine Manual; M: Male; F: Female.

The second group of individuals, whose parents are involved in high and medium skilled occupations, is also divided further based on the education level of parents. One group includes parents whose educational qualification is that of graduate and above, while the other comprises those with secondary and higher secondary level education, below secondary level education, and no formal education. The second group is further categorized into two broad types of regions: NNESW and EC. The NNESW category is further split by rural and urban sector. The last nodes indicate that the individuals whose parents are engaged in medium and high-skilled jobs and have a graduate degree or above, demonstrate the highest average income.

5.2.3 Conditional Forests

As discussed earlier, to address the sensitivity or high variance inherent in the conditional inference trees approach, a more robust approach of conditional inference forests has been proposed by Hothorn et al. (2006) and Brunori et al., (2023). Conditional inference forests employ bootstrapping within the ML framework. In this approach, multiple conditional inference trees are generated, and the final prediction is obtained by averaging the predictions of all the trees. The use of subsamples ensures that each tree provides an independent estimate (Salas-Rojo & Rodriguez, 2022, p.36). Similar to the conditional inference tree approach, an endogenous level of alpha has been obtained by the Grid Search Cross Validation method to determine the appropriate combination of the number of trees for the analysis. The level of alpha with the lowest RMSE is 0.06. The results obtained at 0.06 per cent level are also compared with standard measures of alpha at 1 per cent and 5 per cent, the details of which are given in Appendix 2.2.

Table 5.3

Conditional Inference Forest: Ex-Ante IOp Results

Type	*Overall Gini*	*Absolute Gini*	*Relative IOp*
125	0.408	0.248	0.608

Source: Author's Calculation from Periodic Labour Force Survey 2018-19

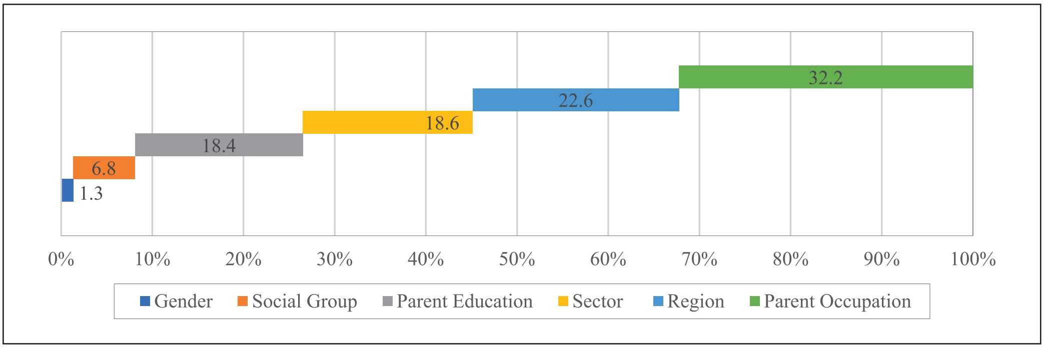
Table 5.3 shows the results based on the conditional inference forest using an endogenously chosen alpha level. The opportunity Gini coefficient for income IOp is estimated to be 0.248. This means that the difference in average income among the 125 subgroups of the sample population is significantly less than the overall income inequality, but slightly higher than the result obtained from the conditional inference tree. The relative IOp measured using the conditional inference forest approach is estimated to be 0.608. This suggests that around 61 per cent of the overall income inequality is attributed to various circumstances such as sector, gender, caste, parents' occupation, parents' education, and region. This relative IOp estimate obtained from the conditional inference forest method is marginally higher than that obtained from the conditional inference tree method. This difference can be attributed to the bootstrapping within the ML algorithm used in the conditional inference forest method. This technique helps to address the sensitivity or high variance inherent in the conditional inference tree. Consequently, this method provides a more robust measure of relative IOp compared to the conditional inference tree and parametric approaches.

5.3.4 Ex-Ante Shapley Value Decompositions

The ex-ante decomposition exercise presented in Figure 2 provides insights into the importance of different circumstance variables in contributing to income IOp. The analysis reveals the relative significance of each factor in explaining the observed variations in income IOp. Among the circumstance variables, the occupation of parents emerges as the most important factor, accounting for the largest share of income IOp at 32.2 per cent. This suggests that the employment opportunities and earnings across occupations based on skill levels have a substantial impact on income IOp. The geographical location of individuals, with a contribution of 22.6 per cent, is the second key factor, indicating that regional disparities in economic development and access to resources can significantly impact income levels. The circumstance variables of sector (18.6%) and parents' education (11.5%) also play a vital role in shaping income IOp. This indicates that the disparities in employment opportunities between rural and urban areas, as well as the educational levels of parents, play an important role, and reflect the influence of educational opportunities and qualifications on income disparities. Social groups and gender disparities, although relatively less influential, also contribute to income IOp.

Figure 2

Decomposition of Factors Contributing to Ex-Ante IOp (in %)



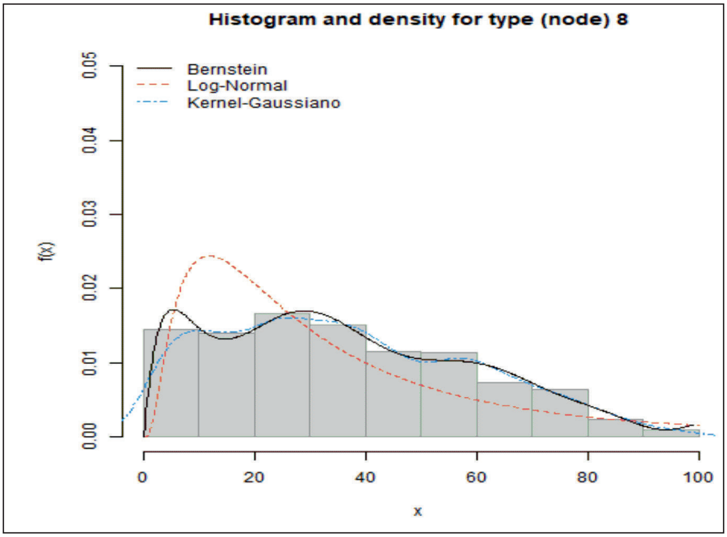
Source: Author's Calculation from Periodic Labour Force Survey 2018-19

5.3 Ex-Post Inequality of Opportunity

As previously discussed, the estimation of IOp using conventional parametric, non-parametric, and data-driven conditional inference tree ML techniques is primarily based on the mean difference between types. These methods do not take into account the higher moments of the within-type distribution. To address this limitation, different methods are employed to approximate the distribution of the outcome variable. These methods include the log-normal, kernel-Gaussian, and Bernstein polynomials, as shown in Figure 3 and Appendix 3.

Figure 3

Log-Normal, Kernel-Gaussian, and Bernstein Polynomials Distribution of MPCI



Source: Author's Calculation from Periodic Labour Force Survey 2018-19

Among these methods, the Bernstein polynomials method is found to be more flexible in predicting the distribution of outcomes within each type. These polynomials enable a more accurate representation of the underlying distribution, capturing higher moments beyond the mean difference. By utilizing the Bernstein polynomials method, the degree of effort or ex-post IOp can be measured, following the approach proposed by Brunori and Neilofer (2021). The ex-ante approach focuses on the mean of each type, while the ex-post approach examines the distribution functions of each type. Instead of examining for statistically significant differences between means, the ex-post approach identifies the most statistically significant differences between the full expected conditional distribution functions.

5.3.1 Transformation Tree

As mentioned previously, the conditional inference trees and conditional inference forests select partitions based on differences in a single statistic of interest within each type; specifically, the mean of the conditional outcome distribution (Brunori et al., 2023). In contrast, the transformation tree (TrT) approach utilizes splits or partitions based on differences across multiple functions of the distribution, including variance, skewness, and kurtosis (Hothorn & Zeilis, 2021). In this paper, the TrT approach is employed to analyze the effects of different variables on the conditional outcome (MPCI or income) distribution. It reveals the configuration of variables that strongly influence the distribution and provides insights into specific conditional outcome distributions (Hothorn, 2018). The TrT demonstrates the distributions obtained after applying the Bernstein polynomial transformation. Hence, the Bernstein polynomial of order 5 has been used to transform the outcome variables, and a transformation tree model is employed to predict the types for each data point for ex-post income IOp analysis. The model predicts the income quantile position of each individual within each type to determine the 'degree of effort'. Based on these income quantile positions, the mean outcome value for each quantile, as well as the population mean, are determined. An individual's outcome value is adjusted using the ratio of the population mean to the quantile mean, enabling the measurement of ex-post IOp.

As shown in Table 6, the number of types generated by the transformation trees is 16, which is higher than the number of types generated by the conditional inference tree, and significantly lower than the number of types generated by the conditional inference forest approach. The overall Gini inequality in the ex-post approach is 0.348, indicating a moderate level of inequality among the sample population. However, IOp measures for income in the ex-post approach yield relatively smaller values compared to the ex-ante approach. The estimated opportunity Gini coefficient is 0.160, and the relative IOp value is 0.460. This indicates that around 46 per cent of the overall income inequality can be attributed to the differences in degree of effort. These results suggest that the ex-post IOp measures, which consider the entire distribution function obtained through the transformation trees or contribution of effort in explaining the IOp, are lower than the measures based on mean differences in the ex-ante approach or contribution of circumstances.

Table 6
Transformation Tree: Ex-Post IOp Results

Type	*Overall Gini*	*Absolute Gini*	*Relative IOp*
16	0.348	0.160	0.460

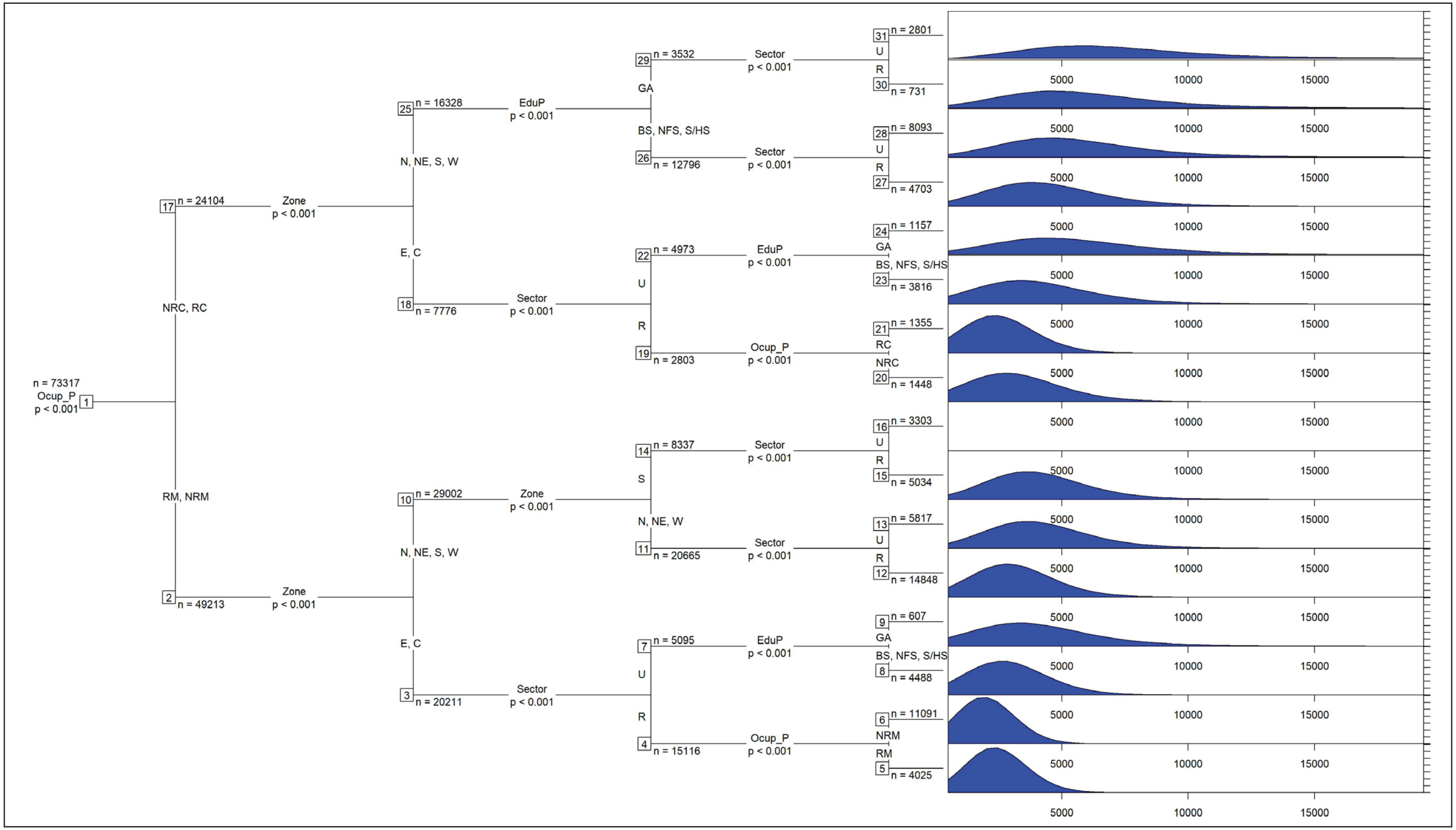
Source: Author's Calculation from Periodic Labour Force Survey 2018-19

The transformation tree depicted in Figure 4 highlights the significant factors influencing ex-post IOp. It reveals that parents' occupation emerges as the most important factor, exhibiting statistically significant variations in average income between two groups. The first group consists of individuals whose parents are involved in non-routine cognitive and routine cognitive occupations, characterized by high and medium skill levels. The second group comprises individuals whose parents are engaged in routine manual and non-routine manual occupations, which are considered low-skilled and unskilled occupations.

The first group of high and medium-skilled individuals is further subdivided into two broad regions: North, North-East, South, West (NNESW) and East, Central (EC). For those located in the NNESW regions, there is an additional split into two broad groups based on their parents' education: one group consists of those whose parents' educational qualification is that of graduate and above, and the other group comprises individuals whose parents have secondary and higher secondary level education, below secondary level education, and no formal education. These groups are further sub-divided by rural and urban areas. On the other hand, for those located in the EC region, a subsequent split is made on the basis of rural and urban areas. In rural areas, there is an additional split into two groups based on parents' occupation: one is of those whose parents have medium

Figure 4

Transformation Tree for MPCI



Source: Author's Calculation from Periodic Labour Force Survey 2018-19

Note: R: Rural; U: Urban; N: North; NE: North East; S: South; W: West; E: East; C: Central; Sec/HS: Secondary/Higher Secondary; GradAbv: Graduate and Above; NoEdu: Illiterate or No Formal Schooling; BS: Below Secondary; NR: Non Routine Cognitive; RC: Routine Cognitive; NRM: Non Routine Manual; RM: Routine Manual; M: Male; F: Female.

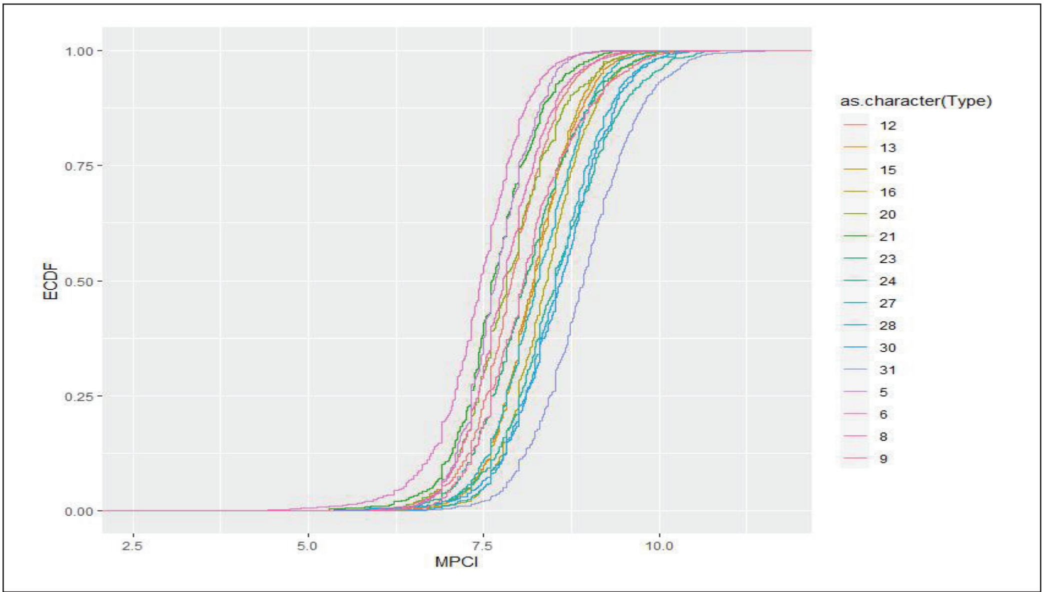
and high-skilled jobs, and the other group is of those individuals whose parents have low-skilled and unskilled jobs. In urban areas, a further split is made based on parents' education: one group includes those whose parents have a graduate degree and above, and the other group includes those whose parents have secondary and higher secondary level education, below secondary level education, and no formal education.

The first group of high and medium-skilled individuals is further divided into two broad regions: North, North-East, South, West (NNESW) and East, Central (EC). For those located in the NNESW regions, an additional split is made based on their parent's education: one group consists of individuals whose parents have a graduate degree and above, while the other group includes individuals whose parents have secondary and higher secondary level education, below secondary level education, and no formal education. These groups are further subdivided into rural and urban areas. On the other hand, individuals located in the EC region are split based on rural and urban areas. In rural areas, an additional split is made based on parents' occupation, creating two groups: one with individuals whose parents have medium and high-skilled jobs, and the other with individuals whose parents have low-skilled and unskilled jobs. In urban areas, further divisions are made based on parents' education, with one group consisting of individuals whose parents have a graduate degree and above, and the other consisting of individuals whose parents have secondary and higher secondary level education, below secondary level education, and no formal education.

The second group, consisting of individuals whose parents are in low-skilled and unskilled jobs, is subdivided into two broad regions: North, North-East, South, West (NNESW) and East, Central (EC). For those located in the NNESW regions, an additional split is made based on geographical regions or zones. One group is categorized as the South, while the other group comprises the North, North-East, and West (NNEW). These groups are further divided into rural and urban areas. Similarly, individuals located in the EC region are split based on rural and urban areas. In rural areas, an additional split is made based on parents' occupation, into low-skilled and unskilled jobs. In urban areas, a further split is made based on parents' education, with one group consisting of individuals whose parents have a graduate degree and above, and the other group comprising individuals whose parents have secondary and higher secondary level education, below secondary level education, and no formal education.

The final nodes of the transformation tree confirm the results of the conditional inference tree, which indicates that the lowest income distribution is observed among individuals whose parents are involved in low-skilled and unskilled (non-routine manual and routine manual) jobs in rural areas of the central and eastern regions. Meanwhile, the highest income distribution is observed among individuals whose parents are involved in high and medium-skilled occupations, and those whose parents have an educational qualification of graduate and above. These individuals reside in urban areas and are located in the North, North-East, South, and South-West regions. Similar results can also be seen from the Expected Conditional Distribution Function (ECDF)[3], as depicted in Figure 5. The lowest average income (MPCI) can be clearly seen at the left-most part of the figure, while the highest can be seen at the right-most part, representing the two groups discussed above.

Figure 5
Expected Cumulative Distribution Functions for MPCI



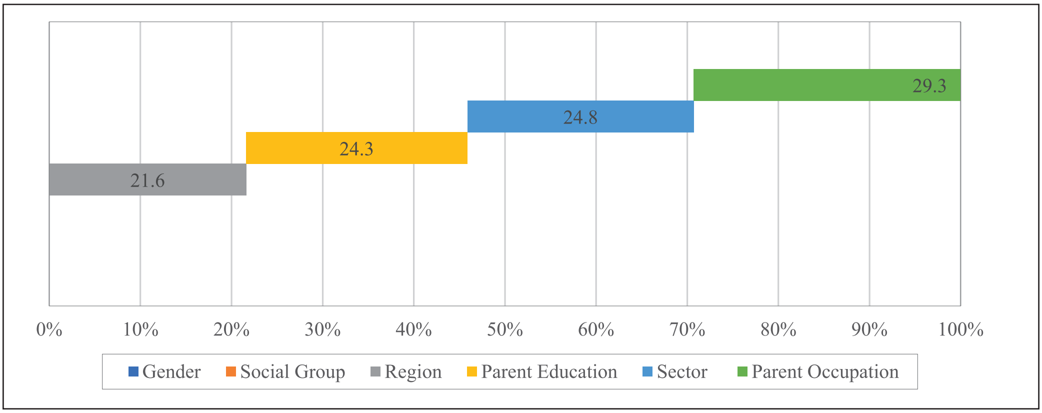
Source: Author's Calculation from Periodic Labour Force Survey 2018-19

3. A Conditional Distribution Functions (CDF) is a function of the form i.e., the probability of Y is *j* given for a given value of X (James et al., 2013, p.37). A type specific ECDF, as used in Brunori et al (2023), describes the probability distribution of a random variable given certain conditions, in the context of the paper, ECDF's give us about the probability distribution of the MPCI given a circumstance type.

5.3.2 Ex-Post Shapley Value Decompositions

The final step in the ex-post analysis, similar to the ex-ante approach, is to assess the relative importance of individual circumstance variables using the Shapley value decomposition, as shown in Figure 6. The results are quite similar to those obtained in the ex-ante analysis. Parents' occupation (29.3%) emerges as the most influential factor, indicating that differences in parents' occupation based on skill type, significantly contributes to income IOp. This is closely followed by the sector to which an individual belongs (24.8%), parents' education level (24.3%), and geographical location (21.6%), in demonstrating considerable importance in explaining income IOp. However, social groups and gender have a minimal role in explaining income IOp, with their contribution being almost negligible.

Figure 6
Decomposition of Factors Contributing to Ex-Post IOp (in %)



Source: Author's Calculation from Periodic Labour Force Survey 2018-19

VI. CONCLUSION

The study provides both ex-ante and ex-post income IOp estimates at the national level. This work is also the first attempt to determine the types and represent the structure of opportunities in Indian society through conditional inference trees, conditional inference forests, and transformation trees. The results based on trees allow for graphical representations of the opportunities provided by society, which can be easily communicated to policymakers and other stakeholders.

The ex-ante income IOp is relatively higher than the ex-post income IOp, which shows the differences in interpretation and understanding of IOp in society. Using the ex-ante approach, approximately 58-61 per cent of the total income-based inequality of opportunity can be explained by differences between circumstances, while the ex-post method paints a different picture, with around 46 per cent of the total income IOp being explained by within-tranche differences or differences in effort levels.

The tree-based analysis reveals that parents' occupation, areas of residence (rural or urban), and region (geographical location) are the most important variables, followed by parental education and social group, in determining the income IOp in Indian society. The ex-ante and ex-post Shaley decomposition exercise further confirms that parents' occupation, geographic location, sector (rural or urban), and parents' education are the most important circumstances contributing to income IOp. In particular, individuals in central and eastern regions, those residing in rural areas, those whose parents are employed in low-skilled and unskilled occupations, those whose parents have below secondary level education and no formal education, and those belonging to marginalized social groups, exhibit significantly lower average income. Hence, there is an urgent need for regional-level development policies that focus on marginalized groups to create a just society and reduce overall inequality in India.

Acknowledgement: The authors are grateful to Prof. S.Madeshwaran, Prof. Arup Maitra, Dr. Pedro Salas Rojo, and the conference audience at the 63rd ISLE conference for their valuable comments and suggestions on the draft of this paper.

Funding: This paper is the outcome of a study titled 'INEQUALITREES – A Novel Look at Socio-Economic Inequalities using Machine Learning Techniques and Integrated Data Sources' funded by the Volkswagen Stiftung, Germany. The authors gratefully acknowledge the financial support received from the VW-Stiftung.

References

Anand, I., & Thampi, A. (2016). *Recent Trends in Wealth Inequality in India*. *51*(50). https://www.epw.in/journal/2016/50/special-articles/recent-trends-wealth-inequality-india.html

Arneson, R. (1989). Equality and Equal Opportunity for Welfare. *Philosophical Studies*, *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. https://doi.org/10.1016/j.worlddev.2011.11.008

Bourguignon, F. (2004). The Poverty-growth-inequality triangle. *Indian Council for Research on International Economic Relations Working Paper*, *131*, 35.

Brunori, P., Ferreira, F., & Neidhöfer, G. (n.d.). *UNU-WIDER: Working Paper : Inequality of opportunity and intergenerational persistence in Latin America* [Working Paper]. UNU_Wider. http://www.wider.unu.edu/publication/inequality-opportunity-and-intergenerational-persistence-latin-america

Brunori, P., & Neidhöfer, G. (2021a). Inequality of opportunity in comparative perspective: Recent advances and challenges. In *Handbook of Income Distribution* (Vol. 3B, pp. 1393–1479). Elsevier.

Brunori, P., & Neidhöfer, G. (2021b). The Evolution of Inequality of Opportunity in Germany: A Machine Learning Approach. *Review of Income and Wealth*, *67*(4), 900–927. https://doi.org/10.1111/roiw.12502

Brunori, P., Palmisano, F., & Peragine, V. (2016). *Inequality of Opportunity in Sub-Saharan Africa* (Working Paper No. 7782). World Bank. https://documents1.worldbank.org/curated/en/522541470317691539/pdf/WPS7782.pdf

Brunori, P., Palmisano, F., & Peragine, V. (2019). Inequality of opportunity in sub-Saharan Africa. *Applied Economics*, *51*(60), 6428–6458. https://doi.org/10.1080/00036846.2019.1619018

Chancel, L., & Piketty, T. (2019). Indian Income Inequality, 1922-2015: From British Raj to Billionaire Raj? *Review of Income and Wealth*, *65*(S1), S33–S62. https://doi.org/10.1111/roiw.12439

Chancel, L., Piketty, T., Alvaredo, F., Saez, E., & Zucman, G. (2018). The Elephant Curve of Global Inequality and Growth. *AEA Papers and Proceedings*, *108*, 103–108. https://doi.org/10.1257/pandp.20181073

Chancel, L., Piketty, T., Saez, E., & Zucman, G. (2022). *World Inequality Report 2022*. UNDP, World Inequality Lab.

Cohen, G. A. (n.d.-a). On the currency of egalitarian justice. *Ethics*.

Cohen, G. A. (n.d.-b). On the currency of egalitarian justice. *Ethics*, *99*(4), 906–944.

Das P, Biswas S (2022). Social Identity, Gender and Unequal Opportunity of Earning in Urban India: 2017-2018 to 2019-2020. *Indian J Labour Econ*. 2022;65(1):39-57.

Deaton, A., & Stone, A. A. (2013). *Grandpa and the Snapper: The Wellbeing of the Elderly who Live with Children* (Working Paper No. 19100). National Bureau of Economic Research. https://doi.org/10.3386/w19100

Dworkin, R. (n.d.). What is equality? Part 2: Equality of resources. *Philosophy & Public Affairs*, *10*(4), 283–345.

Dworkin, R. (1981). What is equality? Part 1: Equality of welfare. *Philosophy & Public Affairs*, *10*(3), 185–246.

Ferreira, F. H. G., & Gignoux, J. (2008). *The Measurement Of Inequality Of Opportunity: Theory and An Application To Latin America*. The World Bank. https://doi.org/10.1596/1813-9450-4659

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. https://doi.org/10.1111/j.1475-4991.2011.00467.x

Ferreira, F. H. G., Gignoux, J., & Aran, M. (2011). Measuring inequality of opportunity with imperfect data: The case of Turkey. *The Journal of Economic Inequality*, *9*(4), 651–680. https://doi.org/10.1007/s10888-011-9169-0

Fleurbaey, M. (1995). Equal opportunity or equal social outcome? *Economics and Philosophy*, *11*(1), 25–55.

Fleurbaey, M. (2008). *Fairness, responsibility, and welfare*. Oxford University Press.

Fleurbaey, M., & Pergaine, V. (2013). Ex post inequalities and ex ante inequalities. In *Justice, political liberalism, and utilitarianism: Themes from Harsanyi and Rawls* (pp. 59–77). Cambridge University Press.

Ghose, A. K. (2021). Structural Change and Development in India. *Indian Journal of Human Developmen*t, 15(1), 7–29. https://doi.org/10.1177/09737030211005496

Hothorn, T. (Director). (2018). *Video: Torsten Hothorn, "Transformation Forests."* http://www.birs.ca/events/2018/5-day-workshops/18w5054/videos/watch/201801161507-Hothorn.html

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, *15*(3), 651–674. https://doi.org/10.1198/106186006X133933

Hothorn, T., & Zeileis, A. (2015). partykit: A Modular Toolkit for Recursive Partytioning in R. *Journal of Machine Learning Research*, *16*(118), 3905–3909.

Hothorn, T., & Zeileis, A. (2021). Predictive Distribution Modeling Using Transformation Forests. *Journal of Computational and Graphical Statistics*, *30*(4), 1181–1196. https://doi.org/10.1080/10618600.2021.1872581

Institute of Applied Manpower Research. (2011). *India Human Development Report 2011: Towards Social Inclusion*. Oxford University Press. https://www.im4change.org/docs/240India-Human-Development-Report.pdf

Kuznets, S. (1955). Economic Growth and Income Inequality. *The American Economic Review*, *45*(1), 1–28.

Lefranc, A., & Kundu, T. (n.d.). *Inequality of Opportunity in Indian Society*. 37.

Palomino, J. C., Marrero, G. A., Nolan, B., & Rodríguez, J. G. (2022). Wealth inequality, intergenerational transfers, and family background. *Oxford Economic Papers*, *74*(3), 643–670. https://doi.org/10.1093/oep/gpab052

Piketty, T., & Zucman, G. (2015). Wealth and Inheritance in the Long Run. In A. B. Atkinson & F. Bourguignon (Eds.), *Handbook of Income Distribution* (Vol. 2B, pp. 1303–1368). Elsevier. https://doi.org/10.1016/B978-0-444-59429-7.00016-9

Plassot, M., Ramos, X., & Van de gaer, D. (2022). The ex-ante and ex-post measurement of inequality of opportunity: A normative framework. *Review of Income and Wealth*, *68*(1), 4–31.

Ravallion, M., & Chen, S. (2003). Measuring pro-poor growth. *Economics Letters*, *78*(1), 93–99. https://doi.org/10.1016/S0165-1765(02)00205-7

Rawls, J. (1958a). A Theory of Justice: Original Edition. In *A Theory of Justice*. Harvard University Press. https://doi.org/10.4159/9780674042605

Rawls, J. (1958b). Justice as fairness. *The Philosophical Review*, *67*(2), 164–194.

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

Roemer, J. E. (1998). *Equality of Opportunity*. Harvard University Press. https://link.springer.com/referenceworkentry/10.1057/978-1-349-95121-5_2223-1

Roemer, J. E., & Trannoy, A. (2016). Equality of Opportunity: Theory and Measurement. *Journal of Economic Literature*, *54*(4), 1288–1332. https://doi.org/10.1257/jel.20151206

Salas-Rojo, P., & Rodríguez, J. G. (2022). Inheritances and wealth inequality: A machine learning approach. *The Journal of Economic Inequality*, *20*(1), 27–51.

Sen, A. (1980). Equality of What? In *Tanner Lectures on Human Values* (Vol. 1). Cambridge University Press.

APPENDIX A1
SAMPLE SELECTION AND CONSTRUCTION OF VARIABLES

Variable Selection

From the PLFS survey 2018-19, six variables have been selected. Of these, three variables, namely sector, caste, and gender are used in their existing forms, while the other three variables, namely states, parents' education, and parents' occupation are modified into a new form. The sector variable is categorized as rural or urban; gender as male or female; and caste as General Caste (Gen), Scheduled Castes (SC), Scheduled Tribes (ST), and Other Backward Classes (OBC). In the gender variable, the category of transgender has been dropped before the analysis.

The state variable is categorized into 36 states/union territories of India, which have been modified and classified into the following six broad geographical regions:

1. North: Jammu & Kashmir, Himachal Pradesh, Punjab, and Haryana
2. East: Bihar, Jharkhand, Orissa, and West Bengal
3. Central: Uttar Pradesh, Uttarakhand, Rajasthan, Madhya Pradesh, and Chhattisgarh
4. North-East: Sikkim, Arunachal Pradesh, Assam, Nagaland, Meghalaya, Manipur, Mizoram, and Tripura
5. South: Karnataka, Andhra Pradesh, Tamil Nadu, Pondicherry, Kerala, and Lakshadweep
6. West: Gujrat, Daman & Diu, Dadra & Nagar Haveli, Maharashtra, and Goa.

The education variable is classified into the following four broad categories:

1. Illiterate or no education: (code 1, Illiterate)
2. Below secondary: (code 2-7, literate to up to middle school)
3. Secondary and above secondary: (code 8-10, secondary to higher secondary)
4. Graduate and above: (code 12-13, graduate and post-graduate)

The occupation/skill level is classified into the following four broad categories using the National Classification of Occupations (NCO) (as per the *OECD* Employment Outlook 2014; National Classification of Occupations, 2015, Ministry of Labour and Employment, Government of India).

1. Unskilled or routine manual tasks: Typically involves the performance of simple and routine physical or manual tasks (NCO code 9- Elementary occupations or unskilled occupations such as domestic helpers, cleaners, street vendors, garbage collectors, etc.)
2. Low-skilled or non-routine manual tasks: Typically involves the performance of tasks such as operating machinery and electronic equipment; driving vehicles; maintenance and repair of electrical and mechanical equipment; and manipulation, ordering, and storage of information (NCO code 4-8- low-skilled jobs such as clerical workers, service workers, shop and market sales workers, craft and related trade workers, etc.)
3. Medium-skilled or non-routine cognitive tasks: Typically involves the performance of complex technical and practical tasks that require an extensive body of factual, technical, and procedural knowledge in a specialized field (NCO code 3- occupations such as professional and technical associates)
4. High-skilled or cognitive tasks: Typically involves the performance of tasks that require complex problem solving, decision making, and creativity based on an extensive body of theoretical and factual knowledge in a specialized field (NCO code 2- occupations such as professionals and technicians).

The concept of skill level is not applied in the case of NCO code 1 for occupations such as legislators, managers, etc. as the skills required for executing the tasks and duties of these occupations vary to such an extent that it was not feasible to link them with any of the four broad skill levels.

Sample Selection

For the selection of the sample, the following multi-stage procedure was adopted.

In the first stage, the parent of each respondent was identified using the relation to the head variable in the data. For an individual identified as self (code 1), the household member with code 7 (labelled Father/Mother/Father-in-Law/Mother-in-Law) was treated as the parent, and the first set of data with children and parents was prepared.

In the second stage, individuals were identified as unmarried children (code 5) and married children (code 3), and further, the parents of these children were identified as household heads, labelled as self (code 1) in the data. The respondent labelled self was identified as the parent and the second set of data with children and parents was prepared.

In case of duplicate records (or multiple parental information), the duplicate cases were deleted after carefully looking at the unit records.. Once both the files were cleaned, they were merged along with the key variables in the data, as discussed above.

APPENDIX A2
GRID SEARCH CV PROCESS FOR CONDITIONAL INFERENCE TREE AND CONDITIONAL INFERENCE FOREST

In the Grid Search CV process, the data is divided into training and test sets. Different combinations of min-split (minimum number of observations required to perform a split) and alpha values are tested, and the combination that yields the lowest root mean squared error (RMSE) for the test set is selected. The RMSE is a measure of the model's prediction accuracy. For the conditional inference tree model with MPCI as dependent variables, the Grid Search CV was conducted. After evaluating various combinations, an alpha value of 0.07 and a min-split value of 10000 were found to provide the lowest RMSE. The robustness of the endogenously chosen alpha is examined by comparing the results with the alpha values of 0.01 and 0.05, as given in Table A.2.1. This comparison is done following the approach outlined by Salas-Rojo and Rodriguez (2022).

Table A.2.1
Ctree Results MPCI

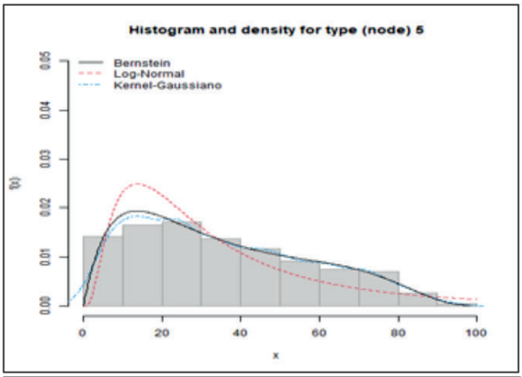
Alpha	*Types*	*Overall Inequality (Gini)*	*Absolute Gini*	*IoP Gini*
0.07	14	0.408	0.239	0.584
0.01	14	0.408	0.239	0.584
0.05	14	0.408	0.239	0.584

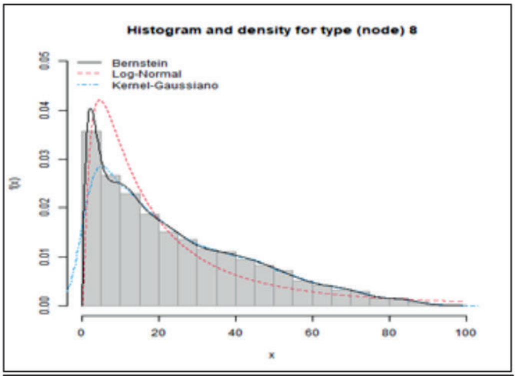
Similarly, after evaluating various combinations, an alpha value of 0.06 and number of trees value of 200 were found to provide the lowest RMSE for the conditional inference forest model. The robustness of the endogenously chosen alpha is examined by comparing the results with the alpha values of 0.01 and 0.05, as given in Table A.2.2.

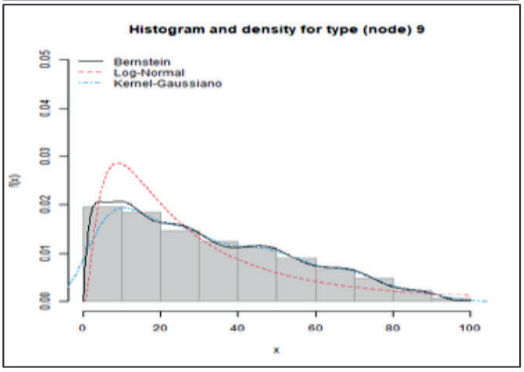
Table A.2.2
Cforest Results MPCI

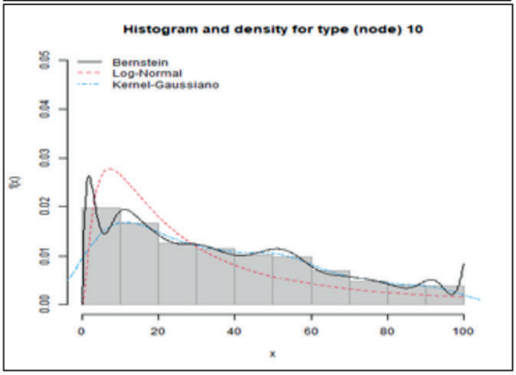
Alpha	*Types*	*Overall Inequality (Gini)*	*Absolute Gini*	*IoP Gini*
0.06	125	0.4084	0.2481	0.6076
0.01	87	0.4084	0.2479	0.6072
0.05	103	0.4084	0.2495	0.6109

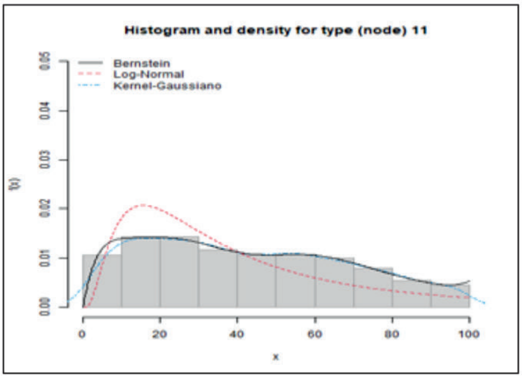
APPENDIX A3
PLOTS FOR MPCI

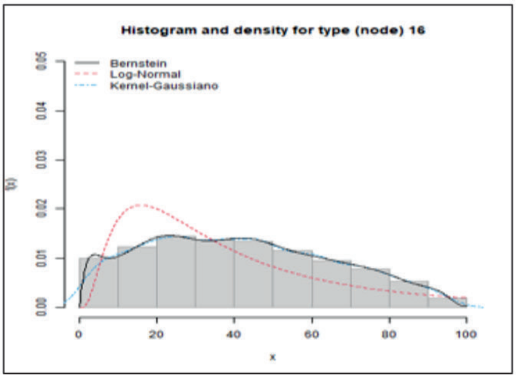


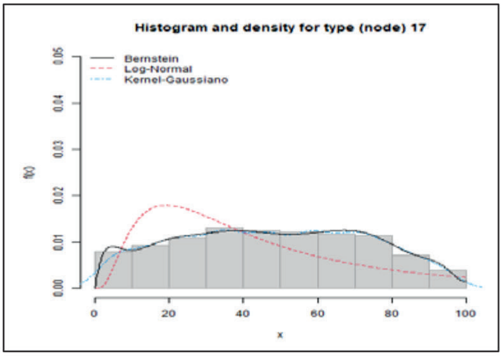


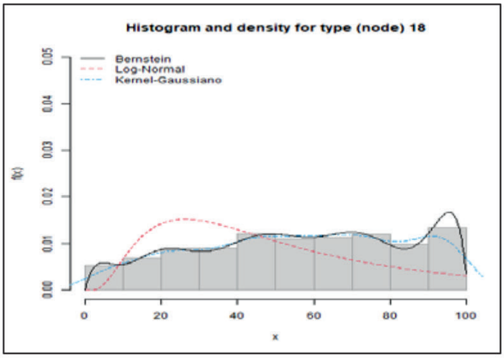


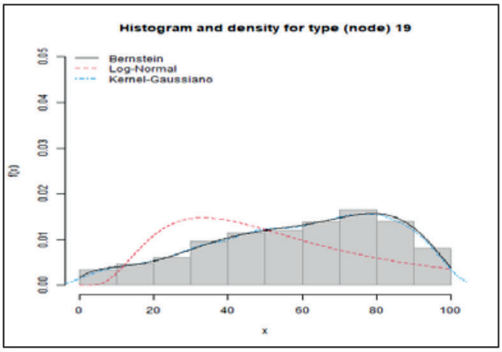


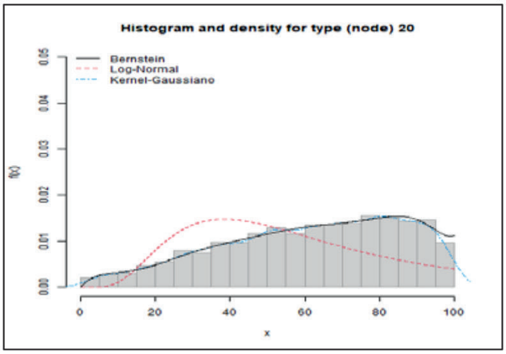


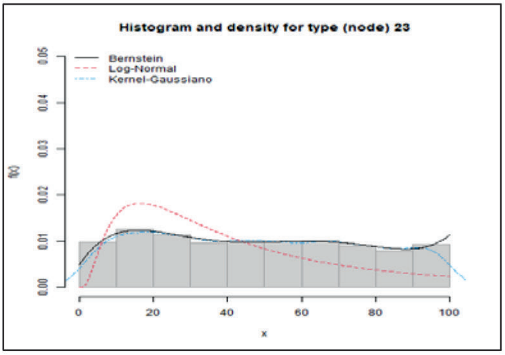


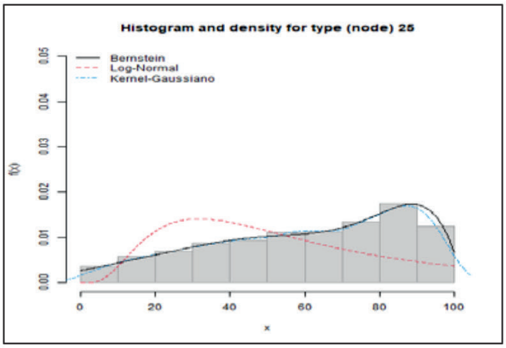

Histogram and density for type (node) 17
Bernstein
Log-Normal
Kernel-Gaussiano
f(x)
x

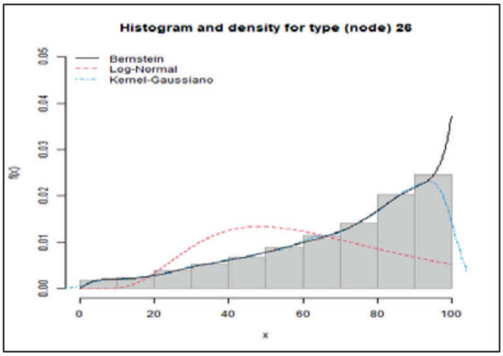

Histogram and density for type (node) 18
Bernstein
Log-Normal
Kernel-Gaussiano
f(x)
x

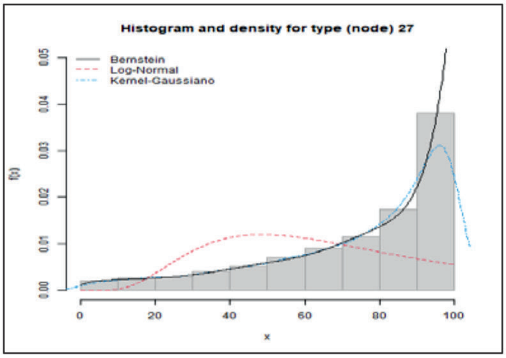

Histogram and density for type (node) 19
Bernstein
Log-Normal
Kernel-Gaussiano
f(x)
x


Histogram and density for type (node) 20
Bernstein
Log-Normal
Kernel-Gaussiano
f(x)
x


Histogram and density for type (node) 23
Bernstein
Log-Normal
Kernel-Gaussiano
f(x)
x


Histogram and density for type (node) 25
Bernstein
Log-Normal
Kernel-Gaussiano
f(x)
x


Histogram and density for type (node) 26
Bernstein
Log-Normal
Kernel-Gaussiano
f(x)
x


Histogram and density for type (node) 27
Bernstein
Log-Normal
Kernel-Gaussiano
f(x)
x

IHD WORKING PAPER SERIES

No.	*Authors*	*Title*
WP 07/2022	Balwant Singh Mehta, Siddharth Dhote, and Ravi Srivastava	Decomposition of Inequality of Opportunity in India: An Application of Data-Driven ML Approach
WP 06/2022	Balwant Singh Mehta and Siddharth Dhote	Inequality of Opportunity in India: Concept and Measurement
WP 05/2022	Charu C. Garg and Roopali Goyanka	Economic Costs for Outpatient Care in Public and Private Facilities in Delhi, India: Implications for Healthcare Policy
WP 04/2022	D Narasimha Reddy	Agroecology and Sustainable Smallholder Agriculture: An Exploratory Analysis with Some Tentative Indications from the Recent Experience of 'Natural Farming in Andhra Pradesh'
WP 03/2022	Dev Nathan	Knowledge and Global Inequality: Monopoly, and Monopsony Capitalism
WP 02/2022	Tanuka Endow	Female Workforce Participation and Vulnerability in Employment: Evidence from Rural Jharkhand
WP 01/2022	Ajit K. Ghose and Balwant S. Mehta	New Technologies, Employment and Inequality in the Indian Economy
WP 02/2021	Dev Nathan	Knowledge and Inequality: An Exploration
WP 01/2021	Ajit K. Ghose and Abhishek Kumar	India's Deepening Employment Crisis in the Time of Rapid Economic Growth
WP 02/2020	Sarthi Acharya and Santosh Mehrotra	The Agricultural Market Reforms: Is There a Trade-off Between Efficiency and Equality?
WP 01/2020	Ajit K. Ghose	Structural Transformation of India's Economy
IHD-CENTRE FOR GENDER STUDIES (CGS) WORKING PAPER SERIES		
IHD-CGS WP 03/2022	Govind Kelkar	Patriarchal Discrimination and Capitalist Relations: The Gender Question in the Gig Economy
IHD-CGS WP 01/2020	Aasha Kapur Mehta	Union Budget 2020-21 and the Gender Budget Statement: A Critical Analysis from a Gender Perspective
IHD-CENTRE FOR EMPLOYMENT STUDIES (CES) WORKING PAPER SERIES		
IHD-CES WP 04/2020	Ravi Srivastava	Understanding Circular Migration in India; Its Nature and Dimensions, the Crisis under Lockdown and the Response of the State

IHD-CES	Ravi Srivastava	Integrating Migration and Development Policy in
WP 03/2020		India:A Case Study of Three Indian States
IHD-CES	Ravi Srivastava	Vulnerable Internal Migrants in India and
WP 02/2020		Portability of Social Security and Entitlements
IHD-CES	Ravi Srivastava and B.	Collapse in Wage/Salary Income Growth in India,
WP 01/2020	Padhi	2011-12 to 2017-18
WP 01/2019	Ajit K. Ghose	Economic Development in China and India: A Tale of Great Divergence
WP 01/2017	Tanuka Endow	Urban Development and Rural – Urban Linkages: Case Study of Six Towns in Bihar
WP 01/2016	Tanuka Endow, Sunil K. Mishra and Abhay Kumar	Urban Development and Rural-Urban Linkages: Case Study of Two Towns in Bihar
WP 04/2015	Alexandre de Freitas Barbosa, Maria Cristina Cacciamali, Nandita Gupta, Ian Prates, Gerry Rodgers and Priscila Vieira	Vocational Education and Training, Inequality and the Labour Market in Brazil and India: A policy Review
WP 03/2015	Maria Cristina Cacciamali Gerry Rodgers Vidya Soundararajan Fabio Tatei	Wage Inequality in Brazil and India: A Quantitative Comparative Analysis
WP 02/2015	Maria C. Cacciamali, Taniya Chakraborty, Gerry Rodgers, Fabio Tatei	Minimum Wage Policy in Brazil and India and its Impact on Labour Market Inequality
01/2015	Ajit K. Ghose	India Needs Rapid Manufacturing – Led Growth
07/2014	Ann George and Pravez Alam	After-Life of Mobile Phones: Waste or Return to Production?
06/2014	B.N. Goldar	Globalisation, Growth and Employment in the Organised Sector of the Indian Economy
05/2014	Ajit K. Ghose	Globalisation, Growth and Employment in India
04/2014	Alexandre de Freitas Barbosa, Maria Cristina Cacciamali, Gerry Rodgers, Vidhya Soundarajan	Comparative Analysis of Labour Market Inequality in Brazil and India: Concepts and Methods of Analysis (CEBRAP-IHD) Research Project on Labour Market Inequality in Brazil and India

03/2014	Alexandre de Freitas Barbosa, Maria Cristina Cacciamali, Gerry Rodgers, Vidhya Soundarajan, Fabio Tatei, Rogerio Barbosa, J. Krishnamurty	Data Sources for the Analysis of Labour Market Inequality in Brazil and India (CEBRAP-IHD) Research Project on Labour Market Inequality in Brazil and India
02/2014	Sheila Bhalla	Scarce Land: Issues, Evidence and Impact
01/2014	Ajit K. Ghose	India's Services-Led Growth
06/2013	Gerry Rodgers, Sunil K. Mishra and Alakh N. Sharma	Four Decades of Village Studies and Surveys in Bihar
05/2013	Sumit Mazumdar Alakh N. Sharma	Poverty and Social Protection in Urban India: Targeting Efficiency and Poverty Impacts of the Targeted Public Distribution System
04/2013	Dev Nathan, Govind Kelkar and Shivani Satija	Witches: Through Changing Contexts Women Remain the Target
03/2013	Shivani Satija	Violence Against Women in Rural Bihar – A case of four villages
02/2013	Sheila Bhalla	Behind the Post 'Challenge to The Functional Efficiency of India's Established Statistical Institutions
01/2013	Preet Rustagi, Dev Nathan, Amrita Datta and Ann George	Women and Work in South Asia: Changes and Challenges
05/2012	Amrita Datta, Gerry Rodgers, Janine Rodgers and B.K.N. Singh	A Tale of Two Villages: Contrasts in Development in Bihar
04/12	Janine Rodgers	Labour Force Participation in Rural Bihar: A Thirty-year Perspective based on Village Surveys
03/12	K.P. Kannan	How Inclusive is Inclusive Growth in India?
02/12	Sheila Bhalla	Notes on Land, Long Run Food Security and the Agrarian Crisis in India
01/2012	Gerry Rodgers	Understanding Unequal Economic and Social Outcomes in Rural Bihar: The Importance of Caste, Class and Landholding
03/2011	Dev Nathan and Sandip Sarkar	Global Inequality, Rising Powers and Labour Standards
02/2011	Preet Rustagi and Rajini Menon	Gender Asset Gaps and Land Rights in the Context of the Asia-Pacific Region

01/2011	D. Narasimha Reddy	NREGS and Indian Agriculture: Opportunities and Challenges
03/2010	Gerry Rodgers and Janine Rodgers	Inclusion or Exclusion on the Periphery? Rural Bihar in India's Economic Growth
02/2010	R Radhakrishna, C Ravi and B Sambi Reddy	Can We Really Measure Poverty and Identify Poor When Poverty Encompasses Multiple Deprivations?
01/2010	Alakh N. Sharma	Political Economy of Poverty in Bihar: Nature, Dimensions and Linkages
03/2009	Dev Nathan and Govind Kelkar	Markets, Technology and Agency: Indigenous People and Change
02/2009	Aseem Prakash	Towards Understanding the Nature of Indian State and the Role of Middle Class
01/2009	C Upendranadh and Rukmini Tankha	Institutional and Governance Challenges in the Social Protection: Designing Implementation Models for the Right to Work Programme in India
06/2008	Dipak Mazumdar and Sandip Sarkar	The Employment Problem in India and the Phenomenon of the Missing Middle
05/2008	Aseem Prakash	Social Conflict, Development and NGOs: An Ethnographic Study
04/2008	Balwant Singh Mehta and Kerren Sherry	Wages and Productivity of Child Labour: A Case Study of Zardosi Industry
03/2008	Sheila Bhalla	Scarce Land: The Cases of All India and West Bengal
02/2008	T.S. Papola and R.P. Mamgain	Market Access to Small Primary Producers: A Study of Vegetable Growers in the Supply Chain
01/2008	Preet Rustagi	Rural Child Labour Markets in India: Nature of Child Work Participation and Role of the Family
03/2007	Alakh N Sharma	Flexibility, Employment and Labour Market Reforms in India
02/2007	Preet Rustagi	Rural Child Work, Labour and Daily Practices: A Time Use Survey-based Analysis
01/2007	R.P. Mamgain	Growth, Poverty and Employment in Uttarakhand
04/2006	Sheila Bhalla	Common Issues and Common Concerns in the SAARC Region: Employment Generation and Poverty Reduction
03/2006	Rajendra P. Mamgain and Balwant Singh Mehta	Employment and Earnings in Uttaranchal: Trends and Policy Issues
02/2006	Preet Rustagi	Women and Poverty: Rural-Urban Dimensions

01/2006	Dipak Mazumdar and Sandip Sarkar	Growth of Employment and Earnings in the Tertiary Sector, 1983-2000
05/2005	Sheila Bhalla	Recent Developments in the Unorganised Rural Non-Farm Sector
04/2005	Preet Rustagi	Challenges for Economic Empowerment of Women In South Asia
03/2005	Alakh N Sharma	Agrarian Relations and Socio-economic Change in Bihar
02/2005	Preet Rustagi	The Deprived, Discriminated and Damned Girl Child: Story of Declining Child Sex Ratios in India
01/2005	Dipak Mazumdar and Sandip Sarkar	Agricultural Productivity, Off-Farm Employment and Rural Poverty: An Analysis Based on NSS Regions

Institute for Human Development

Institute for Human Development (IHD) is an Indian Council of Social Science Research (ICSSR) Recognised category Institute which undertakes research in the themes relating to employment, livelihood and human development. The Institute engages in analytical and policy research, teaching and training, academic and policy debates, networking with other institutions and stakeholders, and publication and dissemination of the result of its activities. The major themes of the current work of IHD are: growth and employment; education and capabilities; health and nutrition; gender and development; security and vulnerability and governance and institutions.

INSTITUTE FOR HUMAN DEVELOPMENT
256, 2nd Floor, Okhla Industrial Estate, Phase-III
New Delhi - 110020
Tel: +91 11 41064676, +91 9871177540
E-mail: mail@ihdindia.org; Website: www.ihdindia.org

Eastern Regional Centre
C-1, Patel Park, Harmu Housing Colony, Ranchi-834012
Phone/Fax: +91-651-2242874
Email: ihd.ranchi@ihdindia.org