Research Article

Trajectory of inequality of opportunity in child height growth: Evidence from the Young Lives study

Toshiaki Aizawa

Contents

Trajectory of inequality of opportunity in child height growth: Evidence from the Young Lives study

Toshiaki Aizawa[1]

Abstract

BACKGROUND
Socioeconomic circumstances during infancy and in early childhood shape later developmental opportunities. Understanding whether inequality in children's health becomes stronger with age from the evidence of longitudinal data is an important first step in revealing the mechanism by which the intergenerational transmission of poverty takes place and evolves.

OBJECTIVE
This study investigates the 15-year trajectory of inequality in height growth associated with early-life circumstances among children in Vietnam, Peru, Ethiopia, and India.

METHODS
This study uses the datasets of the Young Lives study project, which is a large-scale cross-country longitudinal study on childhood poverty conducted in Vietnam, Peru, Ethiopia, and India since 2002. Machine learning approaches are employed to estimate the relationship between early-life circumstances and child height.

RESULTS
The inequality in height stemming from the difference in early-life circumstances persists even after children reach early adolescence. The proportion of inequality peaks when children are 5 years old. Our prediction using the random forest model shows that, if we were able to fully compensate for early-life socioeconomic disadvantages, we could increase the lower percentile of the height distribution and reduce the inequality in height at age 15 by half.

CONCLUSIONS
The results suggest that children from marginalised households should be supported at the earlier developmental stage.

---

[1] University of York, UK. Email: ta812@york.ac.uk.

CONTRIBUTION
This study shows the life-course evolution of inequalities in child growth, explores the dynamic link between early-life circumstances and later consequences, and identifies optimal timing in terms of when circumstances and events matter the most.

1. Introduction

The environments in which children grow up play vital roles in forming their health and human capital development. Children's health is thus heavily influenced by the characteristics of the families into which they are born (Bradley and Corwyn 2002; Case and Paxson 2002). A large volume of the literature has shown that children whose parents have lower socioeconomic backgrounds tend to experience worse health conditions than those with parents who have more advantaged socioeconomic backgrounds (Graham and Power 2004; Bradley and Corwyn 2002; Howe, Lawlor, and Propper 2013). Restricted development in childhood has long-lasting and profound effects on adult health, educational achievement, individual productivity, and life-course income (Strauss and Thomas 2007; Almond and Currie 2011; Alderman, Hoddinott, and Kinsey 2006; Case, Fertig, and Paxson 2005) because childhood is a period of rapid development in key dimensions of physical, cognitive, and psychosocial skills as well as health (Dercon and Sánchez 2013; Himaz 2018; Grantham-McGregor et al. 2007).

Exposure to poor circumstances during developmentally salient periods may alter the trajectory of health over the life-course. It has been repeatedly emphasised that the first few years of life are crucial for child development, so a poor start in life severely limits children's opportunities to achieve their full health potential. Children from disadvantaged backgrounds tend to be at high risk for a lack of or slow human capital development throughout their entire lives (Haas 2008), thereby forming a vicious cycle of the intergenerational transmission of poverty, making it harder for marginalised people to find a way out of poverty. Although catch-up in the periods from middle childhood to early adolescence is reported (Fink and Rockers 2014; Crookston et al. 2013; Georgiadis et al. 2017), the chance of later catch-up is not necessarily equal for all children (Contoyannis and Li 2011). For example, Schott et al. (2019), who examine trajectories in malnutrition among youth, find that children who belong to the richest wealth quartile and have more advantaged maternal educational achievement have an advantage in recovering from stunting.

Inequality of opportunity (IOp) is an illegitimate form of inequality whereby causes are not attributable to individual choices (Roemer 1998). Equality of opportunity is achieved when all children share the same opportunities to enjoy better health, regardless of the circumstances in which they grow up. Around the world, various studies on

child health inequality associated with parental socioeconomic status (SES) show evidence that poorer SES is associated with poor child health (Wagstaff and Watanabe 2000; Van de Poel et al. 2008). However, most of the existing evidence is based on cross-sectional data, and little is known about how the trajectory of inequality in children's development associated with early-life circumstances diverges from early in life through to early adulthood. In particular, what is little known is at what age such inequality peaks and whether and how it changes across the developmental stages. Examining the period when the inequality stemming from early-life circumstances emerges and becomes strongest is important to design policy interventions to optimally reduce life-course health disparities.

Development trajectory may, in part, be forged very early in life, and as the environments in which children grow up influence multiple factors in later life, their effects are expected to linger over time (Ben-Shlomo and Kuh 2002; Lynch, Kaplan, and Salonen 1997; Lynch and Smith 2005). A large volume of studies shows that the socioeconomic inequalities in health widen across childhood with no diminution (Case, Lubotsky, and Paxson 2002; Currie and Stabile 2003; Chen, Matthews, and Boyce 2002; Lundeen et al. 2014), but this finding is not universal (Chen, Martin, and Matthews 2006; Currie and Moretti 2007; Khanam, Nghiem, and Connelly 2009; Propper, Rigg, and Burgess 2007). Up till now, there has been no consensus partially because the results in previous cross-sectional studies depend on the geographic setting and the birth year of the study participants (Howe, Lawlor, and Propper 2013). Understanding if the inequality of opportunity in children's health becomes stronger with age from the evidence of longitudinal data is an important first step in revealing the mechanism by which the intergenerational transmission of poverty takes place and evolves.

This study contributes to the literature by estimating the IOp in child height, which is a proxy of child nutritional development, by using cross-country cohort data that track children born in the 21st century in Vietnam, Peru, Ethiopia, and India. A better understanding of IOp in the life-course is paramount to tackle health inequality and restricted growth. Life-course perspectives on the evolution of inequalities allow us to explore the dynamic link between early-life circumstances and later consequences and to identify the optimal timing in terms of when circumstances and events matter the most.

This paper comprises two parts. In the first part, we examine whether there is significant inequality in height growth across different parental and household socioeconomic backgrounds through the clustering approach. Then, in the second part, we estimate how much of the overall inequality in child height is associated with early-life circumstances using the machine learning approach. We then examine whether the inequality attributable to early-life circumstances diminishes over time or is persistent/growing until early adolescence. After this, we predict the child height distribution in the most desirable hypothetical situation, where disadvantages with respect to household and parental SES in early childhood are fully compensated for.

2. Data

2.1 Young Lives

This study uses the datasets of the Young Lives study project, which is a large-scale cross-country longitudinal study on childhood poverty conducted since 2002 in four countries, namely Vietnam, Peru, Ethiopia, and India. These four countries represent key regions in the low- and middle-income countries and reflect a broad range of social, cultural, political, and geographical contexts (Barnett et al. 2013). These four countries differ with respect to economic and socioeconomic development over the study period. Figure 1 illustrates per-capita GDP of each country from 2002 to 2016. At the start of the project, Peru had the highest GDP per capita (6,659 USD) and Ethiopia had the lowest GDP per capita (645 USD) of the four countries.

Figure 1: **GDP per-capita transition (PPP, 2011 international $)**

Note: GDP per-capita based on purchasing power parity (PPP). PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates.
Source: World Bank (2018).

The Young Lives projects follow two age cohorts in each country born in 1994–1995 and 2000–2001 over 15 years. This study focuses on only the younger cohort, which consists of about 8,000 children. The key strength of the cohort study is that it follows the experiences of the same children as they grow up, which allows us to identify dynamic trends of child growth. The Young Lives study currently has five survey rounds, conducted in 2001–2002, 2006–2007, 2009–2010, 2013–2014, and 2016, when children in the younger cohort were around 1, 5, 8, 12, and 15 years old, respectively. In each sur-

vey round, trained field personnel collected detailed data on a wide range of child and household characteristics.

The Young Lives project employed the multi-stage sampling strategy to select children in the respective countries. The first stage selected 20 geographic clusters per country, at random in Peru and based on a number of predetermined criteria in the other countries. The second stage selected about 100 eligible households with a child aged between 6 and 18 months at random within each cluster. In each household, only one child of the target group was recruited. In Peru, Vietnam, and Ethiopia, a nationwide sampling frame was used, whereas in India samples were collected in the state of Andhra Pradesh. Although the samples were not intended to be nationally representative, their wide geographical and ethnic coverage makes them informative of the living conditions faced by the population in each country, making the data reflect the ethnic, geographic, agroclimatic, and religious diversity of the population (Outes and Sanchez 2008; Nguyen 2008; Kumra 2008; Escobal and Flores 2008). The advantage of the Young Lives data is the low attrition rates in each survey round thanks to extensive tracking (Outes-Leon and Dercon 2008; Barnett et al. 2013). In addition, the questionnaire of the Young Lives project has been standardised across populations, which allows us to compare the estimation results across countries.

2.2 Child health outcome

The health outcome in this study is child height (cm), which has been used as a proxy of early-life health and nutrition status (Floud 1990). Different from weight, height reflects long-term nutritional status, and its significant correlation with adult health, development of cognitive and non-cognitive skills, and educational outcomes is reported in various studies (Case and Paxson 2010, 2008; Glewwe, Jacoby, and King 2001; Dercon and Sánchez 2013; Sánchez 2017; Duc 2011). For example, in the case of Guatemala, shorter height in childhood is significantly associated with lower schooling achievement, lower test scores for reading and nonverbal cognitive skills, and lower future economic well-being (Hoddinott et al. 2013). Although it is not perfect, thanks to its ease of measurement, height works as a useful proxy for nutrition and health in the early years of life (Currie and Vogl 2013). Child height in the Young Lives project is based on actual measurement of supine length (Round 1) and standing height (Rounds 2–5) using measurements precise to 1 mm. Hence the data do not suffer self-reporting bias. Prior to the analysis, implausible measurements with more or less than four standard deviations from the mean for gender and age in each country were recoded as missing.[2]

---

[2] Approximately 0.1% of the original information was recoded.

2.3 Early-life circumstance variables

It is important to consider multidimensional household and parental characteristics to capture various aspects of circumstances in which children grew up. In this study, we consider the following eight circumstances that are related to parental and household characteristics during the infant period surveyed in all four countries (Table 1): (1) housing quality, (2) access to services, (3) consumer durables, (4) maternal education, (5) paternal education, (6) maternal height, (7) household monthly expenditure, and (8) economic and social shock history. As the information about maternal height and household expenditure was not collected in the first survey round, we obtain the information from the second survey round. For the other circumstances, we use the information from the first survey round. As children cannot choose their parents or households, parental and household characteristics are undeniably beyond their control, and so these variables are exogenous in nature from the viewpoint of children.

Table 1:   Circumstance variables in childhood

	Circumstance factors	Components
1	Housing quality	Materials of roofs, walls, and floors, and housing density
2	Access to services	Electricity, safe drinking water, sanitation service, and adequate fuel
3	Consumer durables	Radio, refrigerator, bike, TV, motorbike, motor vehicle, mobile phone, land phone, etc.*
4	Maternal education	Years of education
5	Paternal education	Years of education
6	Maternal height	Mother's height
7	Household consumption	Food and non-food expenditure per person in 2006 local currency**
8	Negative events	Experiences of negative economic, physical, and social events

Note: * Number of items slightly differs across countries. ** In Ethiopia, expenditure per adult equivalent is calculated.

Housing quality, access to services, and consumer durables reflect different aspects of the household's material well-being. Following the Young Lives project guideline (Briones 2017), we calculate the composite indices for each category. Housing quality measures the welfare of household members related to housing-related comfort. In terms of child health, poor, dilapidated, and crowded housing has been considered to be one of the most important factors in child illness (Bradley et al. 2001). The housing quality index is calculated as a simple average of the following four indicators: (1a) material of walls, (1b) material of roof, (1c) material of floors, and (1d) household density.[3] Good-quality walls are those made of good-quality materials such as brick, concrete, and stone. Good-quality roofs are those made of concrete/cement, galvanised/corrugated iron, tiles/slate,

---

[3] A simple average assumes that these four factors are equally important (Briones 2017).

etc. Good-quality floors are those made of concrete/tile, laminated material, stone, parquet, etc. Housing density is defined by the rescaled value of room per person, ranging from 0 to 1.

Access to services measures the household's ability to satisfy the functional requirements of sound shelter. Again following Briones (2017), we create its index as a simple average of the following four indicators: (2a) access to electricity, (2b) access to safe drinking water, (2c) access to a safely managed sanitation service, and (2d) access to adequate fuel for cooking. A managed sanitation service means that a household has a flush toilet/septic tank, pit latrine, etc. Adequate fuel for cooking is gas/electricity or kerosene/paraffin.

Consumer durables measures the household's ownership of items commonly found in typical households in each country. The index is a simple average of a set of variables indicating ownership of a long list of items such as a radio, refrigerator, bike, TV, motorbike, motor vehicle, mobile phone, landline, and so on (Briones 2017). The number of items considered slightly varies across the four countries. In the survey, 9, 12, 10, and 9 items are surveyed in Vietnam, Peru, Ethiopia, and India, respectively.

To characterise the parental financial situation, we use the monthly household food and non-food expenditure per person.[4] Household affluence has been regarded as one of the significant socioeconomic factors creating inequality in health (Huda et al. 2017; Chalasani 2012). Expenditure/consumption is regarded as a more suitable indicator for measuring household affluence than income because consumption is less susceptible to temporary economic shocks and seasonal events such as drought. The Young Lives project team calculated household consumption, using adult respondents' reports of food and non-food items with a recall period ranging from 15 days for food to 12 months for clothing (Behrman et al. 2013). The total expenditures were first converted to real monthly expenditures in 2006 local currency and divided by household size. We use the logarithmic of the monthly expenditure.

We also consider parental educational backgrounds as circumstantial precursors to inequalities. For the relationship between parental educational background and child development, a number of empirical studies report the existence of education-related health inequality (Levin et al. 2000; Currie and Moretti 2003; Chen and Li 2009). We measure the maternal and paternal educational backgrounds by years of education completed by a mother and a father, respectively.

We use maternal height (cm) to take account of both the heritability of height (Silventoinen et al. 2003) and its potential effect on child care.[5] The significant correlation

---

[4] Only in Ethiopia, the adult equivalent is calculated. This is because the Young Lives country team calculated consumption independently, and there is a slight methodological difference for Ethiopia (Briones 2017).

[5] Certainly, paternal height could also be an important circumstance factor. However, information about the height of the father was not surveyed in the Young Lives study.

between maternal height and child health is observed across the countries (Li et al. 2009; Özaltin, Hill, and Subramanian 2010; Beard 2001; Gillespie 1997).

Finally, we consider the history of negative economic, physical, and social events that households experienced. First, we make a continuous index as a simple average of a set of experiences that the household experienced when the children were infants. We consider the following eight shock histories: destruction/theft of crops, destruction/theft of livestock, becoming a victim of crime, crop failure, natural disaster, death of livestock, loss of a source of income, and decrease in food availability. After calculating the sample average, we reverse the scale so that a higher index value will indicate lesser shock experience. The descriptive statistics of these early-life circumstance variables in each country are shown in Table 2.

Table 2: **Descriptive statistics of early-life circumstances (Round 1)**

	Vietnam		Peru		Ethiopia		India	
	mean	sd	mean	sd	mean	sd	mean	sd
Housing quality	0.56	0.32	0.40	0.25	0.22	0.19	0.50	0.30
Service availability	0.42	0.27	0.60	0.35	0.31	0.31	0.55	0.28
Comsumer durables	0.35	0.21	0.28	0.22	0.10	0.13	0.18	0.17
Maternal education years	5.64	4.16	7.25	4.88	2.28	3.62	3.05	4.32
Paternal education years	6.14	4.20	8.47	4.65	3.06	4.15	4.58	5.08
Maternal height	152.18	5.45	150.06	5.29	158.38	5.33	151.73	5.39
ln (HH exp. per-capita)*	5.78	0.62	5.01	0.67	4.75	0.62	6.69	0.53
Negative shock history	0.96	0.07	0.97	0.05	0.81	0.18	0.91	0.13
Observations	1806		1456		1427		1823	

Note: * ln(HH exp. per-capita) stands for the natural logarithm of per-capita household expenditure.

3. Method

The main analysis is composed of the following two steps. First, we use the K-means clustering approach to partition the children in each country into a few distinctive groups and investigate if there is a significant difference in height distribution across different socioeconomic groups. Second, we estimate how much of the total variation in height at each development stage is associated with early-life circumstances. We model the relationship between circumstances in childhood and height in each survey period with the random forest algorithm. These two steps are independent each other and all the analyses are conducted separately by country. More detailed procedure in each step is described in the following subsections.

3.1 Testing the existence of inequality of opportunity

In the first step, we examine whether there is a difference in height across different socioeconomic groups. We partition the children in respective countries into a few types, i.e., children who grew up in the same circumstances (Roemer 1998). However, defining the types is not always easy when we take into account the multidimensional circumstance variables. In this paper, following the approach taken by Aizawa (2019), we cope with the multidimensionality of circumstances using the K-means clustering approach. The advantage of the clustering approach is that types defined by the clustering take account of the differences in unobservable characteristics as well as observed circumstance factors. The following explanation of the K-means clustering approach follows Aizawa (2019).

Clustering is a data-driven technique to partition the observations into a pre-specified number of non-overlapping clusters so that the observations within each cluster group become similar to each other, while at the same time the observations in different groups differ from one another. We first specify the desired number of clusters K. Then, the K-means algorithm assigns each observation to one of the K clusters in such a way that the within-cluster variation, summed over all clusters, becomes as small as possible. In the K-means clustering, the within-cluster variation for the Kth cluster is calculated by the sum of all of the pairwise squared Euclidean distances between the observations within the Kth cluster, divided by the total number of observations in the Kth cluster.[6]

Deciding the optimal number of clusters, K^*, is an essential process in the K-means clustering. We seek the optimal number of clusters with the average silhouette method (Kaufman and Rousseeuw 1990), which computes the average silhouette of observations for various numbers of K. As the average silhouette calculates how well each observation lies within its cluster, its higher value indicates a better clustering. The optimal number of clusters, K^*, is the one that gives the highest average silhouette over a range of the cluster numbers (Kaufman and Rousseeuw 1990).

After the K-means algorithm assigns each observation to exactly one of the K^* clusters, we compare the height distribution across clusters. If a significant difference across clusters exists, it is a sign of the existence of inequality in child height related to the circumstances (Lefranc, Pistolesi, and Trannoy 2008, 2009).

---

[6] Before applying the K-means clustering, we standardise the circumstance factors so that their values will range from 0 to 1 in order to calculate the Euclidean distances between the observations.

3.2 Measuring the inequality of opportunity

3.2.1 Empirical life-course frameworks

Empirically we consider the following life-course model in equation (1), where the child height at period t is determined by the investments that parents make and the circumstances. The investments can be interpreted as the inputs that influence the child height but are not regarded as the illegitimate sources of inequality.[7] Investments are influenced by circumstance factors, and circumstances at period t are path-dependent in the sense that they are partially dependent on the circumstances in a preceding period. Following Roemer (1998), we regard the part of the inputs that are correlated with the circumstance as part of the effect attributable to the circumstance. This means that any variation of investments attributable to the circumstances is regarded as the part induced by the circumstances.

$$\begin{cases} H_{i,t} = H_t(E_{i,t}, C_{i,t}) \\ E_{i,t} = E_t(C_{i,t}) \\ C_{i,t} = C_t(C_{i,t-1}), \end{cases} \tag{1}$$

where $H_{i,t}$ is the child height in household i at period $t = \{1, 2, \ldots, T\}$. $E_{i,t}$ and $C_{i,t}$ denote vectors of inputs and circumstance variables, respectively. The reduced form of equation (1) is simply described as

$$H_{i,t} = f_t(C_{i,1}), \tag{2}$$

where $C_{i,1}$ is the circumstance factors in early life.

3.2.2 Machine learning approach to measure inequality of opportunity

It is of paramount importance to estimate the health production function in equation (2) so that it can yield the most reliable measures of inequality related to early-life circumstances. According to Brunori, Peragine, and Serlenga (2019), estimating the IOp can be understood as a prediction problem, and the authors suggest incorporating some of the machine learning techniques. According to Murphy (2012), machine learning is "a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision-making under uncertainty." Brunori, Peragine, and Serlenga (2019) discuss the importance of incorporating the perspective of a trade-off between bias and variance in estimating the IOp from the

---

[7] They are referred to as "efforts" in the literature of inequality of opportunity (Roemer 1998).

samples at hand. Bias is caused by under-fitting the model to the data, which is usually caused by a function that is too simple. Variance is caused when a model fits the data at hand too well. In general, the good predictive performance of a certain model in the samples does not necessarily guarantee that it also exhibits good predictive performance for the different samples. A very flexible and complex model can explain very well variations of the outcome in the sample at hand, but it may have poor predictive performance when out of sample, which is recognised as an over-fitting problem.

Brunori, Peragine, and Serlenga (2019) especially emphasise the importance of selecting a model that maximises the ability of circumstances to explain the distribution of the outcome; they recommend the cross-validation method for model selection in order to balance the tension between under-fitting and over-fitting. The machine learning approach provides the conditional expectation function with high prediction accuracy and at the same time the least bias and sample variance.

This study employs the random forest algorithm. Random forest is one of the most popular supervised machine learning algorithms, and its high predictive performance is well recognised (Lantz 2015). Random forest uses multiple regression trees, and they are thereby known as ensembles of trees. Each regression tree stratifies the explanatory variable space into distinct and non-overlapping regions. At each internal node in the tree, the algorithm makes a binary decision as to whether we move to either the left or the right sub-branch of the tree. Eventually, at a leaf node, the regression tree makes a prediction, which is simply the mean of the outcome values for the observations in each region.

The random forest builds a number of regression trees on bootstrapped samples, in which a random subset of explanatory variables is chosen each time as split candidates, and only the selected subset of circumstance variables is used to build a regression tree (Breiman 2001). Repeating this process many times grows a forest of regression trees, and then the prediction of the random forest is achieved by finding the average of the predictions made by each regression tree. The random forest is non-parametric and can handle outliers well. Although the model is usually difficult to summarise visually because of multiple trees, it shows good predictive performance, especially with highly nonlinear data (Lantz 2015).

In general, more flexible and complicated models can capture detailed variations in the data, but they tend to work poorly for out-of-sample forecasts in comparison to much simpler models. As we add more flexibility to models, the bias will decrease but the variance will increase. This is known as a bias-variance trade-off, which derives from the model's over-fitting to the in-sample observations. The machine learning approach is designed to find such a model that controls the bias-variance trade-off. We split the entire sample group into training samples and testing samples. The former is used to train a model and the testing samples are used for the estimation and prediction. We use 70% of

the entire sample as the training samples and the remaining 30% as the testing samples. Our estimation of the IOp and prediction are based on these testing samples.

3.2.3 Parameter tunings with tenfold cross-validation

Machine learning entails tuning the model's parameters, which control the complexity of a model. It is essential in machine learning to tune model parameters in order to make sure the model has high out-of-sample predictive performance. The objective of parameter tuning is to allow the model to demonstrate the best balance between under-fitting and over-fitting by seeking the optimal degree of model complexity. The parameter of the random forest is the number of randomly selected trees. In tuning the model's parameters, we apply the tenfold cross-validation method, which is a de facto standard in machine learning (Lantz 2015).

The tenfold cross-validation first divides the set of observations in the training samples into 10 approximately equally sized folds. The first nine folds are used to construct/train the model, and the last fold is used as a validation set to measure the model's predictive performance. The root mean squared error, $RMSE = \sqrt{\frac{1}{N^*} \sum_{i=1}^{N^*} (H_i - \widehat{H_i})^2}$, is then computed on the N^* observations in the validation set. This procedure is repeated 10 times; on each occasion, a different group of observations is treated as a validation set. Finally, we obtain 10 estimates of the test error, $RMSE_1, \ldots, RMSE_{10}$, and the tenfold cross-validation estimate is then computed by averaging these 10 values. After repeating this process for various parameter values, we choose the best parameter value with the highest performance, namely the lowest average RMSE. We consider five parameter values for each parameter.

3.2.4 Prediction and inequality of opportunity measurement

After modelling the relationship between height in each period of time and circumstances during the infancy, we predict the height with the estimated model. When the modelled relationship at period t is expressed by $\widehat{f_t}(.)$, the predicted value, $\widehat{H_{i,t}} = \widehat{f_t}(C_{i,1})$, yields the part of the height that is attributable solely to individual circumstance variables during the infancy. Hence, the inequality in $\widehat{H}$ can be considered the one associated with any variation in the circumstances in early life, which corresponds to the IOp that we seek to estimate. Then we measure the proportion of the IOp to the overall observed inequality.

The prediction error, $\widehat{u_{i,t}} = H_{i,t} - \widehat{H_{i,t}}$ reflects individual unobservable circumstance factors as well as individual idiosyncratic characteristics, including individual efforts. Note that the inequality of this remaining part of the outcome unexplained by C_1 does not necessarily mean legitimate inequality, as the difference, $H - \widehat{H}$, reflects not
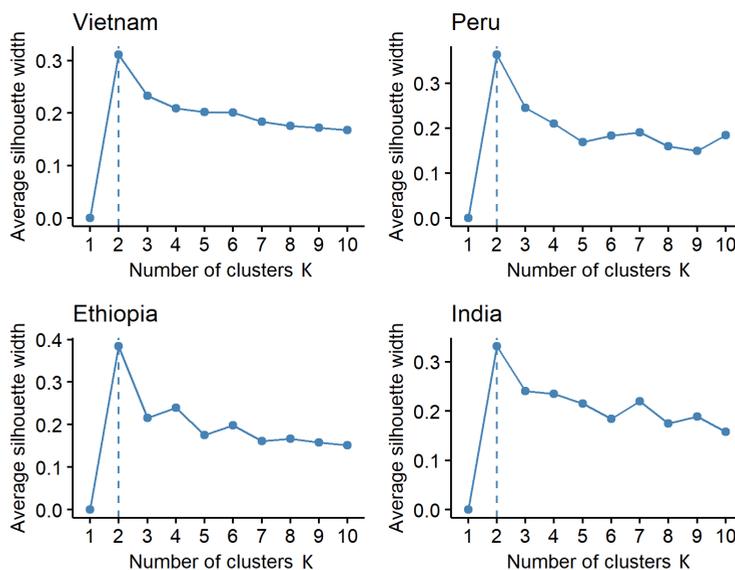
only individual efforts but also circumstances not included in the regression model. In this sense, inequality in $\widehat{H}$ should be regarded as the subset of the entire IOp associated with the circumstances in the model.

4. Results

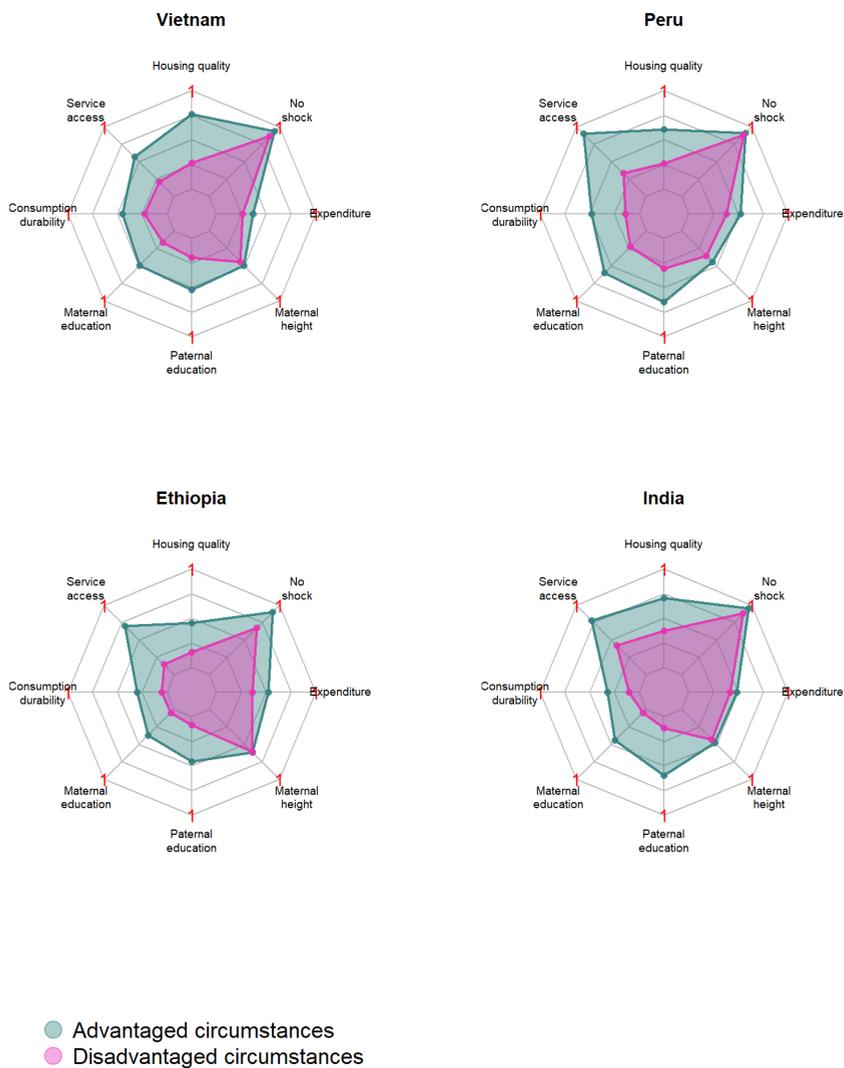## 4.1 Clustering of circumstances in childhood

First, we seek the optimal number of clusters, K^*, for each country. Figure 2 plots the average silhouette width over a range of the cluster numbers, where the number of clusters that maximises the average silhouette is the optimal number of clusters. We find that for all countries, the average silhouette width method chooses two clusters. Figure 3 illustrates the mean values of eight circumstance factors for each cluster in each country. We find the clear between-cluster difference in the early-life circumstance variables; children belonging to the first cluster have more advantaged parental and household SES than those belonging to the second cluster with respect to housing quality, service access, consumption durability, maternal and paternal education, and household expenditure ($p < 0.01$). On the other hand, for maternal height and experience of shocks, the two clusters show relatively small mean differences, but they are statistically significant ($p < 0.05$) except in Ethiopia for maternal height.

Figure 2: **Average silhouette and optimal cluster numbers**

Next, the box plot in Figure 4 illustrates the between-cluster difference in the distribution of height at each developmental stage. The t-test shows that children belonging to the advantaged early-life circumstances cluster are taller on average than those belonging to the disadvantaged early-life circumstances cluster in all the points in time ($p < 0.01$). Significant between-cluster disparities in the proportion of stunted children are also found in every survey period in all four countries (Figure 5), where being stunted is based on whether the height is two standard deviations below the international reference median height for the same sex and age (WHO 2006). The largest between-cluster disparity in the stunted population is found among Ethiopian boys at the age of 15 years as 30 percentage points ($p < 0.01$). The smallest between-cluster disparity is observed among Vietnamese girls as 5 percentage points ($p < 0.05$).

Figure 3: **Differences in circumstances between clusters**

Advantaged circumstances
Disadvantaged circumstances

Note: Each circumstance variable is rescaled so that each ranges from 0 to 1.

Figure 4: **Difference in height trajectory between clusters**
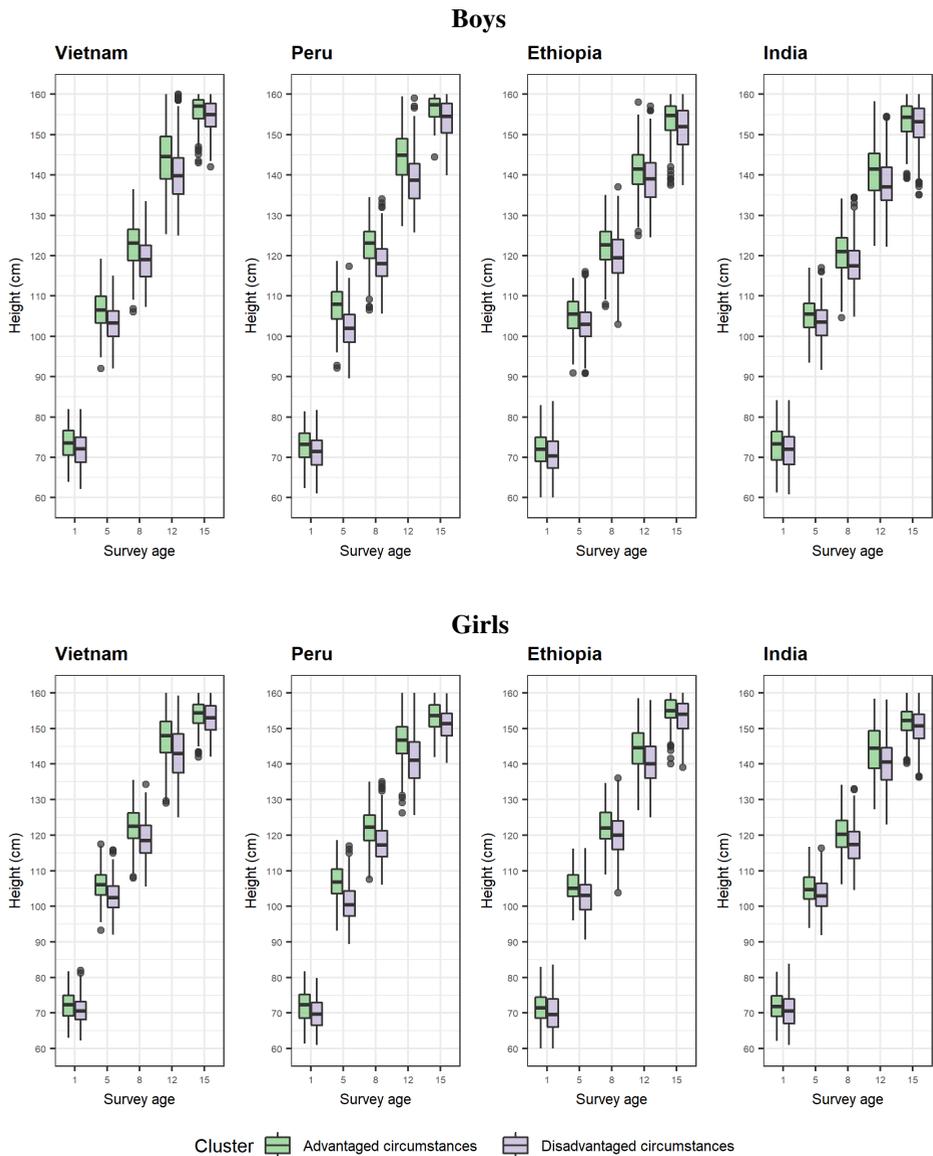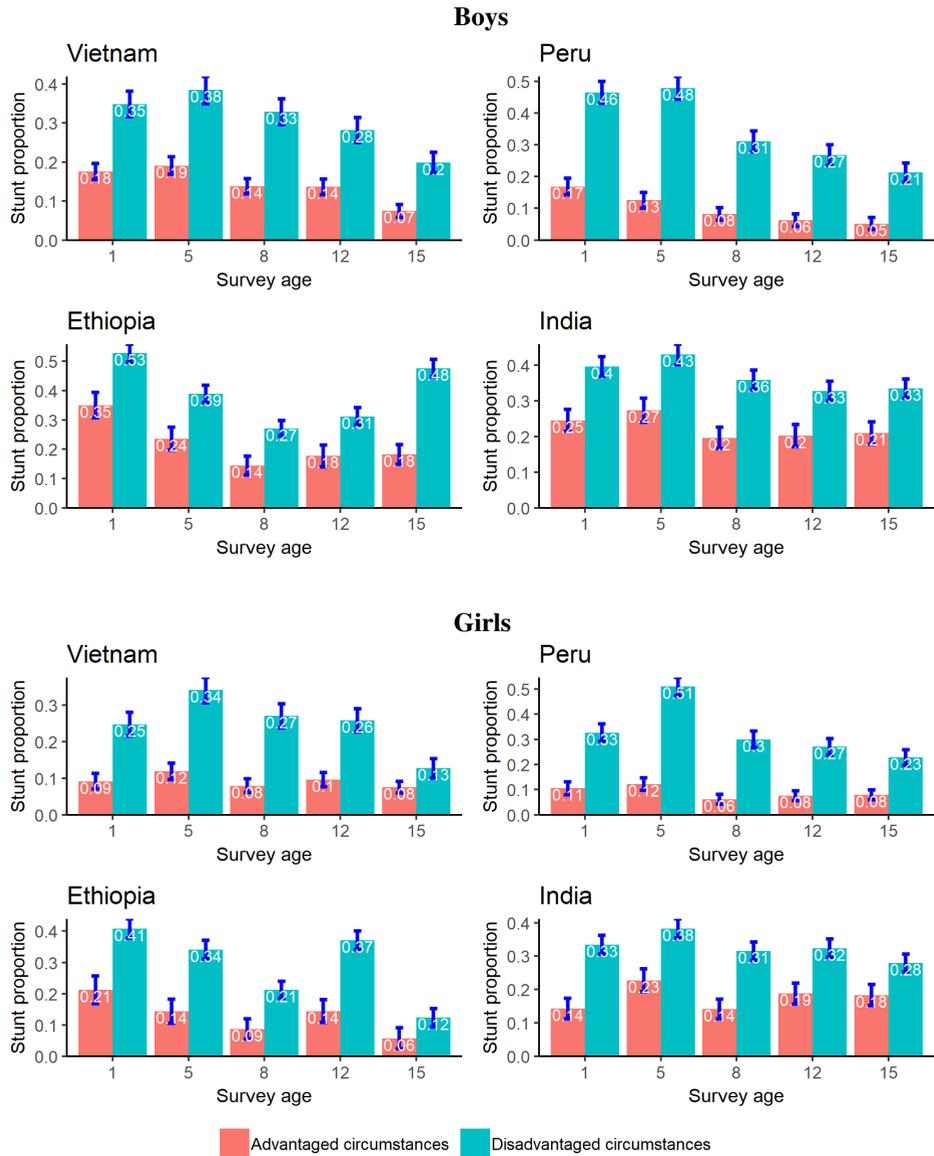
Boys

Girls

Figure 5:   **Difference in the proportion of stunted children between clusters**

More detailed analysis can be made by looking at the entire distribution of height. Looking at the entire distribution allows us to analyse the difference in the distribution that does not show up as the difference in mean. Figure 6 shows the estimated cumulative distribution functions at the age of 15. The Kolmogorov-Smirnov test rejects the equality of these two distributions in each country ($p < 0.01$). As well as the significantly higher means among children belonging to the advantaged household cluster, we observe that the cumulative distribution of the disadvantaged cluster lies above that of the advantaged cluster among boys in Vietnam, Peru, Ethiopia, and India and among girls in Peru, Ethiopia, and India, indicating the first-order stochastic dominance of the advantaged cluster over the disadvantaged cluster in respective samples. For girls in Vietnam, though we do not find the first-order stochastic dominance, the second-order stochastic dominance is observed.

4.2 Measurement of inequality of opportunity

In the second part, we first model the relationship between circumstances in childhood and height in each survey period with the random forest algorithm. In the random forest, an overall summary of the importance of each circumstance is measured by a reduction in total residual sum of squares (RSS) (James et al. 2013). Random forest records total RSS reduction due to splits over a given circumstance, of which information can be used as a measure of the total decrease in node impurity that results from splits over that circumstance variable, averaged over all trees. A large value thereof indicates that a corresponding circumstance variable is an important determinant.

Figure 7 shows the reduction in the RSS in the first survey round. The results for the other survey rounds are available in Table A-2 in the appendix. A larger value indicates that the corresponding circumstance variable made a large contribution in explaining any variation in child height. In all four countries, maternal height and household expenditure show higher reductions in RSS, followed by housing quality. Shock experience, service availability, and consumer durables indicate relatively smaller importance in explaining the variation in height. Also, we observe that both maternal and paternal educational backgrounds are important predictors for child height during infancy.

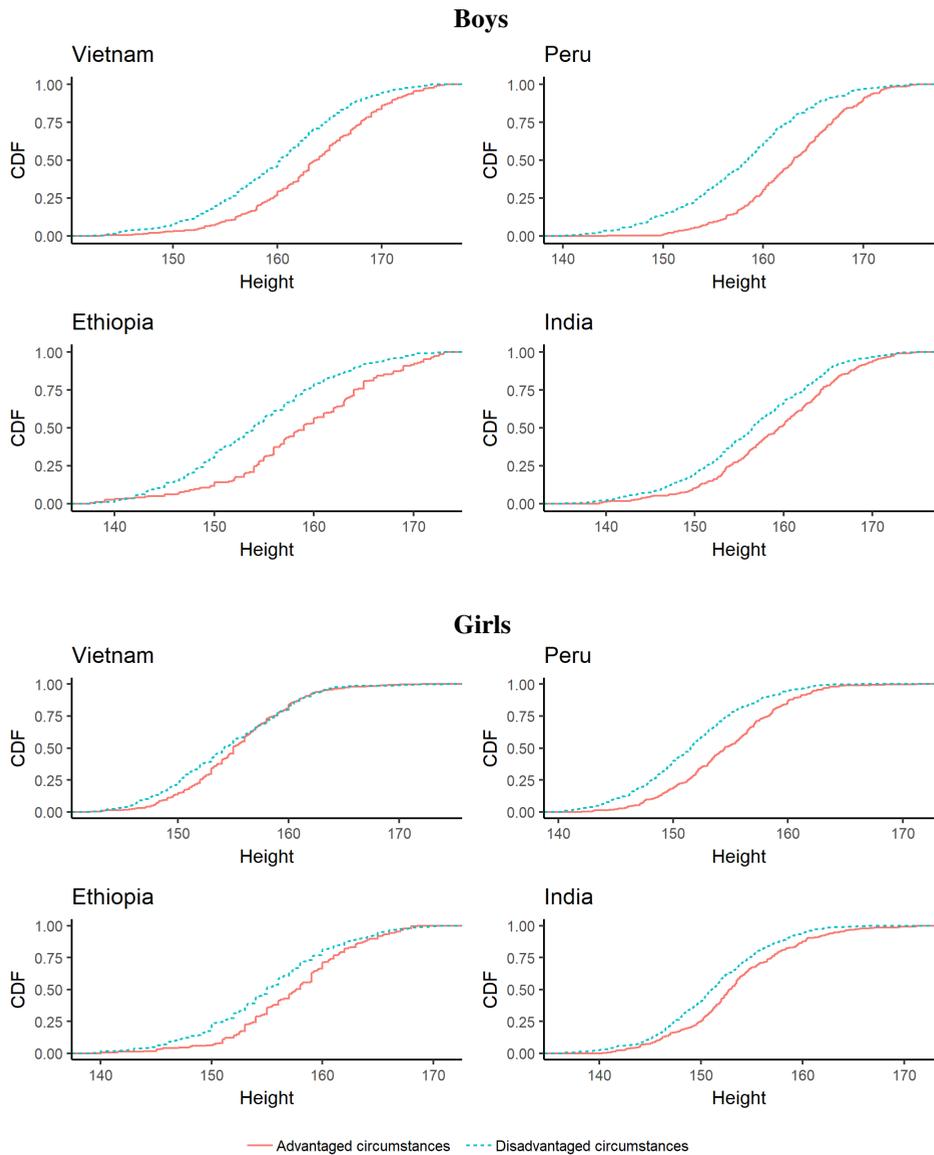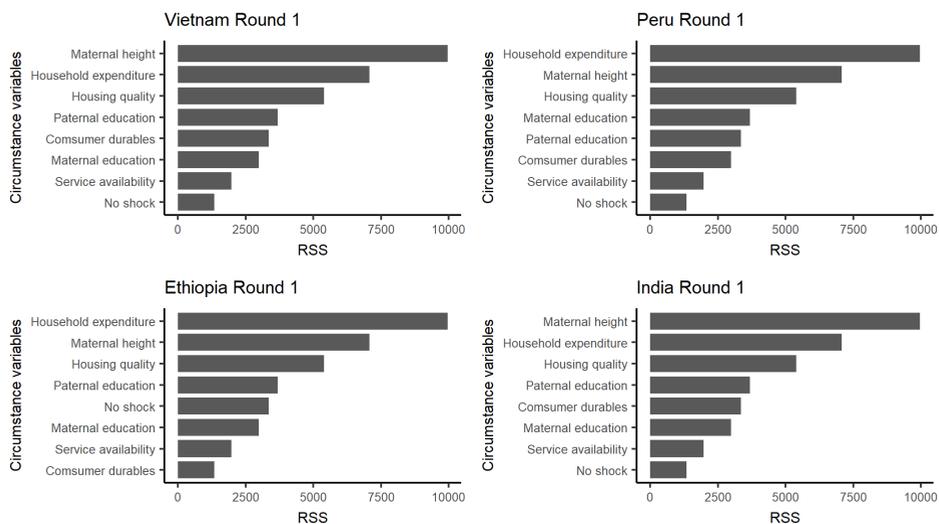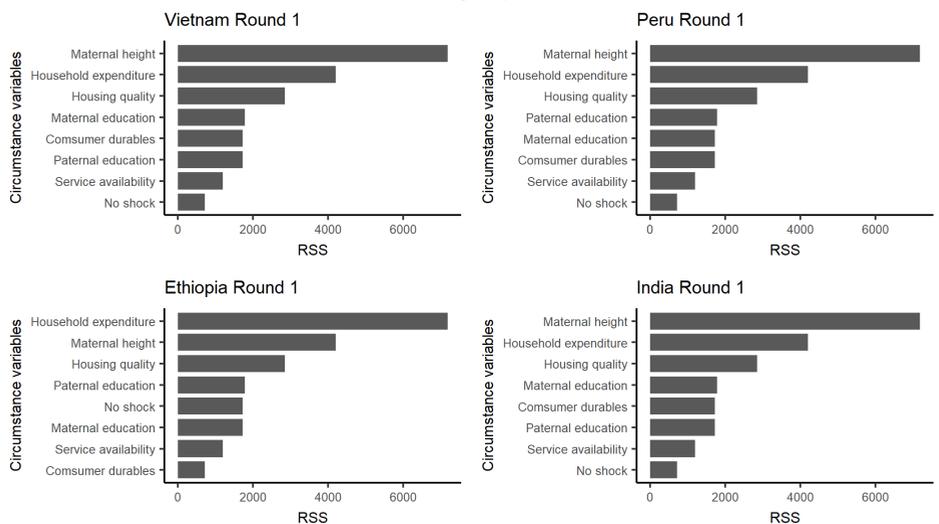**Figure 6:** **Difference in the proportion of stunted children between clusters**

Figure 7:   **Importance of each circumstance factor during infancy**
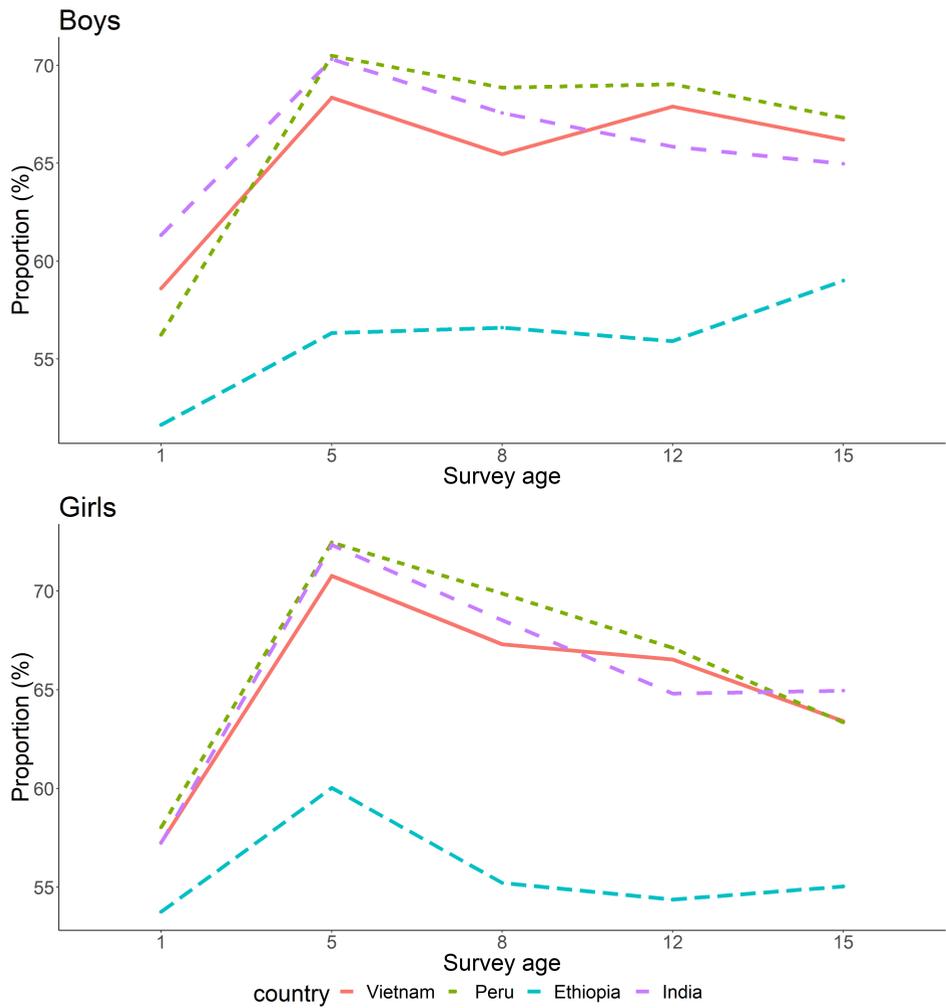
Boys

Girls

Then we investigate how much of the total variation in height at each development stage is associated with early-life circumstances. Intuitively, if the circumstances during the infancy period have only a short-term contemporaneous effect, the proportion of IOp attributable to early-life circumstances should decay over time. On the other hand, if circumstances during the infancy period are forming long-term inequalities, the IOp proportion should be persistent until the early adolescence period.

Figure 8 illustrates the trajectory of the IOp proportion among the testing samples, in which standard deviation is used as an inequality statistic. We find that in Peru, the part of inequality associated with circumstances is the lowest for both boys and girls across time. The heap in the IOp proportion is found in the post-infancy period (after age 5) in each country. For example, in Peru, the IOp proportion increased from 56.0% to 70.6% between the infancy period (age 1) and the pre-school period (age 5). In the early adolescence period, the IOp proportions show gradual declines but their speeds are very slow, thus resulting in persistent IOp. The result shows that even in the early adolescence period, around two-thirds of the total variation is attributed to the parental SES in childhood in Vietnam, Peru, and India. As a robustness check, we use other inequality statistics such as variance, Gini coefficient, coefficient of variation, and Theil index; we obtained consistent findings, albeit with variation in estimated magnitudes. These results are available in Figure A-1 in the appendix.

Figure 8:   **Trajectory of the proportion of inequality of opportunity**

Note: Only the testing samples are used for the calculation of the IOp proportion.
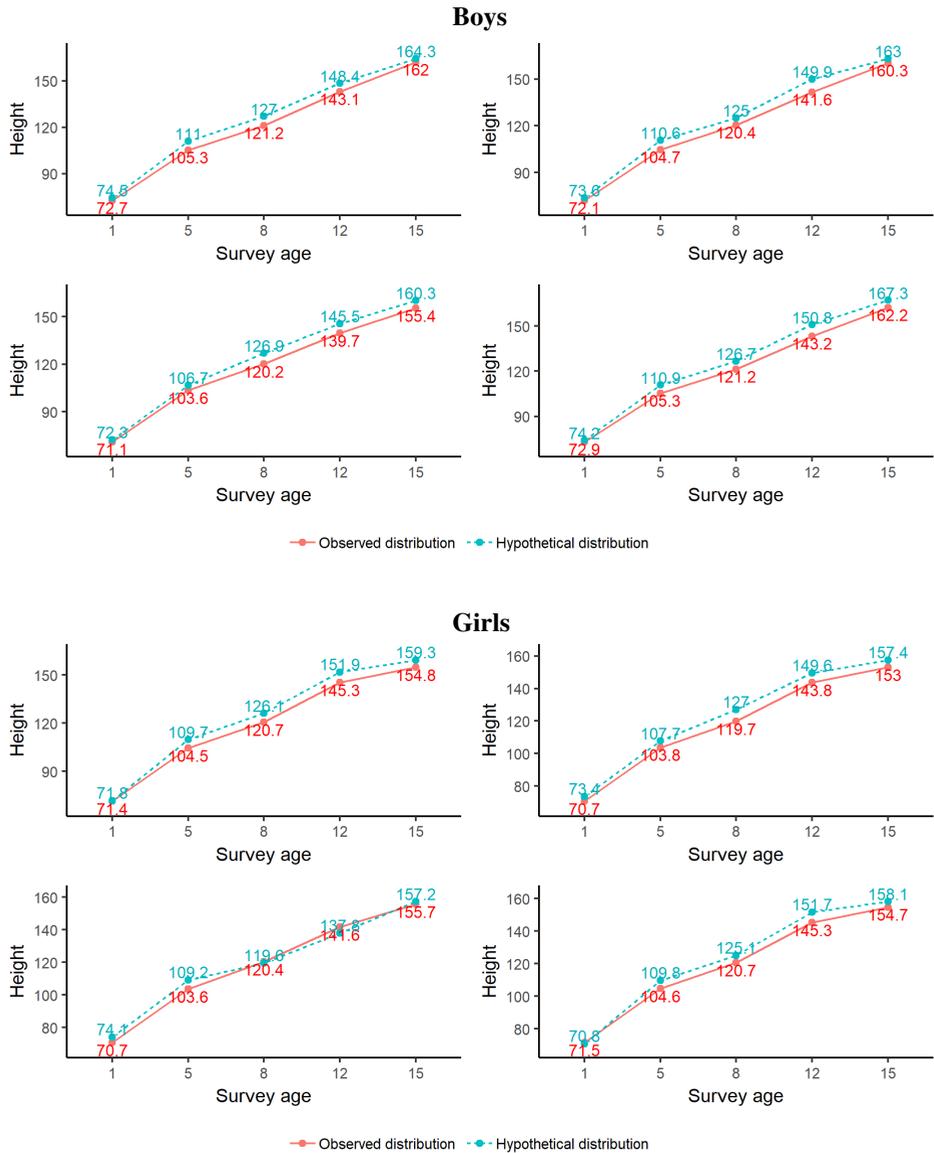
4.3 Height growth trajectory prediction

In the last step, we explore how much fully compensating for the circumstantial disadvantages during infancy could potentially improve child height growth and mitigate its inequality if we were able to totally remove the disadvantages of parental SES by levelling the playing field. This counterfactual exercise predicts the hypothetical height growth under the most desirable situation, where every household enjoyed the most advantaged circumstances during infancy. This hypothetical situation assumes that every child lives in adequate housing and has access to electricity, safe drinking water, improved sanitation, and safe cooking fuel. Also, we assume that every child has a mother and a father with university education attainments and experiences no negative economic and social shock. However, we do not make any assumption regarding maternal height because, in contrast to the other circumstance variables, maternal height cannot be changed by policy interventions. Figure 9 shows the trajectory of height growth, which plots the observed height levels and the predicted mean height levels under the hypothetical situation of the testing samples. The predicted outcome clearly shows the greater height growth under the hypothetical scenario for both boys and girls in all four countries.

More detailed exploration is possible by taking a look at the entire distribution of the counterfactual outcome. Figure 10 plots the estimated cumulative distribution functions for the observed outcome and the predicted outcome at age 15. First, we find that the predicted outcome has smaller variance compared with the observed outcome. This is totally intuitive, as our counterfactual exercise, in essence, suppresses the distribution of the circumstance variables. For boys, standard deviations are herein reduced by 55.4%, 57.3%, 51.1%, and 54.8% in Vietnam, Peru, Ethiopia, and India, respectively. For girls, reductions by 55.5%, 53.9%, 49.3%, and 55.6% are predicted in the respective countries.

Second, we observe a relatively larger increase in height in the lower percentile of the distribution among boys in Vietnam and Ethiopia, and among girls in Vietnam, Peru, and India, suggesting that the compensation for the disadvantage with respect to the early-life circumstance environments could help shorter children catch up. In contrast, we observe relatively smaller or no improvement in the higher percentile of the distribution, implying that the compensation would be more effective for stunted children without exacerbating the existing inequality. Among boys in Ethiopia, for example, while full compensation would increase the 0.25th quantile of the distribution by 12.3 cm, it would increase the 0.75th percentile of the distribution by just 2.0 cm.

Figure 9: **Trajectory of height under the hypothetical situation**

Note: Only the testing samples are used for the trajectory prediction.

Figure 10: **Cumulative distributions of height at age 15**

5. Discussion

Socioeconomic environments in the very early stage of life are of great importance for children's sound development, and the differentials in circumstances during the infancy period have a long-term implication for inequality in health status in the post-infancy period. In this sense, disadvantaged circumstances in early childhood create unjustifiable inequality in children's development. This study explores the IOp in child development across multiple socioeconomic domains in early childhood as precursors to inequalities in later life, exploiting the large-scale cohort study that contemporaneously tracks Vietnamese, Peruvian, Ethiopian, and Indian children over a 15-year period.

Our main findings are summarised as follows. First, the K-means clustering of the circumstance variables shows that there is a significant difference in height trajectory between clusters. Children with advantaged parental socioeconomic backgrounds are significantly taller than those with disadvantaged socioeconomic backgrounds in early childhood for both boys and girls in all four countries. The difference in height associated with early-life circumstances is persistent over time, and the significant difference remains even after children enter into the early adolescence period. The influence of early-life circumstances lingers even in the early adolescence period, forming long-lasting inequality in growth. Second, we measure the proportion of total inequality associated with the early-life circumstances at each development stage by the random forest. We find that the highest IOp proportion is observed when children are 5 years old, during which more than two-thirds of the observed standard deviation of height distribution are explained by the variation of early-life circumstances in Vietnam, Peru, and India. In particular, maternal height, household consumption, and housing quality turned out to be the top three most important predictive factors. Finally, we predict the height growth in the hypothetical situation where any circumstantial disadvantages in very early childhood are fully compensated for. This counterfactual prediction indicates that were the playing field to be levelled, child inequality at the age of 15 would be reduced to about half for both boys and girls.

The heap in the proportion of the IOp during the pre-school period could be attributed to the large amount of time that children spent exclusively with parents at home (West 1997). During this period, children are the most exposed to parental controls and housing conditions. After this period, as the children begin to be exposed more to wider physical and cultural environments, we can attribute the gradual decline of the IOp after the pre-school period to the correspondingly diminished influence of parental and housing characteristics. Also, in the early adolescence period, as food choices are more diversified, the parental and housing characteristics in early childhood become less relevant. Certainly, we find the reduction in the proportion of the IOp after the pre-school period, but more than half of the observed variation in height after the pre-school period

is still associated with the variation in circumstances in early childhood, suggesting that mitigating inequality at an early stage of development could have a long-lasting pay-off.

The results suggest that mitigating the IOp should require policies incorporating a life-course perspective and that policy-makers should make sure that children can equally enjoy better health irrespective of parental socioeconomic backgrounds. Earlier intervention targeting children with lower socioeconomic backgrounds must be a top intervention priority because the proportion of inequality attributable to early-life circumstances becomes the largest in early childhood at age 5. As the benefits of intervention accumulate over time, earlier intervention can be more cost-effective than interventions at a later point in time. Our prediction suggests that policies designed to compensate for disadvantages in earlier life can possibly enhance the long-term well-being of children in the form of improved health trajectories and mitigate inequality over the entire life-course. Alleviation of health disparity over the life-course would then contribute to ameliorating intergenerational persistence of social and economic disadvantages.

6. Conclusion

As a concluding remark, it is worth discussing the limitations of this study and future research extension. The first limitation is the possible underestimation of the overall inequality in child height. Although the Young Lives study reflects the various ethnic, geographic, and religious diversity of the population, its semi-purposive sampling over-sampled the poor households (Barnett et al. 2013). Given the positive socioeconomic gradient in child health, over-sampling the poor households could have resulted in the underestimation of overall inequality. Hence, the overall inequality estimated in this study should be interpreted with caution as the lower limit of the inequality.

Second, it is important to keep in mind that the IOp estimates in this study illustrate the part of inequality that is associated purely with the particular circumstances considered in the model. In our case, there are eight categories of circumstantial factors. Certainly, in the real world, circumstances beyond one's control are much more enormous. Speaking in the extreme, tens of thousands of genetic factors can also be considered to be non-ignorable circumstances from the perspective of children. Therefore, the estimated IOp shown in this study should be recognised as part of the entire IOp associated just with parental and household characteristics in early childhood.

Future research would be well served by exploring a wider array of outcomes beyond height development. Investigating the inequality in child cognitive skills and educational outcomes would complement our initial finding and give us a broader picture of the life-course IOp trajectory. Also, considering even wider circumstance factors is another promising extension. For example, given the argument that inequality in child health begins even at the prenatal stage (Almond 2006; Almond and Mazumder 2013) and in utero

(Barker 1990, 1995; Robinson 2001), considering parental behaviours before childbirth would also contribute to obtaining more dynamic and fruitful policy implications.

References

Aizawa, T. (2019). Ex-ante inequality of opportunity in child malnutrition: New evidence from ten developing countries in Asia. *Economics and Human Biology* 35: 144–161. doi:10.1016/j.ehb.2019.06.003.

Alderman, H., Hoddinott, J., and Kinsey, B. (2006). Long term consequences of early childhood malnutrition. *Oxford Economic Papers* 58(3): 450–474. doi:10.1093/oep/gpl008.

Almond, D. (2006). Is the 1918 influenza pandemic over? Long term effects of in-utero influenza exposure in the post–1940 US population. *Journal of Political Economy* 114(4): 672–712. doi:10.1086/507154.

Almond, D. and Currie, J. (2011). Human capital development before age five. *Handbook of Labor Economics* 4b: 1315–1486.

Almond, D. and Mazumder, B. (2013). Fetal origins and parental responses. *Annual Review of Economics* 5: 37–56.

Barker, D. (1990). The fetal and infant origins of adult disease. *British Medical Journal* 301(6761): 375–376.

Barker, D. (1995). Fetal origins of coronary heart disease. *British Medical Journal* 311(6998): 171–174.

Barnett, I., Ariana, P., Petrou, S., Penny, M.E., Duc, L.T., Galab, S., Woldehanna, T., Escobal, J.A., Plugge, E., and Boyden, J. (2013). Cohort profile: The Young Lives study. *International Journal of Epidemiology* 42(3): 701–708. doi:10.1093/ije/dys082.

Beard, J. (2001). Functional consequences of nutritional anemia during pregnancy and early childhood. In: Ramakrishnan, U. (ed.). *Nutritional Anemias*. Boca Raton: CRC Press: 111–128.

Behrman, J.R., Schott, W., Mani, S., Crookston, B.T., Dearden, K., Duc, L.T., Fernald, L.C.H., Stein, A.D., and the Young Lives determinants and consequences of child growth project team (2013). Intergenerational transmission of poverty and inequality: Young Lives. Oxford: University of Oxford (Young Lives technical note 117).

Ben-Shlomo, Y. and Kuh, D. (2002). A life course approach to chronic disease epidemiology: Conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology* 31(2): 285–293. doi:10.1093/ije/31.2.285.

Bradley, R.H. and Corwyn, R.F. (2002). Socioeconomic status and child development. *Annual Review of Psychology* 53(1): 371–399. doi:10.1146/annurev.psych.53.100901.135233.

Bradley, R.H., Corwyn, R.F., McAdoo, H.P., and Garcia-Coll, C. (2001). The home environments of children in the United States part I: Variations by age, ethnicity, and poverty status. *Child Development* 72(6): 1844–1867. doi:10.1111/1467-8624.t01-1-00382.

Breiman, L. (2001). Random forests. *Machine Learning* 45(1): 5–32. doi:10.1023/A:1010933404324.

Briones, K. (2017). How many rooms are there in your house? Constructing the Young Lives wealth index. Oxford: University of Oxford (Young Lives technical note 43).

Brunori, P., Peragine, V., and Serlenga, L. (2019). Upward and downward bias when measuring inequality of opportunity. *Social Choice and Welfare* 52(4): 635–661. doi:10.1007/s00355-018-1165-x.

Case, A., Fertig, A., and Paxson, C. (2005). The lasting impact of childhood health and circumstance. *Journal of Health Economics* 24(2): 365–389. doi:10.1016/j.jhealeco.2004.09.008.

Case, A., Lubotsky, D., and Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient. *The American Economic Review* 92(5): 1308–1334.

Case, A. and Paxson, C. (2002). Parental behavior and child health. *Health Affairs* 21(2): 164–178. doi:10.1377/hlthaff.21.2.164.

Case, A. and Paxson, C. (2008). Stature and status: Height, ability, and labor market outcomes. *Journal of Political Economy* 116(3): 499–532. doi:10.1086/589524.

Case, A. and Paxson, C. (2010). Causes and consequences of early-life health. *Demography* 47(1): S65–S85. doi:10.1353/dem.2010.0007.

Chalasani, S. (2012). Understanding wealth-based inequalities in child health in India: A decomposition approach. *Social Science and Medicine* 75(12): 2160–2169. doi:10.1016/j.socscimed.2012.08.012.

Chen, E., Martin, A.D., and Matthews, K.A. (2006). Socioeconomic status and health: Do gradients differ within childhood and adolescence? *Social Science and Medicine* 62(9): 2161–2170. doi:10.1016/j.socscimed.2005.08.054.

Chen, E., Matthews, K.A., and Boyce, W.T. (2002). Socioeconomic differences in children's health: How and why do these relationships change with age? *Psychological Bulletin* 128(2): 295–329.

Chen, Y. and Li, H. (2009). Mother's education and child health: Is there a nurturing effect? *Journal of Health Economics* 28(2): 413–426. doi:10.1016/j.jhealeco.2008.10.005.

Contoyannis, P. and Li, J. (2011). The evolution of health outcomes from childhood to adolescence. *Journal of Health Economics* 30(1): 11–32. doi:10.1016/j.jhealeco.2010.10.007.

Crookston, B.T., Schott, W., Cueto, S., Dearden, K.A., Engle, P., Georgiadis, A., Lundeen, E.A., Penny, M.E., Stein, A.D., and Behrman, J.R. (2013). Postinfancy growth, schooling, and cognitive achievement: Young Lives. *The American Journal of Clinical Nutrition* 98(6): 1555–1563. doi:10.3945/ajcn.113.067561.

Currie, J. and Moretti, E. (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *The Quarterly Journal of Economics* 118(4): 1495–1532. doi:10.1162/003355303322552856.

Currie, J. and Moretti, E. (2007). Biology as destiny? Short and long-run determinants of intergenerational transmission of birth weight. *Journal of Labor Economics* 25(2): 231–264. doi:10.1086/511377.

Currie, J. and Stabile, M. (2003). Socioeconomic status and child health: Why is the relationship stronger for older children? *American Economic Review* 93(5): 1813–1823. doi:10.1257/000282803322655563.

Currie, J. and Vogl, T. (2013). Early-life health and adult circumstance in developing countries. *Annual Review of Economics* 5: 1–36.

Dercon, S. and Sánchez, A. (2013). Height in mid childhood and psychosocial competencies in late childhood: Evidence from four developing countries. *Economics and Human Biology* 11(4): 426–432. doi:10.1016/j.ehb.2013.04.001.

Duc, L.T. (2011). Height and cognitive achievement of Vietnamese children. *World Development* 39(12): 2211–2220. doi:10.1016/j.worlddev.2011.04.013.

Escobal, J. and Flores, E. (2008). An assessment of the Young Lives sampling approach in Peru. Oxford: University of Oxford (Young Lives technical note 3).

Fink, G. and Rockers, P.C. (2014). Childhood growth, schooling, and cognitive development: Further evidence from the Young Lives study. *The American Journal of Clinical Nutrition* 100(1): 182–188. doi:10.3945/ajcn.113.080960.

Floud, R. (1990). *Height, health, and history: Nutritional status in the United Kingdom, 1750–1980*. Cambridge: Cambridge University Press.

Georgiadis, A., Benny, L., Galab, S., Reddy, P., and Woldehanna, T. (2017). Growth recovery and faltering through early adolescence in low- and middle-income countries: Determinants and implications for cognitive development. *Social Science and Medicine* 179: 81–90. doi:10.1016/j.socscimed.2017.02.031.

Gillespie, S. (1997). Improving adolescent and maternal nutrition: An overview of bene-

fits and options. New York: UNICEF (staff working paper 97-002).

Glewwe, P., Jacoby, H.G., and King, E.M. (2001). Early childhood nutrition and academic achievement: A longitudinal analysis. *Journal of Public Economics* 81(3): 345–368. doi:10.1016/S0047-2727(00)00118-3.

Graham, H. and Power, C. (2004). Childhood disadvantage and health inequalities: A framework for policy based on lifecourse research. *Child: Care, Health and Development* 30(6): 671–678. doi:10.1111/j.1365-2214.2004.00457.x.

Grantham-McGregor, S., Cheung, Y.B., Cueto, S., Glewwe, P., Richter, L., Strupp, B., and the International Child Development Steering Group (2007). Developmental potential in the first 5 years for children in developing countries. *The Lancet* 369(9555): 60–70. doi:10.1016/S0140-6736(07)60032-4.

Haas, S. (2008). Trajectories of functional health: The 'long arm' of childhood health and socioeconomic factors. *Social Science and Medicine* 66(4): 849–861. doi:10.1016/j.socscimed.2007.11.004.

Himaz, R. (2018). Stunting later in childhood and outcomes as a young adult: Evidence from India. *World Development* 104: 344–357. doi:10.1016/j.worlddev.2017.12.019.

Hoddinott, J., Behrman, J.R., Maluccio, J.A., Melgar, P., Quisumbing, A.R., Ramirez-Zea, M., Stein, A.D., Yount, K.M., and Martorell, R. (2013). Adult consequences of growth failure in early childhood. *The American Journal of Clinical Nutrition* 98(5): 1170–1178. doi:10.3945/ajcn.113.064584.

Howe, L.D., Lawlor, D.A., and Propper, C. (2013). Trajectories of socioeconomic inequalities in health, behaviours and academic achievement across childhood and adolescence. *Journal of Epidemiology & Community Health* 67(4): 358–364. doi:10.1136/jech-2012-201892.

Huda, T.M., Hayes, A., El Arifeen, S., and Dibley, M.J. (2017). Social determinants of inequalities in child undernutrition in Bangladesh: A decomposition analysis. *Maternal and Child Nutrition* 14(1): 1–12. doi:10.1111/mcn.12440.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer texts in statistics. New York: Springer.

Kaufman, L. and Rousseeuw, P. (1990). *Finding groups in data: An introduction to cluster analysis*. New York: Wiley.

Khanam, R., Nghiem, H.S., and Connelly, L.B. (2009). Child health and the income gradient: Evidence from Australia. *Journal of Health Economics* 28(4): 805–817. doi:10.1016/j.jhealeco.2009.05.001.

Kumra, N. (2008). An assessment of the Young Lives sampling approach in Andhra

Pradesh, India. Oxford: University of Oxford (Young Lives technical note 2).

Lantz, B. (2015). *Machine learning with R: Expert techniques for predictive modeling to solve all your data analysis problems, 2nd Edition*. Birmingham: Packt Publishing.

Lefranc, A., Pistolesi, N., and Trannoy, A. (2008). Inequality of opportunities vs. inequality of outcomes: Are western societies all alike? *Review of Income and Wealth* 54(4): 513–546. doi:10.1111/j.1475-4991.2008.00289.x.

Lefranc, A., Pistolesi, N., and Trannoy, A. (2009). Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France. *Journal of Public Economics* 93(11): 1189–1207. doi:10.1016/j.jpubeco.2009.07.008.

Levin, C.E., Ruel, M.T., Morris, S.S., Maxwell, D.G., and Armar-Klemesu, M. (2000). Poor maternal schooling is the main constraint to good child care practices in Accra. *The Journal of Nutrition* 130(6): 1597–1607. doi:10.1093/jn/130.6.1597.

Li, L., Law, C., Conte, R.L., and Power, C. (2009). Intergenerational influences on childhood body mass index: The effect of parental body mass index trajectories. *The American Journal of Clinical Nutrition* 89: 551–557.

Lundeen, E.A., Behrman, J.R., Crookston, B.T., Dearden, K.A., Engle, P., Georgiadis, A., Penny, M.E., and Stein, A.D. (2014). Growth faltering and recovery in children aged 1–8 years in four low- and middle-income countries: Young lives. *Public Health Nutrition* 17(9): 2131–2137. doi:10.1017/S1368980013003017.

Lynch, J. and Smith, G.D. (2005). A life course approach to chronic disease epidemiology. *Annual Review of Public Health* 26: 1–35.

Lynch, J., Kaplan, G., and Salonen, J. (1997). Why do poor people behave poorly? Variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic lifecourse. *Social Science and Medicine* 44(6): 809–819. doi:10.1016/S0277-9536(96)00191-8.

Murphy, K. (2012). *Machine learning: A probabilistic perspective*. Adaptive Computation and Machine Learning Series. Cambridge: MIT Press.

Nguyen, N. (2008). An assessment of the Young Lives sampling approach in Vietnam. Oxford: University of Oxford (Young Lives technical note 4).

Outes, I. and Sanchez, A. (2008). An assessment of the Young Lives sampling approach in Ethiopia. Oxford: University of Oxford (Young Lives technical note 1).

Outes-Leon, I. and Dercon, S. (2008). Survey attrition and attrition bias in Young Lives. Oxford: University of Oxford (Young Lives technical note 5).

Özaltin, E., Hill, K., and Subramanian, S. (2010). Association of maternal stature

with offspring mortality, underweight, and stunting in low- to middle-income countries. *The Journal of the American Medical Association* 303(15): 1507–1516. doi:10.1001/jama.2010.450.

Propper, C., Rigg, J., and Burgess, S. (2007). Child health: Evidence on the roles of family income and maternal mental health from a UK birth cohort. *Health Economics* 16(11): 1245–1269. doi:10.1002/hec.1221.

Robinson, R. (2001). The fetal origins of adult disease. *British Medical Journal* 322(7283): 375–376.

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

Sánchez, A. (2017). The structural relationship between early nutrition, cognitive skills and noncognitive skills in four developing countries. *Economics and Human Biology* 27: 33–54. doi:10.1016/j.ehb.2017.04.001.

Schott, W., Aurino, E., Penny, M.E., and Behrman, J.R. (2019). The double burden of malnutrition among youth: Trajectories and inequalities in four emerging economies. *Economics and Human Biology* 34: 80–91. doi:10.1016/j.ehb.2019.05.009.

Silventoinen, K., Sammalisto, S., Perola, M., Boomsma, D.I., Cornes, B.K., Davis, C., Dunkel, L., De Lange, M., Harris, J.R., and Hjelmborg, J.V. (2003). Heritability of adult body height: A comparative study of twin cohorts in eight countries. *Twin Research* 6(5): 399–408. doi:10.1375/twin.6.5.399.

Strauss, J. and Thomas, D. (2007). Health over the life course. *Handbook of Development Economics* 4: 3375–3474.

Van de Poel, E., Reza Hosseinpoor, A., Speybroeck, N., Van Ourti, T., and Vegab, J. (2008). Socioeconomic inequality in malnutrition in developing countries. *Bulletin of the World Health Organization* 86(4): 282–291.

Wagstaff, A. and Watanabe, N. (2000). Socioeconomic inequalities in child malnutrition in the developing world. Washington, D.C.: World Bank (policy working paper 2434).

West, P. (1997). Health inequalities in the early years: Is there equalisation in youth? *Social Science and Medicine* 44(6): 833–858. doi:10.1016/S0277-9536(96)00188-8.

WHO (2006). *WHO Child Growth Standard: Length/height-for-age, Weight-for-age, Weight-for-length, Weight-for-height and Body Mass Index-for-age. Methods and Development*. Geneva: WHO.

World Bank (2018). World development indicators [electronic resource]. Washington, D.C.: World Bank. data.worldbank.org/data-catalog/world-development-indicators.

Appendix: Supplementary results

Table A-1:   **Proportion of the inequality of opportunity**

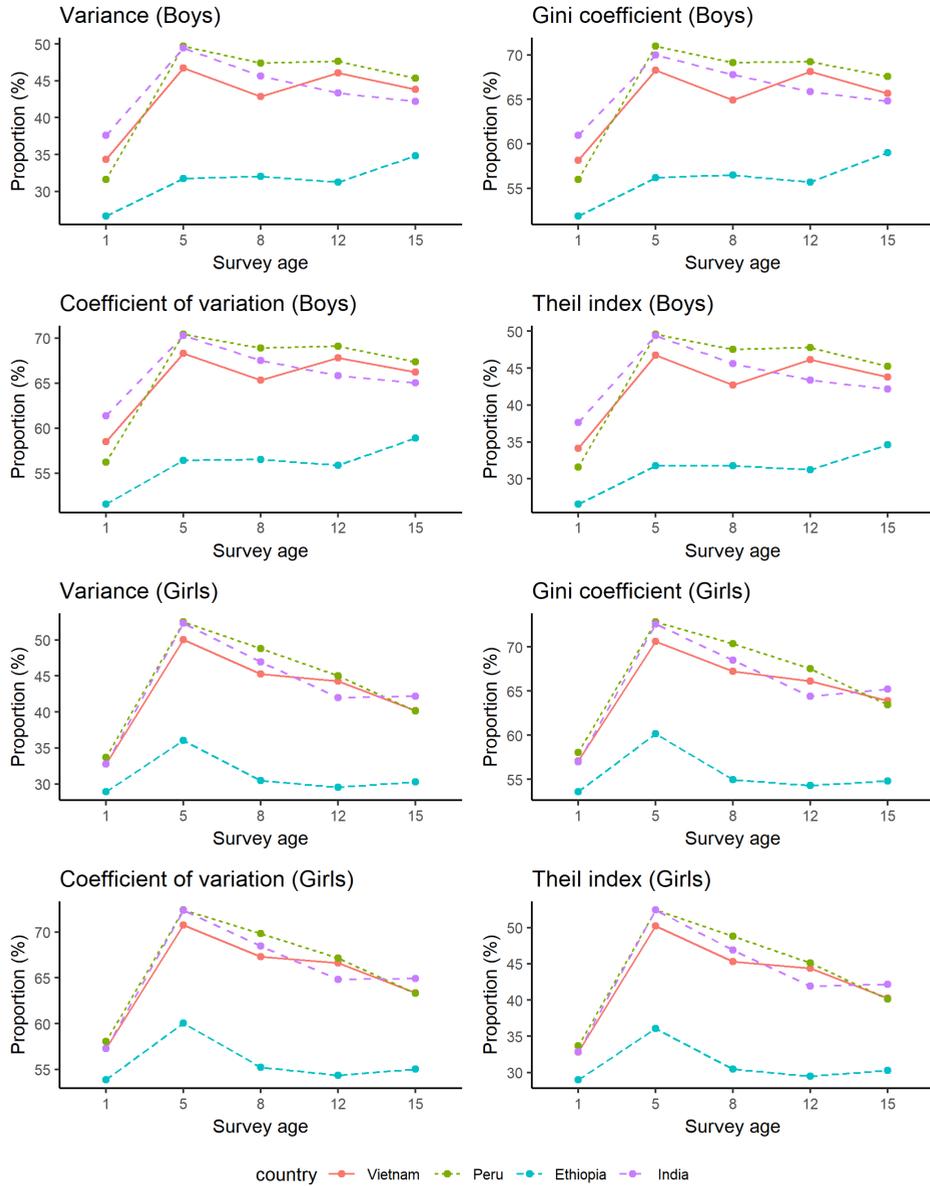|	Boys			Girls		
	Observed	**IOp**	**IOp percentage**	**Observed**	**IOp**	**IOp percentage**
Vietnam						
Round 1	4.1	2.4	58.6	3.7	2.1	57.3
Round 2	5.3	3.6	68.3	4.5	3.2	70.8
Round 3	5.9	3.8	65.5	5.5	3.7	67.3
Round 4	8.1	5.5	67.9	7.4	4.9	66.5
Round 5	6.7	4.4	66.2	5.5	3.5	63.4
Peru						
Round 1	4.1	2.4	58.6	3.7	2.1	57.3
Round 2	5.3	3.6	68.3	4.5	3.2	70.8
Round 3	5.9	3.8	65.5	5.5	3.7	67.3
Round 4	8.1	5.5	67.9	7.4	4.9	66.5
Round 5	6.7	4.4	66.2	5.5	3.5	63.4
Ethiopia						
Round 1	4.1	2.4	58.6	3.7	2.1	57.3
Round 2	5.3	3.6	68.3	4.5	3.2	70.8
Round 3	5.9	3.8	65.5	5.5	3.7	67.3
Round 4	8.1	5.5	67.9	7.4	4.9	66.5
Round 5	6.7	4.4	66.2	5.5	3.5	63.4
India						
Round 1	4.1	2.4	58.6	3.7	2.1	57.3
Round 2	5.3	3.6	68.3	4.5	3.2	70.8
Round 3	5.9	3.8	65.5	5.5	3.7	67.3
Round 4	8.1	5.5	67.9	7.4	4.9	66.5
Round 5	6.7	4.4	66.2	5.5	3.5	63.4

Note: Only the testing samples are used.

Table A-2: **Importance of each circumstance factor in each developmental stage**

	Round – Boys					Round – Girls				
	1	2	3	4	5	1	2	3	4	5
Vietnam										
Housing quality	19.2	15.2	16.5	15.0	15.4	17.0	14.0	15.4	14.9	13.6
Service availability	6.2	7.5	6.4	6.7	5.6	5.9	7.7	7.3	6.5	5.6
Comsumer durables	9.0	9.1	10.0	10.8	9.1	9.6	9.0	8.4	10.0	8.2
Maternal education	8.4	9.6	9.2	9.3	8.5	9.8	10.2	11.2	10.0	8.0
Paternal education	10.3	11.0	10.1	10.4	10.2	9.5	8.2	8.6	9.2	8.6
Maternal height	22.3	22.5	23.2	20.4	27.8	23.6	26.4	25.0	24.4	33.6
Household expenditure	21.4	21.8	21.4	24.2	19.8	21.1	21.9	21.2	21.8	19.1
No shock	3.3	3.3	3.2	3.1	3.7	3.6	2.7	2.9	3.2	3.3
Peru										
Housing quality	16.4	16.6	15.3	16.2	15.4	16.3	14.1	12.7	14.2	13.6
Service availability	6.3	11.2	6.6	7.3	8.2	6.2	10.3	8.5	7.8	5.7
Comsumer durables	9.8	11.1	10.1	10.8	10.1	10.3	12.9	10.3	11.0	8.8
Maternal education	13.3	16.9	17.4	14.7	15.4	12.0	16.1	13.1	12.8	10.4
Paternal education	11.6	9.4	9.8	9.7	9.9	12.0	10.2	11.8	10.8	9.7
Maternal height	19.7	18.2	23.1	21.6	22.1	21.1	18.6	23.0	22.8	32.2
Household expenditure	19.9	14.6	15.4	17.2	16.2	18.9	15.8	17.6	18.0	17.2
No shock	3.0	2.1	2.1	2.5	2.7	3.2	2.1	3.0	2.6	2.4
Ethiopia										
Housing quality	17.7	17.9	16.9	15.9	15.4	15.8	16.4	16.9	16.2	15.6
Service availability	6.6	6.5	6.7	7.1	8.2	7.5	9.9	8.1	9.0	8.0
Comsumer durables	6.3	7.1	7.0	7.1	7.5	6.7	7.3	6.6	8.0	6.1
Maternal education	7.4	7.2	7.1	7.5	8.0	7.6	7.0	7.0	7.5	6.8
Paternal education	9.1	8.9	8.5	8.0	8.5	9.0	9.1	8.9	9.2	8.4
Maternal height	20.9	22.7	23.7	26.3	22.9	20.6	22.4	21.6	21.4	26.0
Household expenditure	23.6	22.0	22.9	21.7	22.1	24.1	20.5	23.2	21.3	20.9
No shock	8.4	7.7	7.2	6.5	7.4	8.8	7.5	7.8	7.4	8.3
India										
Housing quality	19.3	15.1	16.4	15.2	15.1	16.7	14.3	15.7	14.8	13.4
Service availability	6.2	7.7	6.4	6.6	5.5	5.9	8.3	6.9	6.5	5.6
Comsumer durables	8.9	8.7	10.1	11.2	9.4	9.5	9.0	8.4	10.2	8.0
Maternal education	8.4	9.7	9.4	9.1	8.3	10.0	9.7	11.4	10.0	8.1
Paternal education	10.3	11.1	9.9	10.5	10.3	9.4	8.6	8.5	9.0	8.3
Maternal height	22.3	22.5	23.0	20.8	27.8	23.8	25.6	24.8	24.2	33.6
Household expenditure	21.5	21.8	21.6	23.3	19.8	21.1	21.8	21.3	21.9	19.7
No shock	3.2	3.3	3.2	3.3	3.8	3.5	2.8	3.0	3.3	3.3

Note: Only the testing samples are used.

Figure A-1: **Trajectory of inequality of opportunities (other inequality statistics)**