International Inequalities Institute

WORKING PAPER 167

Measuring Inequality of Opportunity in Asia and the Pacific

Gaurav Datt
Monash University

John Nguyen
Monash University

Pedro Salas-Rojo
CUNEF University and International Inequalities Institute, LSE

Francisco H.G. Ferreira
International Inequalities Institute and IZA

Paolo Brunori
Università di Firenze and International Inequalities Institute, LSE

Vito Peragine
University of Bari

Albert Park
Asian Development Bank

Arturo Martinez Jr.
Asian Development Bank

Joseph Albert Nino Bulan
Asian Development Bank

APRIL 2026

Gaurav Datt
Monash University

John Nguyen
Monash University

Pedro Salas-Rojo
CUNEF University and International Inequalities Institute, LSE

Francisco H.G. Ferreira
International Inequalities Institute and IZA

Paolo Brunori
Università di Firenze and International Inequalities Institute, LSE

Vito Peragine
University of Bari

Albert Park
Asian Development Bank

Arturo Martinez Jr.
Asian Development Bank

Joseph Albert Nino Bulan
Asian Development Bank

All publications in our working paper series are available to download free from our website: www.lse.ac.uk/III

International Inequalities Institute
The London School of Economics and Political Science, Houghton Street, London WC2A 2AE

E Inequalities.institute@lse.ac.uk
W www.lse.ac.uk/III

Measuring Inequality of Opportunity in Asia and the Pacific

Gaurav Datt, John Nguyen, Pedro Salas-Rojo, Francisco H.G. Ferreira, Paolo Brunori, Vito Peragine, Albert Park, Arturo Martinez Jr., Joseph Albert Nino Bulan *†

1 April, 2026

Abstract

This paper aims to contribute to an understanding of the extent, nature and persistence of unfair inequality in the Asia Pacific region, building on a rich literature on the measurement of inequality of opportunity (IOp). As part of a project to build a global database of IOp, the paper uses microdata from 39 nationally representative household surveys to present IOp estimates for 14 countries that account for about three-quarters of the region's population. We use consistent data protocols to ensure a high degree of cross-country comparability of IOp estimates. A distinguishing feature of the exercise is the use of machine learning methods to construct IOp estimates, which efficiently balances the risks of potential under- or over-fitting. The results show that, on average, nearly two-fifths of income or consumption inequality across the Asia-Pacific region represents inequality of opportunity attributable to inherited circumstances, though with wide variation across countries, ranging from about a quarter to over half. The cross-country variation in IOp is consistent with a Great Gatsby curve for the Asia-Pacific. A decomposition analysis assesses the relative contributions of different circumstances to IOp.

JEL codes: D31, D63, O15

Keywords: inequality of opportunity, economic mobility, Asia-Pacific, machine learning

Disclaimer: The paper represents the views of the authors which should not be attributed to the respective institutions they belong to.

---

* Gaurav Datt (gaurav.datt@monash.edu) and John Nguyen (johnny.nguyen3@monash.edu) are at Monash University; Pedro Salas-Rojo (pedro.salas@cunef.edu) is at CUNEF University, also affiliated with III (LSE); Francisco Ferreira (F.D.Ferreira@lse.ac.uk) is at the International Inequalities Institute (III) at the London School of Economics (LSE), also affiliated with IZA; Paolo Brunori (P.Brunori@lse.ac.uk) is at Università di Firenze and the III at LSE; Vito Peragine (vitorocco.peragine@uniba.it) is at the University of Bari; Albert Park (afpark@adb.org), Arturo Martinez Jr. (amartinezjr@adb.org) and Joseph Albert Nino Bulan (jbulan@adb.org) are at the Asian Development Bank. Salas-Rojo acknowledges financial support from the Velez Reyes Foundation and the Cesar Nombela grant 2025-T1/PH-HUM-36094.

† For their helpful comments or other forms of help, the authors would like to thank Domenico Moramarco and Pedro Torres López.

1. Introduction

The literature on the measurement of inequality of opportunity (IOp) has come a long way since the early contributions of Roemer (1998) to the conceptual foundations of the idea of IOp (as distinct from the inequality of outcomes) and the "first-generation" empirical studies of IOp exemplified by Bourguignon, Ferreira and Menendez (2007) and Ferreira and Gignoux (2011).[1] Starting with the fundamental notion that not all inequality of outcomes necessarily represents inequality of opportunity, the literature on the measurement of IOp has sought to isolate the component of outcome inequality that could be considered ethically reprehensible, as the component that arises from circumstances beyond individual control rather than from responsible choices and effort. A large and growing number of contributions to the empirical IOp literature have not only enhanced our understanding of the key challenges in the measurement of IOp, but also given us an empirical sense of how much of observed inequality could be deemed to signify IOp. It would be fair to say that the literature has now matured to a point where it is meaningful to talk about cross-country estimates of IOp.

This paper is a contribution to the cross-country measurement of IOp, focusing on countries in the Asia-Pacific region. It is part of a larger (ongoing) project, Global Estimates of Opportunity and Mobility (GEOM), which is building a public data repository of comparable estimates of IOp around the world. Currently, the database contains IOp estimates for 72 countries (drawing upon nearly 200 household surveys) accounting for about two-thirds of the global population.[2] This paper presents findings on IOp for 14 countries in the Asia-Pacific region, accounting for about three-quarters of the region's population. Despite the growth in the empirical IOp literature, to date the coverage of countries for this region has been relatively limited. The paper contributes to filling this gap. For each country, we present IOp estimates for the latest year for which relevant data is available. For some countries where we have several rounds of data, the paper also presents estimates tracking IOp at multiple points of time.

Besides providing new estimates of IOp for Asia-Pacific countries, a distinguishing feature of the paper (and more generally of the GEOM database) is its use of machine learning methods to construct more robust estimates of IOp that mitigate the twin problem of under- or over-

[1] For two excellent surveys of the IOp literature, see Ramos and Van de Gaer (2015) and Roemer and Trannoy (2016).

[2] The Global Estimates of Opportunity and Mobility (GEOM) database is an initiative of researchers from the London School of Economics, University of Bari, Monash University and the Asian Development Bank. See Ferreira et al. (2026) for a detailed introduction to the GEOM database and key findings.

fitting a set of circumstance categories to the data. A second distinguishing feature of the current exercise is its aim and attempt to construct IOp estimates that are as comparable as possible across countries. This translates into consistent protocols for selection and definition of outcome and circumstance variables, and a consistent application of the machine learning algorithm across countries. These methodological and operational features, which underlie the results presented in this paper, set it apart from the existing empirical IOp literature.

The organization of the paper is as follows. Section 2 introduces our methodological approach to IOp measurement. Section 3 discusses the data for the Asia-Pacific region and the estimation protocols. We then discuss our results in three parts. Section 4 presents aggregate results on IOp estimates for the Asia-Pacific countries. Section 5 focuses on the relative contributions of different circumstance variables to measured IOp based on a decomposition analysis. Section 6 uses the structure of conditional inference trees (CIT) derived from the machine learning algorithm discussed in Section 2.2, to drill further into the specific typology of circumstance groups for a couple of countries on an illustrative basis.

2. Measurement methodology

2.1. Conceptual foundation

Central to the measurement of IOp is the idea that inequality of opportunity represents the inequality of outcomes that is attributable to inherited circumstances or more generally circumstances beyond individual control (hereafter, simply referred to as circumstances). The ethical foundations of this idea rest on both the compensation and reward principles. The compensation principle asserts that outcome differences arising from circumstances are unfair and should be compensated, whereas the reward principle requires that outcome differences due to effort be respected. A critical step in the measurement of IOp therefore is the identification of a set of circumstances that could be deemed as outside individual control. Unsurprisingly, a major part of the effort in the empirical IOp literature has accordingly been devoted to the identification of a plausible set of circumstances for any given population or country.

The compensation principle however can pave the way for several possible measures of IOp. We focus in this paper on what has come to be known in the literature as the class of *ex-ante* IOp measures, which correspond to the *ex-ante* version of the compensation principle (Fleurbaey and Peragine 2013). In simple terms, ex-ante compensation requires outcome

differences due to circumstances to be compensated, *prior* to the realization of different levels of effort by individuals. Ex-ante IOp thus obviates the need to observe individual effort levels. Empirical implementation proceeds by first partitioning the population into a finite number of "types", where each type is a category representing a unique combination of circumstances. By construction, the types are mutually exclusive and exhaustive so that each individual in the population of size n belongs to one of, say, K types in the population, with $\sum_1^K n_k = n$ where n_k is the number of individuals belonging to type $k \in \{1, \dots, K\}$.

Following van de Gaer (1993), the value of the opportunity set available to an individual is given by the expected outcome for the circumstance-type the individual belongs to, or in other words, by the expected outcome conditional on type. *Equality* of opportunity based on the ex-ante compensation principle thus demands equality of outcomes conditional on type, across all types (Put differently, it requires expected outcomes to be independent of circumstances). Conversely, the extent to which conditional expected outcomes differ across types represents ex-ante IOp.

This leads to a simple measurement approach. Taking income to be the outcome variable of interest and starting with an actual income distribution, $y = \{y_1, y_2, \dots y_n\}$ for a population comprising n individuals, we construct a "smoothed" counterfactual distribution, obtained by replacing individual incomes with their "smoothed" value, i.e. their type means or expected values conditional on type:

(1) $y^{sm} = \{\mu_1 1_{n_1}, \mu_2 1_{n_2}, \dots \mu_k 1_{n_K}\}$

where μ_k is the mean income of type k, and 1_{n_k} is a 1-vector of size n_k.

This amounts to eliminating within-type inequality in the income distribution, so that y^{sm} only reflects inequality between types. Ex-ante IOp then amounts to the remaining inequality in the smoothed distribution. If $I(y)$ denotes an inequality measure defined on distribution y, then absolute and relative ex-ante IOp are evaluated as:

(2) $IOp_{Abs}^{ex\_ante} = I(y^{sm})$

(3) $IOp_{Rel}^{ex\_ante} = I(y^{sm})/I(y)$

The empirical implementation of this approach requires three critical choices: (i) the selection of a set of observable circumstance variables for the population, (ii) the identification of a plausible set of distinct types based on observable circumstance variables, and (iii) the choice

of a preferred measure of inequality. The following subsections discuss issues pertaining to these elements of empirical implementation and the approach we take to address them.

2.2. The Machine Learning (ML) algorithm

The first-generation empirical IOp studies copped the criticism that they underestimate IOp due to the partial observability of circumstances (Kanbur and Wagstaff 2016). Since the full set of individual circumstances are arguably not observed in survey data, it was reasoned that the estimates of IOp in the literature should be considered a lower bound. This led to efforts to expand the set of observable circumstances to include, for instance, a set of childhood circumstances before the individual reaches the "age of consent" (Hufe et al. 2017).

At the other end, the ensuing empirical IOp literature also raised the possibility of overfitting circumstances to the data. For instance, Brunori, Ferreira and Peragine (2013) survey of IOp estimates for 41 countries noted the number of circumstance types identified ranged from a low of 6 to as high as 7680. It is obvious that the share of IOp in overall inequality is increasing in the number of types. In the limiting case, when the number of types equals the number of individuals in the sample, and each type is a singleton, the share of IOp in overall inequality is 100% by construction. However, the finer the partition of the population into types, the fewer observations each type contains, and the greater the imprecision in the estimated moments of type outcomes. Thus, a larger number of types, while potentially reducing the downward bias due to omitted circumstances, can introduce an upward bias by increasing the sampling variance of the counterfactual distribution of circumstance-specific outcomes (Chakravarty and Eichhorn 1994; Brunori, Peragine and Serlenga 2019). With such overfitting, improvement in the within-sample predictions of expected counterfactual incomes comes at the expense of poorer predictions out-of-sample. This raises the likelihood that the estimated relationship between types and opportunities would not be verified in alternative samples representative of the same population, leading to a potential misattribution of the higher sampling variance that comes with more types to the calculated IOp.

Constructing an appropriate partition of the population into types that mitigates the twin perils of underfitting and overfitting is thus a key empirical challenge in IOp measurement. Recent IOp literature addresses this issue with a data-driven approach by applying machine learning (ML) tools to the task of identification of types. The ML approach is based on the idea of a conditional inference tree (CIT) proposed by Hothorn et al. (2006) and first implemented for

IOp estimation by Brunori, Hufe and Mahler (2023). It proceeds by constructing a series of binary splits or partitions of the data (branches of the tree) by using a non-parametric independence test to detect the circumstance most correlated with the outcome, and then by looking for the circumstance category that yields the maximum difference in expected outcome between the partitions (minimum p-value for a difference in means). The algorithm grows the tree by repeating the process at each node yielded by a previous binary split. The algorithm stops when it is no longer possible to reject the hypothesis of independence across the outcome and the circumstances considered in all splits. The resulting set of terminal nodes (leaves) of the tree identifies the final set of circumstance types in the data. The detailed steps of the ML algorithm to construct CITs for ex-ante IOp are described in Appendix A.

CITs though unbiased can themselves be subject to a degree of sample variability.[3] The ML approach addresses this by constructing random forests or a collection of trees using resampling techniques. We use the conditional inference random forest algorithm proposed in Hothorn, et al. (2006), which involves drawing multiple subsamples from the original data and constructing a tree for each subsample. Random forest estimates of IOp are then obtained by averaging the structure across all trees (Brunori, Hufe, and Mahler 2023).

2.3. Measure of inequality and "between-within" inequality decomposition

Estimation of IOp also requires us to specify a measure of inequality. There are potentially several candidates here. The empirical IOp literature has often preferred additively decomposable inequality measures belonging to the Generalized Entropy class, especially Mean Log Deviation (MLD) or the Theil index. This is related to the interpretation of ex-ante IOp in terms of inequality decomposition, as representing the between-type component of inequality. Additive decomposability ensures that overall inequality is an exact sum of (i) between-group inequality obtained using a smoothed distribution that suppresses all variation within groups and (ii) within-group inequality obtained using a standardized distribution that suppresses variation in group means but preserves all variation within groups. Decompositions of inequality measures such as the Gini index that do not satisfy this property will in general also have a non-negative residual term.[4]

[3] Such variability implies that if estimated on different samples representative of the same population, they may produce different partitions.
[4] For the Gini index, this residual term is strictly positive if there is any overlap of incomes across groups, and zero for non-overlapping groups (Foster and Sen 1999).

However, the conceptual foundation of ex-ante IOp as between-type inequality (as discussed in section 2.1) does not necessitate an exact decomposition of the inequality measure. For the results presented in this paper, our preferred measure of inequality for IOp estimates is the Gini index. The preference for the Gini index is for several reasons. First, the Gini index is the most widely recognized inequality measure, which has a familiar representation in terms of the Lorenz curve. Second, it has an intuitively straightforward interpretation: twice its value gives the expected difference in two randomly picked incomes as a proportion of mean income.[5] Third, the Gini index is bounded between 0 and 1, which eases interpretation; MLD on the other hand is not bounded above. Fourth, MLD is more sensitive to extreme values than the Gini, which makes it a relatively blunt measure to identify the degree of inequality between groups. Noting that ex-ante IOp is based on a smoothed distribution with type means replacing individual incomes, MLD will typically register a larger reduction in inequality in moving from the actual to the smoothed distribution, and hence a smaller relative IOp than the Gini (Brunori, Palmisano and Peragine 2019).

2.4. Alternative decompositions of the Gini index and measures of ex-ante IOp

The conceptual development of ex-ante IOp as the between-type component of inequality also links up with the general issue of decomposition of inequality measures. While the Gini index does not permit an exact decomposition, two alternative decompositions of the Gini into between and within components are possible. One of these is based on a "smoothed" distribution, which was already introduced above in (1).

The other is based on a "standardized" distribution, which is obtained by replacing individual incomes by their "standardized" value, i.e. individual incomes multiplied by the ratio of overall mean to their type mean:

$$(4) \quad y^{st} = \{y_1 . \left(\frac{\mu}{\mu_1}\right), y_2 . \left(\frac{\mu}{\mu_2}\right), \dots y_K . \left(\frac{\mu}{\mu_K}\right)\}$$

This amounts to eliminating between-type inequality, so that y^{st} only reflects inequality within types.

To introduce the two alternative decompositions of the Gini index, the following notation is helpful. Let $L(p)$ represent the Lorenz curve for distribution y, ranked by ascending income

[5] For instance, a Gini of 0.4 implies that two randomly picked incomes differ by 80% of the mean.

levels, where $p = p(y)$ is the cumulative proportion of population with incomes less than or equal to y. Let $G(y)$ denote observed inequality measured by the Gini index defined on y.

(5) $G(y) = 1 - 2\int_0^1 L(p)dp$

In addition, following Fleurbaey et al. (2024), let's also define the ***lexicographic parade of y***, denoted y^{lex}, which is obtained by sorting y first by groups in ascending order of their group means and then within each group sorting by ascending level of income.

(6) $y^{lex} = \{y_1^1, y_2^1, .. y_{n_1}^1; y_1^2, y_2^2, .. y_{n_2}^2; ... y_1^K, y_2^K, .. y_{n_K}^K\}$

where $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_K$ and $y_1^k \leq y_2^k \leq .. \leq y_{n_k}^k$ for all $k \in \{1,2, ... K\}$.

Let $C^{st}(p)$ and $C^{lex}(p)$ denote the concentration curves of y^{st} and y^{lex} respectively.

We can now write the two decompositions of the Gini.

First, the *sm-decomposition* of Gini using the smoothed distribution y^{sm} is written as:

(7) $G = G_{BET}^{sm} + G_{WIT}^{sm} + G_{RES}$

where:

(8) $G_{BET}^{sm} = G(y^{sm})$

G_{BET}^{sm} is twice the area between the perfect equality line and the Lorenz curve corresponding to y^{sm}, and

(9) $G_{WIT}^{sm} = \sum_k p_k s_k G_k$ where G_k is within-group Gini index for type k, and p_k, s_k denote the population share (n_k/n) and income share $((n_k\mu_k)/(n\,\mu))$ of type k respectively.

G_{WIT}^{sm} is twice the area between the Lorenz curve corresponding to y^{sm} and $C^{lex}(p)$. This Gini decomposition is the same as in Bhattacharya and Mahalanobis (1967).

Second, the *st-decomposition* of Gini using the standardized distribution y^{st} is written as:

(10) $G = G_{BET}^{st} + G_{WIT}^{st} + G_{RES}$

where:

(11) $G_{WIT}^{st} = G(y^{st}) = \sum_k p_k^2 G_k$

G_{WIT}^{st} is twice the area between the perfect equality line and $C^{st}(p)$, and

(12) $G_{BET}^{st} = 2\int_0^1 \Big(C^{st}(p) - C^{lex}(p)\Big)dp.$

G^{st}_{BET} is twice the area between $C^{st}(p)$ and $C^{lex}(p)$.

Note the residual is twice the area between $C^{lex}(p)$ and $L(p)$:

(13) $G_{RES} = 2\int_0^1 \Big(C^{lex}(p) - L(p)\Big)dp$

with the following properties:

(i) Both *sm*- and *st*-decompositions, (7) and (10), are not exact insofar as $G_{RES} \geq 0$

(ii) The residual is the same in both decompositions (Fleurbaey et al. 2024).

(iii) $G_{RES} = 0$ if there is no overlap between groups, in which case $C^{lex}(p)$ and $L(p)$ coincide. It is positive otherwise.

From (ii), it also follows that $G^{sm}_{BET} + G^{sm}_{WIT} = G^{st}_{BET} + G^{st}_{WIT}$.

As noted above, ex-ante IOp has been identified in the literature as between-type inequality. The above discussion however implies two alternative measures of between-type inequality for the Gini index: G^{sm}_{BET} and G^{st}_{BET}, leading to two potential measures of ex-ante IOp:

(14) $IOp^{sm}_{ex\_ante} = G^{sm}_{BET}$

(15) $IOp^{st}_{ex\_ante} = G^{st}_{BET}$

These two measures are not equal to each other even when $G_{RES} = 0$, i.e., when there is no overlap between types. Since the residual is the same in both *sm*- and *st*-decompositions,

(16) $G^{sm}_{BET} - G^{st}_{BET} = G^{st}_{WIT} - G^{sm}_{WIT} = \sum_k p_k(p_k - s_k)G_k \gtreqless 0$

The sign of $G^{sm}_{BET} - G^{st}_{BET}$ is in general indeterminate. It is likely to be positive if poorer (lower-mean) types also have larger population shares and higher within-type inequality (i.e. higher p_k and G_k).

The choice between G^{sm}_{BET} and G^{st}_{BET} can be viewed as a choice between two alternative characterizations of the within/between inequality. $IOp^{sm}_{ex\_ante}$, sometimes also referred to as the ***direct*** measure of ex-ante IOp, has often been preferred in empirical applications (Ferreira and Guignoux 2011, Brunori, Hufe and Mahler 2023). Since G_{RES} is positive (for overlapping distributions), it has been argued that $IOp^{sm}_{ex\_ante}$ gives a lower bound for ex-ante IOp.

However, this is not a valid argument. For the same reason (viz., $G_{RES} \geq 0$), $IOp^{st}_{ex\_ante}$ can also be interpreted as a lower bound.[6]

However, there is another potential argument in favour of $IOp^{sm}_{ex\_ante}$. This follows from the properties of the between and within components in the two Gini decompositions. From (7)-(9), it is easily checked that in the *sm*-decomposition, the between component, G^{sm}_{BET}, is independent of the within component, but the within component, G^{sm}_{WIT} is not independent of the between component because it depends on income shares s_k (and hence μ_k). Similarly, from (10)-(12), it can be checked that in the *st*-decomposition, the within component, G^{st}_{WIT}, is independent of the between component as it only depends on p_k and G_k, but the between component, G^{st}_{BET} is not independent of the within component. Hence, if the independence of a measure of ex-ante IOp from within-type inequality is deemed a desirable property, one could construct an argument for preferring the use of $IOp^{sm}_{ex\_ante}$. The argument however rests on the notion of independence rather than the idea of a lower bound.

How much these two ex-ante IOp measures differ from each other is ultimately an empirical matter. Later in section 4.1, we compare the two measures and find them to be very close to each other. In light of this, we will use $IOp^{sm}_{ex\_ante}$ as our preferred measure of IOp, given also its property of independence from within-type inequality.

2.5. Shapley-Shorrocks decompositions of the contributions of circumstances to IOp

Understanding the contribution of each circumstance variable to the outcome's variability requires the joint evaluation of expected effects and the distribution of characteristics in the population. We use Shapley-Shorrocks decompositions to provide an evaluation of the relative contribution of each circumstance to measured IOp (Shorrocks 2013). These assessments offer descriptive insights and should not be interpreted causally due to the intrinsically descriptive nature of the analysis. To improve the robustness of our estimates we perform the decomposition across 100 trees on different subsamples, and then average across iterations, mitigating the potentially high variance associated with a single tree.

In simple terms, the Shapley-Shorrocks decomposition of IOp assesses how much each of the core set of circumstance variables, such as c_k, contributes to the IOp for an outcome (e.g., income y). Specifically, the contribution of c_k is determined by the reduction in $I(y^{sm})$

[6] Put differently, interpreting $IOp^{sm}_{ex\_ante}$ as the preferred lower bound implicitly assumes that the within-component should be defined as $\sum_k p_k s_k G_k$ (and not $\sum_k p_k^2 G_k$).

resulting from its elimination from the set of circumstance variables. Since circumstance variables can be eliminated in varying order, Shapley-Shorrocks decompositions evaluate a variable's contribution as the average reduction in $I(y^{sm})$ over all possible sequences of that variable's elimination. Appendix A provides details of the computational steps for Shapley-Shorrocks decompositions.

3. Data and estimation protocols

3.1. Household survey data and circumstance variables

Our data for the Asia-Pacific region come from nationally representative household surveys since 1999 that contain the necessary information for constructing measures of IOp. Typically, the core set of observable circumstance variables in the empirical IOp literature has included variables related to parental education, parental occupation, gender, race or ethnicity, and place of birth. All ascriptive characteristics that cannot be influenced by individual choices are, by definition, outside of individual control. At a practical level, what is included is driven by the available data, and there are notable variations in the scope of surveys for different countries. Since the aim of the global GEOM project as well as the IOp estimates for the Asia-Pacific is to construct estimates that are as comparable as possible across countries, in our empirical implementation, we stick for the most part to the above core set of circumstance variables. Specifically, we consider the following seven circumstance variables:

- Area of birth
- Sex
- Father's education
- Mother's education
- Father's occupation
- Mother's occupation
- Ethnicity, race or religion[7]

These are all categorical variables. Other than sex, which for our data is a binary variable, the other variables consist of a varying number of subcategories that are also diverse across countries. For instance, the subcategories for area of birth, ethnicity or race will naturally differ across countries. This raises the issue of how disaggregated these categorical variables ought

[7] In most cases, we use the religion of the parent. If that is not available, we use the religion of the household head. In particular cases, this variable can be extended to language spoken at home or similar attribute.

to be: for example, should we distinguish five or fifty areas of birth within a country, and should this number be the same across countries. For such circumstance variables, there is no a priori justification for imposing a uniform level of aggregation, since more or less disaggregated categories may reflect legitimate and consequential socio-ethnic-demographic diversity within a country. We instead follow a data-driven approach using the ML algorithm discussed above to determine the suitable level of aggregation in each case.

The inclusion of a survey in our database is subject to several criteria.

- The survey should be nationally representative, should pertain to year 2000 or more recent years, and have a sample size of at least 1,000 observations with non-missing data.
- The survey should contain information on income or consumption of the household and household size.
- The survey should contain information on at least five of the aforementioned seven circumstance variables for household members.
- As a pragmatic rule, each of these categorical variables is allowed a maximum of 25 categories, merging smaller sub-categories if needed. Any further aggregation is data-driven as determined by the machine learning algorithm.
- The survey should also have information on the age of the household head (or primary respondent if household head is not identified) needed for age adjustment to the outcome variable described below.

There are 14 countries for which we have survey data satisfying the above criteria. Since these include the more populous countries of China, India and Indonesia, the included countries account for 74% of the total population of the Asia-Pacific region and 41% of the global population. For seven of the 14 countries, we have comparable survey data available at more than one point in time. Altogether, the dataset relates to 39 country-years. The list of included countries and years, together with survey data sources are shown in Appendix B.

3.2. Income or consumption as the outcome variable

A distinguishing focus of this exercise is the estimation of IOp in the income or consumption space. We use equivalized household income as our preferred measure of well-being, but if this is not available, then we use equivalized household consumption expenditure as an alternative. For 12 of the 14 Asia-Pacific countries, the outcome measure is income-based; for

two (Indonesia and Timor-Leste) it is consumption-based. Equivalized household income (consumption) is calculated as total household income (consumption) divided by the square root of household size. The use of equivalized rather than per capita measure is intended to account for economies of scale in household consumption related to the widely-recognized idea that two can live more cheaply than twice as one.

The focus on the income or consumption space, as distinct from related literature on intergenerational mobility (IGM) in terms of education outcomes,[8] has two (methodological) advantages. First, though largely reflecting data limitations, educational mobility and IGM have mostly focused on years of schooling, even as there is now a growing recognition of large disparities in the quality of education and their impact on learning outcomes. Second, a related issue is that the same years of schooling can be associated with very different income opportunities in different settings, not only due to the varying quality of schooling but also due to several other differences in the learning environment and labor market conditions. A direct focus on IOp in the income space sidesteps these difficulties in inferring income mobility from educational mobility.

The national household surveys only allow us a measure of equivalized or total income/ consumption for each sample household. The use of equivalized income or consumption helps make an allowance for economies of scale. But it does not address inequality of income or consumption within the household, the measurement of which is fraught with conceptual and practical difficulties due to joint production (e.g., in family farms) or consumption (e.g., common kitchen or other household public goods) by several family members. In light of this, there is not much option other than to assign the equivalized income of the household to each household member. This is what we do, though subject to the following age adjustment.

3.3. Age adjustment to the outcome variable

It is well-known that individual income trajectories are influenced by life cycle effects. Typically, individual incomes increase during prime working age before plateauing and declining in old age. Thus, in a cross-section sample of adult household members of different ages, a part of the observed income dispersion is on account of life cycle effects that arguably

[8] See, for instance, van der Weide et al. (2021) and World Bank (2021) on the Global Database on Intergenerational Mobility (GDIM) with estimates of intergenerational mobility in terms of years of schooling for 153 countries.

do not represent inequality of opportunity.

To address this, prior to estimation (but after determining the analysis sample), we adjust our outcome variable (equivalised income or consumption) by age. This adjustment controls for systematic correlations between age and inequality. For this, we employ a regression where the income or consumption of individual j depends on their age and age squared.

$$lny_j = \alpha + \beta age_j + \gamma age_j^2 + \varepsilon_j$$

We restrict the regression to household heads (or primary respondents, if household head is not identified in the survey data) as this better approximates life cycle effects. We use the estimated parameters to predict lny_i for all individuals and obtain residuals $\widehat{\varepsilon_i} = lny_i - \widehat{lny_i}$. The adjusted income or consumption (net of the effect of age) for all individuals is then obtained as: $y_i^{adj} = \exp(\hat{\alpha} + \widehat{\varepsilon_i})$.[9]

4. Aggregate IOp results

4.1. Comparison of smoothed and standardized distribution-based estimates of IOp

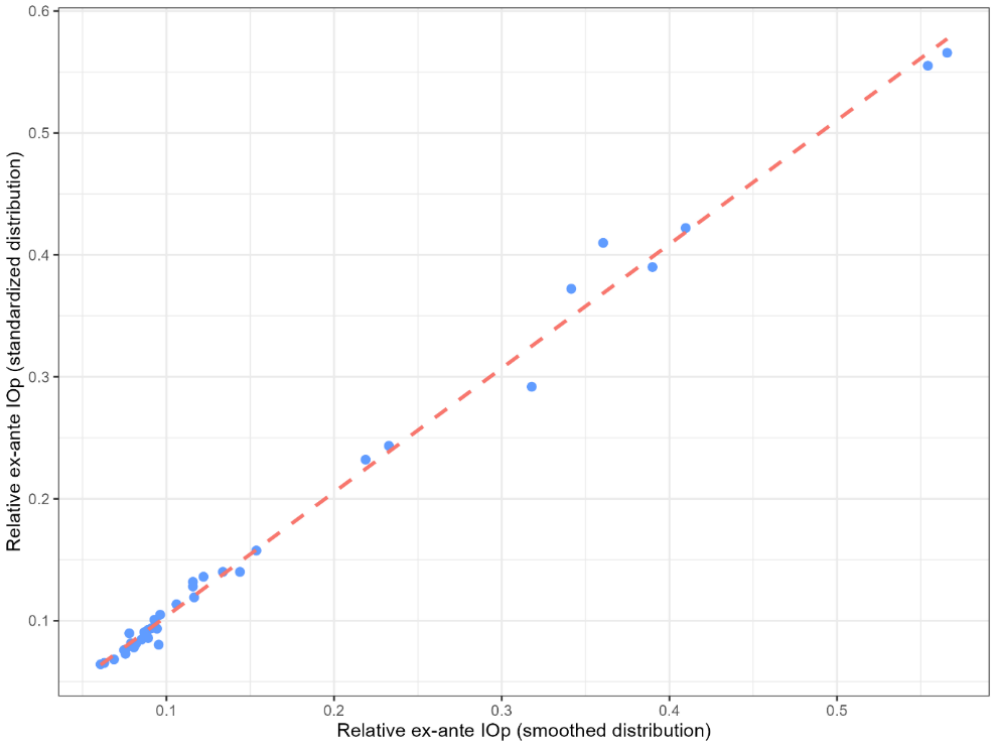
As discussed in Section 2.4, corresponding to the alternative decompositions of the Gini index, two alternative measures of IOp can be constructed corresponding to the smoothed and standardized distributions respectively. The former replaces individual incomes by mean income of their respective type. This removes within-type inequality, so the resulting ex-ante IOp measure captures inequality between types. The latter scales individual incomes by the ratio of the overall mean income to the mean income of their type. This eliminates between-type inequality while retaining within-type inequality; the ex-ante IOp measure in this case is calculated as the increase in inequality obtained when the original type means are reintroduced. The standardized-distribution approach therefore yields a different measure of between-type inequality from the smoothed-distribution method, and the magnitude of this difference is an empirical matter.

Figure 1 compares the IOp estimates generated by these two approaches for the Asia–Pacific region. It plots the CIT estimates of relative IOp derived from the standardized distribution

[9] This effectively amounts to using predicted income at age 0. One could have predicted incomes at another fixed age, for instance, say 40 years. But given the scale invariance property of most inequality measures, including the Gini, this makes no difference to the estimates of inequality or IOp.

method against those derived from the smoothed distribution approach for 38 country-years for the Asia-Pacific region.[10] It turns out that the two sets of estimates are very close to each other, with a correlation coefficient of 0.997. In light of this, and the property that between-type inequality derived from the smoothed distribution is independent from within-type inequality, we use the smoothed distribution-based measures as our preferred estimates of ex-ante IOp.

Figure 1: Comparison of relative IOp estimates based on smoothed and standardized distributions



Note: The Figure plots conditional inference tree (CIT) estimates of relative IOp derived from the standardized distribution against those derived from the smoothed distribution. Each point represents one of the 38 country-years in the Asia-Pacific dataset for which CITs could be constructed.

4.2. Efficacy of machine learning-based estimates of IOp

Even with a limited number of circumstance variables, each admitting several categories, the potential combinations increase rapidly. For instance, for the six circumstance variables used in this study, we could have two categories for sex, seven categories each for mother's and

[10] The Figure is based on 38 (not 39) country-years because CIT estimates for Kazakhstan could not be constructed due to its small sample size (though random forest estimates of IOp were feasible).

father's education, six categories each for mother's and father's occupation, up to ten categories for ethnicity and up to 25 categories for birth area.[11] The total number of potential combinations of these categories thus could be as large as 2×7×7×6×6×10×25 or 882,000. This is orders of magnitude greater than the sample size of household surveys. For the Asia-Pacific countries, the analysis samples range from a little over 1,000 for several Central Asian countries to the largest sample of about 98,000 for India (column 4 in Table 1).

Table 1: Sample sizes and the number of types in fully saturated models and conditional inference trees

				Fully saturated (FS) model		Conditional Inference Tree (CIT)	
Country	**Year**	**Outcome**	**Sample Size**	**# of FS Types**	**Sample size /FS Type**	**# of CIT Types**	**Sample size /CIT Type**
Armenia	2016	Income	1,221	753	2	3	407
Australia	2019	Income	8,934	2,330	4	7	1,276
China	2018	Income	15,915	4,363	4	9	1,768
Georgia	2016	Income	1,317	710	2	2	659
Indonesia	2014	Cons.	7,481	2,317	3	8	935
India	2012	Income	97,828	6,002	16	53	1,846
Kazakhstan	2016	Income	1,098	661	2	-	-
Kyrgyzstan	2016	Income	1,268	534	2	2	634
South Korea	2019	Income	9,275	3,414	3	10	928
Mongolia	2016	Income	1,394	628	2	3	465
Nepal	2011	Income	11,672	1,258	9	11	1,061
Tajikistan	2016	Income	1,110	484	2	2	555
Timor-Leste	2014	Cons.	12,618	599	21	13	971
Uzbekistan	2016	Income	1,139	634	2	2	570

Note: Column 5 shows the number of circumstance types generated by a fully saturated (FS) model with the maximal number of types identifiable in the sample. Column 6 shows the number of types identified with the machine learning algorithm to construct conditional inference trees. For Kazakhstan, no CIT types were found due to the small sample size, though random forest estimates of IOp (reported later in Table 2) were feasible.

Of course, not all possible combinations are empirically observed; a large number of combinations are empty cells. Column 5 of Table 1 shows the maximum number of distinct circumstance categories observable in the sample data, constituting what would be a fully saturated (FS) model for identifying circumstance types. The number of FS types is nonetheless still very large, ranging from 484 (for Tajikistan) to 6,002 (for India), and the average number of sample observations per FS type is as low as 2 for a number of countries to a maximum of 21 (column 6, Table 1). Clearly, there is little hope of estimating FS type means reliably with so few observations.

[11] The number of categories varies across countries; we use maximal numbers here to illustrate what is at stake in robust estimation of IOp.

The key empirical challenge therefore is how to abridge the FS model to arrive at a more parsimonious, meaningful and robust subset of circumstance types from available data. This is where the strength of the machine learning (ML) approach comes into play. Column 7 of Table 1 reports the number of circumstance types identified through CITs. This ranges from as low as just 2 types for some countries to 53 for India.[12] Column 8 shows that the numbers of sample observations per CIT type are much higher, ranging between 85 and 6,309. This is what allows mean incomes across types, and hence the counterfactual smoothed distribution for evaluating IOp, to be estimated with a much higher degree of precision relative to the FS model.

A detailed comparison of ML methods (conditional inference trees and random forests) with extant parametric and non-parametric methods for the Asia-Pacific countries is beyond the scope of this paper. We refer to Brunori, Hufe and Mahler (2023) for evidence on such comparisons for 31 European countries. Their evidence indicates that relative to other benchmark methods random forests (i) deliver higher prediction accuracy (have lower mean squared error) both for simulated (researcher-specified) data generating processes as well as representative survey data, (ii) provide estimates of IOp that have the lowest expected bias, and (iii) tend to be less sensitive to sample size. The last point is particularly important in a context where some of the countries have a limited sample size (seven countries have a sample size between 1,000 and 1,500). In such cases, IOp estimates obtained with a traditional method, such as standard regression, are likely to be upward biased, while IOp estimated with machine learning methods (in particular conditional inference trees) is likely to be downward biased.

4.3. IOp estimates for the most recent year

Table 2 reports our preferred estimates of absolute and relative ex-ante IOp using the smoothed distribution approach. It shows estimates for the most recent year for which survey data are available for each of the 14 countries in the Asia-Pacific region. (The full set of IOp estimates for all 39 country-years are presented in Table C1 in Appendix C.)

Table 2 presents both the conditional inference trees (CIT) and the random forest (RF) estimates. The RF estimates of IOp are usually, though not always, higher than the CIT

[12] The number of CIT types should nonetheless be interpreted with caution. As noted in section 2.4, the identification of CIT types is subject to sample variability, which is why the random forest estimates are our preferred estimates of IOp. An extreme example of this is Kazakhstan. Owing to its small sample size, while no CIT tree for the given confidence level of the Bonferroni-adjusted p-value (Appendix A) could be identified, random forest estimates based on 100 subsample draws were still feasible.

estimates. The reason is that despite their flexible approach to identifying types, CITs are nonetheless based on a single sample. The binary splitting algorithm of a CIT could thus miss the informational content of a particular circumstance category which could have been picked up with repeated sampling. By using resampling methods to construct many trees, random forests on the other hand increase the likelihood that all circumstance categories with informational content will feature in one or another CIT (Brunori, Hufe and Mahler 2023). This also makes the RF estimates of IOp more robust in the presence of sample variability and is the reason why we focus on them in the following discussion.

Table 2: Ex-ante Inequality of Opportunity in the Asia-Pacific (using the Gini index)

					Absolute Ex-ante IOp		Relative Ex-ante IOp		
	Country	**Year**	**Outcome variable**	**Total inequality (Gini)**	**RF**	**CIT**	**RF (%)**	**CIT (%)**	**Number of types (CIT)**
	Low inequality countries								
TIM	Timor-Leste	2014	Cons.	0.282	0.101	0.117	36	41	13
TJK	Tajikistan	2016	Income	0.309	0.087	0.061	28	20	2
KAZ	Kazakhstan	2016	Income	0.338	0.081	-	24	-	-
KOR	South Korea	2019	Income	0.351	0.121	0.106	35	30	10
AUS	Australia	2019	Income	0.355	0.101	0.094	28	27	7
	Average			**0.327**	**0.098**	**0.095**	**30**	**30**	**8**
	High inequality countries								
ARM	Armenia	2016	Income	0.412	0.179	0.116	44	28	3
IDN	Indonesia	2014	Cons.	0.428	0.126	0.116	29	27	8
KGZ	Kyrgyzstan	2016	Income	0.448	0.143	0.078	32	17	2
UZB	Uzbekistan	2016	Income	0.46	0.182	0.095	40	21	2
GEO	Georgia	2016	Income	0.468	0.211	0.122	45	26	2
MNG	Mongolia	2016	Income	0.47	0.181	0.144	39	31	3
CHN	China	2018	Income	0.497	0.219	0.194	44	39	9
IND	India	2012	Income	0.527	0.279	0.292	53	55	53
NPL	Nepal	2011	Income	0.538	0.216	0.219	40	41	11
	Average			**0.472**	**0.193**	**0.153**	**41**	**32**	**10**
Average for 14 countries				**0.420**	**0.159**	**0.135**	**37**	**31**	**10**

Note: The Table shows for each country the overall Gini index of income or consumption, the ex-ante random forest (RF) and conditional inference tree (CIT) measures of IOp. Columns 6 and 7 show the absolute IOp measures, while columns 8 and 9 show measures of relative IOp (i.e. absolute IOp as a percentage of the overall Gini index). Low (High) inequality countries are those with Ginis of income or consumption below (above) 0.4. For Kazakhstan, no CIT types were found due to the small sample size.

Several key findings stand out from the IOp estimates in Table 2.

1. The headline finding is that *on average* IOp represents 37% of observed in-country inequality for the Asia-Pacific region. This corresponds to an average Gini of 0.42 across

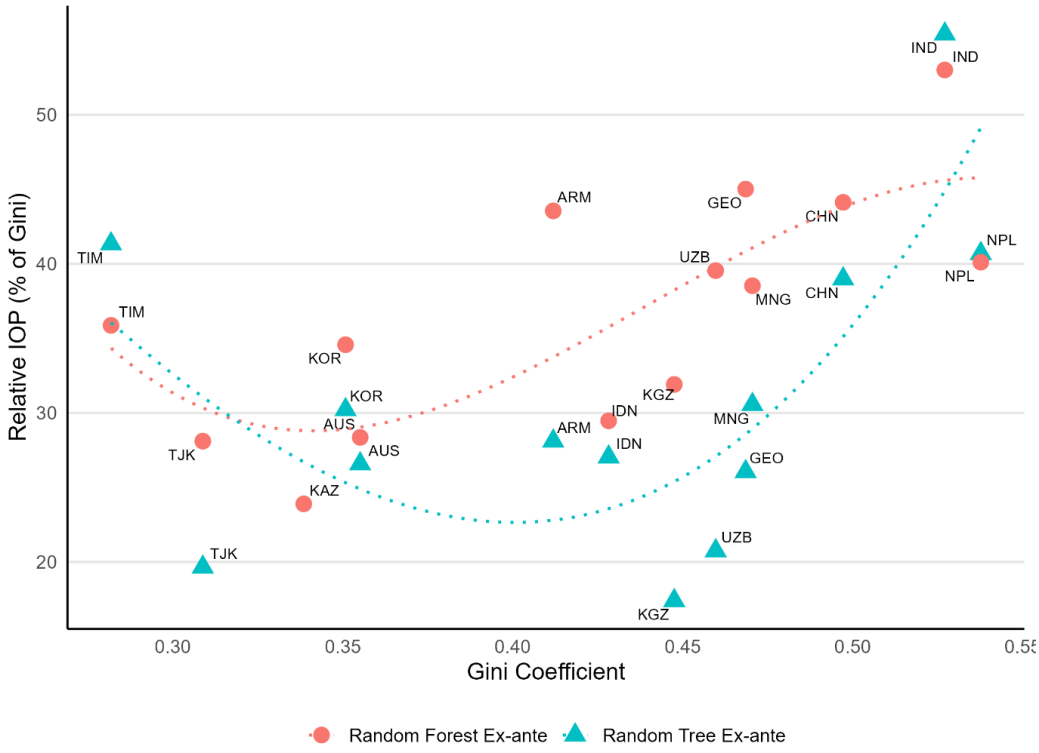
the 14 countries, of which the average between-type inequality representing ex-ante IOp is 0.16.

2. There is a large variation in IOp across countries. Absolute IOp varies in line with the total (observed) inequality. It is thus more helpful to look at relative IOp, which itself varies widely across countries, ranging from the lowest value of 24% for South Korea to the highest of 53% for India. Other countries at the lower end of relative IOp are Australia, Tajikistan, Kazakhstan and Indonesia with relative IOp of under 30%, while countries at the higher end include China, Armenia, India, Nepal, Georgia and Uzbekistan with relative IOp above 40%.

3. Observed levels of total inequality vary also widely across the 14 countries with income Ginis ranging from 0.31 for Tajikistan to 0.54 for Nepal, and consumption Ginis ranging from 0.28 for Timor-Leste to 0.43 for Indonesia. Broadly grouping them into "low" and "high" inequality countries, as those with Gini indices of less or greater than 0.4 respectively (while recognizing the arbitrariness of this threshold), is suggestive of a pattern: low (high) inequality countries tend to have low (high) relative IOp's. Thus, while low-inequality countries with an average Gini of 0.33 have an average relative IOp of 30%, the high-inequality countries with an average Gini of 0.47 have a relative IOp of 41%. In other words, the share of inequality of opportunity in overall inequality of outcomes in high-inequality countries is about 37% higher than in low-inequality countries.

4. This positive association between income inequality and the share of inequality attributable to circumstances holds across the 14 countries in the sample, indicating the presence of an Asia–Pacific Great Gatsby Curve (Figure 2). The fitted RF curve for relative IOp is generally upward sloping, although it exhibits a slight dip at low Gini coefficient values due to Timor-Leste, which has low overall inequality but moderate levels of relative IOp. On the other hand, the fitted CIT estimates produce a more pronounced U-shaped pattern. This curvature is driven primarily by Kyrgyzstan and Uzbekistan, both of which have high overall inequality but unusually low levels of relative IOp. Figure 2 also suggests that the Great Gatsby Curve holds between those the low-inequality and high-inequality countries, and within the high-inequality group, but not within the low-inequality group.

Figure 2: The Asia-Pacific Great Gatsby Curve



Note: The dotted lines show a third-degree polynomial fitted to the data for the 14 countries in the Asia-Pacific.

4.4. Asia-Pacific in global perspective

How does inequality of opportunity in the Asia-Pacific region stack up against the rest of the world? Drawing upon the full GEOM database for 72 countries worldwide, Figures 3A and 3B place inequality of opportunity for the Asia-Pacific in comparative perspective. Figure 3A presents the ex-ante (random forest) relative IOp estimates, while Figure 3B presents the corresponding estimates of absolute IOp.

Figure 3A: Relative Inequality of opportunity in the Asia-Pacific in global perspective (Random Forest ex-ante IOp estimates)

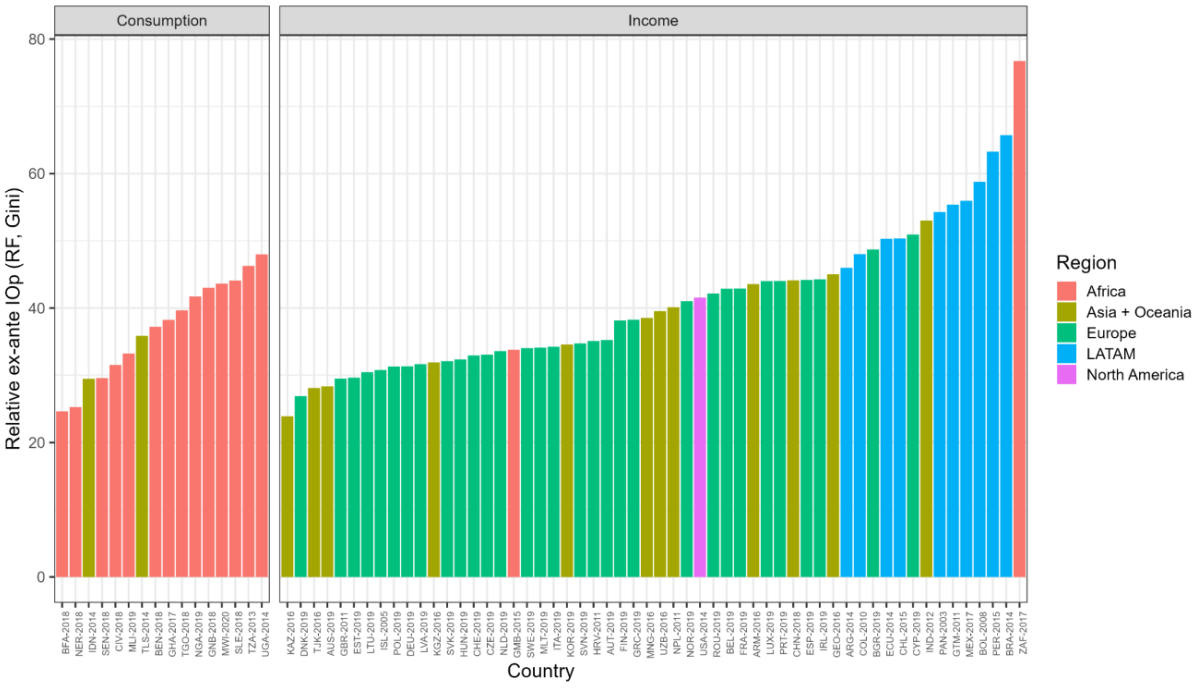
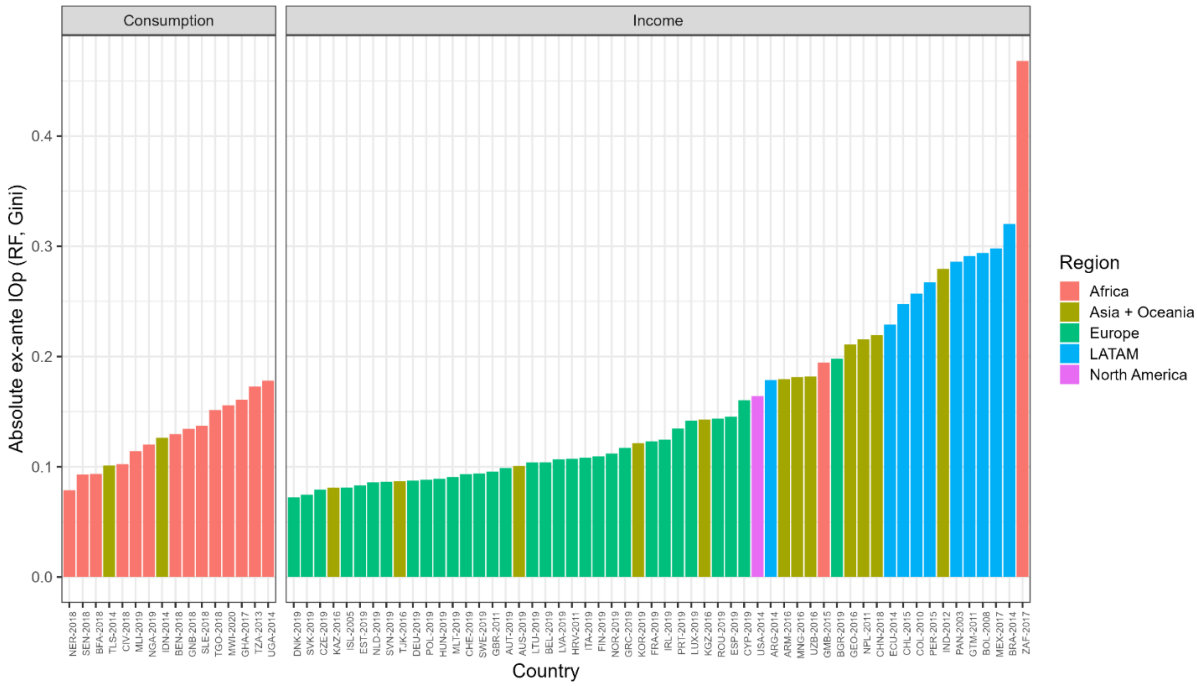


Figure 3B: Absolute Inequality of opportunity in the Asia-Pacific in global perspective (Random Forest ex-ante IOp estimates)



Source: Figure 5 in Ferreira et al. (2026).

In general, the Asia-Pacific countries are spread widely across the global distribution of national IOps. Focusing on relative IOp (Figure 3A), there are three clusters of countries:

(i) Countries with relatively low IOp levels below 30% of overall inequality: Australia, Indonesia, Tajikistan and Kazakhstan.

(ii) Countries in the middle range of IOp levels between 30-40% of overall inequality: Timor-Leste, South Korea, Kyrgyzstan, Mongolia and Nepal.

(iii) Countries with high IOp levels of around 40% or more of overall inequality: Uzbekistan, Georgia, China, Armenia and India. The world's two most populous countries, India and China, are notable in this group, with relative IOps of 44% and 53% respectively. India, in fact, has the eighth largest relative IOp amongst the 72 countries worldwide.

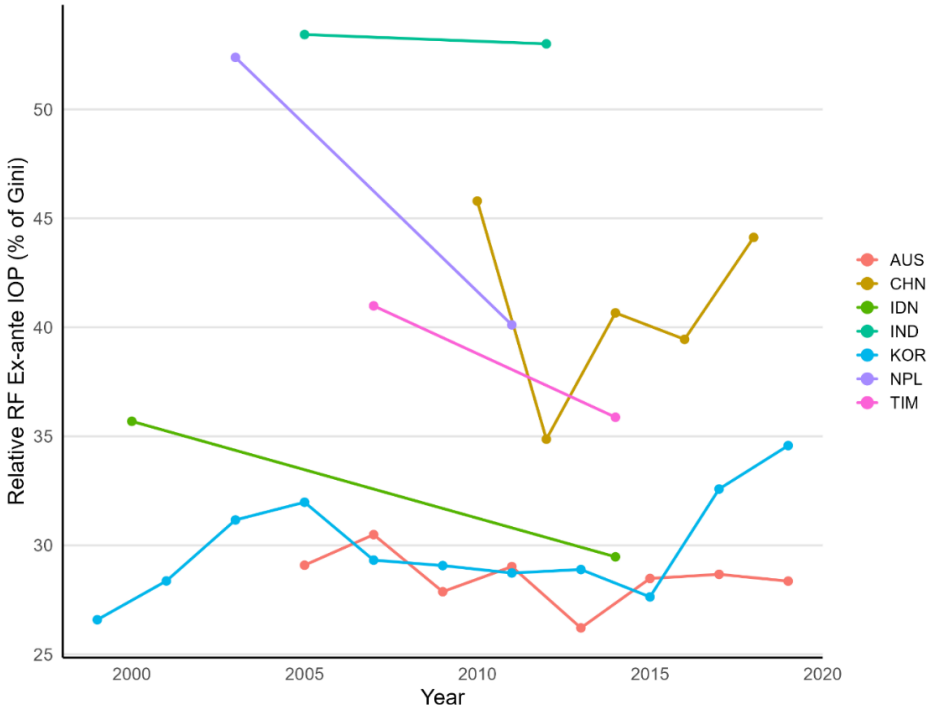
In light of this evidence, representation of the Asia-Pacific as a region of low inequality of opportunity does not seem tenable. IOp in Asia-Pacific is generally lower than in Latin America, but the same cannot be said with respect to other parts of the world. Moreover, comparisons with Africa should be interpreted with caution, as most African IOp estimates are based on consumption-expenditure outcomes, whereas IOp estimates for Asia-Pacific countries rely predominantly on income-based measures.

4.5. Changes over time in IOp

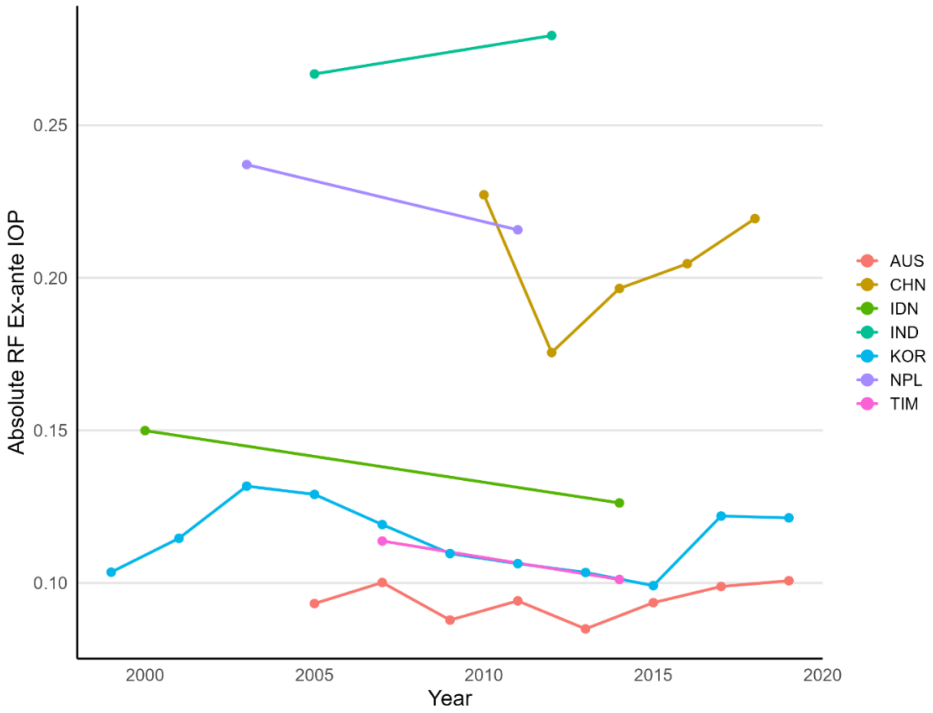
For a subset of seven countries in the region, we can estimate IOp at multiple points of time. These estimates are broadly intertemporally comparable, since for each country they are based on data from different rounds of the same underlying household survey. Figure 4 plots the random forest estimates of ex-ante IOp for these countries.

Figure 4: Changes in Absolute and Relative Ex-ante IOp (RF estimates)

Panel A: Relative IOp



Panel B: Absolute IOp



What can we learn about changes in IOp in the region from these estimates? We focus on the relative IOP estimates (Panel A, Figure 4). The time pattern of absolute IOp in Panel B largely mirrors relative IOp in Panel A.

1. For two of these countries—Australia and South Korea—we have high-frequency data: eight observations spanning 2005–2019 for Australia, and eleven observations for South Korea between 1999–2019. For neither country do we observe a clear trend in relative IOp over the entire period (Panel A, Figure 4). In Australia, relative IOp is between 26-30%; in South Korea, it ranges between 27-35%. In the case of South Korea, however, the estimates reveal an upward trend in relative IOp until 2005, followed by a decline between 2005–2015 and then increasing again following 2015.

2. For China, where we have five data points over a shorter period from 2010–2018, there is likewise no clear or consistent trend. Comparing the two endpoints of the period suggests a decline in relative IOp. However, if we exclude the high IOp share of 46% in 2010, the remaining period from 2012–2018 indicate an upward trend, with relative IOp rising from about 35% to 44%.

3. For the other four countries, India, Indonesia, Nepal and Timor-Leste, there is a decline in relative IOp over time. But with only two data points for these countries, it is difficult to be sure if this represents a definite negative trend.

Overall, the evidence from the available data points to the absence of a strong or consistent trend in IOp across these countries. However, aside from Australia, China, and South Korea, the data are too limited in terms of the number of observations to definitively discern time trends. Arguably, the key message here is the need to augment national surveys to collect such data and enable more regular monitoring of IOp in the region over time.

5. Contribution of different circumstances to IOp

This section presents results on the relative contribution of different circumstance variables to ex-ante IOp for the Asia-Pacific countries. We note some prefatory caveats prior to discussing these results. First, a caveat is warranted on the contribution of sex as a circumstance variable. As seen below in Table 3, sex as a circumstance variable only makes a very small contribution to overall inequality of opportunity, on average about 5% for the 14 countries taken together. The minor contribution of sex is, however, an artifact of the nature of the underlying household

survey data. As noted earlier, the data only give us a measure of total household or equivalized income (or consumption), which therefore does not allow intra-household variation in the outcome variable. The circumstance variable for sex thus only reflects variations in the gender composition of the adult members of the household. If data on individual incomes or consumption within the household were available, we would likely see a much higher contribution of sex to IOp.

A second important caveat concerns the relative importance of variables with very little variability. For example, in Australia in 2019, 97.1% of respondents in the final sample reported being of non-Indigenous origin. Although ethnicity, and being of Indigenous origin, is well known to represent a disadvantage with a substantial penalty in income attainment in Australia, this characteristic is predictive only for the small minority. For the large majority of highly heterogenous non-Indigenous individuals, their ethnicity is not a particularly predictive circumstance with respect to their differential income outcomes, and hence ethnicity does not appear among the final set of types identified by the ML algorithm.

Third, information on father's and/or mother's occupation was not available for some countries, in particular, Nepal, Timor-Leste and India. In the case of India, while father's and mother's occupation were included in the questionnaire, the survey data had a very large number of missing values, and hence the parental occupation variables were dropped for the analysis dataset.[13]

Subject to these caveats, Table 3 presents the Shapley-Shorrocks decompositions of the random forest IOp estimates for the most recent year for each country. For the purposes of this discussion and also to highlight some interesting differences across countries, we group the circumstance variables into two main categories: (i) ethno-spatial background, which includes area of birth and ethnicity; and (ii) and family human capital, which includes parental (mother's and father's) education and occupation.[14] The row for each country in Table 3 shows the percentage contributions of these sets of circumstances to that country's IOp. The last row shows the average contributions for all 14 countries; countries are arranged in ascending order of the contribution of family human capital variables. (The full set of estimates for all country-years are shown in Appendix Table C2.).

[13] Information on mother's occupation for South Korea was also compromised by a large number of missing values, and hence, it was excluded from the analysis dataset.
[14] In light of the caveat mentioned above, the variable pertaining to sex is kept separate.

Table 3: Contribution of circumstance variables to IOp (Gini) in the Asia-Pacific (%), using RF estimates

Country	Year	Birth area	Ethnicity	**Ethno-spatial background**	Parental education	Parental occupation	**Family human capital**	Sex
		(1)	**(2)**	**(3)=(1)+(2)**	**(4)**	**(5)**	**(6)=(4)+(5)**	**(7)**
Timor-Leste	2014	29	48	**77**	20		**20**	3
China	2018	49	12	**61**	7	29	**36**	3
Kazakhstan	2016	26	31	**57**	29	12	**41**	2
Nepal	2011	43	9	**52**	46		**46**	2
India	2012	29	15	**44**	54		**54**	1
S. Korea	2019	27	10	**37**	35	23	**58**	4
Mongolia	2016	30	5	**35**	40	18	**58**	6
Indonesia	2014	20	8	**28**	51	17	**68**	4
Kyrgyzstan	2016	12	14	**26**	43	27	**70**	4
Armenia	2016	23		**23**	39	32	**71**	6
Uzbekistan	2016	9	9	**18**	33	46	**79**	3
Georgia	2016	11		**11**	50	34	**84**	5
Tajikistan	2016	6	2	**8**	46	39	**85**	6
Australia	2019				38	49	**87**	14
Average		**22**	**12**	**34**	**38**	**23**	**61**	**5**

Note: The Table shows the decomposition of IOp by each of the seven circumstance variables. The reported values indicate the percentage contribution of each circumstance variable to IOp. If a circumstance variable was unavailable for a particular country, then a hyphen was used to identify these cases. On the other hand, a 0 means that the circumstance variable is available but had no contribution to IOp.

The decomposition results in Table 3 show that, on average, family human capital accounts for about 60% of IOp in the Asia-Pacific region, while ethno-spatial background accounts for about one-third. There is however substantial variation across countries. We can categorize the 14 countries into three broad groups: those with contributions of family human capital to IOp of under 50 percent, those in the middle with family human capital contributions of 50 to under 70 percent, and the rest with contributions of 70 percent and above. While inquiring into the underlying reasons for this variation is a topic that warrants further research, the following discussion is limited to a few tentative remarks on the varied experience of the three groups.

The first group comprises a set of four diverse countries: Timor-Leste, Kazakhstan, China and Nepal. In all these countries, ethno-spatial background accounts for the bulk of IOp, while the contribution of family human capital is relatively low. The latter is particularly low for Timor-Leste at only about 20% and possibly reflect the widespread low level of human capital of the

preceding generation in the country. By comparison, the contribution of family human capital is also low in China at a little over one-third relative to a 60% contribution of ethno-spatial attributes, which likely reflects the relatively more equal spread of education already achieved by the preceding generation, alongside the persistence of (well-documented) regional income inequalities that accompanied economic growth.[15] Persistent regional disparities are also an important factor shaping economic inequality, and by extension IOp, in Nepal.

The middle group of countries include India, South Korea, Mongolia and Indonesia – also a diverse group. For this group of countries, family human capital starts to play a larger role as sources of IOp relative to ethno-spatial factors. For India, Mongolia and Indonesia, this seems consistent with relatively high levels of inequality in human capital of the preceding generation, while for South Korea this is likely reflective of a lower degree of regional heterogeneity of preceding economic development.

The third group of countries with the highest contributions of family human capital to IOp consists entirely of ex-Soviet Union countries with the lone exception of Australia. These ex-Soviet countries are relatively small and, though different from each other, are ethnically more homogenous. This may explain the low contribution of ethnic-spatial factors to IOp in these countries relative to the role of parental occupation and education. Australia is somewhat of an outlier in this group, though this owes to limitations of available data on ethno-spatial attributes. The caveat with respect to its indigenous population for Australia was already mentioned above. In addition, the underlying information in the survey data on the place of birth (as different to the place of current residence) is limited to whether the birth place was within or outside Australia. With only a very small proportion of those born outside Australia, the analytical sample with non-missing values of all circumstance variables did not yield enough variation in the place of birth for it to be identified as a discriminating variable for estimating IOp. Had the underlying data, for instance, identified the postcode of birth in Australia, we would have likely identified an appreciable contribution ethno-spatial factors to IOp.

In summary, the evidence is indicative of widely varying importance of particular circumstance variables in contributing to IOp in different countries; relative contributions vary even for

[15] In 2010, the contribution of family human capital variables to IOp in China was however notably more than 50% (Appendix Table C2). This suggests that the relatively limited significance of parental human capital as a source of IOp in China is a more recent phenomenon.

countries with comparable levels of IOp. Generalizations are hard to come by on the decomposition results; the estimates suggest the country context matters a lot to the relative importance of different circumstances.

6. The structure of inequality of opportunity

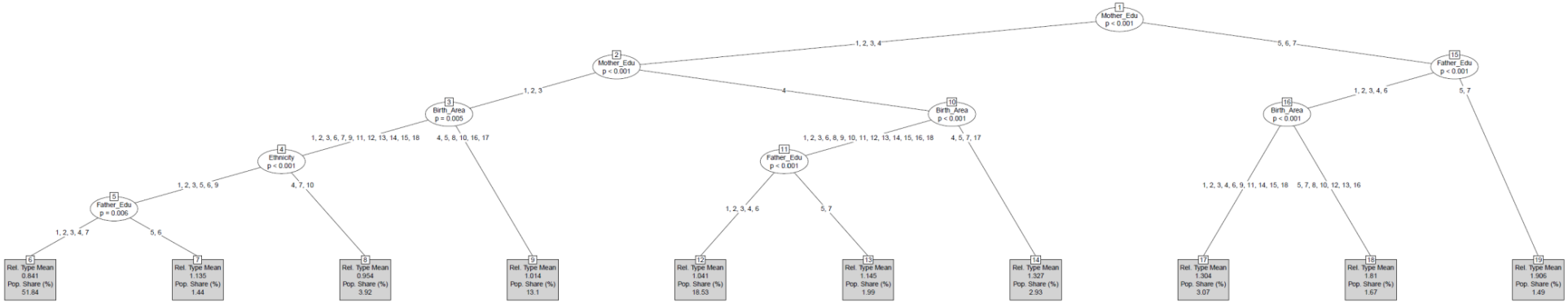
Beyond their value in isolating essential types for evaluating IOp, conditional inference trees have additional informational content. We can gain further insight into the structure of inequality of opportunity in a particular country context by following the structure of that country's CIT. We illustrate this with CITs for two countries: South Korea and India, as representative of low and high levels of IOp in the region. A word of caution is however warranted in interpreting these examples, owing to the variability of CITs (as noted earlier), such that under repeated sampling there could be a degree of variation in the particular tree identified. Nonetheless, subject to this caveat, a country's CIT can be quite revelatory of the prevailing structure of unequal opportunity in that country.

6.1. South Korea, 2019

Figure 5 shows the CIT for South Korea based on data from the Korean Labour Income Panel Study (KLIPS) for 2019. The ML algorithm, with sequential partitioning of the sample that maximizes differences in expected income outcomes across partitions, identifies ten distinct types for the country for this year.

The first split of the South Korea sample is based on mother's education. One branch corresponds to mother's education codes 1, 2, 3, and 4, representing no schooling, elementary school, middle school, and high school, respectively. The second branch includes codes 5, 6, and 7, corresponding to community college, undergraduate degree, and graduate degree. This split is intuitive, as significant differences in earnings are expected between these two educational clusters. At the next level, the tree splits again by both mother's and father's education. At the third level, these education-based partitions are further divided by birth area. From there, one of the birth-area branches splits on ethnicity and then on father's education; another branch splits only on father's education; and the final branch splits once more by birth area. This process results in ten distinct types.

Figure 5: Conditional Inference Tree for South Korea, 2019



As the bottom row of Figure 5 shows, the population shares of the ten types range from 1% (type 19) to 52% (type 6), and their relative means (ratio of type mean to overall mean) range from 0.84 (type 6) to 1.91 (type 19). It is instructive to look at the combination of circumstance categories that different types represent. We illustrate the poorest and the richest types in terms of the relative means.

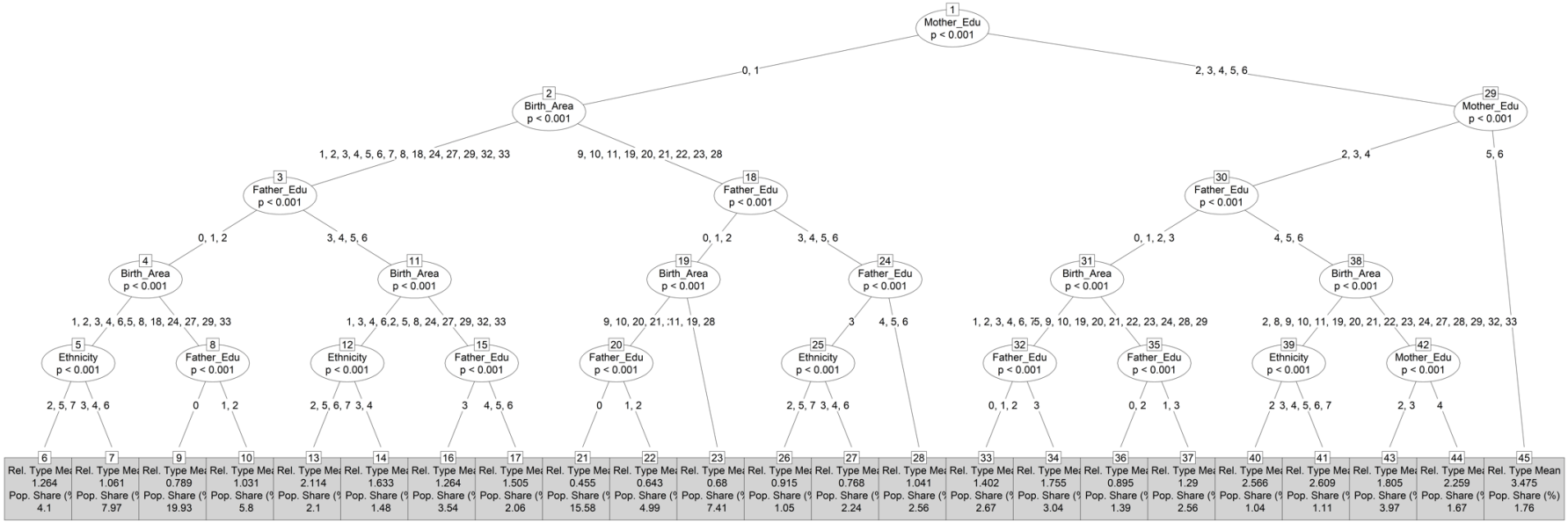
Type 6 has a mean income of about 16% below the overall mean, and represents individuals with the following combination of circumstances: (i) their mothers' highest level of education is middle school or below, (ii) they were born in Seoul, Busan, Daegu, Gwangju, Ulsan, Gangwon, Chungcheongnam, Jeollabuk, Jeollanam, Gyeongsangbuk, Gyeongsangnam or overseas, (iii) they identify as an Atheist, Buddhist, Protestant, Confucian or as followers of the Daesoon Faith, and (iv) their fathers had either no schooling or completed elementary school, middle school, high school or a graduate degree.

By contrast, type 19 has a mean income 91% above the national mean and represents individuals (i) whose mothers have completed community college, an undergraduate degree or a graduate degree, and (ii) whose fathers have completed either community college or a graduate degree. It is also instructive to note that type 6 is the largest group accounting for more than half the population, whereas type 19 represents the smallest comprising only 1.5%. The association of the disparate set of inherited circumstances of these groups with their income outcomes shows how the CIT is informative of the structure of inequality of opportunity in the country.

6.2. India, 2012

India with the highest relative IOp among the 14 Asia-Pacific countries offers an interesting contrast to South Korea which has the lowest. Figure 6 shows the pruned version of the CIT for India for 2012, showing branches and nodes up to the fifth split which identify 23 types. The full CIT for India 2012 is deeper and identifies 53 distinct types which is too large for a readable graphical depiction.

Figure 6: Conditional Inference Tree for India, 2012



The first split in India's CIT is on the basis on mother's education into two branches: branch 1 with individuals whose mothers had completed primary or higher levels of education, and branch 2 whose mothers had none or incomplete primary education.

Following the two branches through successive splits, we find that branch 1 is split further into sub-branches for those whose mothers' education was between primary and completed secondary, and others whose mothers had post-secondary or higher education. There are no further splits of the latter sub-branch. However, the former sub-branch (with primary-secondary education) is split at the next stage on the basis of low (below completed secondary) or high (completed secondary or above) father's education. At the fourth split, birth area comes into play, and at the fifth stage there is re-splitting on the basis of ethnicity and sub-categories of mothers' and fathers' education again.

Similarly, following branch 2, we find a split based on birth area at the second stage, splits based on fathers' education at third stage, further splits based on birth area and sub-categories of fathers' education at the fourth stage, and at the fifth stage splits based on ethnicity and further sub-categories of fathers' education.

The tree for India clearly has a complex structure that would be practically undiscoverable through parametric or standard non-parametric approaches to identifying types. To illustrate the range of types identified by the CIT, we focus again on the richest and poorest ones. As seen in Figure 5, these range from type 45 with a relative mean of 3.48 and a small population share of 1.8% to type 21 with a relative mean of 0.46 (less than half the national mean) and a significant population share of 15.6%. It is striking that type 45 (the richest type) consists of individuals distinguished by a *single* characteristic: they had mothers with post-secondary or higher education. At the other end, type 21 (the poorest type) is distinguished by the combination of several characteristics; it comprises (i) individuals whose mothers had none or less than primary education, (ii) who were born in either of the states of Uttar Pradesh, Bihar, Jharkhand or Odisha, which are known to be among the poorest states of India, *and* (iii) whose fathers had no education.

India's CIT also identifies several types that highlight the role of ethnicity (caste/religion) in shaping IOp in the country. For instance, one of the poorer types – type 27 with mean income about three-quarters of the national mean – comprises a group that emanates from branch 2 (mothers with none or below primary education), but further conditions it based on birth in one

of the relatively poorer states, fathers with incomplete secondary education *and* belonging to the *dalit* (scheduled caste), other backward caste or Muslim community.

Conditional inference trees for other Asia-Pacific countries (available as Supplementary Online Material) can be interpreted along similar lines. As observed for South Korea and India, these trees can be quite illuminative of the prevailing structure of inequality of opportunity in individual countries, beyond their instrumental value in quantifying the overall levels of IOp. Again, recall that even though they are appealing due to their simplicity and ability to reveal interesting interactions, CITs are subject to variability across samples and should therefore be interpreted with caution. Basic insights from CITs also ought to be complemented with those from Shapley value decompositions.

7. Conclusion

This paper on inequality of opportunity in the Asia-Pacific region, as well as the larger global GEOM project of which it is a part, has a clear message: while not all inequality is necessarily ethically reprehensible, a disconcertingly large part of it is. As a simple average across 14 Asia-Pacific countries, nearly two-fifths of observed inequality in income or consumption represents inequality of opportunity attributable to a set of inherited circumstances beyond individual control. This is an important reminder that notwithstanding relatively rapid growth in the region, large fractions of the population battle or savor destinies written in advance.

A second message relates to the large variation in IOp estimates across the region, with shares of IOp in total inequality ranging from 24 to 53 percent. While economic and social policy likely have an important role, we are yet to adequately comprehend the factors underlying this variation. This remains an important agenda for further research.

The IOp literature has matured to the point where we have more refined tools for the measurement of IOp. The machine learning methods deployed in this paper are efficient in accounting for a large fraction of observed inequality with a limited number of circumstance variables. While we seem to have reasonable tools for estimation of IOp, we fall short on the requisite data. For instance, with the available data this paper could construct IOp estimates for only 14 of the 49 Asia-Pacific countries, and only for a handful of them could do so at multiple points of time.

A third message of the paper thus relates to the need to collect IOp-relevant data. Wider collection of such data is not only a desirable but also an achievable goal. For many countries that already conduct household income or consumption surveys, the marginal cost of collecting additional data on IOp-relevant variables like ethnicity, place of birth, parental education or occupation is low. Efforts in augmenting such data collection will likely have high returns in richer analysis of IOp to inform public policy in the Asia-Pacific region and beyond.

References

Bhattacharya, N., and B. Mahalanobis. 1967. Regional disparities in household consumption in India. *Journal of the American Statistical Association*, 62(317): 143–161.

Bourguignon, F., Ferreira, F. H. G., and M. Menendez. 2007. Inequality of opportunity in Brazil. *Review of Income and Wealth*, 53(4): 585-618.

Brunori, P., Ferreira, F. H. G., and V. Peragine. 2013. Inequality of opportunity, income inequality, and economic mobility: Some international comparisons. In *Getting Development Right* (pp. 85-115). New York: Palgrave Macmillan.

Brunori, P., Ferreira, F. H. G. and P. Salas-Rojo. 2023. Inherited inequality: a general framework and an application to South Africa. International Inequalities Institute Working Paper 107, The London School of Economics and Political Science.

Brunori, P., Hufe, P., and D. G. Mahler. 2023. The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees. *Scandinavian Journal of Economics*, 125 (4): 900-932.

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

Brunori, P., Peragine, V., and L. Serlenga. 2019. Upward and Downward Bias when Measuring Inequality of Opportunity. *Social Choice and Welfare*, 52 (4): 635–661.

Chakravarty, S., & Eichhorn, W. 1994. Measurement of income inequality: Observed versus true data. In W. Eichhorn (Ed.), Models and measurement of welfare and inequality. Springer, Berlin.

Ferreira, F.H.G., and P. Brunori. 2024. Inherited Inequality, Meritocracy and the Purpose of Economic Growth. International Inequalities Institute Working Paper 147, The London School of Economics and Political Science.

Ferreira, F.H.G., and J. Gignoux. 2011. The measurement of inequality of opportunity: Theory and an application to Latin America. *Review of Income and Wealth*, 57, 622-657.

Ferreira, Francisco H. G., Peragine, Vito, Brunori, Paolo, Salas Rojo, Pedro, Moramarco, Domenico, Barajas Prieto, Luis, Barbieri, Teresa, Daza Baez, Nancy, Datt, Gaurav, de Sandi, Vito, Farella, Fabio, Martinez Jr., Arturo, Nguyen, John, Park, Albert, Simeone, Enza, Sirugue, Louis, Torres Lopez, Pedro, Zotti, Giorgia. 2026. Global Estimates of Opportunity and Mobility: A Database. III Working Paper 158. International Inequalities Institute, London School of Economics and Political Science.

Fleurbaey, M., Lambert, P., Moramarco, D., and V. Peragine. 2024. Within-group inequality: a comparison of different definitions and a new proposal of decomposition. ECINEQ Working Paper 2024 673.

Fleurbaey, Marc, and Vito Peragine (2013): "Ex ante versus ex post equality of opportunity," *Economica*, 80: 118-130.

Foster, J. E. and A. Sen. 1999. *On Economic Inequality*. Third Impression. New Delhi: Oxford India Press. Oxford University Press.

Hothorn, T., Hornik, K., and A. Zeileis. 2006. Unbiased Recursive Partitioning: Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15: 651-674.

Hufe, P., Peichl, A., Roemer, J., and M. Ungerer. 2017. Inequality of income acquisition: the role of childhood circumstances. *Social Choice and Welfare*, 49: 499–544.

Kanbur, R. and A. Wagstaff, A. 2016. How useful is inequality of opportunity as a policy construct? In K. Basu & J. E. Stiglitz (Eds.), *Inequality and Growth: Patterns and Policy* (International Economic Association Series). Palgrave Macmillan.

Ramos, X., and D. Van de Gaer. 2015. Approaches to Inequality of Opportunity: Principles, Measures and Evidences. *Journal of Economic Surveys*, 30 (5): 855–883.

Roemer, J. E. 1998. *Equality of Opportunity*. Cambridge, MA: Harvard University Press.

Roemer, J. E., and A. Trannoy. 2016. Equality of Opportunity: Theory and Measurement. *Journal of Economic Literature*, 54 (4): 1288–1332.

Shorrocks, A. 2013. Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. *The Journal of Economic Inequality*, 11: 99-126.

Van de Gaer, D. 1993. Equality of opportunity and investment in human capital. Ph.D. diss., Katholieke Universiteit Leuven: Belgium.

Van der Weide, R., Lakner, C., Mahler, D. G., Narayan, A., and R. Ramasubbaiah. 2021. Intergenerational Mobility around the World. World Bank Policy Research Working Paper No. 9707.

World Bank. 2021. *Global Database on Intergenerational Mobility.* Washington, DC: World Bank Group.

Appendix A: Details of the machine learning algorithm for conditional inference trees and Shapely-Shorrocks decomposition

Following the original proposal by Hothorn et al. (2006), the steps below describe the procedure for how the sample data of individuals is partitioned circumstance types to construct a conditional inference tree:

1. Set a confidence level (1- α).
2. Test the correlation between the dependent variable (outcome) and each regressor (circumstance). If the Bonferroni-adjusted p-value of the correlation test is higher than the chosen critical value α, exit. The value of α was set at 0.01.
3. Among those regressors where the null hypothesis of independence is rejected, select the variable whose correlation with the outcome has the smallest adjusted p-value as splitting variable [c].
4. Consider how circumstance [c] can be used to partition the sample into two subsamples $[s_i, s_{-i}]$. Let S_c denote the set of all possible binary splits of the sample based on [c]. For each possible binary partition, compute the p-value for the null hypothesis that the mean in two sub-samples is the same ($p^{[s_i, s_{-i}]}$).
5. Choose $[s_i, s_{-i}]^* = argmin_{S_c}\ p^{[s_j, s_{-j}]}$ as the most appropriate partition.
6. Repeat steps 2 – 5 for each resulting node until the null hypothesis of step 2 cannot be rejected in all resulting sub-sample.

We follow Brunori, Ferreira and Salas-Rojo (2023) to evaluate the Shapley-Shorrocks decompositions of IOp as below:

1. Draw a subsample, indexed [i], of the full sample. To favor computational speed, the sub-sample is capped at 5,000 observations or 90% of the original sample size if that amounts to less than 5,000.
2. Estimate $IOp[i]$ for subsample [i] by constructing a conditional inference tree for the subsample data, allowing the tree to over-fit (setting α = 0.9, and minimum sample size = 0.1% of the sample size).
3. Re-estimate $IOp[i\backslash k]$ in the subsample for all possible elimination sequences of a circumstance c_k. Elimination of a circumstance is obtained by replacing their values with a vector of value 1.

4. Estimate the difference between $IOp[i]$ and $IOp[i\backslash k]$ for each sequence of elimination of c_k. Evaluate the contribution of c_k as the weighted average of these differences for all possible sequences of elimination of c_k.

The final estimate of the contribution of c_k to IOp is the average contribution across 100 draws of the subsample.

Appendix B : Description of household survey data

Country	Years	Circumstance variables	Outcome	Survey
Australia	2005	Ethnicity, sex, father's education, mother's education, father's occupation, mother's occupation	Income	Household Income Labour Dynamics in Australia Survey (HILDA)
	2007			
	2009			
	2011			
	2013			
	2015			
	2017			
	2019			
China	2010	Birth area, ethnicity, sex, father's education, mother's education, father's occupation, mother's occupation	Income	China Family Panel Studies (CFPS)
	2012			
	2014			
	2016			
	2018			
Indonesia	2000	Birth area, ethnicity, sex, father's education, mother's education; father's occupation, mother's occupation	Consumption	Indonesian Family Life Survey (IFLS)
	2014			
India	2005	Birth area, ethnicity, sex, father's education, mother's education	Income	Indian Human Development Studies (IHDS)
	2012			
South Korea	1999	Sex, birth area, ethnicity, father's education, father's occupation, mother's education	Income	Korean Labour Income Panel Study (KLIPS)
	2001			
	2003			
	2005			
	2007			
	2009			
	2011			
	2013			
	2015			
	2017			
	2019			
Nepal	2003	Birth area, sex, ethnicity, father's education, mother's education	Income	Nepal Living Standards Survey (NLSS)
	2011			
Timor-Leste	2007	Sex, birth area, ethnicity, father's education, mother's education	Consumption	Timor-Leste Survey of Living Standards (TLSLS)
	2011			

Armenia	2016	Sex, birth area, ethnicity, father's education, father's occupation, mother's education, mother's occupation	Income	Life in Transition Survey (LITS)
Georgia	2016	Sex, birth area, ethnicity, father's education, father's occupation, mother's education, mother's occupation	Income	Life in Transition Survey (LITS)
Kazakhstan	2016	Birth area, ethnicity, sex, father's education, father's occupation, mother's education, mother's occupation	Income	Life in Transition Survey (LITS)
Kyrgyzstan	2016	Sex, birth area, ethnicity, father's education, father's occupation, mother's education, mother's occupation	Income	Life in Transition Survey (LITS)
Mongolia	2016	Sex, birth area, ethnicity, father's education, father's occupation, mother's education, mother's occupation	Income	Life in Transition Survey (LITS)
Tajikistan	2016	Sex, birth area, ethnicity, father's education, father's occupation, mother's education, mother's occupation	Income	Life in Transition Survey (LITS)
Uzbekistan	2016	Sex, birth area, ethnicity, father's education, father's occupation, mother's education , mother's occupation	Income	Life in Transition Survey (LITS)

Appendix C : IOp estimates for all country-years in the Asia-Pacific

Table C1: Ex-ante Inequality of Opportunity in the Asia-Pacific (using the Gini index)

	Country	Year	Outcome variable	Total inequality (Gini)	Absolute		Relative		Number of types (CIT)
					IOp Ex-ante RF	IOp Ex-ante CIT	IOp Ex-ante RF	IOp Ex-ante CIT	
ARM	Armenia	2016	Income	0.412	0.179	0.116	43.554	28.114	3
AUS	Australia	2005	Income	0.320	0.093	0.082	29.089	25.531	5
AUS	Australia	2007	Income	0.328	0.100	0.081	30.490	24.581	4
AUS	Australia	2009	Income	0.315	0.088	0.076	27.873	24.000	5
AUS	Australia	2011	Income	0.324	0.094	0.085	29.025	26.249	7
AUS	Australia	2013	Income	0.324	0.085	0.076	26.212	23.310	6
AUS	Australia	2015	Income	0.328	0.094	0.081	28.480	24.551	5
AUS	Australia	2017	Income	0.345	0.099	0.089	28.671	25.856	8
AUS	Australia	2019	Income	0.355	0.101	0.094	28.358	26.584	7
CHN	China	2010	Income	0.496	0.227	0.203	45.8	41.0	10
CHN	China	2012	Income	0.503	0.176	0.172	34.9	34.1	12
CHN	China	2014	Income	0.483	0.197	0.174	40.7	36.1	11
CHN	China	2016	Income	0.519	0.205	0.165	39.5	31.8	12
CHN	China	2018	Income	0.497	0.219	0.194	44.1	39.0	9
GEO	Georgia	2016	Income	0.469	0.211	0.122	45.016	26.062	2
IDN	Indonesia	2000	Cons.	0.420	0.150	0.154	35.690	36.571	16
IDN	Indonesia	2014	Cons.	0.428	0.126	0.116	29.472	27.043	8
IND	India	2005	Income	0.499	0.267	0.282	53.4	56.6	55
IND	India	2012	Income	0.527	0.279	0.292	53.0	55.4	53
KAZ	Kazakhstan	2016	Income	0.339	0.081	-	23.900	-	-
KGZ	Kyrgyzstan	2016	Income	0.448	0.143	0.078	31.911	17.408	2
KOR	South Korea	1999	Income	0.389	0.104	0.079	26.586	20.267	8
KOR	South Korea	2001	Income	0.404	0.115	0.089	28.366	22.054	8
KOR	South Korea	2003	Income	0.423	0.132	0.093	31.164	22.030	7
KOR	South Korea	2005	Income	0.403	0.129	0.091	31.978	22.583	5
KOR	South Korea	2007	Income	0.406	0.119	0.093	29.321	22.846	7
KOR	South Korea	2009	Income	0.377	0.110	0.075	29.072	19.814	6
KOR	South Korea	2011	Income	0.370	0.106	0.069	28.730	18.568	5
KOR	South Korea	2013	Income	0.358	0.103	0.087	28.891	24.253	8
KOR	South Korea	2015	Income	0.359	0.099	0.063	27.635	17.540	4
KOR	South Korea	2017	Income	0.374	0.122	0.096	32.585	25.742	7
KOR	South Korea	2019	Income	0.351	0.121	0.106	34.578	30.217	10
MNG	Mongolia	2016	Income	0.471	0.181	0.144	38.533	30.563	3
NPL	Nepal	2003	Income	0.453	0.237	0.233	52.386	51.414	14
NPL	Nepal	2011	Income	0.538	0.216	0.219	40.115	40.692	11
TJK	Tajikistan	2016	Income	0.309	0.087	0.061	28.109	19.657	2
TLS	Timor-Leste	2007	Cons.	0.277	0.114	0.134	40.988	48.198	8
TLS	Timor-Leste	2014	Cons.	0.282	0.101	0.117	35.877	41.341	13
UZB	Uzbekistan	2016	Income	0.460	0.182	0.095	39.548	20.753	2

Note: For Kazakhstan, no CIT types were found due to the small sample size, though random forest estimates of IOp were feasible.

Table C2: Shapley-Shorrocks decompositions: contribution of circumstance variables to IOp (Gini) in the Asia-Pacific (%), using RF estimates

	Country	Year	Birth area	Ethnicity	Father's education	Mother's education	Father's occupation	Mother's occupation	Sex
ARM	Armenia	2016	16	0	21	23	16	16	7
AUS	Australia	2005	-	0	28	17	26	26	7
AUS	Australia	2007	-	0	30	14	25	26	8
AUS	Australia	2009	-	0	32	14	24	23	9
AUS	Australia	2011	-	0	27	19	21	25	9
AUS	Australia	2013	-	0	28	19	22	24	8
AUS	Australia	2015	-	0	28	17	23	24	8
AUS	Australia	2017	-	0	26	18	26	24	8
AUS	Australia	2019	-	0	25	18	27	20	12
CHN	China	2010	39	4	11	17	9	17	2
CHN	China	2012	36	0	10	21	11	19	3
CHN	China	2014	52	4	5	15	5	16	3
CHN	China	2016	60	6	3	15	2	9	5
CHN	China	2018	49	12	4	16	3	13	3
GEO	Georgia	2016	10	0	25	24	21	15	7
IDN	Indonesia	2000	19	6	27	24	10	10	4
IDN	Indonesia	2014	16	7	21	21	12	19	5
IND	India	2005	26	21	30	20	-	-	2
IND	India	2012	29	15	31	23	-	-	1
KAZ	Kazakhstan	2016	53	9	10	10	5	9	4
KGZ	Kyrgyzstan	2016	10	5	21	37	13	10	4
KOR	South Korea	1999	30	11	17	14	23	-	5
KOR	South Korea	2001	23	12	19	18	21	-	6
KOR	South Korea	2003	28	11	20	19	14	-	8
KOR	South Korea	2005	30	15	17	15	19	-	5
KOR	South Korea	2007	29	4	25	15	20	-	6
KOR	South Korea	2009	21	9	27	14	23	-	7
KOR	South Korea	2011	27	9	22	18	17	-	6
KOR	South Korea	2013	19	17	21	19	19	-	5
KOR	South Korea	2015	22	8	25	19	20	-	6
KOR	South Korea	2017	20	13	25	23	16	-	4
KOR	South Korea	2019	24	7	24	25	15	-	5
MNG	Mongolia	2016	16	6	25	16	13	17	7
NPL	Nepal	2003	37	9	37	12	-	-	5
NPL	Nepal	2011	39	7	33	16	-	-	5
TJK	Tajikistan	2016	1	4	36	11	17	20	10
TLS	Timor-Leste	2007	33	51	9	3	-	-	4
TLS	Timor-Leste	2014	31	45	14	7	-	-	3
UZB	Uzbekistan	2016	11	5	17	19	25	20	3

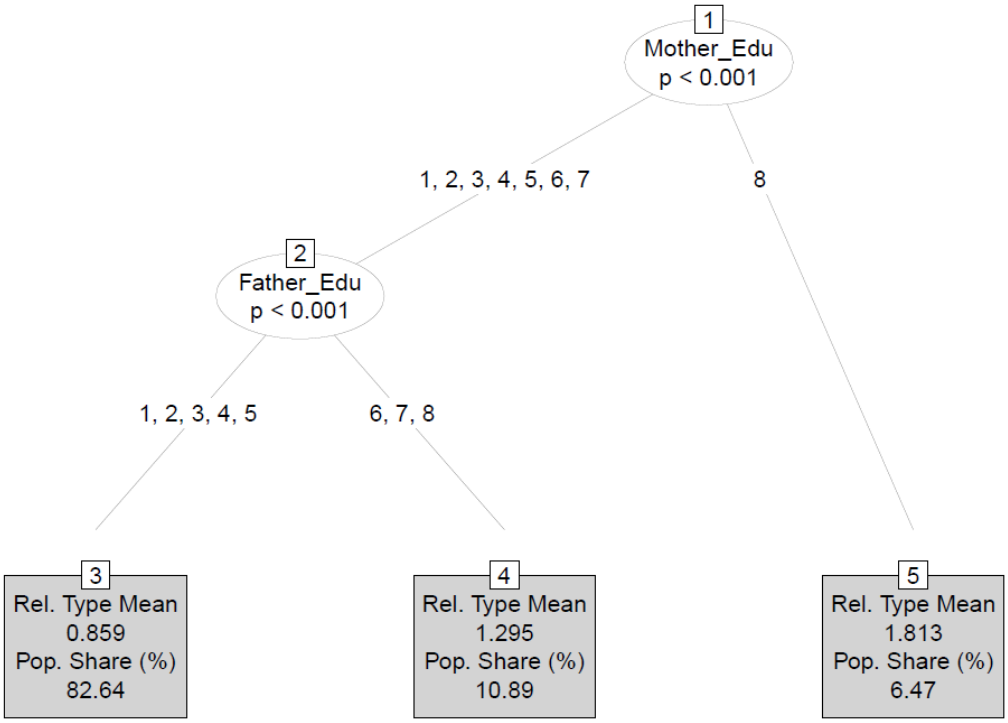
Measuring Inequality of Opportunity in Asia and the Pacific

Gaurav Datt, John Nguyen, Pedro Salas-Rojo, Francisco H.G. Ferreira, Paolo Brunori, Vito Peragine, Albert Park, Arturo Martinez Jr., Joseph Albert Nino Bulan

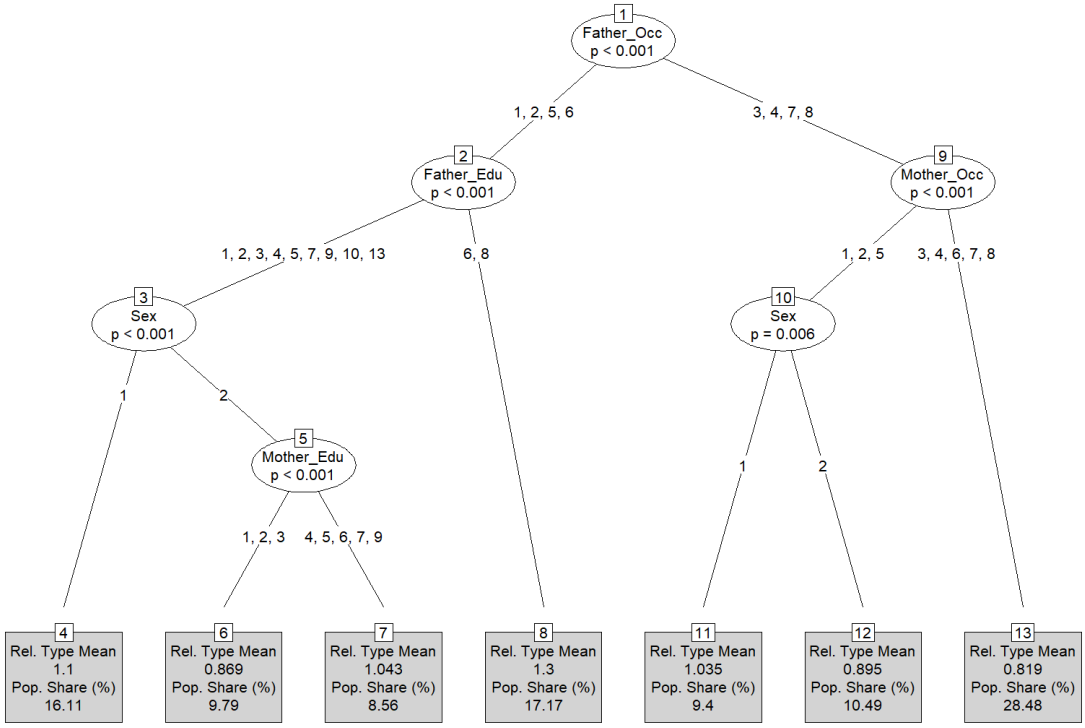
1 April, 2026

Supplementary Online Material:
Ex-ante conditional inference trees for Asia-Pacific countries
(for the most recent survey year)

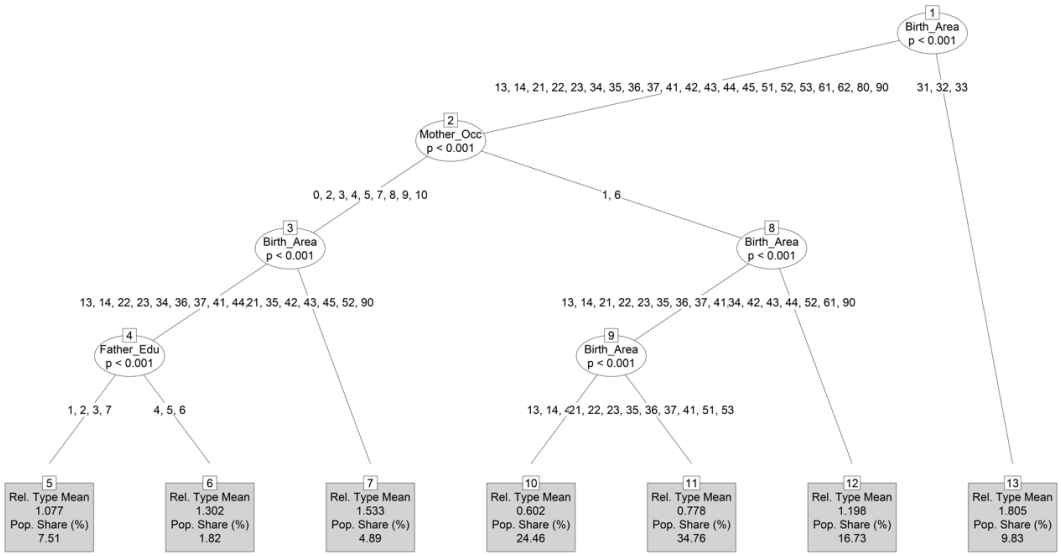
Armenia (2016)



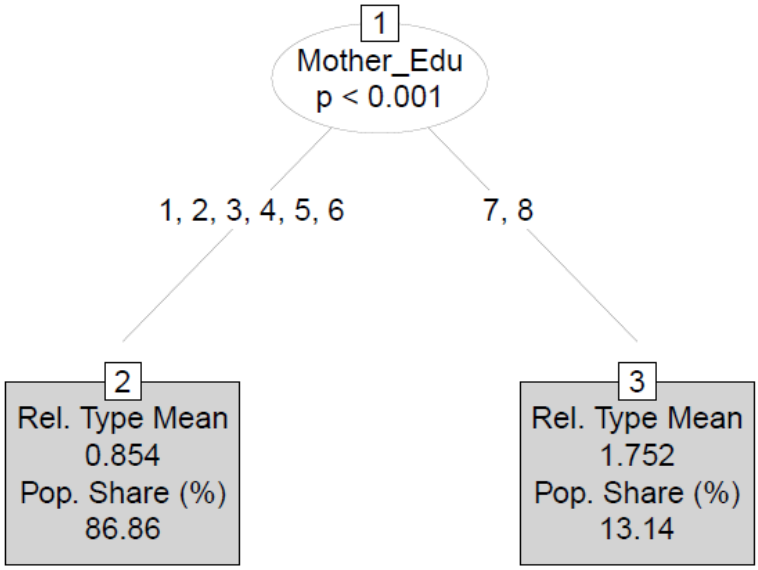
Australia (2019)



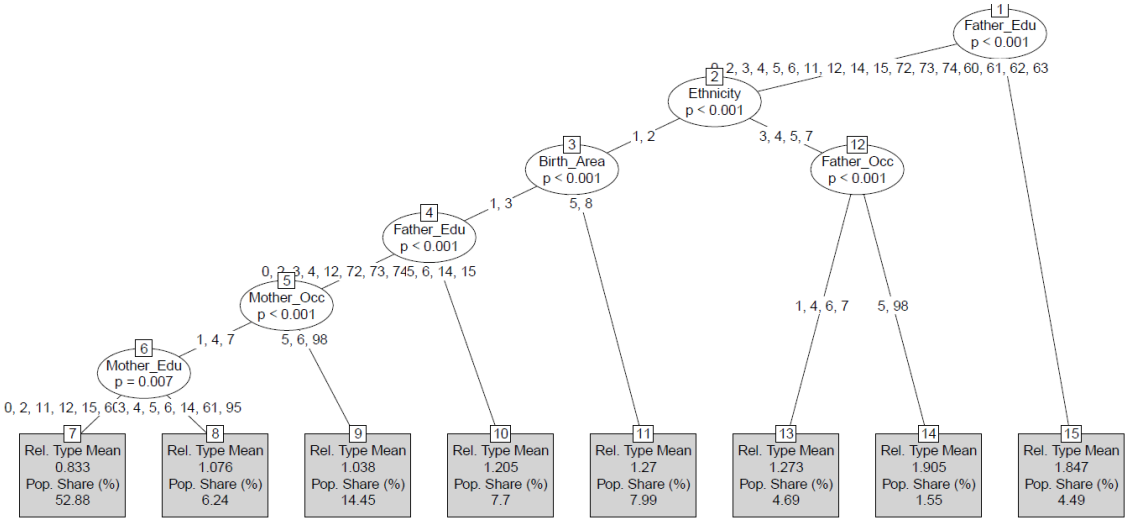
China (2018)


1
Birth_Area
p < 0.001
13, 14, 21, 22, 23, 34, 35, 36, 37, 41, 42, 43, 44, 45, 51, 52, 53, 61, 62, 80, 90
31, 32, 33
2
Mother_Occ
p < 0.001
0, 2, 3, 4, 5, 7, 8, 9, 10
1, 6
3
Birth_Area
p < 0.001
8
Birth_Area
p < 0.001
13, 14, 22, 23, 34, 36, 37, 41, 44
21, 35, 42, 43, 45, 52, 90
13, 14, 21, 22, 23, 35, 36, 37, 41
34, 42, 43, 44, 52, 61, 90
4
Father_Edu
p < 0.001
9
Birth_Area
p < 0.001
1, 2, 3, 7
4, 5, 6
13, 14, 21, 22, 23, 35, 36, 37, 41, 51, 53
5
Rel. Type Mean
1.077
Pop. Share (%)
7.51
6
Rel. Type Mean
1.302
Pop. Share (%)
1.82
7
Rel. Type Mean
1.533
Pop. Share (%)
4.89
10
Rel. Type Mean
0.602
Pop. Share (%)
24.46
11
Rel. Type Mean
0.778
Pop. Share (%)
34.76
12
Rel. Type Mean
1.198
Pop. Share (%)
16.73
13
Rel. Type Mean
1.805
Pop. Share (%)
9.83

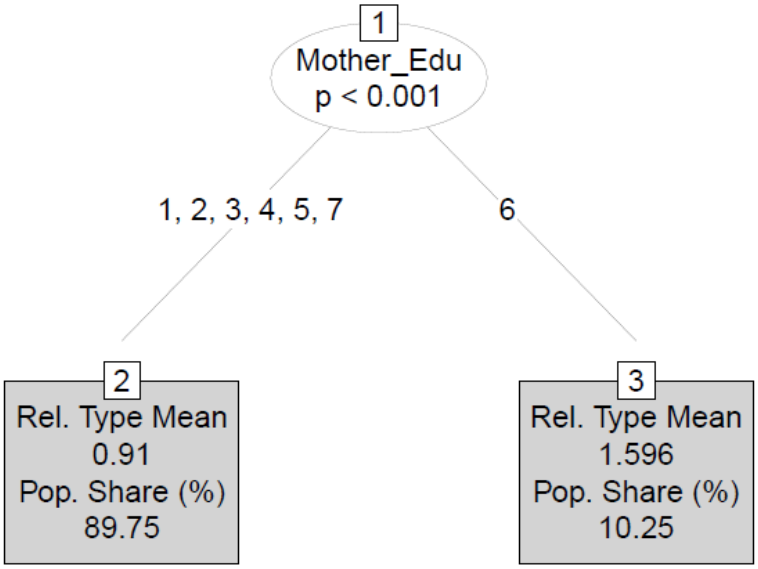
Georgia (2016)



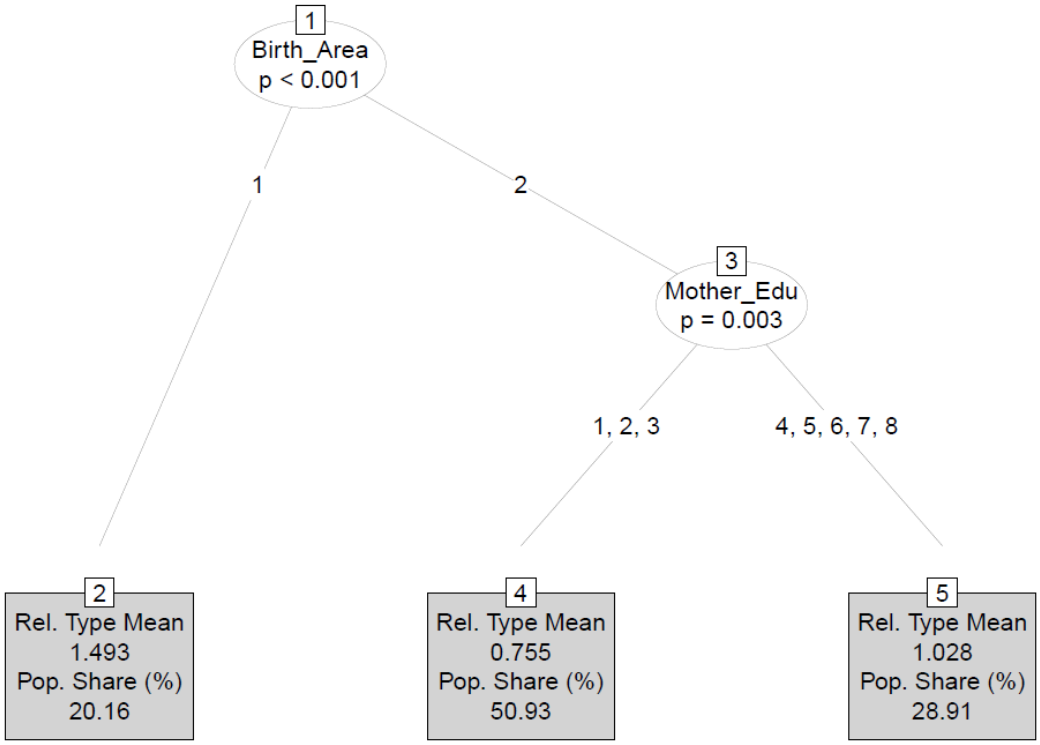
Indonesia (2014)



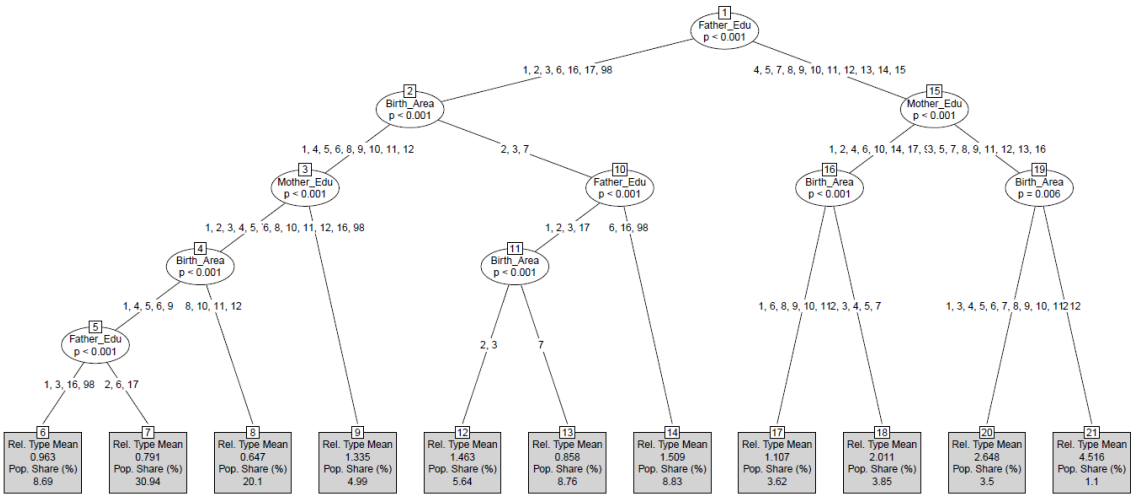
Kyrgyzstan (2016)



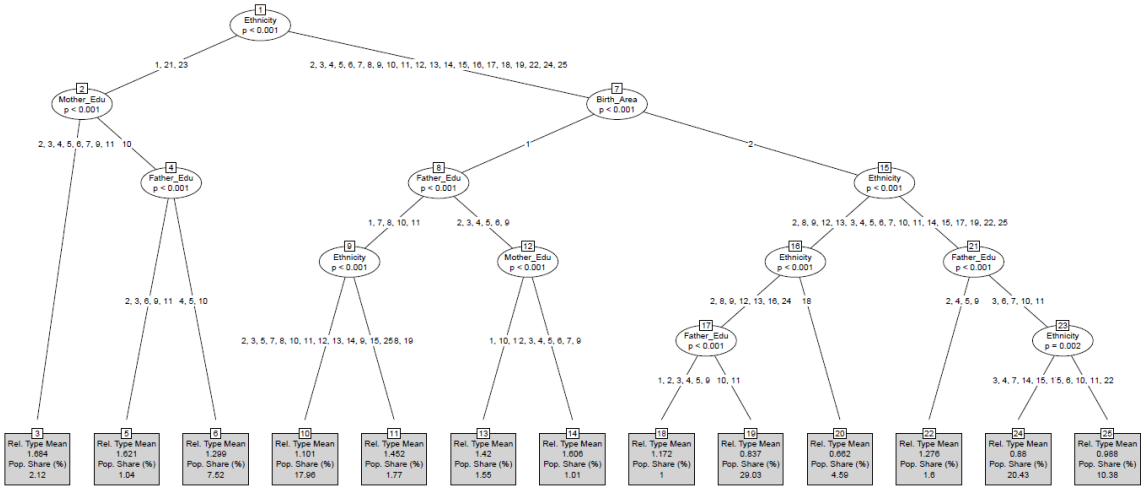
Mongolia (2016)



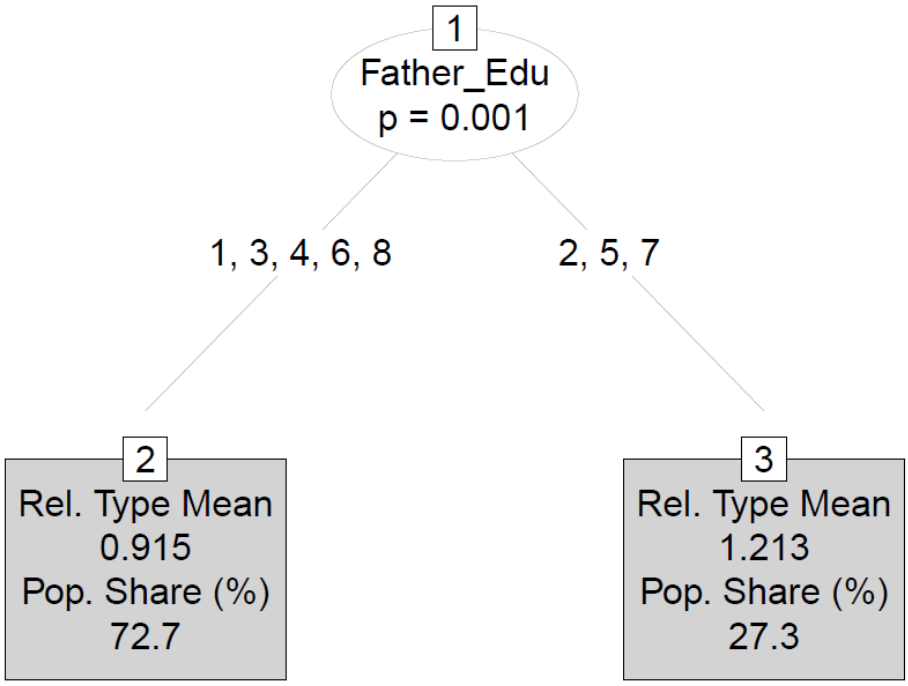
Nepal (2011)


1
Father_Edu
p < 0.001
1, 2, 3, 6, 16, 17, 98
4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15
2
Birth_Area
p < 0.001
15
Mother_Edu
p < 0.001
1, 4, 5, 6, 8, 9, 10, 11, 12
2, 3, 7
1, 2, 4, 6, 10, 14, 17
3, 5, 7, 8, 9, 11, 12, 13, 16
3
Mother_Edu
p < 0.001
10
Father_Edu
p < 0.001
16
Birth_Area
p < 0.001
19
Birth_Area
p = 0.006
1, 2, 3, 4, 5, 6, 8, 10, 11, 12, 16, 98
1, 2, 3, 17
6, 16, 98
4
Birth_Area
p < 0.001
11
Birth_Area
p < 0.001
1, 4, 5, 6, 9
8, 10, 11, 12
1, 6, 8, 9, 10, 11
2, 3, 4, 5, 7
1, 3, 4, 5, 6, 7, 8, 9, 10, 11
2 12
5
Father_Edu
p < 0.001
2, 3
7
1, 3, 16, 98
2, 6, 17
6
Rel. Type Mean
0.963
Pop. Share (%)
8.69
7
Rel. Type Mean
0.791
Pop. Share (%)
30.94
8
Rel. Type Mean
0.647
Pop. Share (%)
20.1
9
Rel. Type Mean
1.335
Pop. Share (%)
4.99
12
Rel. Type Mean
1.463
Pop. Share (%)
5.64
13
Rel. Type Mean
0.858
Pop. Share (%)
8.76
14
Rel. Type Mean
1.509
Pop. Share (%)
8.83
17
Rel. Type Mean
1.107
Pop. Share (%)
3.62
18
Rel. Type Mean
2.011
Pop. Share (%)
3.85
20
Rel. Type Mean
2.648
Pop. Share (%)
3.5
21
Rel. Type Mean
4.516
Pop. Share (%)
1.1

Timor-Leste (2014)



Tajikistan (2016)



Uzbekistan (2016)