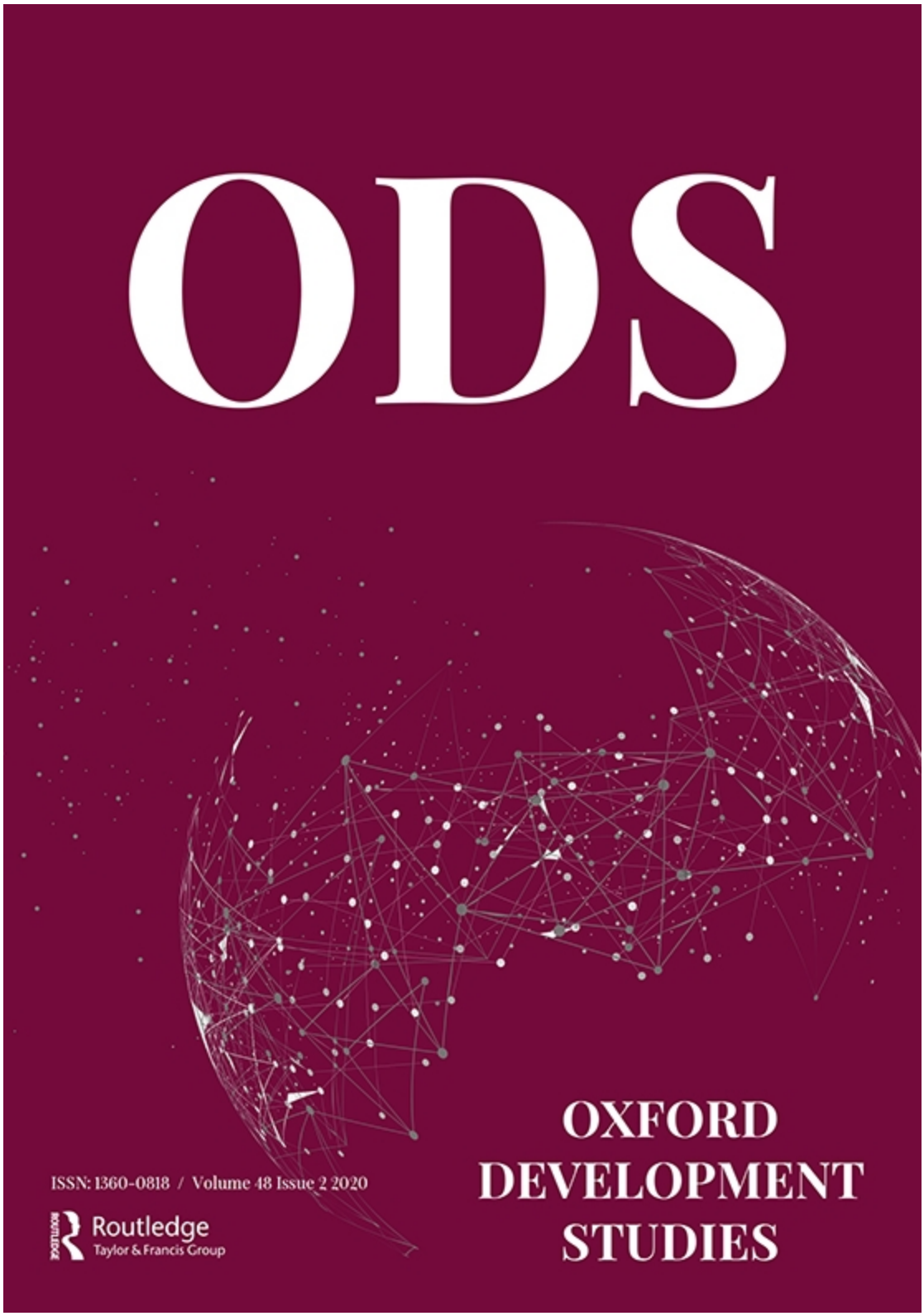

ODS
ISSN: 1360-0818 / Volume 48 Issue 2 2020
Routledge
Taylor & Francis Group
OXFORD
DEVELOPMENT
STUDIES

# Unequal from the start? Children's life histories and their employment outcomes in India.


| | |
|---|---|
| Journal: | *Oxford Development Studies* |
| Manuscript ID | CODS-2025-0034.R1 |
| Manuscript Type: | Original Article |
| Keywords: | early-life poverty, gendered wage gap, youth employment, life history trajectories, Young Lives |
| | |




# Unequal from the start? Children's life histories and their employment outcomes in India

**Abstract**

Five waves of the Young Lives Indian survey are analysed using sequence and regression analysis. Four clusters of individuals sharing similar life histories are identified. The life histories reflect aspects of childhood and family circumstances measured by living conditions, health and exposure to family and economic shocks. Regression analysis investigates how different life histories and gender impact employment outcomes at age 22. The results show that accumulation of (dis)advantages over children's lives helps to explain inequalities in employment and earnings. At age 22, those who have led wealthier, healthier, more stable lives have higher earnings but are less likely to be employed. Those from the poorest but most stable cluster are the most likely to be employed, but among women, this group earns the least. For young men, earnings are lowest for those exposed to the most shocks. Policy to equalise employment opportunities in India should therefore aim to improve the living conditions, health and security of disadvantaged young children.

**Keywords**: Early-life poverty, gender wage gap, youth employment, life history trajectories

For Peer Review Only

## 1. Introduction

This study analyses how early-life conditions and gender influence young adults' employment outcomes in India. As one of the world's most populous nations, with significant regional and gender disparities, India provides a unique context for understanding how early-life experiences and attributes shape employment outcomes of men and women. Despite significant economic growth in recent decades, female labour force participation in India remains among the lowest in emerging economies and has shown a declining trend over time (Chatterjee et al., 2015). This persistent gender disparity is driven by a combination of cultural norms, societal expectations, and institutional barriers that limit women's access to and participation in the labour market. These factors underscore the importance of studying the intersections between early-life conditions, gender, and employment outcomes in this setting.

The research contributes to sociological and economic literature exploring inequality of opportunity, social mobility and gender gaps in career progression. Prior studies have demonstrated that early-life socioeconomic, biosocial and demographic factors, including living environments, health and gender, play a critical role in shaping employment opportunities, social mobility and later life economic outcomes (Devine et al., 2022; Crivello & Morrow, 2019; Himaz, 2018;Vauhkonen et al, 2017). By focusing on the specific case of India, this study extends this literature, offering insights into how inequalities in employment and earnings are rooted in early-life experiences and perpetuated across the life course.

The study uses longitudinal panel data tracking children's lives from age 8 into early adulthood in India. Employing regression and sequence analysis, it investigates how different life course histories are linked to employment participation and earnings at age 22. The analysis identifies four clusters of life trajectories characterised by different experiences of household

For Peer Review Only

living conditions, health and exposure to family and economic shocks. In combination, these socioeconomic, well-being and environmental factors have significant and longer-term impacts conferring advantages and disadvantages in later life. At age 22, those with more stable and advantageous backgrounds had higher earnings, reflecting better job quality, despite being less likely to be employed. Young women, already disadvantaged in employment participation and pay compared with men, had the highest employment participation but lowest earnings in the poorest but most stable cluster. For young men, earnings were lowest for those experiencing the most shocks. These results highlight how early-life advantages amplify inequalities with employment failing to offset early disadvantage.

This life course approach provides new evidence of the cumulative effects of early-life circumstances and gendered experiences on employment outcomes, contributing to a deeper understanding of social and economic mobility in the Indian context. In a setting characterised by pronounced gender inequality, the key pathways identified can inform policies aiming to equalize opportunities and reduce gender inequality in employment. In doing so, this research enhances understanding of labour market dynamics in India and offers broader implications for addressing inequities in employment globally. It underscores the importance of examining early-life determinants within diverse socioeconomic contexts, providing a valuable contribution to debates on human capital development, mobility and gender equity.

The next section overviews previous research on early-life conditions, gender and adult outcomes. Subsequent sections describe the data, methodology and results. The final two sections discuss the findings and conclude.

For Peer Review Only

## 2. Research Background

Previous research on employment in India has focused more on macroeconomic or institutional determinants of male and female employment participation and earnings, often overlooking the role of early-life socioeconomic attributes (Afridi et al., 2018). Our study addresses this gap by emphasising the role of childhood circumstances and gender-specific socialisation processes in later life employment outcomes. The approach recognises that early life circumstances including exposure to adverse events, such as economic shocks or parental unemployment and gendered allocations of household work can have long-term effects on human capital development, labour market prospects and well-being (Carmichael et al., 2023; Chuta, 2014).

Gender is a significant determinant of employment outcomes, including career progression and earnings. The persistence of gender participation and wage gaps is notable in both developed countries, including OECD countries, and developing countries, including India (Agrawal, 2021; World Bank, 2026). Discrimination and job segregation contribute to these gender differentials that can arise early in careers (Boye and Grönlund, 2018)) and persist (and widen) over long-term career trajectories (Nutz and Lersch, 2020). Gender inequality in employment is deeply ingrained in India's socio-cultural framework. Patriarchal norms often restrict women's mobility, educational attainment, and access to economic opportunities (Klasen & Pieters, 2015). These constraints are compounded by early-life disadvantages, including limited access to quality education, healthcare, and nutrition that significantly affect women's ability to participate in the labour market (Agrawal, 2021; Kingdon & Unni, 2001). The gender gap in employment in India is further intensified by the dual burden of unpaid care work and limited access to formal employment opportunities. As noted by Desai and Joshi (2019), women in India spend disproportionately more time on domestic responsibilities, reducing their ability to engage in paid

For Peer Review Only

employment. This imbalance begins in early life, as young girls are allocated caregiving roles, while boys pursue education and career-oriented activities. By the time they reach adulthood, young women in India already have lower income aspirations than males, particularly if they are from more disadvantaged castes (Sarkar et al. 2026). Investigating how gendered early-life experiences impact employment outcomes in adulthood is crucial for designing policy interventions to reduce gender inequalities.

Previous studies, albeit predominantly in the context of developed countries, have shown how early-life socioeconomic circumstances (including parental income and social class) life events and interventions can be important determinants of adult outcomes including learning (Zou et al., 2020; Conley et al, 2007) and employment and earnings (Bukodi et al., 2016; Bargain and Zeidan, 2017; García et al., 2021). For example, Wiborg and Moberg (2010) show that in Nordic countries a disadvantaged social background is associated with receipt of social assistance and higher risks of unemployment in later life. Vauhkonen et al (2017) find that indicators of household disadvantage (school drop-out, receipt of social assistance and unemployment) in Finland are intergenerationally inherited. Lower family income is associated with negative outcomes such as school dropout, lower educational attainment and health problems (Duncan, et al., 2011).

Health in childhood and early-life health indicators including environmental and socioeconomic conditions have been associated with education, employment and well-being outcomes (Canning, 2012). Bharadwaj et al. (2013) link improved early life health care to educational attainment. Devine et al. (2022) find early-life cancer is a risk factor for poor educational and employment outcomes. Hamil-Luker and O'Rand (2007) find that early exposure to adverse socioeconomic conditions is associated with higher risk for heart attack among women. Van Den Berg et al. (2007) find adult life expectancy is lower for children born during economic

For Peer Review Only

downturns. Melchior et al. (2007) find that children's socioeconomic status and living conditions are related to a range of health problems in adulthood. Conley and Bennett (2001) report links between socioeconomic environment and birthweight.

Early-life environments and gender roles are likely to be just as (if not more) important for adult outcomes in developing counties, where children are more vulnerable to shocks such as famine and conflict (Currie and Vogl, 2013). Studies on early-life environments and adult outcomes (including employment, education and well-being) in developing countries have examined the effects of circumstances and conditions in childhood. These have included children's health and physical development (Currie, 2009; Vogl, 2012) and specific causes of ill-health such as malnutrition (Linnemayr and Alderman, 2011). In Himaz's (2018) study of children from age 8 to 19 in India, children's physical development is correlated with psychosocial outcomes and has intergenerational consequences. Later life effects are associated with exposure to shocks including parental death (Crivello and Morrow, 2019) famine (Pathania, 2009) and armed conflict (Akresh et al., 2012). Early-life effects may vary according to gender and socioeconomic status. For example, higher early-life rainfall in Indonesia affects adult outcomes for women but not men (Maccini and Yang, 2009) whereas Carillo (2019) reports no gender differences for Colombia.

Theoretical links underpinning these relationships suggest a life course approach which recognises that historical context shapes individual lives (Elder, 1998). Theories of opportunity equalisation further highlight how aspects of an individual's childhood and family circumstances that are beyond their control can influence outcomes, reinforcing inequalities (Roemer and Trannoy, 2016). While the contribution of circumstances to outcome inequality is complicated by multidimensional aspects of equality of opportunity (Roemer and Trannoy, 2016:1309) previous research provides guidance. In developing economy contexts research has shown that early-life

For Peer Review Only

conditions, including parental investments in health and education and the development of both physical and cognitive skills, have consequences for labour market success (Currie and Vogl, 2013). Channels for these links are both economic and nutritional (León, 2010) and the impact of shocks may be compensated for or reinforced by social supports, beliefs and values, including gender norms.

Our study contributes to this literature by focusing on the impact of earlier-life conditions and gender on employment outcomes. The former are incorporated by constructing life history records using sequence analysis with optimal matching and clustering. The method has been widely used in research on social mobility and careers (Bukodi et al, 2016) and family formation (Potârcă, et al., 2013). Here sequence analysis is used to group young people sharing similar social circumstances over their life courses (reflected in living conditions, health and exposure to shocks) from age 8 to age 22. Regression analysis is then used to investigate whether inequalities in employment outcomes are linked to relatively more or less advantageous life course trajectories. By analysing how early-life conditions and gender influence young adults' labour market outcomes, this study contributes to research on inequality of opportunity and outcomes in developing countries.

## 3. Data and sample

The analysis uses data from the Young Lives project, a longitudinal cohort study of childhood poverty for India and three other low-middle income countries: Ethiopia, Peru and Vietnam. In India, Young Lives followed 1000 children from an older cohort from age 8 and 2000 from a younger cohort from age 1. 20 local area survey sites in Coastal Andhra, Rayalaseema and Telangana were selected as part of a semi-purposive sampling strategy that oversamples poor

For Peer Review Only

households and ensures a mix of urban and rural regions. Details of the survey design and sampling are available from Young Lives (2018) and Favara (2017).

This study uses the first five rounds of data conducted in 2002, 2006, 2009, 2013 and 2016.[1] In the fifth round, the older cohort were 22 and the younger cohort 15. As our focus is on employment outcomes in early adulthood, we restrict the analysis to the older cohort since the majority (87.8%) of the younger cohort were in school at age 15. A limitation is that because the older cohort were first surveyed at age eight the analysis does not account for factors affecting children before age eight.

There were marginally fewer males (48.64%) in the sample (Supplementary Appendix Table A1). 61.90% of the young people were employed at age 22 but there is already wide variation in the employment participation of men and women: 76.4% of males were employed and only 48.2% of females. Figure 1 shows that at age 22, there is wide wage variation in earnings among those in employment, both within and between genders; women earned on average 67.25% of male hourly earnings and only 63.25% of male daily earnings. The lower employment rate and earnings of women are in line with evidence for India where low rates of female employment participation (World Bank, 2026) are combined with lower productivity and lower paying jobs (Duraisamy and Duraisamy, 2016) reinforcing gender inequality.

***Figure 1: Hourly earnings***

[1]The 6th round in 2020/21 was a limited phone survey conducted during, and focusing on the impact of the Covid pandemic.

For Peer Review Only

## 4. Methods

### *4.1 Sequence analysis and life history clusters*

Sequence analysis is used to identify four clusters of life history trajectories for 913 sample members from age 8. The trajectories record salient aspects of childhood and family environments which are largely beyond the children's control and theory suggests can impact opportunities in later life: living conditions (quantified by access to amenities), health and exposure to family and economic shocks. Living conditions reflect parental income and ability to provision resources, an important channel through which parents can affect children's opportunities (Roemer and Trannoy, 2015:260-261). Health is condition that "matters for a lifetime" (Roemer and Trannoy, 2015:261) and links between childhood health and adult income are highlighted by Canning (2012). Crivello and Morrow (2019) show how children's fortunes can be transformed by family difficulties and shocks such as a parental death. Our analysis examines how the combined and cumulative effect of these aspects of young peoples' lives impacts their early employment.

Living conditions are captured by states recording whether households have full, some (i.e. limited) or no access to four basic amenities; electricity, piped drinking water, their own (not shared) toilet and adequate fuels (paraffin, kerosene, gas or electricity) for cooking. Health is recorded subjectively (caregiver or self-reported) as either very good/good (better than average), average or very poor/poor (worse than average). Subjective measures of health are well-used indicators and strong predictors of morbidity and mortality (Jylha, 2009). Family shocks include divorce, separation, family death or illness and economic shocks record loss of employment, a source of income or family enterprise. The sequence states record whether the individual experienced no family or economic shocks, a family shock, and/or an economic shock since last surveyed, or in the last three years when first surveyed. The coding process resulted in 27 mutually

For Peer Review Only

exclusive states (3x3x3) recording living conditions, health and exposure to shocks in a single dimension.

The 913 coded sequences were analysed using optimal matching and clustering using the Stata SQ package (Brzinsky-Fay, Kohler, & Luniak, 2006). The optimal matching uses the non-parametric optimal matching algorithm to compare all pairs of sequences by computing minimum distances needed, to turn one sequence into another through substitution or combinations of insertions or deletions (as explained in detail elsewhere e.g. Brzinsky-Fay et al., 2006; Potârcă, et al., 2013, Bukodi et al., 2016). As suggested by Brzinsky-Fay et al. (2006) substitution costs were set at double the insertion/deletion (indel) costs which were set at the default lower value of one as in sequences of equal length indel operations have a limited role (Bukodi et al., 2016). The resulting dissimilarity matrix was subjected to cluster analysis using the commonly used Ward's cluster linkage algorithm. The Duda-Hart Je(2)/Je(1) index suggested four distinct clusters of individuals who had experienced similar pathways in terms of the duration of, and transitions between different states.[2]

Sequence analysis is particularly suited to identifying patterns over life courses as it takes into account the time element (order and duration) in longitudinal data. Other methods such as event history analysis or conditional inference trees and forests either decompose the life-course into specific events or are static, focusing on absolute time points rather than relative timing over life-courses. In optimal matching, sequences are treated as complete entities that capture aspects

[2]A low pseudo-T-squared value, T=0.38, for the four-cluster grouping and higher T-values for the 3 and 5-cluster groupings indicated a jump in the hierarchical clustering. The Calinski/Harabasz pseudo-F index was inconclusive. Further consideration of the cluster hierarchy using cluster dendograms and tests of independence between the clusters and the categorical sequence variables indicated that the four-cluster solution was sufficiently distinct (Figures 2 and 3, Table 1).



For Peer Review Only

of the totality of life-courses. In this study, these aspects are characterised by experiences relating to living conditions, health and shocks.

### *4.2 Regression analysis of employment outcomes*

We conduct regression analysis of employment and earnings in the fifth wave of the survey when the participants were 22. Since the sample for the earnings regressions consisted of only individuals for whom earnings were observed (in the fifth round) estimating the earnings equation without taking account of selection into employment could bias the estimates if those employed are not a random subset of the population. We address sample selection using the Heckman two-stage formulation (Heckman, 1979).

The employment participation equation (1) is estimated as probit model using maximum likelihood. The dependent variable, *Employed,* takes the value one if the individual is employed at age 22 (self-employment or an employee) and takes the value zero otherwise. The estimation is specified as:

$$Pr(Employed_i = 1|X, C,S) = \Phi(\alpha + \sum_{j=1}^{k} \beta_j x_{ji} + \sum_{j=1}^{l} \gamma_j c_{ji} + \sum_{j=1}^{m} \theta_j s_{ji}) \quad (1)$$

$Pr(Employed_i=1)$ is the probability that individual $i$ is employed. $\Phi$ denotes the standard normal cumulative distribution function. $\alpha$ is a constant. The $x_j$, $c_j$, $s_j$ are explanatory variables in the vectors $X, C,S$, the $\beta_j, \gamma_j$, and $\theta_j$ are the respective coefficients. In the first stage of the Heckman formulation, the selection term, the Inverse Mills Ratio (IMR) or Heckman's lambda, is derived from the estimation of (1). In the second stage it is included as an additional explanatory variable in the earnings equation (2) in which the dependent variable, $lny_i$ is the log of either individual hourly or daily earnings in Indian Rupees. The earnings equations are estimated using OLS and specified as:

For Peer Review Only

$$lny_i = \alpha + \sum_{j=1}^{k} \beta_j x_{ji} + \sum_{j=1}^{l} \gamma_j c_{ji} + \mu_i IMR + \varepsilon_i \quad (2)$$

We estimate both equations for the pooled sample and for men and women separately.[3] Average marginal effects are reported for the probit estimation of equation (1) and coefficients for equation (2). We estimate robust variances to account for any intragroup correlation within the 20 local area survey sites.

The $x_j$ are demographic and geographic variables recorded at age 22. The former indicate female gender (in the pooled estimations), age in months, educational attainment (grade completed) and whether the individual has children. In line with human capital theory, educational attainment is expected to be positively associated with earnings but the effect on employment participation is difficult to predict at this young age. Having children is likely to impose a time constraint, particularly for the main carer, and this could affect employment participation and possibly earnings. The geographical controls record employment participation in the 20 local area survey sites, rural or urban location and residence in either Telangana, Coastal Andhra or Rayalaseema. Local area employment rates are indicators of community economic development and individuals in communities with a higher employment should have a higher likelihood of employment (Pagán, 2002) which may impact earnings. Employment opportunities and earnings are also likely to vary between urban and rural areas and across regions and inclusion of these controls helps to account for regional disparities post the 1990s market liberalization reforms (Reddy and Bantilan, 2012).

[3]Three separate estimates of the IMR derived from the pooled, male and female estimations are included in the corresponding estimations of (2).

For Peer Review Only

The $c_j$ explanators record aspects of childhood and family circumstance. First, the categorical variable recording membership of the four different life history clusters derived from the sequence analysis. The clusters group individuals with similar life histories in terms of living conditions, health and experience of shocks. Second, following Sarkar et al. (2026). a control for caste, recording membership of the two most disadvantaged groups, Scheduled Castes (SC) and Scheduled Tribes (ST). Third, variables recording average daily hours of household work and work outside the household at age 12. Household work includes family caring, domestic chores (e.g. cooking) and tasks related to family farming, herding or other family businesses. Work outside the household is undertaken for non-household members. At age 12, most of the participants, 73.47%, undertook some household related work. Fewer, 5.36%, were engaged in work outside the household. Engagement in both kinds of work is expected to raise the propensity for employment by providing work related experience and may impact earnings although the direction of the effect is difficult to predict; child labour may enhance work related skills but also constrains time in activities such as education that could enhance employment opportunities.

The $s_j$ are variables recording other aspects of children's lives that are expected to impact employment participation but not directly impact earnings. Their inclusion in equation (1) but not equation (2) is part of the sample selection specification that defines the selection term, *IMR*, satisfying the exclusion restrictions for the correction. They record the age at which the child started school in grade 1, the number of siblings in the household aged between 0 and 12 when the participants were themselves age 12 and BMI at ages 12 and 8. A later school start age is likely to be associated with later entry into the labour force. The number of young children in the home will impact household resources which could influence employment decisions (Conley et al, 2007). BMI in childhood reflects nutritional disparities. While these variables could influence future

For Peer Review Only

employment participation, they would not be expected to have a direct effect on earnings once in employment. Table A1 provides summary statistics for all variables.

## 5. Results

### *5.1 Life history clusters*

Figure 2 shows the modal sequences for each of the four life history trajectories identified in the optimal matching and clustering. These comprise the most frequent state in each period and show the modal trajectory between the different states over time. Figures 3(a)-3(c) shows how access to amenities, health and shocks evolve over the five waves within the four clusters. Table 1 shows summary statistics for each cluster.

Figures 2-3 and Table 1 identify Cluster 1 (*Better-off*) as the most advantaged. The individuals in this group experienced mostly full access to all four basic amenities (85.86% of observations over all rounds, and 94.88% of cluster members by age 22 compared with 77.21% at age 8). They experienced mostly good health and relatively few shocks (only 28.19% of states recorded shocks). In the modal sequence there is continuous full access to amenities, good health and no experience of shocks. Cluster 2 (*Poorer and unstable)* is the largest cluster and is characterised by the second lowest incidence of full access to amenities (10.42% of observations overall and only 30.92% of cluster members by age 22). The participants in cluster 2 experienced the highest incidence of shocks (47.8% of observations). The high incidence of shocks is highlighted in the modal sequence where family shocks are recorded in three of the five surveyed years. This cluster experienced the lowest incidence of good health (56.66% of observations). Cluster 3 (*Transitioning to better-off*) is characterised by the steepest rise in full access to amenities (from 2.83% of cluster members in round 1 to 33.96% in round 4 and 58.96% by round 5). This transition is clearly illustrated in the trajectory for cluster 3 in Figure 3(a) and in the modal

For Peer Review Only

sequence (Figure 2) which records full access only in the final period. Cluster 4 (*Poorest and stable*) is characterised by the lowest incidence of full access to amenities (5.35% of observations overall and only 17.32% of cluster members by age 22). However, cluster 4 members experienced the lowest incidence of shocks and the highest reported incidence of good health. This is illustrated in the trajectories in Figure 3(c) and in the modal sequence (Figure 2) where there is continuous access to some amenities, good health and no experience of shocks.

Further analysis indicates that employment outcomes and some demographic characteristics vary across the clusters (Supplementary Table A2) Employment participation is highest in the poorer clusters 2 and 4 and lowest in the *Better-off* and *Transitioning to better off* clusters 1 and 3. Earnings are highest in Cluster 1 (*Better-off*). While females are marginally overrepresented in cluster 3 (55.66%) the gender distribution does not differ significantly. However, rural-urban residence is strongly differentiated across the clusters with those in the *Better-off* cluster more likely to reside in urban areas. Region is also a differentiator with those in cluster 3 (*Transitioning to better-off*) significantly more likely to live in Rayalaseema, the poorest and least urbanised region, although cities within the region such as Chittoor are experiencing significant urban growth (Reddy and Bantilan, 2012; Malakar et al. 2022).[4] Educational attainment also varies across the clusters, with cluster 1 the most highly educated. In the next section regression analysis is used to examine whether the differences in employment participation and earnings remain significant when these factors are controlled for.

[4]Urban/rural residence was the primary differentiator of cluster membership in conditional tree analysis (using the Chi-square automated interaction detection decision tree algorithm with Bonferroni adjusted significance testing) where the independent/splitting variables were all those used in the regression analysis After segmenting the data, educational attainment was a differentiator within the urban sample and region a differentiator in the rural sub-sample. Urban/rural residence was also the strongest predictor of cluster membership in a multinomial logit specification. Gender was insignificant in both analyses and gender-cluster interactions were also insignificant in both the employment participation and earnings estimations.

For Peer Review Only

**Figure 2:*Modal Sequence by Cluster***
**Figures 3(a),3(b) 3(c):*Trajectories by cluster***
**Table 1: *Cluster characteristics***

### *5.2 Regression analysis*

Table 2 reports average marginal effects for the probit estimations of the employment participation equation (1). Table 3 reports results for the OLS estimations of the earnings equation (2). For brevity Tables 2 and 3 report results only for gender and the life-history clusters, the main variables of interest (Supplementary Appendix tables A3-A4 show results for the controls).

***Table 2. Employment participation***

#### *5.2.1 Employment*

The average marginal effects in Table 2 confirm that after allowing for other factors, females are still less likely to be employed. The clusters grouping the respondents' life histories all take positive signs. The reference group is cluster 1 (*Better-off)* implying that individuals in less advantaged clusters are more likely to be in paid work at age 22 and this effect is larger and more significant for those in the poorest cluster, cluster 4 (*Poorest but stable)*; on average and compared with cluster 1, the probability of employment is 0.1923 higher for an individual in cluster 4. As the comparator group, cluster 1, is the most advantaged group, these results may reflect limited social protection such that only better-off individuals can remain out of work while waiting for high-paying opportunities (Arbache et. al., 2010). Mitra and Verick (2013) report that in India educated youth from relatively well-off households queue for jobs that match their aspirations, while poor and uneducated youths cannot afford to remain unemployed for long. This is consistent with a kind of "waithood" that Honwana (2012:5) argues is only available to youths able to afford a good

For Peer Review Only

education while depending on their parents. This interpretation is also consistent with the negative significance of grade attained (Table A3).

The results for estimations 2 and 3 in Table 2 show that relative to the *Better-off* cluster, the probability of female employment is significantly higher in each of the other clusters. In the male sample, this is only true for the *Poorest but stable* cluster. This suggests that lower employment participation for those in the *Better-off* cluster is more salient for women which has parallels with evidence from African countries that female unemployment is higher in the richest households while this is less true for men (Arbache et al., 2010).

Results for some included controls are worth noting (Table A3). A later school start age is negatively related to employment participation as predicted. Male but not female employment participation is positively related to having more younger siblings, membership of the more disadvantaged castes (ST and SC) and urban location. Already having a child by age 22 is predictably negatively related to women's employment (90.41% of those with children are female, Table A1).

***About here- Table 3. Earnings***

### *5.2.2 Earnings*

Table 3 shows the earnings estimates for the whole sample and women and men separately. The dependent variables measure either earnings per hour or day of work. The results for the pooled estimations indicate that when employed, young women earn significantly less per hour (30% less in unlogged Rupees) and per day (38% less) than young men after controlling for other factors. In relation to the respondents' life histories, those in clusters 2, 3 and 4 earn significantly less (hourly

For Peer Review Only

and daily) relative to cluster 1 (*Better-off*). Even though the results in Table 2 show that young people in these clusters were more likely to be in work than those in cluster 1 (*Better-off*) once in work, they earn less, reinforcing the earlier conclusion that employment is driven by lack of choice or need.

In the separate estimations for women, the cluster effects on hourly earnings are all negatively significant but the cluster effects on daily earning are only (marginally) significant for cluster 3 (*Transitioning to better-off*) and 4 (*Poorest but stable*). Earnings in cluster 4 (*Poorest but stable*) are predicted to be the lowest (hourly earnings are 63% lower than cluster 1, daily earnings are 57% lower). In contrast, in the estimations for men, only hourly earnings in cluster 2 (*Poorer & unstable*) are lower relative to the *Better-off* cluster while daily earnings are significantly lower in cluster 2 (34% lower) and also(marginally) in clusters 3 and 4.

The IMR is marginally negatively significant in the female estimation predicting hourly earnings (estimation 3). This implies a negative correlation between female employment participation and hourly earnings, consistent with need being a strong driver of employment among women. There is no sample selection into paid employment for men perhaps because the majority, 76.5%, were employed. More generally, links between employment and productivity and, therefore, wages may be less apparent among young workers (McKenzie, 2017).

Some results for the controls are worth highlighting (Table A4). Hours of household work in childhood are negatively linked to women's but not men's hourly and daily earnings consistent with previous research (Carmichael et al., 2023). However, having children has no impact on women's wages once employed but daily earnings are somewhat higher for young fathers. Men, but not women, in the more disadvantaged castes earn significantly less.

For Peer Review Only

Together, these results suggest that earlier life conditions and experiences exert a strong influence on the employment and earnings of young men and women. Differences in employment and earnings are predicted by different life histories, particularly for women. However, it is notable that the daily earnings of young men in the most unstable although not the poorest cluster (*Poorer & unstable*) are the lowest relative to those in the *Better-off* cluster. This suggests that experience of family and economic shocks could particularly disadvantage young men. In contrast, the earnings of young women in the poorest cluster (*Poorest but stable*) are the lowest suggesting that the cumulative impacts of poverty are more damaging for women. Male and female outcomes are both tied to earlier lives but in different ways.

## 6. Discussion

The research makes significant empirical and methodological contributions to the literature analysing the transition to adulthood and employment in developing countries. It does this by examining the different experiences of young men and women in India very early on in their lives and their careers using sequence analysis and regression analysis.

The analysis identifies four clusters grouping individuals sharing similar life trajectories reflecting childhood and family circumstances: *Better-off, Poorer and unstable, Transitioning to better-off* and *Poorest but stable*. The trajectories track differences in living conditions (measured by access to basic amenities e.g. electricity, piped drinking water), health and exposure to family and economic shocks. Young people in the *Better-off* cluster shared the most advantageous life histories characterised by mainly full access to amenities, good health and very few shocks. Those in the *Transitioning to better-off* cluster experienced a sharp increase in access to amenities over their life courses. In contrast, those in the *Poorer and unstable* and *Poorest but stable* clusters had limited access to basic amenities, particularly the latter but with the former experiencing the

For Peer Review Only

highest number of shocks of any cluster. The analysis found that, relative to the most advantaged *Better-off* cluster, other young adults were significantly more likely to be in employment, particularly those in the poorest cluster (*Poorest but stable*). At these young ages employment appears to be driven to greater extent by lack of choice or need, with better-off young people more able to exercise choice. However, once in employment those in the *Better-off* cluster earned significantly more than those in the less advantaged clusters, consistent with theories linking unequal childhood circumstances to inequality of opportunity (Roemer and Trannoy, 2015, 2016).

Even at very early stages in their adult lives, young women in the sample were much less likely to be in work, 48.2% percent of women, compared with 76.4% of men. The World Bank reports a lower employment rate for all women of 32.4% in 2025 and while this figure represents an improvement compared with previous years it continues to be low globally (World Bank, 2026). The analysis showed that women earn less both hourly and daily than men even after controlling for childhood and family circumstances, educational attainment, caste and other factors. There is significant earnings inequality between women, with hourly earnings of women in the more advantaged *Better-off* cluster significantly higher than those of women in all other clusters, particularly the *Poorest but stable* cluster. There is also earnings inequality among men with those in the *Better-off* cluster earn significantly more and earnings lowest among those experiencing the most shocks in the *Poorer & unstable* cluster.

These results confirm that links found between upbringing conditions and material outcomes in developed economies (Kobayashi et al., 2017) are important in India, a lower-middle income country. Furthermore, the employment outcomes of women in the *Better-off* cluster are more strongly differentiated from those of other women, compared with the differentiation among men. This suggests that the impact of family advantage on life chances is more pronounced for

For Peer Review Only

women. Nevertheless, once in employment, the earnings of less advantaged young people, regardless of gender, are lower compared with those from better-off backgrounds. That this relationship is particularly evident for men who have suffered multiple shocks over their life courses, suggests they are the least likely to be compensated for their less advantageous early-life conditions. The underlying causes of this lack of compensation require further research (Currie and Vogl, 2013).

There are limitations of this study linked to the data and methods. The longitudinal data used to capture early-life conditions and experiences, while spanning fifteen years is derived from five survey years which limits the degree of variability in the data. The data used to construct the life histories is recorded from age eight and research has highlighted the formative importance for later life outcomes of life conditions at ages younger than this (Mancini and Yang, 2009; García et al, 2021). The sequence analysis requires choices regarding the number of clusters and the costs used to calculate the distance matrix (Potârcă et al., 2013). Lastly, the mechanisms and interactions underlying the linkages found between early life conditions and later life consequences require further investigation in a developing country context (Currie and Vogl, 2013).

## 7. Conclusion

Our findings have clear policy implications which can inform measures to support the equalisation of employment opportunities among young people in India by addressing disadvantage in childhood and family circumstances. First, the results link employment outcomes in early adulthood to childhood and family circumstances reflected in living conditions, health and exposure to shocks over children's life courses. Our analysis highlights the role of advantage in terms of socioeconomic circumstance for adult earnings and the ability to delay employment participation, consistent with a kind of voluntary "*waithood*" only available to relatively privileged

For Peer Review Only

young people (Honwana, 2012:5). These links between earlier life advantages and more advantageous early-adult outcomes suggests that differences widen over time in line with theories of cumulative (dis)advantage (Dannefer, 2003). Labour market reforms to improve employment prospects in India would, therefore, likely yield higher (and more inclusive) long-term returns for society if they are accompanied by policies aimed at improving the living conditions, health and security of disadvantaged young children. Such policies could help nullify the impact of circumstantial inequalities on outcomes, a goal noted by Roemer and Trannoy (2016:230)

Second, the results confirm the existence of gender inequalities in employment participation and earnings even at relatively young ages. This aligns with Darko and Vasilakos's (2020) argument and findings showing that the effort to reduce inequality in employment opportunities really need to start in earlier life with interventions that narrow gaps in the educational and employment outcomes of boys and girls from different social classes.

Lastly, the analysis suggests the way lives unfold in India are part of the reason for lower female employment participation and earnings. As such, the research highlights the importance of childhood circumstances and gender in adult outcomes. By implication, policy interventions designed to support the economic outcomes of young adults in India and other lower-middle income countries need to consider children's material living conditions and experiences as well as gender differences in access to employment opportunities.

For Peer Review Only

## References


Afridi, F., Dinkelman, T., & Mahajan, K. (2018). Why are fewer married women joining the work force in India? A decomposition analysis over two decades. *Journal of Population Economics*, 31(3):783–818.

Agrawal, T. (2021), Gender segregation and wage differentials in India: the role of educational attainment and occupational choices, *International Journal of Manpower*, 42(1):1-20.

Akresh, R., Lucchetti L. & Thirumurthy, H. (2012). Wars and child health: evidence from the Eritrean-Ethiopian conflict. *Journal of Development Economics*, 99(2):330–40

Arbache, J. S., Kolev, A. and Filipiak, E. (2010). *Gender disparities in Africa's labor market*, Africa Development Forum Series, Washington DC: The World Bank

Bargain, O. and Zeidan, J. (2017). Stature, skills and adult life outcomes: Evidence from Indonesia. *The Journal of Development Studies*, 53(6):873-890.

Bharadwaj, P, Løken K.V., Neilson, C. (2013). Early life health interventions and academic achievement. *American Economic Review*, 103(5):1862–91.

Boye, K. and Grönlund, A. (2018). Workplace Skill Investments–An Early Career Glass Ceiling? Job Complexity and Wages Among Young Professionals in Sweden. *Work, Employment and Society*, 32(2):368-386.

Brzinsky-Fay, C., Kohler U. & Luniak, M. (2006). Sequence analysis with Stata**,** *The Stata Journal* 6(4):435–460.

Bukodi, E., Goldthorpe, J. H., Halpin, B., Waller, L. (2016). Is Education Now Class Destiny? Class Histories across Three British Birth Cohorts, *European Sociological Review*, 32(6):835–849

Canning, D. (2012). Progress in Health around the World. *Journal of Development Studies*, 48(12):1784-1798.

Carmichael, F., Darko, C., Kanji, S. & Vasilakos, N. (2023) The contribution of girls' longer hours in unpaid work to gender gaps in early adult employment: Evidence from Ethiopia, India, Peru and Vietnam, *Feminist Economics*, 29(1):1-37

Carrillo, B. (2017). Early Rainfall Shocks and Later-Life Outcomes: Evidence from Colombia. *The World Bank Economic Review*. 34(1):179-209

Chatterjee,U., Murgai, R., Rama, M. (2015). Job opportunities along the rural-urban gradation and female labor force participation in India*. Policy Research Working Paper* 7412 Washington D.C.:World Bank Group.

Chuta, N. (2014). Children's Agency in Responding to Shocks and Adverse Events in Ethiopia. Young Lives, University of Oxford, www.younglives.org [accessed 21/12/2024]

Conley, D. and Bennett N.G. (2001). Birth weight and income: interactions across generations. *Journal of Health and Social Behavior*, 42:450–465.

Conley, D., Pfeiffer K.M., Velez, M. (2007). Explaining sibling differences in achievement and behavioral outcomes: The importance of within-and between-family factors, *Social Science Research*, 36(3):1087–1104.

Crivello, G. and Morrow, V. (2019). Against the odds: Why some children fare well in the face of adversity. The Journal of Development Studies, 56(5):1-18.

Currie, J., & Vogl, T. (2013). Early-Life Health and Adult Circumstance in Developing Countries, *Annual Review of Economics*, 5(1):1–36

Darko, C.K. and Vasilakos, N. (2020). Picking winners: An empirical analysis of the determinants of educational outcomes in India. *British Educational Research Journal*, *46*(6):1445-1467.


For Peer Review Only

Dannefer, D. (2003). Cumulative advantage/disadvantage and the life course: cross fertilizing age and social science theory, *Journal of Gerontology: Social Sciences*, 58B(6):S327-337.

Devine, K. A., Christen, S., Mulder, R. L. et al. (2022) Recommendations for the surveillance of education and employment outcomes in survivors of childhood, adolescent, and young adult cancer; A report from the international late effects of childhood cancer guideline harmonization group. *Cancer,*128(13):2405-2419.

Desai, S., & Joshi, O. (2019). The paradox of declining female work participation in an era of economic growth: The case of India. *Indian Journal of Labour Economics*, 62(1):55–71.

Duncan, G. J., Morris, P. A., & Rodrigues, C. (2011). Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments. *Developmental Psychology,* 47(5):1263-1279.

Duraisamy, M., & Duraisamy, P. (2016). Gender wage gap across the wage distribution in different segments of the Indian labour market, 1983–2012: exploring the glass ceiling or sticky floor phenomenon. *Applied Economics*, 48(43):4098–4111.

Elder, G.H. (1998), The Life Course as Developmental Theory. *Child Development*, 69(1): 1-12

Favara, M. (2017). Do dreams come true? Aspirations and educational attainments of Ethiopian boys and girls. *Journal of African Economies*, 26(5):561-583.

García, J.L., Heckman, J.J., Leaf D. E., and Prados. M.J. (2020). Quantifying the life-cycle benefits of an influential early-childhood program. *Journal of Political Economy* 128(7):2502-2541.

Hamil-Luker, J., & O'Rand, A.M. (2007). Gender differences in the link between childhood socioeconomic conditions and heart attack risk in adulthood. *Demography*, *44*(1):137-158.

Heckman, J.J. (1979). Sample bias as a specification error. *Econometrica*, 47(1):153–161.

Himaz, R. (2018). Stunting later in childhood and outcomes as a young adult: Evidence from India. *World Development*, *104*:344-357.

Honwana, A. (2012). *The Time of Youth: Work, Social Change, and Politics in Africa*, Boulder CO:Kumarian Press

Jylhä, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science & Medicine*, 69(3):307–316.

Kingdon, G. G., & Unni, J. (2001). Education and women's labour market outcomes in India. *Education Economics*, 9(2), 173–195.

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

Kobayashi, L. C., Glymour, M. M., Kahn, K., Payne, C. F., Wagner, R. G., Montana, L., ... & Berkman, L. F. (2017). Childhood deprivation and later-life cognitive function in a population-based study of older rural South Africans. *Social Science & Medicine*, 190:20-28.

León G. (2010). Civil Conflict and Human Capital Accumulation: The Long Term Effects of Political Violence in Peru. *BREAD Working Paper*, 256.

Linnemayr, S. & Alderman H. (2011). Almost random: evaluating a large-scale randomized nutrition program in the presence of crossover. *Journal of Development Economics,* 96(1):106–114

Maccini. S., Yang, D. (2009). Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall. *American Economic Review,* 99(3):1006-26.

Malakar, K. D., Kuzur, G., Maity D. K., Yadav, M., and Roy, S. (2022). A Socio-Ecological Study of Population, Migration, Urbanization, and Socio-Climate Variation in Andhra Pradesh and Telangana, India. *Journal of Geography, Environment and Earth Science International* 26(12):1-33.

For Peer Review Only

McKenzie, D. (2017). How effective are active labor market policies in developing countries? a critical review of recent evidence. *The World Bank Research Observer*, *32*(2):127-154.

Melchior, M., Moffitt, T. E., Milne, B. J., Poulton, R., & Caspi, A. (2007). Why do children from socioeconomically disadvantaged families suffer from poor health when they reach adulthood? A life-course study. *American Journal of Epidemiology*, 166(8):966-974.

Mitra, A. and Verick, S. (2013) Youth employment and unemployment: an Indian perspective, *ILO Asia-Pacific Working Paper Series*, DWT South Asia and Country Office for India, New Delhi:ILO [Accessed 12/01/2025]https://webapps.ilo.org/public/libdoc/ilo/2013/113B09_171_engl.pdf

Nutz, T. and Lersch, P.M. (2021). Gendered employment trajectories and individual wealth at older ages in Eastern and Western Germany. *Advances in Life Course Research*, *47*, p.100374.

Pagán, J. A. (2002). Gender differences in labor market decisions in rural Guatemala. *Review of Development Economics*, 6(3), 428-441.

Pathania, V. (2009). The long run impact of drought at birth on height of women in rural India. Working Paper, *University of California,* Berkeley.Reddy, A. A., & Bantilan, M. (2013). Regional disparities in Andhra Pradesh, India. *Local Economy: The Journal of the Local Economy Policy Unit*, *28*(1):123-135.

Potârcă, G., Mills, M. and Lesnard, L. (2013). Family formation trajectories in Romania, the Russia Federation and France: Towards the second demographic transition? *European Journal of Population*, 29(1):65-76.

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

Roemer, J. E., and Trannoy, A. (2015). Equality of opportunity, chapter 4 in Atkinson, A. B. & Bourguignon, F., (eds.) *Handbook of Income distribution*, 2A:217-300 http://dx.doi.org/10.1016/B978-0-*444-59428-0.00005*

Sarkar, S., Chakravorty, B., & Lyonette, C. (2025) Caste, gender and intersectionality in labour market aspirations: evidence from a skill-training programme in India, *Oxford Development Studies*, 54(1):22-40.

Van Den Berg, G. J., Lindeboom, M., Portrait, F. (2006). Economic conditions early in life and individual mortality. *American Economic Review* 96(1):290–302.

Vauhkonen, T., Kallio, J., Kauppinen, T.M. and Erola, J. (2017). Intergenerational accumulation of social disadvantages across generations in young adulthood. *Research in Social Stratification and Mobility*, 48:42-52.

Vogl, T. (2012). Height, skills, and labor market outcomes in Mexico. *NBER* Working Paper 18318 http://www.nber.org/papers/w18318

Wiborg, Ø.N. and Møberg, R.J. (2010). Social origin and the risks of disadvantage in Denmark and Norway:the early life course of young adults. *Work, Employment and Society*, *24*(1):105-125.

World Bank (2026) *India gender landscape*, Gender Data Portal. [Accessed 01/05/2026]: https://genderdata.worldbank.org/en/economies/india

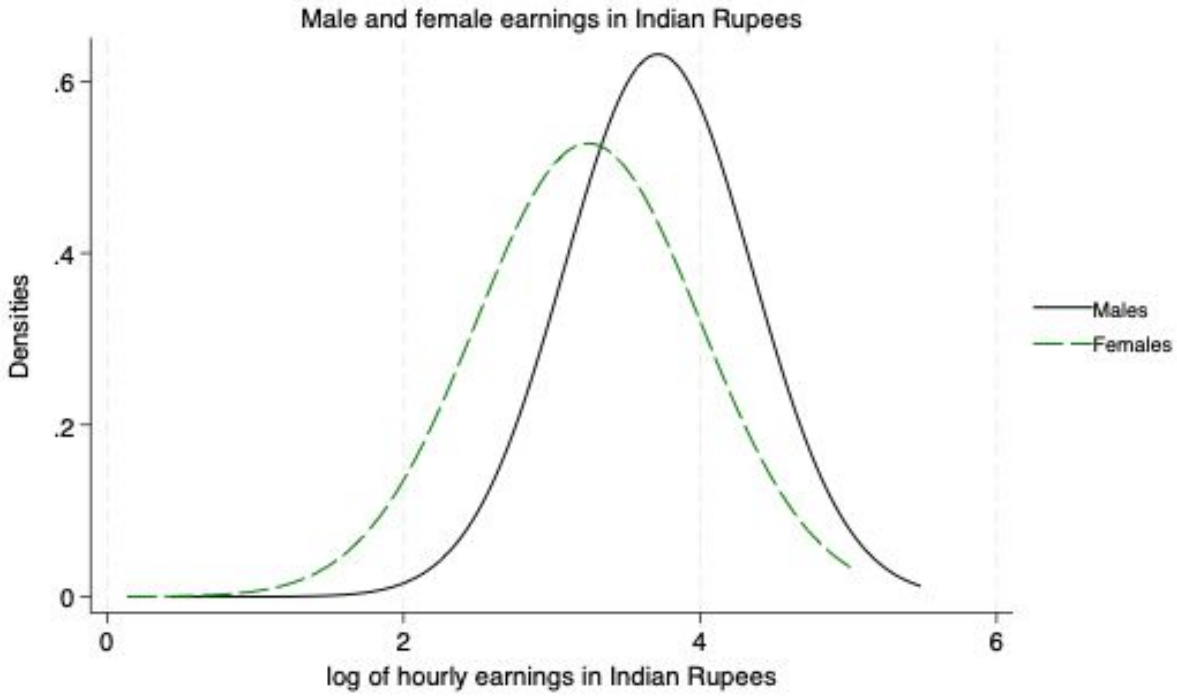
Young Lives (2018) *Survey Design and Sampling Fact sheets*, [Accessed 4/06/18] www.younglives.org.uk

Zou, W., Lybbert, T., Vosti, S. and Abbeddou, S. (2020). Early Childhood Nutrition, Parental Growth Perceptions and Educational Aspirations in Rural Burkina Faso. The Journal of Development Studies, 56(4):798-816.

For Peer Review Only

## Figures and Tables

**Figure 1. Hourly earnings**




For Review Only

**Figure 2: Modal sequence by cluster**

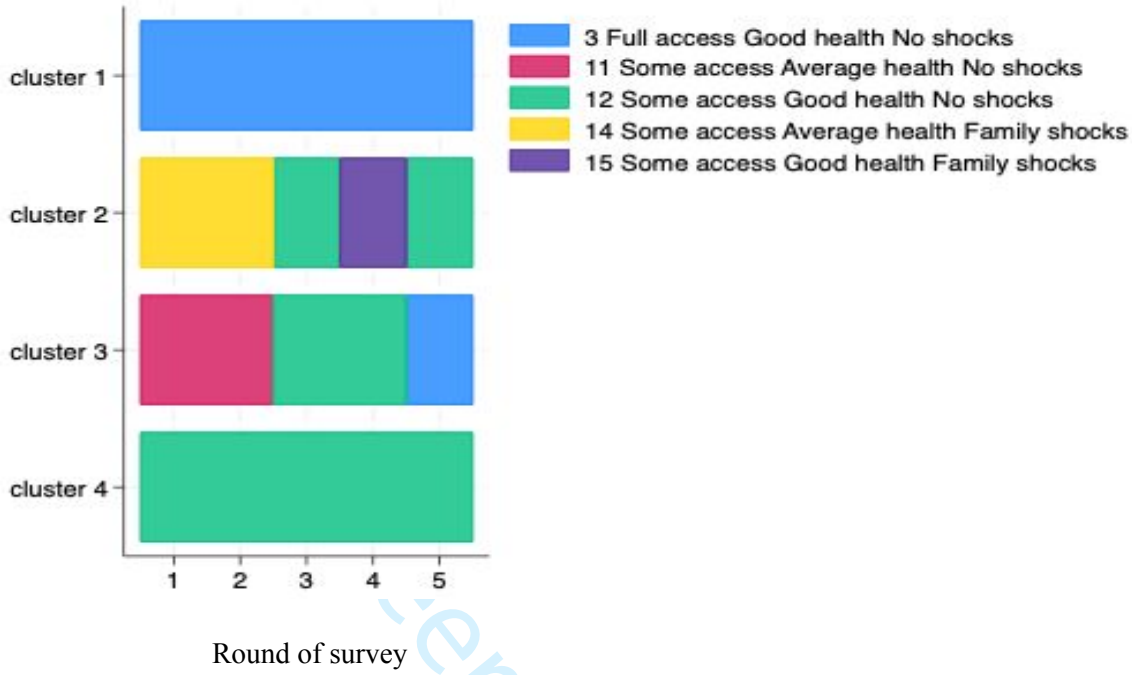



**Figure 3(a): Access to amenities trajectories over the four clusters**

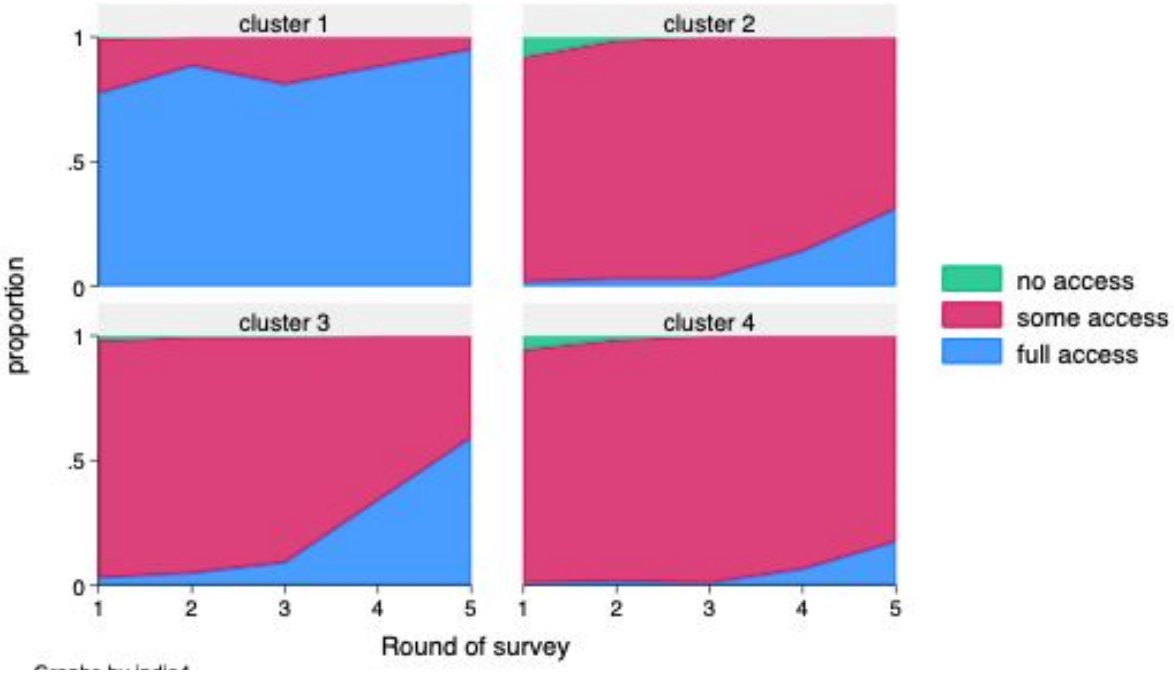


**Figure 3(b): Health trajectories over the four clusters**

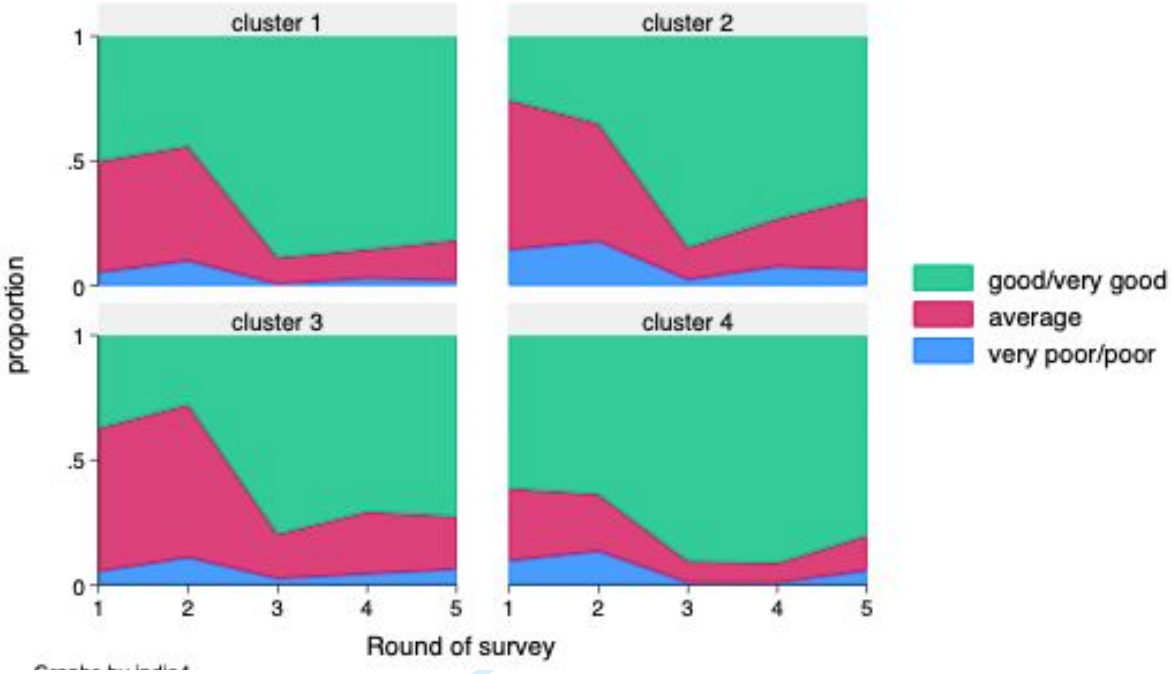



**Figure 3(c): Shock trajectories over the four clusters**

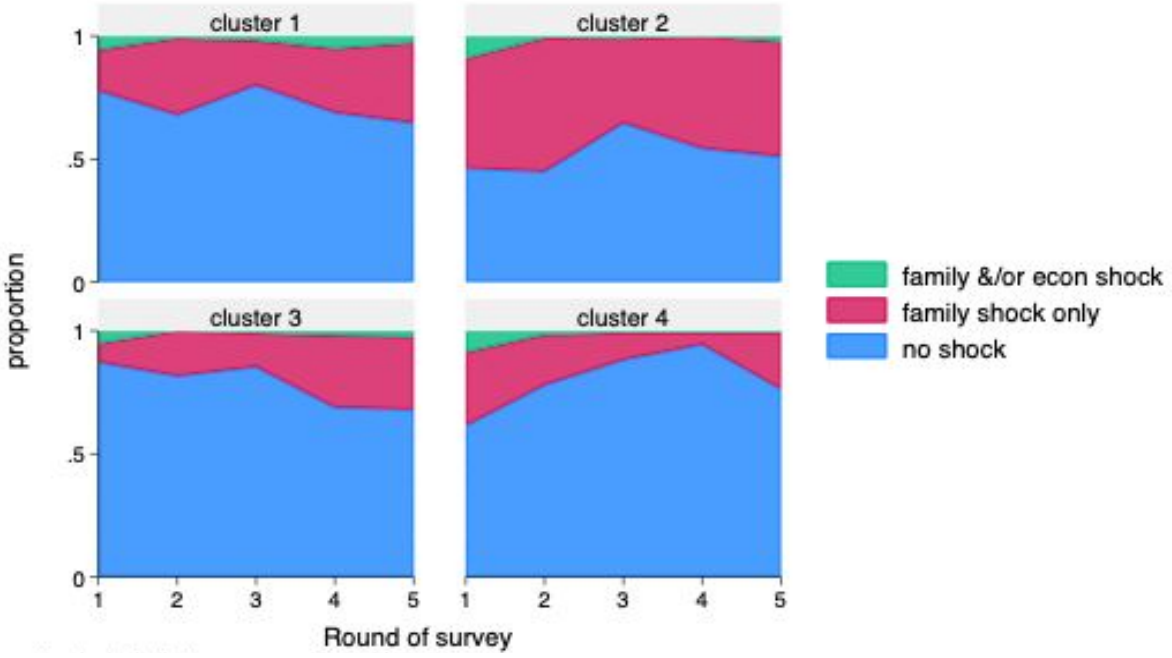


**Table 1. Cluster characteristics**

| *% of sequence states* | Cluster 1 Better-off | Cluster 2 Poorer and unstable | Cluster 3 Transitioning to better off | Cluster 4 Poorest but stable | All |
|---|---|---|---|---|---|
| Access to amenities*** | | | | | |
| % full access | 85.86 | 10.42 | 21.89 | 5.35 | 30.14 |
| % Some access | 14.04 | 87.63 | 77.54 | 93.22 | 68.72 |
| % No access | 0.1 | 1.95 | 0.47 | 1.10 | 1.14 |
| Health*** | | | | | |
| % Very good/good health | 70.14 | 56.66 | 57.64 | 77.48 | 62.96 |
| % Average health | 25.86 | 33.82 | 36.6 | 16.54 | 30.19 |
| % very poor /poor | 4.00 | 9.53 | 5.75 | 5.98 | 6.86 |
| Shocks*** | | | | | |
| % No shocks experienced | 71.81 | 52.2 | 78.21 | 79.69 | 66.68 |
| Number of participants | 215 | 359 | 212 | 127 | 913 |

*** p<0.0: significance in Pearson chi squared tests of independence between clusters and categorical variables recording access to amenities, health and exposure to shocks

**Table 2. Employment participation (probit): Average marginal effects**

| | (1) All | (2) Women | (3) Men |
|---|---|---|---|
| Female | -0.2660*** | | |
| | (0.0302) | | |
| Cluster 1 *Better-off* -reference cluster | | | |
| Cluster 2 *Poorer & unstable* | 0.1087** | 0.1544*** | 0.0862 |
| | (0.0437) | (0.0516) | (0.0807) |
| Cluster 3 *Transitioning to better-off* | 0.1006** | 0.1292*** | 0.1060 |
| | (0.0454) | (0.0491) | (0.0814) |
| Cluster 4 *Poorest but stable* | 0.1923*** | 0.2723*** | 0.1416** |
| | (0.0446) | (0.0596) | (0.0721) |
| Demographic, childhood circumstance & regional controls[a] | YES | YES | YES |
| Observations | 865 | 440 | 425 |
| Pseudo R-squared | 0.262 | 0.18 | 0.296 |
| Wald Chi squared | 1359.72*** | 758.85*** | 465.90*** |
| Log pseudo likelihood | -426.394 | -249.101 | -163.221 |

Figures are average marginal effects on the probability of employment. Robust standard errors in parentheses (clustered at the local area survey site level): *** p<0.01, ** p<0.05, * p<0.1.

[a]Results for controls in Table A3: Controls: age, educational attainment, whether has own children, caste. Employment rate local area site, urban or rural location, region. Hours of work within and outside the household at age 12, age started school, BMI at ages 8 and 12, number of siblings 0-12 years at age 12.

For Peer Review Only

**Table 3. Earnings (OLS): Marginal effects**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Pooled | | Women | | Men | |
| Log of earnings in Rupees | Hourly earnings | Daily earnings | Hourly earnings | Daily earnings | Hourly earnings | Daily earnings |
| Female | -0.361*** | -0.4825*** | | | | |
| | (0.1553) | (0.1595) | | | | |
| Cluster 1 *Better-off* | -reference cluster | | | | | |
| Cluster 2 *Poorer & unstable* | -0.3387** | -0.3997** | -0.6306** | -0.3650 | -0.3176* | -0.4116*** |
| | (0.1569) | (0.1508) | (0.2984) | (0.3644) | (0.1791) | (0.1258) |
| Cluster 3 *Transitioning to better-off* | -0.3546** | -0.3924*** | -0.8696** | -0.7682* | -0.2818 | -0.2885* |
| | (0.1368) | (0.1245) | (0.3357) | (0.4119) | (0.1788) | (0.1581) |
| Cluster 4 *Poorest but stable* | -0.3555** | -0.4097** | -0.9794*** | -0.8539* | -0.2657 | -0.2682* |
| | (0.1440) | (0.1714) | (0.3101) | (0.4432) | (0.1829) | (0.1546) |
| IMR | -0.1551 | -0.2981 | -0.8305* | -0.4742 | -0.2712 | -0.4836 |
| | (0.2747) | (0.2716) | (0.4810) | (0.5289) | (0.2587) | (0.2903) |
| Demographic, childhood circumstance & regional controls[a] | YES | YES | YES | YES | YES | YES |
| Constant | -0.4721 | 1.6122 | -9.4315 | -12.6454 | 3.7098 | 7.0221** |
| | (2.1509) | (2.8830) | (7.3966) | (9.7142) | (2.5948) | (3.3171) |
| Observations | 338 | | 105 | | 233 | |
| R-squared | 0.162 | 0.179 | 0.213 | 0.172 | 0.087 | 0.116 |
| F | 73.53*** | 42.99*** | 33.31*** | 8.92*** | **10.22***** | **7.71***** |

Figures are marginal effects (coefficients). Robust standard errors in parentheses (clustered at the local area level): *** p<0.01, ** p<0.05, * p<0.1.

[a]Results for controls in Table A4: Controls: age, educational attainment, whether has own children. Caste, hours of work within and outside the household at age 12, Employment rate local area site, urban or rural area, region.

For Peer Review Only

# Supplementary Appendix

## Table A1: Summary statistics[a]

| | All | | | Women | | | Men | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | N | Mean | SD | N | Mean | SD | N |
| Women | 0.51 | 0.50 | 916 | | | | | | |
| Employed[b] | 0.62 | 0.49 | 916 | 0.48 | 0.50 | 471 | 0.76 | 0.43 | 445 |
| Hourly earnings (Rupees) | 42.95 | 30.47 | 363 | 32.32 | 26.53 | 118 | 48.06 | 30.97 | 245 |
| Daily earnings (Rupees) | 355.32 | 248.7 | 363 | 255.24 | 228.3 | 118 | 403.53 | 244.1 | 245 |
| Daily earnings USD[c] | 9.261 | 29.98 | 363 | 6.72 | 21.11 | 118 | 10.49 | 33.39 | 245 |
| Age (in months) | 263.98 | 4.20 | 916 | 264.07 | 4.17 | 471 | 263.88 | 4.22 | 445 |
| Educational grade attained | 11.29 | 3.20 | 901 | 10.86 | 3.30 | 465 | 11.76 | 3.02 | 436 |
| SC or ST caste[d] | 0.33 | 0.47 | 916 | 33.33 | 0.47 | 471 | 32.36 | 0.32 | 445 |
| Has own children | 0.24 | 0.43 | 916 | 0.42 | 0.49 | 471 | 0.05 | 0.21 | 445 |
| Daily hours household work age 12 | 1.57 | 1.93 | 915 | 1.87 | 2.00 | 471 | 1.25 | 1.82 | 444 |
| Daily hours work outside home age 12 | 0.37 | 1.70 | 913 | 0.42 | 1.79 | 470 | 0.32 | 1.60 | 443 |
| School age start /grade1 | 5.03 | 0.92 | 907 | 5.00 | 0.92 | 465 | 5.06 | 0.93 | 442 |
| BMI age 12 | 25.74 | 145.4 | 909 | 29.90 | 159.3 | 466 | 21.37 | 129.2 | 443 |
| BMI age 8 | 13.93 | 1.69 | 916 | 13.87 | 1.26 | 471 | 13.99 | 2.04 | 445 |
| No. of siblings age 0-12 (in household) at age 12 | 0.87 | 0.98 | 916 | 0.93 | 0.96 | 471 | 0.81 | 1.00 | 445 |
| Employed rate local area (survey) site | 0.62 | 0.17 | 916 | 0.59 | 0.15 | 471 | 0.65 | 0.18 | 445 |
| Rural location | 0.68 | 0.47 | 905 | 0.65 | 0.48 | 463 | 0.70 | 0.46 | 442 |
| Coastal Andhra | 0.35 | 0.85 | 909 | 34.84 | 0.84 | 465 | 35.14 | 0.85 | 444 |
| Rayalaseema | 0.28 | 0.85 | 909 | 29.25 | 0.84 | 465 | 27.48 | 0.85 | 444 |
| Telangana | 0.37 | 0.85 | 909 | 35.91 | 0.84 | 465 | 37.39 | 0.85 | 444 |

[a]Unless stated, statistics are survey round 5 (age 22).
[b]Either self-employed, in regular, salaried, formal employment, casual, informal employment, a farm servant, housemaid or non-specified employment.
[c]Rupees converted to US Dollars using appropriate daily real exchange rates from Thomson-Reuter Datastream.
[d]Reference is BC (other economically and socially backward classes) or OC (other/open/general castes)

For Peer Review Only

**Table A2 Cluster employment and demographic characteristics (age 22)**

| | Cluster 1 Better-off | Cluster 2 Poorer and unstable | Cluster 3 Transitioning to better off | Cluster 4 Poorest but stable | All |
|---|---|---|---|---|---|
| % employed *** | 39.53 | 68.80 | 63.68 | 76.38 | 61.77 |
| Mean daily earnings (Rupees)*** | 467.406 | 323.292 | 330.066 | 322.896 | 355.287 |
| Mean daily earnings (USD)*** | 17.706 | 6.771 | 6.618 | 6.768 | 9.27 |
| % female participants | 50.23 | 49.86 | 55.66 | 50.39 | 50.08 |
| % living in rural locations*** | 17.6 | 85.67 | 74.76 | 88.89 | 67.52 |
| Regional location*** | | | | | |
| % living in Coastal Andhra | 40.93 | 37.25 | 27.88 | 30.16 | 34.99 |
| % living in Rayalaseema | 22.33 | 20.45 | 45.67 | 31.75 | 28.26 |
| % living in Telangana | 36.74 | 42.30 | 26.44 | 38.10 | 36.75 |
| % post-secondary educational grade*** | 69.77 | 38.44 | 45.75 | 50.39 | 49.18 |

*** p<0.01, ** p<0.05, * p<0.1 significance in Pearson chi squared tests of independence between clusters and categorical variables and ANOVA F and Kruskal-Wallis chi squared tests of differences in cluster means for continuous earnings measures.

**Table A3. Employment participation (probit): Average Marginal effects for control variables not reported in Table 2**

| | (1) All | (3) Women | (2) Men |
|---|---|---|---|
| Age | -0.0009 | 0.0001 | 0.0009 |
| | (0.0040) | (0.0058) | (0.0032) |
| Educational grade attained | -0.0494*** | -0.0381*** | -0.0708*** |
| | (0.0082) | (0.0108) | (0.0134) |
| Has own children | -0.1734*** | -0.1691*** | -0.1700 |
| | (0.0395) | (0.0394) | (0.1994) |
| SC or ST caste | 0.0610** | 0.0219 | 0.1032*** |
| | (0.0243) | (0.0406) | (0.0249) |
| Daily hours household work age 12 | 0.0222* | 0.0280*** | 0.0183 |
| | (0.0123) | (0.0104) | (0.0237) |
| Daily hours work outside household age 12 | -0.0206* | -0.0169 | -0.0222 |
| | (0.0117) | (0.0122) | (0.0206) |
| Age started school/grade 1 | -0.0542*** | -0.0554** | -0.0580*** |
| | (0.0137) | (0.0254) | (0.0138) |
| No of siblings aged 0-12 (age 12) | 0.0259** | 0.0153 | 0.0314** |
| | (0.0113) | (0.0166) | (0.0156) |

For Peer Review Only

| | | | |
|---|---|---|---|
| BMI age 12 | -0.0000 | -0.0001 | -0.0088 |
| | (0.0000) | (0.0000) | (0.0092) |
| BMI age 8 | 0.0037 | 0.0144 | -0.0049 |
| | (0.0099) | (0.0189) | (0.0067) |
| Employed rate local area (survey) site | 0.6617*** | 0.6801*** | 0.7129*** |
| | (0.1174) | (0.1507) | (0.1118) |
| Rural area | -0.0447 | 0.0427 | -0.1442*** |
| | (0.0599) | (0.0620) | (0.0427) |
| Coastal Andhra | -0.0101 | -0.0489 | 0.0419 |
| | (0.0227) | (0.0400) | (0.0379) |
| Rayalaseema | -0.0231 | -0.0262 | -0.0189 |
| | (0.0228) | (0.0429) | (0.0397) |

Robust standard errors in parentheses (clustered at the local area survey site level): *** p<0.01, ** p<0.05, * p<0.1.

**Table A4. Earnings (OLS): Marginal effects for control variables not reported in Table 3**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Pooled | | Women | | Men | |
| Age | 0.0170** | 0.0182 | 0.0726** | 0.0539** | 0.0012 | -0.0022 |
| | (0.0081) | (0.0108) | (0.0346) | (0.0254) | (0.0098) | (0.0124) |
| Educational grade attained | 0.0073 | -0.0120 | -0.0297 | 0.0335 | 0.0136 | 0.0042 |
| | (0.0185) | (0.0184) | (0.0421) | (0.0271) | (0.0212) | (0.0226) |
| Has own children | 0.0170 | 0.0248 | -0.0874 | 0.0997 | 0.2264 | 0.2941* |
| | (0.1260) | (0.1507) | (0.2173) | (0.1719) | (0.1608) | (0.1563) |
| SC or ST caste | -0.1838** | -0.1767* | 0.0578 | -0.0227 | -0.2935** | -0.3364*** |
| | (0.0879) | (0.0925) | (0.1310) | (0.1218) | (0.1059) | (0.1166) |
| Daily hours household work age 12 | -0.0125 | -0.0211 | -0.0390 | -0.0102 | -0.0033 | -0.0094 |
| | (0.0128) | (0.0136) | (0.0256) | (0.0261) | (0.0193) | (0.0192) |
| Daily hours work outside household age 12 | -0.0288 | -0.0392** | -0.1279** | -0.1110** | -0.0094 | -0.0074 |
| | (0.0196) | (0.0184) | (0.0566) | (0.0410) | (0.0213) | (0.0208) |
| Employed rate local area (survey) site | 0.0907 | 0.1088 | 0.4610 | -0.3535 | -0.1817 | -0.2790 |
| | (0.3320) | (0.4623) | (1.1853) | (0.8674) | (0.3404) | (0.3855) |
| Rural area | 0.0535 | -0.0142 | -0.1860 | -0.1834 | 0.1893 | 0.1011 |
| | (0.1192) | (0.0943) | (0.1518) | (0.1578) | (0.1229) | (0.1044) |
| Coastal Andhra | -0.0292 | 0.0479 | -0.1366 | -0.1734 | -0.0300 | 0.0225 |
| | (0.0708) | (0.0827) | (0.2070) | (0.1566) | (0.0684) | (0.0868) |
| Rayalaseema | -0.1152 | -0.1239 | -0.1569 | -0.1290 | -0.1254 | -0.1308 |
| | (0.0852) | (0.0964) | (0.2504) | (0.1490) | (0.0937) | (0.0945) |

Robust standard errors in parentheses (clustered at the local area survey site level): *** p<0.01, ** p<0.05, * p<0.1.

For Peer Review Only

# Unequal from the start? Children's life histories and their employment outcomes in India

**Abstract**

Five waves of the Young Lives Indian survey are analysed using sequence and regression analysis. Four clusters of individuals sharing similar life histories are identified. The life histories reflect aspects of childhood and family circumstances measured by living conditions, health and exposure to family and economic shocks. Regression analysis investigates how different life histories and gender impact employment outcomes at age 22. The results show that accumulation of (dis)advantages over children's lives helps to explain inequalities in employment and earnings. At age 22, those who have led wealthier, healthier, more stable lives have higher earnings but are less likely to be employed. Those from the poorest but most stable cluster are the most likely to be employed, but among women, this group earns the least. For young men, earnings are lowest for those exposed to the most shocks. Policy to equalise employment opportunities in India should therefore aim to improve the living conditions, health and security of disadvantaged young children.

**Keywords**: Early-life poverty, gender wage gap, youth employment, life history trajectories

For Peer Review Only

## 1. Introduction

This study analyses how early-life conditions and gender influence young adults' employment outcomes in India. As one of the world's most populous nations, with significant regional and gender disparities, India provides a unique context for understanding how early-life experiences and attributes shape employment outcomes of men and women. Despite significant economic growth in recent decades, female labour force participation in India remains among the lowest in emerging economies and has shown a declining trend over time (Chatterjee et al., 2015). This persistent gender disparity is driven by a combination of cultural norms, societal expectations, and institutional barriers that limit women's access to and participation in the labour market. These factors underscore the importance of studying the intersections between early-life conditions, gender, and employment outcomes in this setting.

The research contributes to sociological and economic literature exploring inequality of opportunity, social mobility and gender gaps in career progression. Prior studies have demonstrated that early-life socioeconomic, biosocial and demographic factors, including living environments, health and gender, play a critical role in shaping employment opportunities, social mobility and later life economic outcomes (Devine et al., 2022; Crivello & Morrow, 2019; Himaz, 2018;Vauhkonen et al, 2017). By focusing on the specific case of India, this study extends this literature, offering insights into how inequalities in employment and earnings are rooted in early-life experiences and perpetuated across the life course.

The study uses longitudinal panel data tracking children's lives from age 8 into early adulthood in India. Employing regression and sequence analysis, it investigates how different life course histories are linked to employment participation and earnings at age 22. The analysis identifies four clusters of life trajectories characterised by different experiences of household

For Peer Review Only

living conditions, health and exposure to family and economic shocks. In combination, these socioeconomic, well-being and environmental factors have significant and longer-term impacts conferring advantages and disadvantages in later life. At age 22, those with more stable and advantageous backgrounds had higher earnings, reflecting better job quality, despite being less likely to be employed. Young women, already disadvantaged in employment participation and pay compared with men, had the highest employment participation but lowest earnings in the poorest but most stable cluster. For young men, earnings were lowest for those experiencing the most shocks. These results highlight how early-life advantages amplify inequalities with employment failing to offset early disadvantage.

This life course approach provides new evidence of the cumulative effects of early-life circumstances and gendered experiences on employment outcomes, contributing to a deeper understanding of social and economic mobility in the Indian context. In a setting characterised by pronounced gender inequality, the key pathways identified can inform policies aiming to equalize opportunities and reduce gender inequality in employment. In doing so, this research enhances understanding of labour market dynamics in India and offers broader implications for addressing inequities in employment globally. It underscores the importance of examining early-life determinants within diverse socioeconomic contexts, providing a valuable contribution to debates on human capital development, mobility and gender equity.

The next section overviews previous research on early-life conditions, gender and adult outcomes. Subsequent sections describe the data, methodology and results. The final two sections discuss the findings and conclude.

For Peer Review Only

## 2. Research Background

Previous research on employment in India has focused more on macroeconomic or institutional determinants of male and female employment participation and earnings, often overlooking the role of early-life socioeconomic attributes (Afridi et al., 2018). Our study addresses this gap by emphasising the role of childhood circumstances and gender-specific socialisation processes in later life employment outcomes. The approach recognises that early life circumstances including exposure to adverse events, such as economic shocks or parental unemployment and gendered allocations of household work can have long-term effects on human capital development, labour market prospects and well-being (Carmichael et al., 2023; Chuta, 2014).

Gender is a significant determinant of employment outcomes, including career progression and earnings. The persistence of gender participation and wage gaps is notable in both developed countries, including OECD countries, and developing countries, including India (Agrawal, 2021; World Bank, 2026). Discrimination and job segregation contribute to these gender differentials that can arise early in careers (Boye and Grönlund, 2018)) and persist (and widen) over long-term career trajectories (Nutz and Lersch, 2020). Gender inequality in employment is deeply ingrained in India's socio-cultural framework. Patriarchal norms often restrict women's mobility, educational attainment, and access to economic opportunities (Klasen & Pieters, 2015). These constraints are compounded by early-life disadvantages, including limited access to quality education, healthcare, and nutrition that significantly affect women's ability to participate in the labour market (Agrawal, 2021; Kingdon & Unni, 2001). The gender gap in employment in India is further intensified by the dual burden of unpaid care work and limited access to formal employment opportunities. As noted by Desai and Joshi (2019), women in India spend disproportionately more time on domestic responsibilities, reducing their ability to engage in paid

employment. This imbalance begins in early life, as young girls are allocated caregiving roles, while boys pursue education and career-oriented activities. By the time they reach adulthood, young women in India already have lower income aspirations than males, particularly if they are from more disadvantaged castes (Sarkar et al. 2026). Investigating how gendered early-life experiences impact employment outcomes in adulthood is crucial for designing policy interventions to reduce gender inequalities.

Previous studies, albeit predominantly in the context of developed countries, have shown how early-life socioeconomic circumstances (including parental income and social class) life events and interventions can be important determinants of adult outcomes including learning (Zou et al., 2020; Conley et al, 2007) and employment and earnings (Bukodi et al., 2016; Bargain and Zeidan, 2017; García et al., 2021). For example, Wiborg and Moberg (2010) show that in Nordic countries a disadvantaged social background is associated with receipt of social assistance and higher risks of unemployment in later life. Vauhkonen et al (2017) find that indicators of household disadvantage (school drop-out, receipt of social assistance and unemployment) in Finland are intergenerationally inherited. Lower family income is associated with negative outcomes such as school dropout, lower educational attainment and health problems (Duncan, et al., 2011).

Health in childhood and early-life health indicators including environmental and socioeconomic conditions have been associated with education, employment and well-being outcomes (Canning, 2012). Bharadwaj et al. (2013) link improved early life health care to educational attainment. Devine et al. (2022) find early-life cancer is a risk factor for poor educational and employment outcomes. Hamil-Luker and O'Rand (2007) find that early exposure to adverse socioeconomic conditions is associated with higher risk for heart attack among women. Van Den Berg et al. (2007) find adult life expectancy is lower for children born during economic

For Peer Review Only

downturns. Melchior et al. (2007) find that children's socioeconomic status and living conditions are related to a range of health problems in adulthood. Conley and Bennett (2001) report links between socioeconomic environment and birthweight.

Early-life environments and gender roles are likely to be just as (if not more) important for adult outcomes in developing counties, where children are more vulnerable to shocks such as famine and conflict (Currie and Vogl, 2013). Studies on early-life environments and adult outcomes (including employment, education and well-being) in developing countries have examined the effects of circumstances and conditions in childhood. These have included children's health and physical development (Currie, 2009; Vogl, 2012) and specific causes of ill-health such as malnutrition (Linnemayr and Alderman, 2011). In Himaz's (2018) study of children from age 8 to 19 in India, children's physical development is correlated with psychosocial outcomes and has intergenerational consequences. Later life effects are associated with exposure to shocks including parental death (Crivello and Morrow, 2019) famine (Pathania, 2009) and armed conflict (Akresh et al., 2012). Early-life effects may vary according to gender and socioeconomic status. For example, higher early-life rainfall in Indonesia affects adult outcomes for women but not men (Maccini and Yang, 2009) whereas Carillo (2019) reports no gender differences for Colombia.

Theoretical links underpinning these relationships suggest a life course approach which recognises that historical context shapes individual lives (Elder, 1998). Theories of opportunity equalisation further highlight how aspects of an individual's childhood and family circumstances that are beyond their control can influence outcomes, reinforcing inequalities (Roemer and Trannoy, 2016). While the contribution of circumstances to outcome inequality is complicated by multidimensional aspects of equality of opportunity (Roemer and Trannoy, 2016:1309) previous research provides guidance. In developing economy contexts research has shown that early-life

For Peer Review Only

conditions, including parental investments in health and education and the development of both physical and cognitive skills, have consequences for labour market success (Currie and Vogl, 2013). Channels for these links are both economic and nutritional (León, 2010) and the impact of shocks may be compensated for or reinforced by social supports, beliefs and values, including gender norms.

Our study contributes to this literature by focusing on the impact of earlier-life conditions and gender on employment outcomes. The former are incorporated by constructing life history records using sequence analysis with optimal matching and clustering. The method has been widely used in research on social mobility and careers (Bukodi et al, 2016) and family formation (Potârcă, et al., 2013). Here sequence analysis is used to group young people sharing similar social circumstances over their life courses (reflected in living conditions, health and exposure to shocks) from age 8 to age 22. Regression analysis is then used to investigate whether inequalities in employment outcomes are linked to relatively more or less advantageous life course trajectories. By analysing how early-life conditions and gender influence young adults' labour market outcomes, this study contributes to research on inequality of opportunity and outcomes in developing countries.

## 3. Data and sample

The analysis uses data from the Young Lives project, a longitudinal cohort study of childhood poverty for India and three other low-middle income countries: Ethiopia, Peru and Vietnam. In India, Young Lives followed 1000 children from an older cohort from age 8 and 2000 from a younger cohort from age 1. 20 local area survey sites in Coastal Andhra, Rayalaseema and Telangana were selected as part of a semi-purposive sampling strategy that oversamples poor

For Peer Review Only

households and ensures a mix of urban and rural regions. Details of the survey design and sampling are available from Young Lives (2018) and Favara (2017).

This study uses the first five rounds of data conducted in 2002, 2006, 2009, 2013 and 2016.[1] In the fifth round, the older cohort were 22 and the younger cohort 15. As our focus is on employment outcomes in early adulthood, we restrict the analysis to the older cohort since the majority (87.8%) of the younger cohort were in school at age 15. A limitation is that because the older cohort were first surveyed at age eight the analysis does not account for factors affecting children before age eight.

There were marginally fewer males (48.64%) in the sample (Supplementary Appendix Table A1). 61.90% of the young people were employed at age 22 but there is already wide variation in the employment participation of men and women: 76.4% of males were employed and only 48.2% of females. Figure 1 shows that at age 22, there is wide wage variation in earnings among those in employment, both within and between genders; women earned on average 67.25% of male hourly earnings and only 63.25% of male daily earnings. The lower employment rate and earnings of women are in line with evidence for India where low rates of female employment participation (World Bank, 2026) are combined with lower productivity and lower paying jobs (Duraisamy and Duraisamy, 2016) reinforcing gender inequality.

***Figure 1: Hourly earnings***

[1]The 6th round in 2020/21 was a limited phone survey conducted during, and focusing on the impact of the Covid pandemic.

For Peer Review Only

## 4. Methods

### *4.1 Sequence analysis and life history clusters*

Sequence analysis is used to identify four clusters of life history trajectories for 913 sample members from age 8. The trajectories record salient aspects of childhood and family environments which are largely beyond the children's control and theory suggests can impact opportunities in later life: living conditions (quantified by access to amenities), health and exposure to family and economic shocks. Living conditions reflect parental income and ability to provision resources, an important channel through which parents can affect children's opportunities (Roemer and Trannoy, 2015:260-261). Health is condition that "matters for a lifetime" (Roemer and Trannoy, 2015:261) and links between childhood health and adult income are highlighted by Canning (2012). Crivello and Morrow (2019) show how children's fortunes can be transformed by family difficulties and shocks such as a parental death. Our analysis examines how the combined and cumulative effect of these aspects of young peoples' lives impacts their early employment.

Living conditions are captured by states recording whether households have full, some (i.e. limited) or no access to four basic amenities; electricity, piped drinking water, their own (not shared) toilet and adequate fuels (paraffin, kerosene, gas or electricity) for cooking. Health is recorded subjectively (caregiver or self-reported) as either very good/good (better than average), average or very poor/poor (worse than average). Subjective measures of health are well-used indicators and strong predictors of morbidity and mortality (Jylha, 2009). Family shocks include divorce, separation, family death or illness and economic shocks record loss of employment, a source of income or family enterprise. The sequence states record whether the individual experienced no family or economic shocks, a family shock, and/or an economic shock since last surveyed, or in the last three years when first surveyed. The coding process resulted in 27 mutually

For Peer Review Only

exclusive states (3x3x3) recording living conditions, health and exposure to shocks in a single dimension.

The 913 coded sequences were analysed using optimal matching and clustering using the Stata SQ package (Brzinsky-Fay, Kohler, & Luniak, 2006). The optimal matching uses the non-parametric optimal matching algorithm to compare all pairs of sequences by computing minimum distances needed, to turn one sequence into another through substitution or combinations of insertions or deletions (as explained in detail elsewhere e.g. Brzinsky-Fay et al., 2006; Potârcă, et al., 2013, Bukodi et al., 2016). As suggested by Brzinsky-Fay et al. (2006) substitution costs were set at double the insertion/deletion (indel) costs which were set at the default lower value of one as in sequences of equal length indel operations have a limited role (Bukodi et al., 2016). The resulting dissimilarity matrix was subjected to cluster analysis using the commonly used Ward's cluster linkage algorithm. The Duda-Hart Je(2)/Je(1) index suggested four distinct clusters of individuals who had experienced similar pathways in terms of the duration of, and transitions between different states.[2]

Sequence analysis is particularly suited to identifying patterns over life courses as it takes into account the time element (order and duration) in longitudinal data. Other methods such as event history analysis or conditional inference trees and forests either decompose the life-course into specific events or are static, focusing on absolute time points rather than relative timing over life-courses. In optimal matching, sequences are treated as complete entities that capture aspects

[2]A low pseudo-T-squared value, T=0.38, for the four-cluster grouping and higher T-values for the 3 and 5-cluster groupings indicated a jump in the hierarchical clustering. The Calinski/Harabasz pseudo-F index was inconclusive. Further consideration of the cluster hierarchy using cluster dendograms and tests of independence between the clusters and the categorical sequence variables indicated that the four-cluster solution was sufficiently distinct (Figures 2 and 3, Table 1).

of the totality of life-courses. In this study, these aspects are characterised by experiences relating to living conditions, health and shocks.

### *4.2 Regression analysis of employment outcomes*

We conduct regression analysis of employment and earnings in the fifth wave of the survey when the participants were 22. Since the sample for the earnings regressions consisted of only individuals for whom earnings were observed (in the fifth round) estimating the earnings equation without taking account of selection into employment could bias the estimates if those employed are not a random subset of the population. We address sample selection using the Heckman two-stage formulation (Heckman, 1979).

The employment participation equation (1) is estimated as probit model using maximum likelihood. The dependent variable, *Employed,* takes the value one if the individual is employed at age 22 (self-employment or an employee) and takes the value zero otherwise. The estimation is specified as:

$$Pr(Employed_i = 1|X, C,S) = \Phi(\alpha + \sum_{j=1}^{k} \beta_j x_{ji} + \sum_{j=1}^{l} \gamma_j c_{ji} + \sum_{j=1}^{m} \theta_j s_{ji}) \quad (1)$$

$Pr(Employed_i=1)$ is the probability that individual $i$ is employed. $\Phi$ denotes the standard normal cumulative distribution function. $\alpha$ is a constant. The $x_j$, $c_j$, $s_j$ are explanatory variables in the vectors $X, C,S$, the $\beta_j, \gamma_j$, and $\theta_j$ are the respective coefficients. In the first stage of the Heckman formulation, the selection term, the Inverse Mills Ratio (IMR) or Heckman's lambda, is derived from the estimation of (1). In the second stage it is included as an additional explanatory variable in the earnings equation (2) in which the dependent variable, $lny_i$ is the log of either individual hourly or daily earnings in Indian Rupees. The earnings equations are estimated using OLS and specified as:

For Peer Review Only

$$lny_i = \alpha + \sum_{j=1}^{k} \beta_j x_{ji} + \sum_{j=1}^{l} \gamma_j c_{ji} + \mu_i IMR + \varepsilon_i \quad (2)$$

We estimate both equations for the pooled sample and for men and women separately.[3] Average marginal effects are reported for the probit estimation of equation (1) and coefficients for equation (2). We estimate robust variances to account for any intragroup correlation within the 20 local area survey sites.

The $x_j$ are demographic and geographic variables recorded at age 22. The former indicate female gender (in the pooled estimations), age in months, educational attainment (grade completed) and whether the individual has children. In line with human capital theory, educational attainment is expected to be positively associated with earnings but the effect on employment participation is difficult to predict at this young age. Having children is likely to impose a time constraint, particularly for the main carer, and this could affect employment participation and possibly earnings. The geographical controls record employment participation in the 20 local area survey sites, rural or urban location and residence in either Telangana, Coastal Andhra or Rayalaseema. Local area employment rates are indicators of community economic development and individuals in communities with a higher employment should have a higher likelihood of employment (Pagán, 2002) which may impact earnings. Employment opportunities and earnings are also likely to vary between urban and rural areas and across regions and inclusion of these controls helps to account for regional disparities post the 1990s market liberalization reforms (Reddy and Bantilan, 2012).

[3]Three separate estimates of the IMR derived from the pooled, male and female estimations are included in the corresponding estimations of (2).

For Peer Review Only

The $c_j$ explanators record aspects of childhood and family circumstance. First, the categorical variable recording membership of the four different life history clusters derived from the sequence analysis. The clusters group individuals with similar life histories in terms of living conditions, health and experience of shocks. Second, following Sarkar et al. (2026). a control for caste, recording membership of the two most disadvantaged groups, Scheduled Castes (SC) and Scheduled Tribes (ST). Third, variables recording average daily hours of household work and work outside the household at age 12. Household work includes family caring, domestic chores (e.g. cooking) and tasks related to family farming, herding or other family businesses. Work outside the household is undertaken for non-household members. At age 12, most of the participants, 73.47%, undertook some household related work. Fewer, 5.36%, were engaged in work outside the household. Engagement in both kinds of work is expected to raise the propensity for employment by providing work related experience and may impact earnings although the direction of the effect is difficult to predict; child labour may enhance work related skills but also constrains time in activities such as education that could enhance employment opportunities.

The $s_j$ are variables recording other aspects of children's lives that are expected to impact employment participation but not directly impact earnings. Their inclusion in equation (1) but not equation (2) is part of the sample selection specification that defines the selection term, *IMR*, satisfying the exclusion restrictions for the correction. They record the age at which the child started school in grade 1, the number of siblings in the household aged between 0 and 12 when the participants were themselves age 12 and BMI at ages 12 and 8. A later school start age is likely to be associated with later entry into the labour force. The number of young children in the home will impact household resources which could influence employment decisions (Conley et al, 2007). BMI in childhood reflects nutritional disparities. While these variables could influence future

For Peer Review Only

employment participation, they would not be expected to have a direct effect on earnings once in employment. Table A1 provides summary statistics for all variables.

## 5. Results

### *5.1 Life history clusters*

Figure 2 shows the modal sequences for each of the four life history trajectories identified in the optimal matching and clustering. These comprise the most frequent state in each period and show the modal trajectory between the different states over time. Figures 3(a)-3(c) shows how access to amenities, health and shocks evolve over the five waves within the four clusters. Table 1 shows summary statistics for each cluster.

Figures 2-3 and Table 1 identify Cluster 1 (*Better-off*) as the most advantaged. The individuals in this group experienced mostly full access to all four basic amenities (85.86% of observations over all rounds, and 94.88% of cluster members by age 22 compared with 77.21% at age 8). They experienced mostly good health and relatively few shocks (only 28.19% of states recorded shocks). In the modal sequence there is continuous full access to amenities, good health and no experience of shocks. Cluster 2 (*Poorer and unstable)* is the largest cluster and is characterised by the second lowest incidence of full access to amenities (10.42% of observations overall and only 30.92% of cluster members by age 22). The participants in cluster 2 experienced the highest incidence of shocks (47.8% of observations). The high incidence of shocks is highlighted in the modal sequence where family shocks are recorded in three of the five surveyed years. This cluster experienced the lowest incidence of good health (56.66% of observations). Cluster 3 (*Transitioning to better-off*) is characterised by the steepest rise in full access to amenities (from 2.83% of cluster members in round 1 to 33.96% in round 4 and 58.96% by round 5). This transition is clearly illustrated in the trajectory for cluster 3 in Figure 3(a) and in the modal

For Peer Review Only

sequence (Figure 2) which records full access only in the final period. Cluster 4 (*Poorest and stable*) is characterised by the lowest incidence of full access to amenities (5.35% of observations overall and only 17.32% of cluster members by age 22). However, cluster 4 members experienced the lowest incidence of shocks and the highest reported incidence of good health. This is illustrated in the trajectories in Figure 3(c) and in the modal sequence (Figure 2) where there is continuous access to some amenities, good health and no experience of shocks.

Further analysis indicates that employment outcomes and some demographic characteristics vary across the clusters (Supplementary Table A2) Employment participation is highest in the poorer clusters 2 and 4 and lowest in the *Better-off* and *Transitioning to better off* clusters 1 and 3. Earnings are highest in Cluster 1 (*Better-off*). While females are marginally overrepresented in cluster 3 (55.66%) the gender distribution does not differ significantly. However, rural-urban residence is strongly differentiated across the clusters with those in the *Better-off* cluster more likely to reside in urban areas. Region is also a differentiator with those in cluster 3 (*Transitioning to better-off*) significantly more likely to live in Rayalaseema, the poorest and least urbanised region, although cities within the region such as Chittoor are experiencing significant urban growth (Reddy and Bantilan, 2012; Malakar et al. 2022).[4] Educational attainment also varies across the clusters, with cluster 1 the most highly educated. In the next section regression analysis is used to examine whether the differences in employment participation and earnings remain significant when these factors are controlled for.

[4]Urban/rural residence was the primary differentiator of cluster membership in conditional tree analysis (using the Chi-square automated interaction detection decision tree algorithm with Bonferroni adjusted significance testing) where the independent/splitting variables were all those used in the regression analysis After segmenting the data, educational attainment was a differentiator within the urban sample and region a differentiator in the rural sub-sample. Urban/rural residence was also the strongest predictor of cluster membership in a multinomial logit specification. Gender was insignificant in both analyses and gender-cluster interactions were also insignificant in both the employment participation and earnings estimations.

**Figure 2:*Modal Sequence by Cluster***
**Figures 3(a),3(b) 3(c):*Trajectories by cluster***
**Table 1: *Cluster characteristics***

### *5.2 Regression analysis*

Table 2 reports average marginal effects for the probit estimations of the employment participation equation (1). Table 3 reports results for the OLS estimations of the earnings equation (2). For brevity Tables 2 and 3 report results only for gender and the life-history clusters, the main variables of interest (Supplementary Appendix tables A3-A4 show results for the controls).

***Table 2. Employment participation***

#### *5.2.1 Employment*

The average marginal effects in Table 2 confirm that after allowing for other factors, females are still less likely to be employed. The clusters grouping the respondents' life histories all take positive signs. The reference group is cluster 1 (*Better-off)* implying that individuals in less advantaged clusters are more likely to be in paid work at age 22 and this effect is larger and more significant for those in the poorest cluster, cluster 4 (*Poorest but stable)*; on average and compared with cluster 1, the probability of employment is 0.1923 higher for an individual in cluster 4. As the comparator group, cluster 1, is the most advantaged group, these results may reflect limited social protection such that only better-off individuals can remain out of work while waiting for high-paying opportunities (Arbache et. al., 2010). Mitra and Verick (2013) report that in India educated youth from relatively well-off households queue for jobs that match their aspirations, while poor and uneducated youths cannot afford to remain unemployed for long. This is consistent with a kind of "waithood" that Honwana (2012:5) argues is only available to youths able to afford a good

For Peer Review Only

education while depending on their parents. This interpretation is also consistent with the negative significance of grade attained (Table A3).

The results for estimations 2 and 3 in Table 2 show that relative to the *Better-off* cluster, the probability of female employment is significantly higher in each of the other clusters. In the male sample, this is only true for the *Poorest but stable* cluster. This suggests that lower employment participation for those in the *Better-off* cluster is more salient for women which has parallels with evidence from African countries that female unemployment is higher in the richest households while this is less true for men (Arbache et al., 2010).

Results for some included controls are worth noting (Table A3). A later school start age is negatively related to employment participation as predicted. Male but not female employment participation is positively related to having more younger siblings, membership of the more disadvantaged castes (ST and SC) and urban location. Already having a child by age 22 is predictably negatively related to women's employment (90.41% of those with children are female, Table A1).

***About here- Table 3. Earnings***

### *5.2.2 Earnings*

Table 3 shows the earnings estimates for the whole sample and women and men separately. The dependent variables measure either earnings per hour or day of work. The results for the pooled estimations indicate that when employed, young women earn significantly less per hour (30% less in unlogged Rupees) and per day (38% less) than young men after controlling for other factors. In relation to the respondents' life histories, those in clusters 2, 3 and 4 earn significantly less (hourly

For Peer Review Only

and daily) relative to cluster 1 (*Better-off*). Even though the results in Table 2 show that young people in these clusters were more likely to be in work than those in cluster 1 (*Better-off*) once in work, they earn less, reinforcing the earlier conclusion that employment is driven by lack of choice or need.

In the separate estimations for women, the cluster effects on hourly earnings are all negatively significant but the cluster effects on daily earning are only (marginally) significant for cluster 3 (*Transitioning to better-off*) and 4 (*Poorest but stable*). Earnings in cluster 4 (*Poorest but stable*) are predicted to be the lowest (hourly earnings are 63% lower than cluster 1, daily earnings are 57% lower). In contrast, in the estimations for men, only hourly earnings in cluster 2 (*Poorer & unstable*) are lower relative to the *Better-off* cluster while daily earnings are significantly lower in cluster 2 (34% lower) and also(marginally) in clusters 3 and 4.

The IMR is marginally negatively significant in the female estimation predicting hourly earnings (estimation 3). This implies a negative correlation between female employment participation and hourly earnings, consistent with need being a strong driver of employment among women. There is no sample selection into paid employment for men perhaps because the majority, 76.5%, were employed. More generally, links between employment and productivity and, therefore, wages may be less apparent among young workers (McKenzie, 2017).

Some results for the controls are worth highlighting (Table A4). Hours of household work in childhood are negatively linked to women's but not men's hourly and daily earnings consistent with previous research (Carmichael et al., 2023). However, having children has no impact on women's wages once employed but daily earnings are somewhat higher for young fathers. Men, but not women, in the more disadvantaged castes earn significantly less.

For Peer Review Only

Together, these results suggest that earlier life conditions and experiences exert a strong influence on the employment and earnings of young men and women. Differences in employment and earnings are predicted by different life histories, particularly for women. However, it is notable that the daily earnings of young men in the most unstable although not the poorest cluster (*Poorer & unstable*) are the lowest relative to those in the *Better-off* cluster. This suggests that experience of family and economic shocks could particularly disadvantage young men. In contrast, the earnings of young women in the poorest cluster (*Poorest but stable*) are the lowest suggesting that the cumulative impacts of poverty are more damaging for women. Male and female outcomes are both tied to earlier lives but in different ways.

## 6. Discussion

The research makes significant empirical and methodological contributions to the literature analysing the transition to adulthood and employment in developing countries. It does this by examining the different experiences of young men and women in India very early on in their lives and their careers using sequence analysis and regression analysis.

The analysis identifies four clusters grouping individuals sharing similar life trajectories reflecting childhood and family circumstances: *Better-off, Poorer and unstable, Transitioning to better-off* and *Poorest but stable*. The trajectories track differences in living conditions (measured by access to basic amenities e.g. electricity, piped drinking water), health and exposure to family and economic shocks. Young people in the *Better-off* cluster shared the most advantageous life histories characterised by mainly full access to amenities, good health and very few shocks. Those in the *Transitioning to better-off* cluster experienced a sharp increase in access to amenities over their life courses. In contrast, those in the *Poorer and unstable* and *Poorest but stable* clusters had limited access to basic amenities, particularly the latter but with the former experiencing the

For Peer Review Only

highest number of shocks of any cluster. The analysis found that, relative to the most advantaged *Better-off* cluster, other young adults were significantly more likely to be in employment, particularly those in the poorest cluster (*Poorest but stable*). At these young ages employment appears to be driven to greater extent by lack of choice or need, with better-off young people more able to exercise choice. However, once in employment those in the *Better-off* cluster earned significantly more than those in the less advantaged clusters, consistent with theories linking unequal childhood circumstances to inequality of opportunity (Roemer and Trannoy, 2015, 2016).

Even at very early stages in their adult lives, young women in the sample were much less likely to be in work, 48.2% percent of women, compared with 76.4% of men. The World Bank reports a lower employment rate for all women of 32.4% in 2025 and while this figure represents an improvement compared with previous years it continues to be low globally (World Bank, 2026). The analysis showed that women earn less both hourly and daily than men even after controlling for childhood and family circumstances, educational attainment, caste and other factors. There is significant earnings inequality between women, with hourly earnings of women in the more advantaged *Better-off* cluster significantly higher than those of women in all other clusters, particularly the *Poorest but stable* cluster. There is also earnings inequality among men with those in the *Better-off* cluster earn significantly more and earnings lowest among those experiencing the most shocks in the *Poorer & unstable* cluster.

These results confirm that links found between upbringing conditions and material outcomes in developed economies (Kobayashi et al., 2017) are important in India, a lower-middle income country. Furthermore, the employment outcomes of women in the *Better-off* cluster are more strongly differentiated from those of other women, compared with the differentiation among men. This suggests that the impact of family advantage on life chances is more pronounced for

For Peer Review Only

women. Nevertheless, once in employment, the earnings of less advantaged young people, regardless of gender, are lower compared with those from better-off backgrounds. That this relationship is particularly evident for men who have suffered multiple shocks over their life courses, suggests they are the least likely to be compensated for their less advantageous early-life conditions. The underlying causes of this lack of compensation require further research (Currie and Vogl, 2013).

There are limitations of this study linked to the data and methods. The longitudinal data used to capture early-life conditions and experiences, while spanning fifteen years is derived from five survey years which limits the degree of variability in the data. The data used to construct the life histories is recorded from age eight and research has highlighted the formative importance for later life outcomes of life conditions at ages younger than this (Mancini and Yang, 2009; García et al, 2021). The sequence analysis requires choices regarding the number of clusters and the costs used to calculate the distance matrix (Potârcă et al., 2013). Lastly, the mechanisms and interactions underlying the linkages found between early life conditions and later life consequences require further investigation in a developing country context (Currie and Vogl, 2013).

## 7. Conclusion

Our findings have clear policy implications which can inform measures to support the equalisation of employment opportunities among young people in India by addressing disadvantage in childhood and family circumstances. First, the results link employment outcomes in early adulthood to childhood and family circumstances reflected in living conditions, health and exposure to shocks over children's life courses. Our analysis highlights the role of advantage in terms of socioeconomic circumstance for adult earnings and the ability to delay employment participation, consistent with a kind of voluntary "*waithood*" only available to relatively privileged

For Peer Review Only

young people (Honwana, 2012:5). These links between earlier life advantages and more advantageous early-adult outcomes suggests that differences widen over time in line with theories of cumulative (dis)advantage (Dannefer, 2003). Labour market reforms to improve employment prospects in India would, therefore, likely yield higher (and more inclusive) long-term returns for society if they are accompanied by policies aimed at improving the living conditions, health and security of disadvantaged young children. Such policies could help nullify the impact of circumstantial inequalities on outcomes, a goal noted by Roemer and Trannoy (2016:230)

Second, the results confirm the existence of gender inequalities in employment participation and earnings even at relatively young ages. This aligns with Darko and Vasilakos's (2020) argument and findings showing that the effort to reduce inequality in employment opportunities really need to start in earlier life with interventions that narrow gaps in the educational and employment outcomes of boys and girls from different social classes.

Lastly, the analysis suggests the way lives unfold in India are part of the reason for lower female employment participation and earnings. As such, the research highlights the importance of childhood circumstances and gender in adult outcomes. By implication, policy interventions designed to support the economic outcomes of young adults in India and other lower-middle income countries need to consider children's material living conditions and experiences as well as gender differences in access to employment opportunities.

For Peer Review Only

## References


Afridi, F., Dinkelman, T., & Mahajan, K. (2018). Why are fewer married women joining the work force in India? A decomposition analysis over two decades. *Journal of Population Economics*, 31(3):783–818.

Agrawal, T. (2021), Gender segregation and wage differentials in India: the role of educational attainment and occupational choices, *International Journal of Manpower*, 42(1):1-20.

Akresh, R., Lucchetti L. & Thirumurthy, H. (2012). Wars and child health: evidence from the Eritrean-Ethiopian conflict. *Journal of Development Economics*, 99(2):330–40

Arbache, J. S., Kolev, A. and Filipiak, E. (2010). *Gender disparities in Africa's labor market*, Africa Development Forum Series, Washington DC: The World Bank

Bargain, O. and Zeidan, J. (2017). Stature, skills and adult life outcomes: Evidence from Indonesia. *The Journal of Development Studies*, 53(6):873-890.

Bharadwaj, P, Løken K.V., Neilson, C. (2013). Early life health interventions and academic achievement. *American Economic Review*, 103(5):1862–91.

Boye, K. and Grönlund, A. (2018). Workplace Skill Investments–An Early Career Glass Ceiling? Job Complexity and Wages Among Young Professionals in Sweden. *Work, Employment and Society*, 32(2):368-386.

Brzinsky-Fay, C., Kohler U. & Luniak, M. (2006). Sequence analysis with Stata**,** *The Stata Journal* 6(4):435–460.

Bukodi, E., Goldthorpe, J. H., Halpin, B., Waller, L. (2016). Is Education Now Class Destiny? Class Histories across Three British Birth Cohorts, *European Sociological Review*, 32(6):835–849

Canning, D. (2012). Progress in Health around the World. *Journal of Development Studies*, 48(12):1784-1798.

Carmichael, F., Darko, C., Kanji, S. & Vasilakos, N. (2023) The contribution of girls' longer hours in unpaid work to gender gaps in early adult employment: Evidence from Ethiopia, India, Peru and Vietnam, *Feminist Economics*, 29(1):1-37

Carrillo, B. (2017). Early Rainfall Shocks and Later-Life Outcomes: Evidence from Colombia. *The World Bank Economic Review*. 34(1):179-209

Chatterjee,U., Murgai, R., Rama, M. (2015). Job opportunities along the rural-urban gradation and female labor force participation in India*. Policy Research Working Paper* 7412 Washington D.C.:World Bank Group.

Chuta, N. (2014). Children's Agency in Responding to Shocks and Adverse Events in Ethiopia. Young Lives, University of Oxford, www.younglives.org [accessed 21/12/2024]

Conley, D. and Bennett N.G. (2001). Birth weight and income: interactions across generations. *Journal of Health and Social Behavior*, 42:450–465.

Conley, D., Pfeiffer K.M., Velez, M. (2007). Explaining sibling differences in achievement and behavioral outcomes: The importance of within-and between-family factors, *Social Science Research*, 36(3):1087–1104.

Crivello, G. and Morrow, V. (2019). Against the odds: Why some children fare well in the face of adversity. The Journal of Development Studies, 56(5):1-18.

Currie, J., & Vogl, T. (2013). Early-Life Health and Adult Circumstance in Developing Countries, *Annual Review of Economics*, 5(1):1–36

Darko, C.K. and Vasilakos, N. (2020). Picking winners: An empirical analysis of the determinants of educational outcomes in India. *British Educational Research Journal*, *46*(6):1445-1467.


For Peer Review Only

Dannefer, D. (2003). Cumulative advantage/disadvantage and the life course: cross fertilizing age and social science theory, *Journal of Gerontology: Social Sciences*, 58B(6):S327-337.

Devine, K. A., Christen, S., Mulder, R. L. et al. (2022) Recommendations for the surveillance of education and employment outcomes in survivors of childhood, adolescent, and young adult cancer; A report from the international late effects of childhood cancer guideline harmonization group. *Cancer,*128(13):2405-2419.

Desai, S., & Joshi, O. (2019). The paradox of declining female work participation in an era of economic growth: The case of India. *Indian Journal of Labour Economics*, 62(1):55–71.

Duncan, G. J., Morris, P. A., & Rodrigues, C. (2011). Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments. *Developmental Psychology,* 47(5):1263-1279.

Duraisamy, M., & Duraisamy, P. (2016). Gender wage gap across the wage distribution in different segments of the Indian labour market, 1983–2012: exploring the glass ceiling or sticky floor phenomenon. *Applied Economics*, 48(43):4098–4111.

Elder, G.H. (1998), The Life Course as Developmental Theory. *Child Development*, 69(1): 1-12

Favara, M. (2017). Do dreams come true? Aspirations and educational attainments of Ethiopian boys and girls. *Journal of African Economies*, 26(5):561-583.

García, J.L., Heckman, J.J., Leaf D. E., and Prados. M.J. (2020). Quantifying the life-cycle benefits of an influential early-childhood program. *Journal of Political Economy* 128(7):2502-2541.

Hamil-Luker, J., & O'Rand, A.M. (2007). Gender differences in the link between childhood socioeconomic conditions and heart attack risk in adulthood. *Demography*, *44*(1):137-158.

Heckman, J.J. (1979). Sample bias as a specification error. *Econometrica*, 47(1):153–161.

Himaz, R. (2018). Stunting later in childhood and outcomes as a young adult: Evidence from India. *World Development*, *104*:344-357.

Honwana, A. (2012). *The Time of Youth: Work, Social Change, and Politics in Africa*, Boulder CO:Kumarian Press

Jylhä, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science & Medicine*, 69(3):307–316.

Kingdon, G. G., & Unni, J. (2001). Education and women's labour market outcomes in India. *Education Economics*, 9(2), 173–195.

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

Kobayashi, L. C., Glymour, M. M., Kahn, K., Payne, C. F., Wagner, R. G., Montana, L., ... & Berkman, L. F. (2017). Childhood deprivation and later-life cognitive function in a population-based study of older rural South Africans. *Social Science & Medicine*, 190:20-28.

León G. (2010). Civil Conflict and Human Capital Accumulation: The Long Term Effects of Political Violence in Peru. *BREAD Working Paper*, 256.

Linnemayr, S. & Alderman H. (2011). Almost random: evaluating a large-scale randomized nutrition program in the presence of crossover. *Journal of Development Economics,* 96(1):106–114

Maccini. S., Yang, D. (2009). Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall. *American Economic Review,* 99(3):1006-26.

Malakar, K. D., Kuzur, G., Maity D. K., Yadav, M., and Roy, S. (2022). A Socio-Ecological Study of Population, Migration, Urbanization, and Socio-Climate Variation in Andhra Pradesh and Telangana, India. *Journal of Geography, Environment and Earth Science International* 26(12):1-33.

For Peer Review Only

McKenzie, D. (2017). How effective are active labor market policies in developing countries? a critical review of recent evidence. *The World Bank Research Observer*, *32*(2):127-154.

Melchior, M., Moffitt, T. E., Milne, B. J., Poulton, R., & Caspi, A. (2007). Why do children from socioeconomically disadvantaged families suffer from poor health when they reach adulthood? A life-course study. *American Journal of Epidemiology*, 166(8):966-974.

Mitra, A. and Verick, S. (2013) Youth employment and unemployment: an Indian perspective, *ILO Asia-Pacific Working Paper Series*, DWT South Asia and Country Office for India, New Delhi:ILO [Accessed 12/01/2025]https://webapps.ilo.org/public/libdoc/ilo/2013/113B09_171_engl.pdf

Nutz, T. and Lersch, P.M. (2021). Gendered employment trajectories and individual wealth at older ages in Eastern and Western Germany. *Advances in Life Course Research*, *47*, p.100374.

Pagán, J. A. (2002). Gender differences in labor market decisions in rural Guatemala. *Review of Development Economics*, 6(3), 428-441.

Pathania, V. (2009). The long run impact of drought at birth on height of women in rural India. Working Paper, *University of California,* Berkeley.Reddy, A. A., & Bantilan, M. (2013). Regional disparities in Andhra Pradesh, India. *Local Economy: The Journal of the Local Economy Policy Unit*, *28*(1):123-135.

Potârcă, G., Mills, M. and Lesnard, L. (2013). Family formation trajectories in Romania, the Russia Federation and France: Towards the second demographic transition? *European Journal of Population*, 29(1):65-76.

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

Roemer, J. E., and Trannoy, A. (2015). Equality of opportunity, chapter 4 in Atkinson, A. B. & Bourguignon, F., (eds.) *Handbook of Income distribution*, 2A:217-300 http://dx.doi.org/10.1016/B978-0-*444-59428-0.00005*

Sarkar, S., Chakravorty, B., & Lyonette, C. (2025) Caste, gender and intersectionality in labour market aspirations: evidence from a skill-training programme in India, *Oxford Development Studies*, 54(1):22-40.

Van Den Berg, G. J., Lindeboom, M., Portrait, F. (2006). Economic conditions early in life and individual mortality. *American Economic Review* 96(1):290–302.

Vauhkonen, T., Kallio, J., Kauppinen, T.M. and Erola, J. (2017). Intergenerational accumulation of social disadvantages across generations in young adulthood. *Research in Social Stratification and Mobility*, 48:42-52.

Vogl, T. (2012). Height, skills, and labor market outcomes in Mexico. *NBER* Working Paper 18318 http://www.nber.org/papers/w18318

Wiborg, Ø.N. and Møberg, R.J. (2010). Social origin and the risks of disadvantage in Denmark and Norway:the early life course of young adults. *Work, Employment and Society*, *24*(1):105-125.

World Bank (2026) *India gender landscape*, Gender Data Portal. [Accessed 01/05/2026]: https://genderdata.worldbank.org/en/economies/india

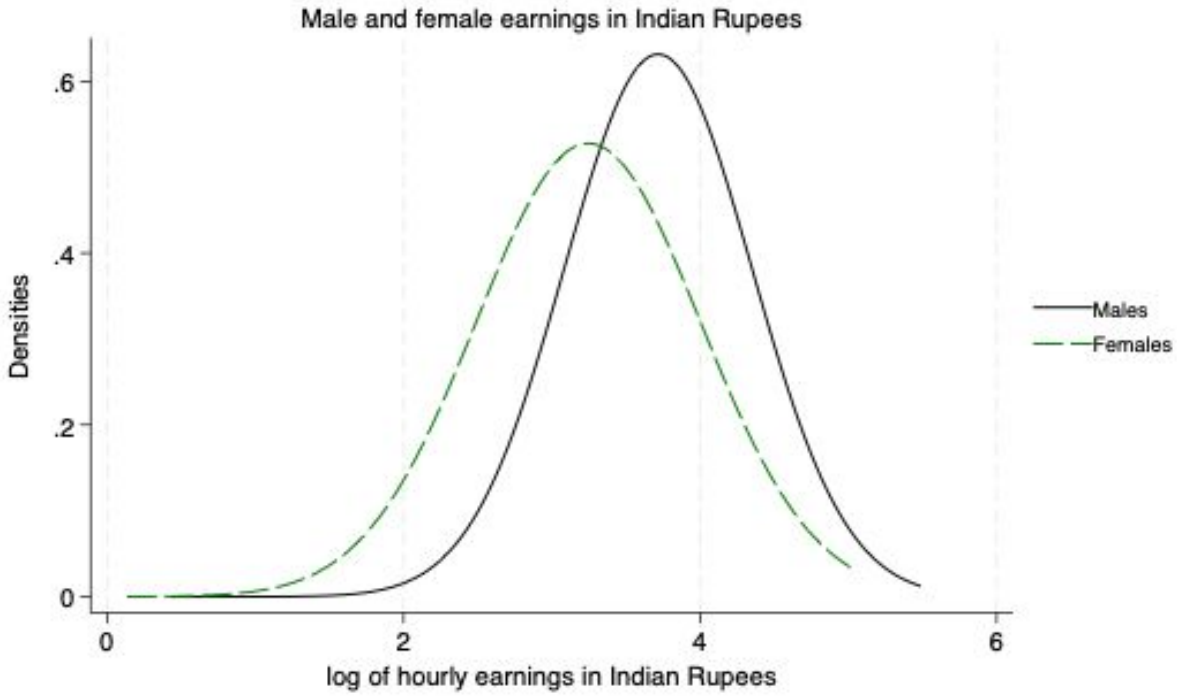
Young Lives (2018) *Survey Design and Sampling Fact sheets*, [Accessed 4/06/18] www.younglives.org.uk

Zou, W., Lybbert, T., Vosti, S. and Abbeddou, S. (2020). Early Childhood Nutrition, Parental Growth Perceptions and Educational Aspirations in Rural Burkina Faso. The Journal of Development Studies, 56(4):798-816.

For Peer Review Only

## Figures and Tables

**Figure 1. Hourly earnings**




For Review Only

**Figure 2: Modal sequence by cluster**

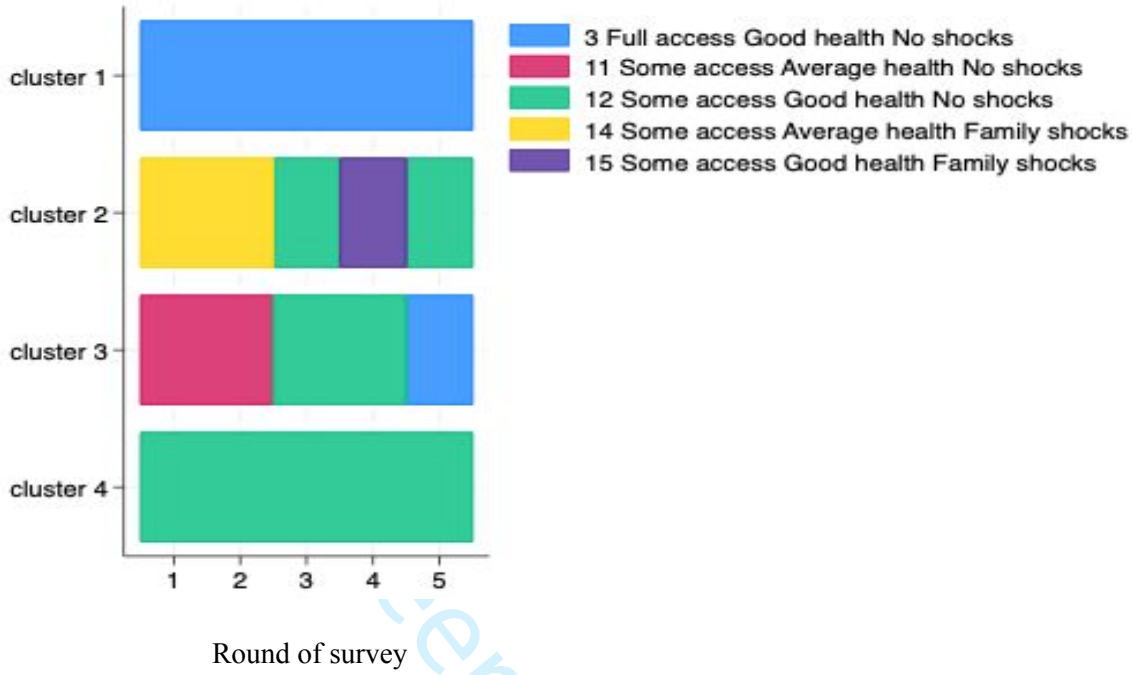



**Figure 3(a): Access to amenities trajectories over the four clusters**

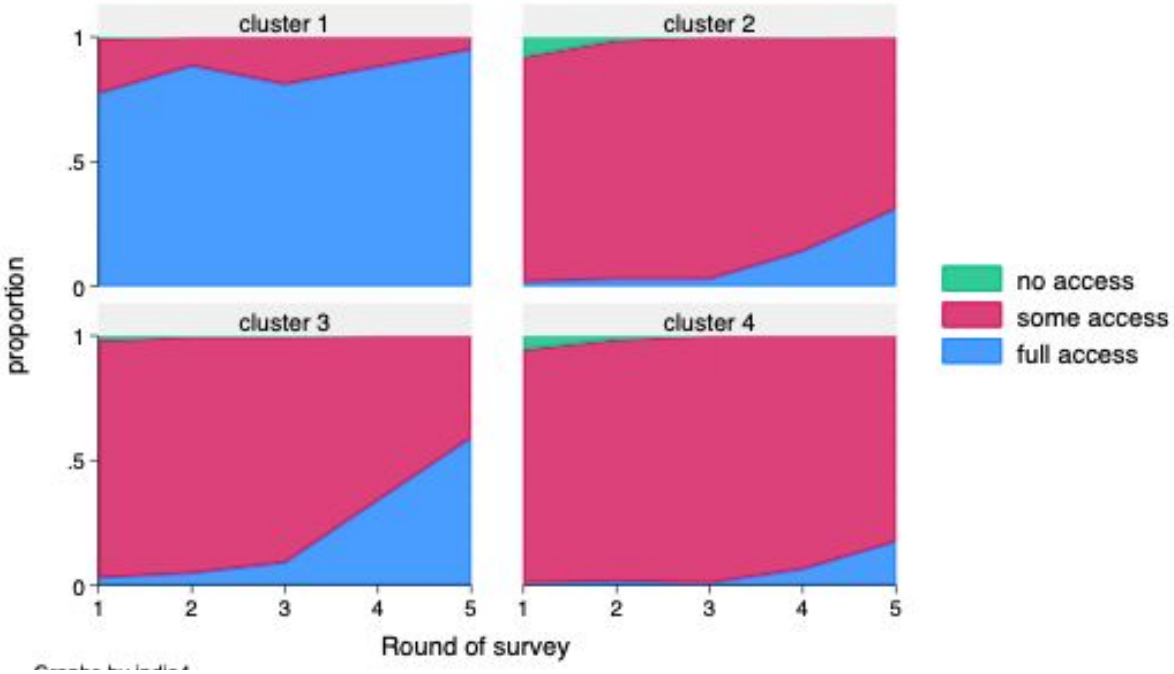


**Figure 3(b): Health trajectories over the four clusters**

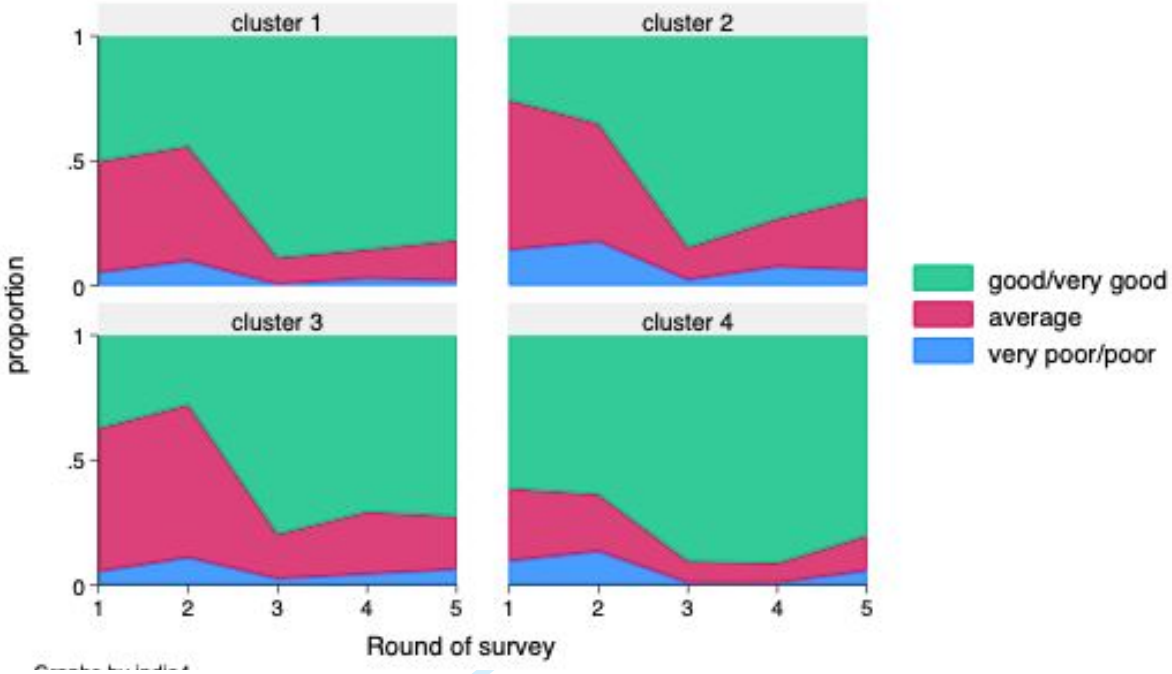



**Figure 3(c): Shock trajectories over the four clusters**

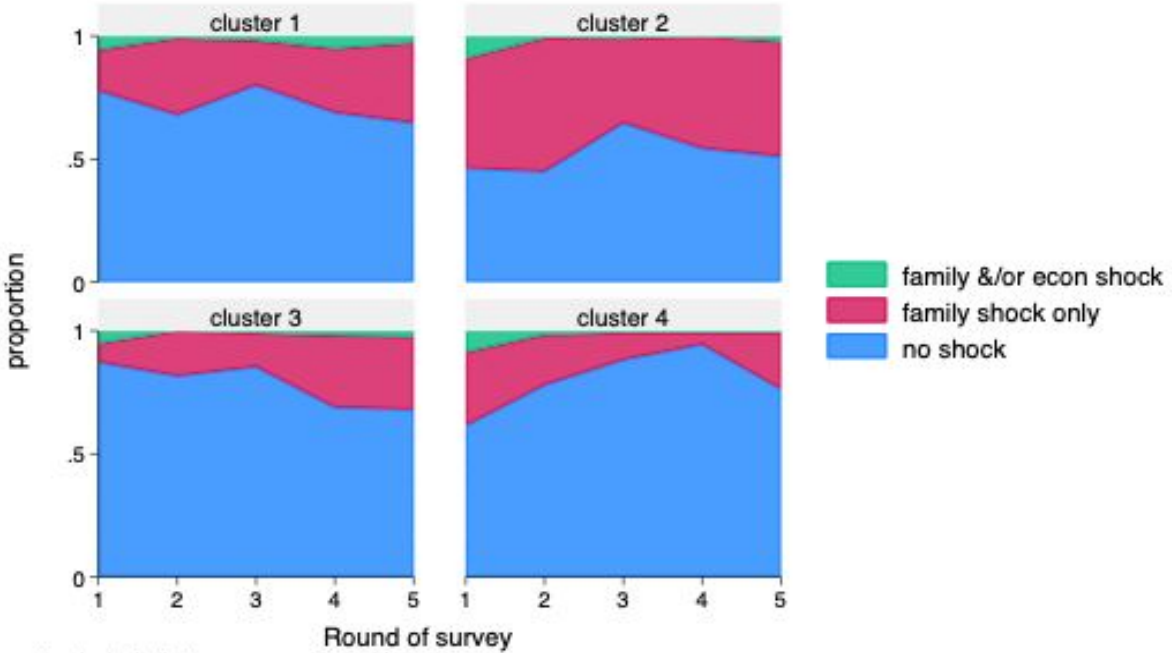


**Table 1. Cluster characteristics**

| *% of sequence states* | Cluster 1 Better-off | Cluster 2 Poorer and unstable | Cluster 3 Transitioning to better off | Cluster 4 Poorest but stable | All |
|---|---|---|---|---|---|
| Access to amenities*** | | | | | |
| % full access | 85.86 | 10.42 | 21.89 | 5.35 | 30.14 |
| % Some access | 14.04 | 87.63 | 77.54 | 93.22 | 68.72 |
| % No access | 0.1 | 1.95 | 0.47 | 1.10 | 1.14 |
| Health*** | | | | | |
| % Very good/good health | 70.14 | 56.66 | 57.64 | 77.48 | 62.96 |
| % Average health | 25.86 | 33.82 | 36.6 | 16.54 | 30.19 |
| % very poor /poor | 4.00 | 9.53 | 5.75 | 5.98 | 6.86 |
| Shocks*** | | | | | |
| % No shocks experienced | 71.81 | 52.2 | 78.21 | 79.69 | 66.68 |
| Number of participants | 215 | 359 | 212 | 127 | 913 |

*** p<0.0: significance in Pearson chi squared tests of independence between clusters and categorical variables recording access to amenities, health and exposure to shocks

**Table 2. Employment participation (probit): Average marginal effects**

| | (1) All | (2) Women | (3) Men |
|---|---|---|---|
| Female | -0.2660*** | | |
| | (0.0302) | | |
| Cluster 1 *Better-off* -reference cluster | | | |
| Cluster 2 *Poorer & unstable* | 0.1087** | 0.1544*** | 0.0862 |
| | (0.0437) | (0.0516) | (0.0807) |
| Cluster 3 *Transitioning to better-off* | 0.1006** | 0.1292*** | 0.1060 |
| | (0.0454) | (0.0491) | (0.0814) |
| Cluster 4 *Poorest but stable* | 0.1923*** | 0.2723*** | 0.1416** |
| | (0.0446) | (0.0596) | (0.0721) |
| Demographic, childhood circumstance & regional controls[a] | YES | YES | YES |
| Observations | 865 | 440 | 425 |
| Pseudo R-squared | 0.262 | 0.18 | 0.296 |
| Wald Chi squared | 1359.72*** | 758.85*** | 465.90*** |
| Log pseudo likelihood | -426.394 | -249.101 | -163.221 |

Figures are average marginal effects on the probability of employment. Robust standard errors in parentheses (clustered at the local area survey site level): *** p<0.01, ** p<0.05, * p<0.1.

[a]Results for controls in Table A3: Controls: age, educational attainment, whether has own children, caste. Employment rate local area site, urban or rural location, region. Hours of work within and outside the household at age 12, age started school, BMI at ages 8 and 12, number of siblings 0-12 years at age 12.

For Peer Review Only

**Table 3. Earnings (OLS): Marginal effects**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Pooled | | Women | | Men | |
| Log of earnings in Rupees | Hourly earnings | Daily earnings | Hourly earnings | Daily earnings | Hourly earnings | Daily earnings |
| Female | -0.361*** | -0.4825*** | | | | |
| | (0.1553) | (0.1595) | | | | |
| Cluster 1 *Better-off* -reference cluster | | | | | | |
| Cluster 2 *Poorer & unstable* | -0.3387** | -0.3997** | -0.6306** | -0.3650 | -0.3176* | -0.4116*** |
| | (0.1569) | (0.1508) | (0.2984) | (0.3644) | (0.1791) | (0.1258) |
| Cluster 3 *Transitioning to better-off* | -0.3546** | -0.3924*** | -0.8696** | -0.7682* | -0.2818 | -0.2885* |
| | (0.1368) | (0.1245) | (0.3357) | (0.4119) | (0.1788) | (0.1581) |
| Cluster 4 *Poorest but stable* | -0.3555** | -0.4097** | -0.9794*** | -0.8539* | -0.2657 | -0.2682* |
| | (0.1440) | (0.1714) | (0.3101) | (0.4432) | (0.1829) | (0.1546) |
| IMR | -0.1551 | -0.2981 | -0.8305* | -0.4742 | -0.2712 | -0.4836 |
| | (0.2747) | (0.2716) | (0.4810) | (0.5289) | (0.2587) | (0.2903) |
| Demographic, childhood circumstance & regional controls[a] | YES | YES | YES | YES | YES | YES |
| Constant | -0.4721 | 1.6122 | -9.4315 | -12.6454 | 3.7098 | 7.0221** |
| | (2.1509) | (2.8830) | (7.3966) | (9.7142) | (2.5948) | (3.3171) |
| Observations | 338 | | 105 | | 233 | |
| R-squared | 0.162 | 0.179 | 0.213 | 0.172 | 0.087 | 0.116 |
| F | 73.53*** | 42.99*** | 33.31*** | 8.92*** | **10.22***** | **7.71***** |

Figures are marginal effects (coefficients). Robust standard errors in parentheses (clustered at the local area level): *** p<0.01, ** p<0.05, * p<0.1.

[a]Results for controls in Table A4: Controls: age, educational attainment, whether has own children. Caste, hours of work within and outside the household at age 12, Employment rate local area site, urban or rural area, region.

For Peer Review Only

# Supplementary Appendix

## Table A1: Summary statistics[a]

| | All | | | Women | | | Men | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | N | Mean | SD | N | Mean | SD | N |
| Women | 0.51 | 0.50 | 916 | | | | | | |
| Employed[b] | 0.62 | 0.49 | 916 | 0.48 | 0.50 | 471 | 0.76 | 0.43 | 445 |
| Hourly earnings (Rupees) | 42.95 | 30.47 | 363 | 32.32 | 26.53 | 118 | 48.06 | 30.97 | 245 |
| Daily earnings (Rupees) | 355.32 | 248.7 | 363 | 255.24 | 228.3 | 118 | 403.53 | 244.1 | 245 |
| Daily earnings USD[c] | 9.261 | 29.98 | 363 | 6.72 | 21.11 | 118 | 10.49 | 33.39 | 245 |
| Age (in months) | 263.98 | 4.20 | 916 | 264.07 | 4.17 | 471 | 263.88 | 4.22 | 445 |
| Educational grade attained | 11.29 | 3.20 | 901 | 10.86 | 3.30 | 465 | 11.76 | 3.02 | 436 |
| SC or ST caste[d] | 0.33 | 0.47 | 916 | 33.33 | 0.47 | 471 | 32.36 | 0.32 | 445 |
| Has own children | 0.24 | 0.43 | 916 | 0.42 | 0.49 | 471 | 0.05 | 0.21 | 445 |
| Daily hours household work age 12 | 1.57 | 1.93 | 915 | 1.87 | 2.00 | 471 | 1.25 | 1.82 | 444 |
| Daily hours work outside home age 12 | 0.37 | 1.70 | 913 | 0.42 | 1.79 | 470 | 0.32 | 1.60 | 443 |
| School age start /grade1 | 5.03 | 0.92 | 907 | 5.00 | 0.92 | 465 | 5.06 | 0.93 | 442 |
| BMI age 12 | 25.74 | 145.4 | 909 | 29.90 | 159.3 | 466 | 21.37 | 129.2 | 443 |
| BMI age 8 | 13.93 | 1.69 | 916 | 13.87 | 1.26 | 471 | 13.99 | 2.04 | 445 |
| No. of siblings age 0-12 (in household) at age 12 | 0.87 | 0.98 | 916 | 0.93 | 0.96 | 471 | 0.81 | 1.00 | 445 |
| Employed rate local area (survey) site | 0.62 | 0.17 | 916 | 0.59 | 0.15 | 471 | 0.65 | 0.18 | 445 |
| Rural location | 0.68 | 0.47 | 905 | 0.65 | 0.48 | 463 | 0.70 | 0.46 | 442 |
| Coastal Andhra | 0.35 | 0.85 | 909 | 34.84 | 0.84 | 465 | 35.14 | 0.85 | 444 |
| Rayalaseema | 0.28 | 0.85 | 909 | 29.25 | 0.84 | 465 | 27.48 | 0.85 | 444 |
| Telangana | 0.37 | 0.85 | 909 | 35.91 | 0.84 | 465 | 37.39 | 0.85 | 444 |

[a]Unless stated, statistics are survey round 5 (age 22).
[b]Either self-employed, in regular, salaried, formal employment, casual, informal employment, a farm servant, housemaid or non-specified employment.
[c]Rupees converted to US Dollars using appropriate daily real exchange rates from Thomson-Reuter Datastream.
[d]Reference is BC (other economically and socially backward classes) or OC (other/open/general castes)

For Peer Review Only

**Table A2 Cluster employment and demographic characteristics (age 22)**

| | Cluster 1 Better-off | Cluster 2 Poorer and unstable | Cluster 3 Transitioning to better off | Cluster 4 Poorest but stable | All |
|---|---|---|---|---|---|
| % employed *** | 39.53 | 68.80 | 63.68 | 76.38 | 61.77 |
| Mean daily earnings (Rupees)*** | 467.406 | 323.292 | 330.066 | 322.896 | 355.287 |
| Mean daily earnings (USD)*** | 17.706 | 6.771 | 6.618 | 6.768 | 9.27 |
| % female participants | 50.23 | 49.86 | 55.66 | 50.39 | 50.08 |
| % living in rural locations*** | 17.6 | 85.67 | 74.76 | 88.89 | 67.52 |
| Regional location*** | | | | | |
| % living in Coastal Andhra | 40.93 | 37.25 | 27.88 | 30.16 | 34.99 |
| % living in Rayalaseema | 22.33 | 20.45 | 45.67 | 31.75 | 28.26 |
| % living in Telangana | 36.74 | 42.30 | 26.44 | 38.10 | 36.75 |
| % post-secondary educational grade*** | 69.77 | 38.44 | 45.75 | 50.39 | 49.18 |

*** p<0.01, ** p<0.05, * p<0.1 significance in Pearson chi squared tests of independence between clusters and categorical variables and ANOVA F and Kruskal-Wallis chi squared tests of differences in cluster means for continuous earnings measures.

**Table A3. Employment participation (probit): Average Marginal effects for control variables not reported in Table 2**

| | (1) All | (3) Women | (2) Men |
|---|---|---|---|
| Age | -0.0009 | 0.0001 | 0.0009 |
| | (0.0040) | (0.0058) | (0.0032) |
| Educational grade attained | -0.0494*** | -0.0381*** | -0.0708*** |
| | (0.0082) | (0.0108) | (0.0134) |
| Has own children | -0.1734*** | -0.1691*** | -0.1700 |
| | (0.0395) | (0.0394) | (0.1994) |
| SC or ST caste | 0.0610** | 0.0219 | 0.1032*** |
| | (0.0243) | (0.0406) | (0.0249) |
| Daily hours household work age 12 | 0.0222* | 0.0280*** | 0.0183 |
| | (0.0123) | (0.0104) | (0.0237) |
| Daily hours work outside household age 12 | -0.0206* | -0.0169 | -0.0222 |
| | (0.0117) | (0.0122) | (0.0206) |
| Age started school/grade 1 | -0.0542*** | -0.0554** | -0.0580*** |
| | (0.0137) | (0.0254) | (0.0138) |
| No of siblings aged 0-12 (age 12) | 0.0259** | 0.0153 | 0.0314** |
| | (0.0113) | (0.0166) | (0.0156) |

For Peer Review Only

| | | | |
|---|---|---|---|
| BMI age 12 | -0.0000 | -0.0001 | -0.0088 |
| | (0.0000) | (0.0000) | (0.0092) |
| BMI age 8 | 0.0037 | 0.0144 | -0.0049 |
| | (0.0099) | (0.0189) | (0.0067) |
| Employed rate local area (survey) site | 0.6617*** | 0.6801*** | 0.7129*** |
| | (0.1174) | (0.1507) | (0.1118) |
| Rural area | -0.0447 | 0.0427 | -0.1442*** |
| | (0.0599) | (0.0620) | (0.0427) |
| Coastal Andhra | -0.0101 | -0.0489 | 0.0419 |
| | (0.0227) | (0.0400) | (0.0379) |
| Rayalaseema | -0.0231 | -0.0262 | -0.0189 |
| | (0.0228) | (0.0429) | (0.0397) |

Robust standard errors in parentheses (clustered at the local area survey site level): *** p<0.01, ** p<0.05, * p<0.1.

**Table A4. Earnings (OLS): Marginal effects for control variables not reported in Table 3**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Pooled | | Women | | Men | |
| Age | 0.0170** | 0.0182 | 0.0726** | 0.0539** | 0.0012 | -0.0022 |
| | (0.0081) | (0.0108) | (0.0346) | (0.0254) | (0.0098) | (0.0124) |
| Educational grade attained | 0.0073 | -0.0120 | -0.0297 | 0.0335 | 0.0136 | 0.0042 |
| | (0.0185) | (0.0184) | (0.0421) | (0.0271) | (0.0212) | (0.0226) |
| Has own children | 0.0170 | 0.0248 | -0.0874 | 0.0997 | 0.2264 | 0.2941* |
| | (0.1260) | (0.1507) | (0.2173) | (0.1719) | (0.1608) | (0.1563) |
| SC or ST caste | -0.1838** | -0.1767* | 0.0578 | -0.0227 | -0.2935** | -0.3364*** |
| | (0.0879) | (0.0925) | (0.1310) | (0.1218) | (0.1059) | (0.1166) |
| Daily hours household work age 12 | -0.0125 | -0.0211 | -0.0390 | -0.0102 | -0.0033 | -0.0094 |
| | (0.0128) | (0.0136) | (0.0256) | (0.0261) | (0.0193) | (0.0192) |
| Daily hours work outside household age 12 | -0.0288 | -0.0392** | -0.1279** | -0.1110** | -0.0094 | -0.0074 |
| | (0.0196) | (0.0184) | (0.0566) | (0.0410) | (0.0213) | (0.0208) |
| Employed rate local area (survey) site | 0.0907 | 0.1088 | 0.4610 | -0.3535 | -0.1817 | -0.2790 |
| | (0.3320) | (0.4623) | (1.1853) | (0.8674) | (0.3404) | (0.3855) |
| Rural area | 0.0535 | -0.0142 | -0.1860 | -0.1834 | 0.1893 | 0.1011 |
| | (0.1192) | (0.0943) | (0.1518) | (0.1578) | (0.1229) | (0.1044) |
| Coastal Andhra | -0.0292 | 0.0479 | -0.1366 | -0.1734 | -0.0300 | 0.0225 |
| | (0.0708) | (0.0827) | (0.2070) | (0.1566) | (0.0684) | (0.0868) |
| Rayalaseema | -0.1152 | -0.1239 | -0.1569 | -0.1290 | -0.1254 | -0.1308 |
| | (0.0852) | (0.0964) | (0.2504) | (0.1490) | (0.0937) | (0.0945) |

Robust standard errors in parentheses (clustered at the local area survey site level): *** p<0.01, ** p<0.05, * p<0.1.

For Peer Review Only

University of East Anglia

22nd of May, 2026

The Editors

Oxford Development Studies

Dear Professor Beall,

**Re: R&R2 Response Letter**

Thank you for the opportunity to revise our paper in response to the reviewers' constructive and helpful comments. We provide detailed responses to each point below. The manuscript has been extensively revised in response to the reviewers' comments, with changes ranging from minor editorial clarifications/additional footnotes to substantial re-estimations, methodological refinements and significant additions to the literature review. To keep the revised manuscript reasonably readable (and preserve our own sanity), we have highlighted only the major substantive revisions in yellow in the tracked-changes version of the paper.

As requested by both reviewers, we have expanded and clarified the discussion of the methodology. In response to Reviewer 2's comments, we now explicitly state that the employment participation equation (1) is estimated using a probit specification and that the reported coefficients are presented as average marginal effects. Following Reviewer 1's suggestion, we have also incorporated the literature on inequality of opportunity and engaged more directly with it. In line with Reviewer 2's recommendations, we have streamlined the regression tables presented in the main text by reporting only the coefficients for the key variables of interest and moving the full set of control-variable estimates to the online supplementary appendix. Finally, in response to Reviewer 1, we have introduced an additional regional control in the analysis, and following Reviewer 2's recommendation, we have also included a control for caste.

To remain within the journal's word limit, we have aimed to incorporate the additional explanations and discussion requested by the reviewers as concisely as possible. For the same reason, we have also kept the number of additional references to a minimum. We hope that the revisions we have made satisfactorily address the reviewers' comments and meet both your and the reviewers' expectations.

Yours sincerely,

Professor Nicholas Vasilakos (corresponding author), University of East Anglia

*(co-signed by Professor Fiona Carmichael and Dr Christian Darko, University of Birmingham).*

For Peer Review Only

**Response to Referee Comments**

**Comment.** Dear Reviewers, thank you for taking the time to review and reflect on our manuscript. Your constructive comments, ideas and suggestions have greatly helped us improve our work. We are very conscious of the time and effort involved in conducting these reviews and would like to express our sincere gratitude for your thoughtful feedback and careful consideration of the manuscript. Please find below our point-by-point responses to the issues raised.

## Referee 1

**Comments to the Author**

*I am very grateful for the opportunity to review the manuscript titled "Unequal from the start? Children's life histories and their employment outcomes in India". I find it very interesting and relevant for the field of development studies.*

**Response:** Thank you very much for your positive comments and helpful suggestions. We are grateful for your review.

1. *One way of strengthening the theoretical framework of the paper is by referring to the literature on inequality of opportunity (see Roemer, 1998, or Roemer and Trannoy, 2015 for a survey). The analysis focuses on shocks that occurred during the respondents' childhood and teenage years and, thus, can be conceptualized as circumstances under this framework. Discussing the results under this approach will strengthen one of the paper's strongest arguments: the cumulative effects of shocks on life trajectories.*

**Response:** Thank you very much for recommending that we refer to the literature on inequality of opportunity. We have edited the introduction, methods and discussion sections and restructured the research background section, drawing on this literature. We refer directly throughout to ideas in this literature linking childhood and family circumstances to inequality of opportunity, and to policies that aim to nullify the effects of circumstances on outcomes. In particular, the research background section cites Roemer and Trannoy (2015, 2016) to strengthen the discussion of the theoretical links between early-life circumstances beyond an individual's control and outcomes in young adulthood. We have reemphasised these arguments in the methods section to strengthen the rationale for the empirical modelling, and we have also respecified the equations to better align with this literature by using similar terminology (e.g. using C for the vector of circumstances, as in equation (10) in Roemer and Trannoy

For Peer Review Only

(2016:1314)). We would have liked to have included additional references to this literature, but we have reached the journal's word limit.

2. *The methodology used to identify life trajectories should be discussed in reference to newer classification methods, such as conditional trees and random forests. The aim of this discussion is for the authors to explain their methodological decisions to the reader in a more transparent way.*

**Response:** We have strengthened the methodology section by discussing the advantages of sequence analysis relative to these and other alternative methods. At the end of the section, we note that sequence analysis is better suited to identifying patterns over the life course than event history analysis or conditional tree/random forests, as sequence analysis takes into account the time element in the data, specifically the order and duration of experiences (in our case, those relating to living conditions, health and shocks). Clusters derived from conditional trees/random forests are static in the sense that they don't model the timing of experiences.

We think that while conditional trees/forests analysis is not a substitute for sequence analysis, it can be a complementary tool for exploring the socio-demographic profiles of the people in the clusters we identified. We would normally use multinomial logit for this, but tree analysis offers an interesting alternative that highlights hierarchical structures. We implemented Luchman's Chi-square automated interaction detection (CHAID) command in Stata, which uses Bonferroni-adjusted *p*-values to control for multiple comparisons. The algorithm uses logistic models and is suitable for tree analysis with a categorical dependent variable with more than two categories, which some of the other algorithms are not (e.g. crtrees). The analysis suggested that our variable defining residential area as either urban or rural was a significant differentiator, with cluster 1 being significantly more urban than the other clusters. Once the rural-urban divide was identified as a differentiator, no further tree branches were identified. However, after segmenting the data, educational attainment was identified as a differentiator within the urban sample, and region as a differentiator in the rural sub-sample. The alternative multinomial logit specification confirmed the importance of the urban-rural divide, education and region, which are all controlled for in all the estimated regressions. We have provided a necessarily brief summary of this exploration in a footnote to the paragraph at the end of section 5.1 where we discuss the demographic characteristics of the people in the four clusters, referencing Appendix Table A2 (footnote 4).

3. *An issue that needs to be more thoroughly discussed is how all the life trajectories identified*

For Peer Review Only

*were also exposed to a similar event (the structural change of the Indian economy after its opening in the nineties) that took different forms depending on the region the person lived in, and altered the composition of demand in the different labor markets.*

**Response:** We previously controlled for residential location by including a variable indicating whether the individual lived in an urban or rural area. The regressions also included a variable controlling for the employment rate in the 20 local area survey sites (Young Lives, 2017). We also employed the clustered sandwich estimator to estimate robust variances, allowing for intragroup correlation across the local area survey sites. In response to your comments, we now also include an additional categorical variable which stratifies the sample at a higher regional level within the previously unified state of Andhra Pradesh: the state of Telangana and two regions within the state of New Andhra Pradesh, Coastal Andhra and Rayalaseema (the Young Lives Indian study samples from the previously unified state of Andhra Pradesh bifurcation took place in 2014). In the methods section, in response to your comment, we note your point that the inclusion of controls for region helps control for regional disparities post-reforms (citing Reddy and Bantilan, 2013). We also highlight regional differences across the clusters in Table A2 and the related discussion.

Additional References

Reddy, A. A., & Bantilan, M. (2013). Regional disparities in Andhra Pradesh, India. *Local Economy: The Journal of the Local Economy Policy Unit*, *28*(1), 123-135.

Roemer, J. E., and Trannoy, A. (2015). Equality of opportunity, chapter 4 in Atkinson, A. B. & Bourguignon, F., (eds.) *Handbook of Income distribution*, 2A:217-300

Young lives (2017) Young Lives Survey Design and Sampling (Round 5) United Andhra Pradesh. Younglives-india.org.uk. https://www.younglives.org.uk/sites/default/files/migrated/INDIA-SurveyDesign-Factsheet-Oct17_0.pdf

**Referee 2**

**Comments to the Author**

1. *My main concern relates to the estimation strategy. Identification is essentially based on OLS-type regressions that control for gender, four life-history clusters (which appear to be derived from all five survey rounds), and a set of additional control variables.*

**Response:** Thank you for bringing this lack of clarity in our model specifications to our attention. We have edited the methods section to clarify that the employment participation

For Peer Review Only

estimate uses a non-linear probit model, with the dependent variable being a binary indicator of employment status. This has been clarified by rewriting equation (1) to highlight the nonlinearity of the estimation. We have also made it clearer in the methods and results sections that the reported results for the probit estimation (Table 2) are average marginal effects not coefficients. We should have made this clearer, especially since the use of a probit specification is integral to our analysis, as we control for selection effects in the earnings equation. This is achieved by including the Inverse Mills Ratio, which can only be derived from a probit estimation of the employment participation equation. Thank you for drawing our attention to this lack of clarity.

We have also made it clearer in the methods section that for both estimations, we use the clustered sandwich estimator to estimate robust variances, allowing for intragroup correlation across the 20 local-area survey sites. In the original version of the paper, this information was only provided in the table notes.

*2. Given the paper's emphasis on gendered life-course effects, it would be useful to allow the effect of gender to vary across life-history clusters by interacting gender with the cluster indicators, rather than estimating pooled gender effects or relying solely on separate regressions by gender.*

**Response:** We are sorry that we did not make it clearer in the paper that employment participation is estimated using a non-linear probit model and the results presented are average marginal effects (as you recommend). We have rectified this lack of clarity as explained in response to your first and third points.

Thank you also for the helpful suggestion to consider how gender effects might vary across clusters. It is interesting to consider how gender differences in employment participation may be moderated by different childhood and family circumstances. While the separate estimations for men and women suggest that linkages between cluster membership and employment vary for men and women, you are right that they do not say anything about within-cluster gender differences in these linkages. To address your point, we took steps to investigate the significance of within-cluster gender differences. In the non-linear probit estimation of the employment participation equation, the marginal effects are conditional on the values of all other included variables remaining constant, and it is not straightforward to test for a moderating influence of cluster membership on gender differences (or vice versa) using interaction terms. Separate marginal effects for these terms are not estimated since the values of the interacted variables cannot both be held constant. To explore the possibility of cluster-

For Peer Review Only

specific gender effects, we first estimated the employment participation equation with the interactions. While marginal effects are not reported for interaction terms, coefficients are reported. None of the reported interaction coefficients was significant, and their inclusion did not affect the reported marginal effects for either gender or the clusters. Second, to test if any of the within-cluster gender differences were significant, we calculated the marginal effects of gender for each cluster independently. The procedure tests for any significant differences in the marginal effect of gender between the clusters. In the pairwise tests, we used the Bonferroni option to adjust for multiple comparisons. The results, as shown below, indicated no significant differences in the marginal effect of gender across the four clusters.

| | Contrast | Delta-method | Bonferroni | | Bonferroni | |
|---|---|---|---|---|---|---|
| | | dy/dx | std. err. | z | P>\|z\| | [95% CI] |
| 1.gender =male (base outcome) | | | | | | |
| 2.gender = female | | | | | | |
| cluster 2 vs cluster 1 | .0681573 | .0722228 | 0.94 | 1.000 | -.122385 | .2586995 |
| cluster 3 vs cluster 1 | .0495913 | .0929414 | 0.53 | 1.000 | -.1956121 | .2947947 |
| cluster 4 vs cluster 1 | .1535275 | .0790196 | 1.94 | 0.312 | -.0549467 | .3620016 |
| cluster 3 vs cluster 2 | -.018566 | .0682873 | -0.27 | 1.000 | -.1987255 | .1615935 |
| cluster 4 vs cluster 2 | .0853702 | .0778229 | 1.10 | 1.000 | -.1199467 | .2906871 |
| cluster 4 vs cluster 3 | .1039362 | .0847436 | 1.23 | 1.000 | -.1196391 | .3275116 |

On the basis of these results, we concluded that within cluster gender effects were not significantly different in the employment participation estimation and have not include these interaction terms in the reported estimations, retaining the separate probit estimations for men and women. We also tried including the cluster-gender interactions in the earnings estimation. They were not significant in this either and therefore we have not included them in the reported results. As a further test, we conducted Oaxaca-Blinder decompositions of the earnings estimations and the results for these indicated that differences in cluster coefficients or membership did not contribute to the gender earnings gap.

The lack of significance of the interactions between gender and the life history clusters in the estimated models is noted in footnote 4. This finding may, in some ways, mirror the lack of significance of gender in the profiles of the cluster members. While the figures in Table A2 show that females are slightly overrepresented in cluster 3, the gender differences between clusters are not statistically significant, as clarified in the paper. We also experimented with conditional decision tree analysis to explore the socio-demographic profiles of people within the clusters. To do this, we used Luchman's Chi-square automated interaction detection (CHAID) command in Stata, which uses Bonferroni-adjusted p-values to control for multiple

For Peer Review Only

comparisons. The algorithm uses logistic models and is suitable for tree analysis with a categorical dependent variable with more than two categories, which some of the other algorithms are not (e.g. crtrees). In this analysis, the dependent variable was the categorical variable identifying cluster membership and the independent/splitting variables were all the other independent variables used in the regression analysis. The analysis did not identify gender as a predictor of cluster membership. Neither was gender significant in a multinomial regression estimated to explore cluster membership (with the categorical cluster variable as the dependent variable). These findings (summarised in Table A2 and footnote 4) suggest that gender is not a significant differentiator between the clusters which could partly explain the lack of difference in the effect of gender within clusters. Unfortunately, we could only include a summary discussion of these investigations in the article due to journal word limits (we are already at the limit) – however, we believe that the level of detail provided is sufficient for readers to understand the methodological steps undertaken in the paper.

*3. Table 2 reports probit coefficients rather than marginal effects. As such, the estimated coefficients cannot be interpreted directly as changes in employment probabilities. Given that the discussion in the text occasionally adopts a probabilistic interpretation, it would be helpful to clarify this distinction and, ideally, to report average marginal effects to facilitate interpretation.*

**Response:** Thank you for identifying this lack of clarity. We have edited the methods section (e.g. equation 1) and the results section to make it clearer that Table 2 reports average marginal effects rather than coefficients. The title to Table 2 now includes the words 'Average marginal effects' and a note to the table provides further clarification.

*4. The empirical strategy relies heavily on the life-history clusters, yet the paper provides limited detail on how these clusters are constructed. Additional information on the clustering methodology (including variable selection, scaling, choice of number of clusters, and robustness checks) is necessary to assess the validity and interpretability of the results.*

**Response:** The methodology section now provides a more detailed, though necessarily concise explanation of how the sequence analysis was conducted (referring to the optimal matching algorithm and the clustering procedures and cluster stopping rules). We also reference earlier papers that describe these procedures in detail and briefly summarise the advantages of using sequence analysis with longitudinal data in analyses where timing and duration are important. Further details regarding the operationalisation of the Duda-Hart Je(2)/Je(1) index stopping

For Peer Review Only

rule and other procedures used to support the four-cluster solution are briefly summarised in footnote 2. We also used conditional tree analysis using Luchman's Chi-square automated interaction detection algorithm to check the distinctiveness of the clusters (the dependent). In this tree analysis, the independent/splitting variables were those used to construct the sequences. The analysis predicted no further partitions in the data, beyond the four clusters identified in the sequence analysis, supporting their distinctiveness. We have not mentioned this analysis in the paper as we are not sure that it adds much information but could briefly summarise in footnote 2 if recommended.

We have additionally restructured the research background section and edited the methods section to provide a stronger rationale for selecting the variables used to construct the sequences, drawing on literature on inequality of opportunity (recommended by another reviewer) and additional evidence. For example, we highlight how theories of opportunity equalisation draw links between adult outcomes and aspects of an individual's childhood and family circumstances, such as health and parental income, that are, crucially, beyond their control, reinforcing inequalities (sections 2 and 4.1). Scaling was not used in the analysis as no variables had a particularly large magnitude. However, earnings were logged as is standard.

*5. While age, educational attainment, and other individual characteristics are relevant covariates, I view these primarily as control variables rather than objects of substantive interest. It would therefore be preferable to keep the main tables focused on the key coefficients of interest (gender and life-history clusters), and to report specifications including additional controls in an appendix as robustness checks.*

**Response:** Thank you very much for this suggestion. We have edited the tables in the main text of the paper as you suggested. They now only include the results for the main variables. The results for the additional control variables are now in the supplementary Appendix. We think this edit has significantly improved the paper's readability.

*6. I also recommend looking at Sarkar, S., Chakravorty, B., & Lyonette, C. (2025), which examines caste, gender, and intersectionality in labour market aspirations using evidence from a skill-training programme in India.*

**Response:** Thank you for this suggestion. The evidence in the paper relating to gender, caste and employment aspirations among young people in India is insightful. We have cited the article and, in response, now include a categorical variable to control for membership of the most disadvantaged castes, SC or ST, in all regressions. In the results section, we have noted

For Peer Review Only

that in our estimations SC/ST caste is only (negatively) significant for young men. We also investigated the inclusion of an interaction between caste and female gender in the estimations, following Sarkar et al. (2025). It was insignificant in the employment participation probit and the hourly earnings estimation, and only marginally (positively) significant at the 10% level (p=0.074) level in the daily earnings estimation, suggesting (following your interpretation above) that the gender wage gap for daily earnings is somewhat narrower within the more disadvantaged cases. We have not reported these results in the current version of the paper. On balance, we thought that including the variable recording SC/ST caste and noting the negative significance of caste in the male estimations and insignificance in the female estimations was sufficiently informative in the context of the paper overall. The results for the other variables are not really affected.

For Peer Review Only