Data Appendix

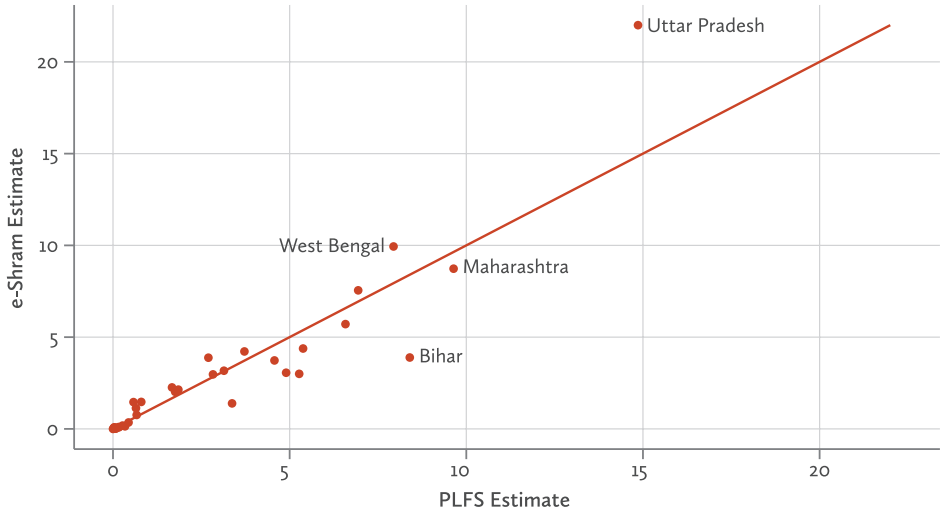
Chapter 2

Comparing e-Shram with PLFS

In contrast to the 29 crore workers recorded in e-Shram, as per estimates using the PLFS, the estimated count of potential e-Shram eligible workers is around 39 crore, implying approximately 75% coverage by the e-Shram portal (notwithstanding inclusion and exclusion errors).

A comparison of state-wise worker shares in PLFS and e-Shram reveals two broad patterns. Most states lie close to the 45-degree line, indicating broad rank-order consistency between the two sources, while a small number of large states emerge as clear outliers. In particular, UP and Bihar account for substantially larger shares in e-Shram than in PLFS, while several southern and smaller states are modestly under-represented.

Figure 1: State wise comparison : PLFS vs e-Shram



Sources and notes: PLFS and e-Shram

Chapter 4

AISHE

The All India Survey on Higher Education (AISHE) is an important nationwide annual survey initiated in 2010-11 by the Ministry of Education, Government of India. The primary objective of this extensive survey is to develop a comprehensive and reliable database on India's diverse and expanding higher education system.

It is important to note that AISHE is **not a sample survey**. Rather, it involves self-reporting by all institutions imparting higher education. Higher education is defined as education obtained after completing 12 years of schooling and of a duration of at least nine months or after completing 10 years of schooling and of a duration of at least three years. The AISHE survey encompasses the following institutions:

- University & University Level Institutions: These are institutions which are empowered to award degrees under some Act of Parliament or State Legislature.

- Colleges: Colleges are affiliated/recognised with a University and are not empowered to provide a degree in its own name.

- Stand-alone Institutions: Stand-alone institutions are institutions which are not affiliated with Universities and are not empowered to provide degrees. They run Diploma Level Programmes (such as polytechnics, teacher training institutes, etc.) offering programs classified as higher education.

AISHE collects information on a range of indicators for each college including number of students enrolled, number of teachers, level of education (UG, PG, PhD, etc.),teaching and non-teaching staff, infrastructure and facilities, programmes offered, caste and gender background of students and teachers etc. Since reporting is on a voluntary basis, information available under AISHE pertains to institutions registered in the AISHE database who report in that year. Typically, a year on year non-response rate of around 9–10% is observed, particularly among affiliated and constituent colleges.There is an improvement in response rate over the years, for example a 61% response rate from the sample of colleges in 2010-11 to 94.2% in the year 2022.[1] To maintain the consistency and reliability of the data, imputation techniques are applied for non-responding institutions especially for enrolment figures at the undergraduate (UG) and postgraduate (PG) levels.

Table 1: Response Rate over the years

Category	University	Colleges	Stand-alone
Total	642	34,908	11,356
Response in AISHE 2011-12	464 (72%)	16,021 (46%)	4,654 (41%)
Total Institutions after pooling (2010-11)	**601 (94%)**	**21,158 (61%)**	**6,702 (59%)**
Total	667	35,525	11,565
Response in AISHE 2012-13	656 (98%)	25,138 (71%)	5,749 (50%)
Total Institutions after pooling (2011-12)	**660 (99%)**	**27,345 (77%)**	**6,880 (59%)**
Total	723	36,634	11,664
Response in AISHE 2013-14	676 (93%)	27,916 (76%)	5,897 (51%)
Total Institutions after pooling (2012-13)	**702 (97%)**	**29,330 (80%)**	**6,860 (59%)**
Listed for AISHE 2014-15	760	38,498	12,276
Response in AISHE 2014-15	724 (95.3%)	33,187 (86.2%)	7,056 (54.5%)
Total Institutions after pooling (2012-13 & 2013-14)	**740 (97.3%)**	**34,452 (89.5%)**	**7,627 (62.1%)**
Listed for AISHE 2015-16	799	39071	11923
Response in AISHE 2015-16	754 (94.4%)	33903 (86.8%)	7154 (60%)
Total number of Institutions after pooling data from AISHE 2013-14 and AISHE 2014-15	**774 (96.9%)**	**35667 (91.3%)**	**7915 (66.4%)**

[Table continued on next page]

Sources and notes: Various AISHE reports

Table 1 [contd.]: Response Rate over the years

Category	University	Colleges	Stand-alone
Listed for AISHE 2016-17	864	40026	11669
Response in AISHE 2016-17	795 (92.01%)	34193 (85.42%)	7496 (64.2%)
Total number of Institutions after pooling data from AISHE 2014-15 and AISHE 2015-16	**835 (96.6%)**	**36852 (92.1%)**	**8453 (72.4%)**
Listed for AISHE 2017-18	903	39050	10011
Actual Response in AISHE 2017-18	828 (91.7%)	34628 (88.7%)	7854 (78.5%)
Total number of Institutions after pooling data from AISHE 2013-14 to AISHE 2016-17	**882 (98%)**	**38061 (97.5%)**	**9090 (90.8%)**
Listed for AISHE 2018-19	993	39931	10725
Actual Response in AISHE 2018-19	944 (95.1%)	36308 (91%)	8354 (77.9%)
Total number of Institutions after pooling data from AISHE 2016-17 to AISHE 2017-18	**962 (96.9%)**	**38179 (95.6%)**	**9190 (85.7%)**
Listed for AISHE 2019-20	1043	42343	11779
Actual Response in AISHE 2019-20	993 (95.2%)	38102 (90.0%)	8631 (73.3%)
Total number of Institutions after pooling data from AISHE 2017-18 to AISHE 2018-19	**1019 (97.7%)**	**39955 (94.4%)**	**9599 (81.5%**
Listed for AISHE 2020-21	1113	43796	11296
Actual Response in AISHE 2020-21	1085 (97.5%)	40212 (91.8%)	8696 (77%)
Total number of Institutions after pooling data from AISHE 2018-19 and AISHE 2019-20	**1099 (98.7%)**	**41600 (95%)**	**10307 (91.2%)**
Listed for AISHE 2021-22	1168	45473	12002
Actual Response in AISHE 2021-22	1154 (98.8%)	38886 (85.5%)	9163 (76.3%)
Total number of Institutions after pooling data from AISHE 2019-20 and AISHE 2020-21	**1162 (99.5%)**	**42825 (94.2%)**	**10576 (88.1%)**

Sources and notes: Various AISHE reports

UDAYA Data

Understanding the Lives of Adolescents and Young Adults (UDAYA) is a large-scale longitudinal survey conducted by the Population Council of India in the states of Uttar Pradesh and Bihar. The first round of the survey was conducted in September 2015 to January 2016 in Uttar Pradesh and January 2016 to July 2016 in Bihar. The survey focused on five groups of adolescents - unmarried adolescent girls and boys aged 10-14 and 15-19 and married girls aged 15-19 - who are the primary respondents. The second round (wave 2) was conducted about three years later between 2018 and 2019 by which time these groups would be aged 13-17 and 18-22, respectively. Along with a household questionnaire administered to the household head or any adult in the household, the identified adolescent within the household is administered an individual level questionnaire. While many themes are covered, we are particularly interested in choices and aspirations for education, vocational training and employment. Individual learning levels are also assessed.

Using the 2011 Census as a frame, the survey followed a multi-stage systematic sampling design with 150 randomly selected primary sampling units comprising villages in rural areas and census wards in urban areas (Paul et al 2023). With the use of the appropriate sampling weights, the survey is intended to be representative at the state level.

Since we are specifically interested in understanding the link between education, aspirations and employment, we focus on 15-19 year olds who are at the school leaving stage during the two waves of the survey. We further introduce an additional categorization to identify initially unmarried girls who get married between the two waves. Since aspirations, educational attainment and school to work transition are likely to systematically differ between married and unmarried girls of the same age, we split the unmarried 15-19 girls to two groups - ones that remained unmarried, and ones that married in the interim.

Table 2: Final sample shares by age groups

Age category	Unweighted sample	Weighted share
UM15-19	2,716	20.33
UF15-19_U	4,617	33.57
UF15-19_M	1,551	12.6
MF15-19	4,257	33.49
Total	**13,141**	**100**

Notes: UM - Unmarried male, UF - unmarried female, UF_M - unmarried female, married between the two waves, MF - married female.

Household Social Consumption: Education Survey

For analysis of stream choices and the costs of enrolling in higher education, we use data from the three rounds of 'Household Social Consumption: Education' conducted by the National Sample Survey Office (NSSO). These surveys collect data on enrolment status, educational level, type of institution, along with household-level indicators such as social groups, religion, consumption expenditure, place of residence etc. Within the households, information about household members' gender, age, and educational enrollment. For enrolled individuals, additional details pertaining to their course enrollment, subject/stream selection, and various educational expenses are available. We use the latest three rounds of the Education consumption – 64th Round (2007-08); 71st Round (2014) and 75th Round (2017-18). Across the three rounds, data for 340,908 individuals aged between 3 and 35 and enrolled in any course is available.

Estimation of Enrolment: The NSS data reports information about household members aged between 3 and 35 years for those who are enrolled in education courses as well as those who are not enrolled. Using this information, the Gross Enrolment Rate (GER) is computed as the proportion of individuals in the 18–23 age group who are enrolled in undergraduate-level courses.

Estimation of Education expenditure for academic year: For each observation, information about the level of enrollment and course are available. Also, respondents report the expenses incurred towards education including the costs related to course fee, books, stationery, transport, exam fee etc. A summation of all these components is reported as the annual cost of the course. We estimate the mean annual education expense incurred from the reported expenses by specific courses and the level of the course.

Gender Gaps in stream choice: To estimate gender gaps in stream choice, we use a regression model where the outcome variable is an indicator for a specific course. The variable takes the value 1 if student *i* from household *h* is enrolled in the specified undergraduate course, and 0 if the student is enrolled in any other course.

The key explanatory variable is an indicator variable for female students (that takes the value 1 if the student is female, 0 otherwise). We also include indicators for social groups that take the value 1 if student *i* belongs to social group *j* ($j \in \{OBC, SC, ST\}$), with students belonging to the General (GEN) category serving as the reference group. In addition, we control for the household's real monthly per capita consumption expenditure (MPCE).

Separate regressions are estimated for each course category, such as engineering, medicine, science, and humanities. The estimated coefficients can be interpreted as the likelihood of female students choosing a particular course relative to male students.

Endnotes

1 Data Appendix Table 1 gives the year on year response rate by type of institutions, as reported by AISHE.

Methods & Results Appendix

Chapter 2

States' salaried employment share growth vis-a-vis economic growth

The first year for which detailed unit-level employment data is available from the employment surveys corresponds to the NSS 38th round: Employment & Unemployment Survey (EUS), 1983. The last EUS round came out in 2011, between 1983 and 2011 there were four more rounds in 1987, 1993, 1999 & 2004. EUS was replaced by the annual Periodic Labour Force Survey in 2017. Since 2017, there have been six more rounds, the latest one for the year 2023-24.

In the analysis below, the age group considered is 25–29. We exclude the 20-24 year olds as some may still be in education. Thus, the estimates reflect the share of the salaried youth population within the 25–29 age group, calculated separately for men and women. The index is computed using the following formula:
Salaried Index for period i = Salaried share in period i / Salaried share in 1983

Net State Domestic Product (NSDP) per capita information is sourced from the RBI's Database on Indian Economy. The numbers are in constant terms, in 2011prices. For the long run analysis states created since 2000 were considered as part of earlier existing states. To calculate per capita NSDP of these states, a weighted average is used where weights are the proportion of the population in earlier existing state and newly formed states. For example, to calculate per capita income for undivided Bihar in 2004, the per capita NSDP of Jharkhand and Bihar is used with respective population proportions (population projections are taken from the Ministry of Health & Family Welfare population projection report).

Figure 1: States' salaried employment share growth vis-a-vis economic growth

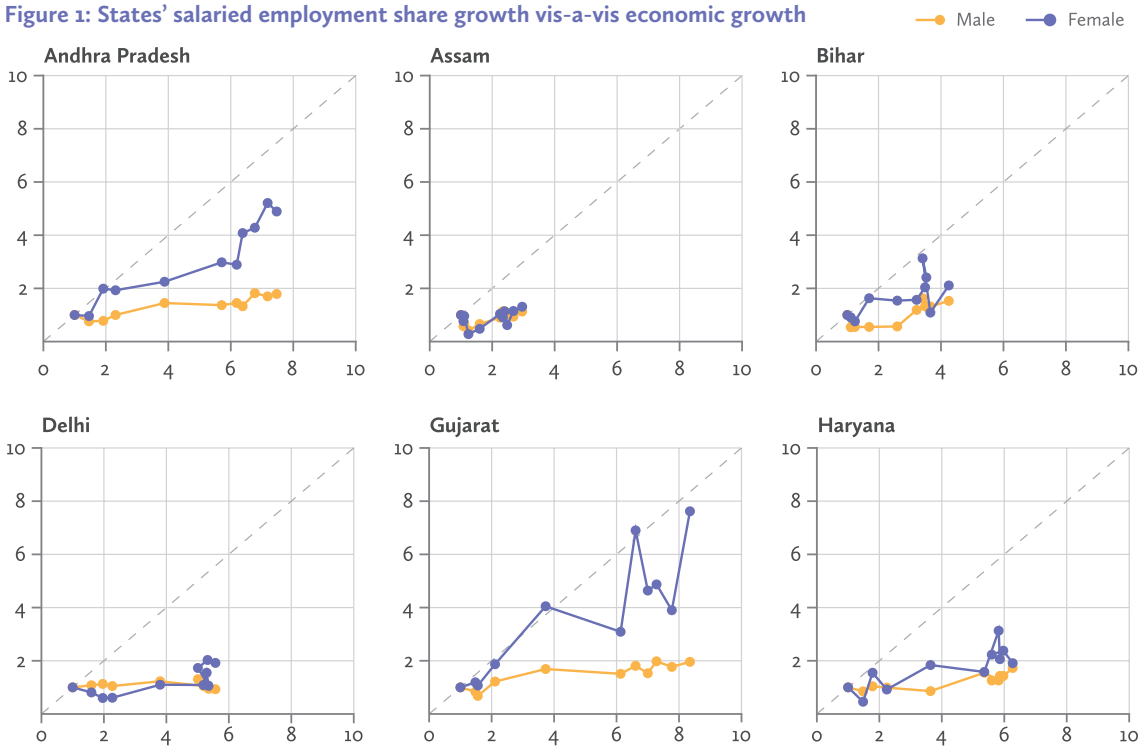
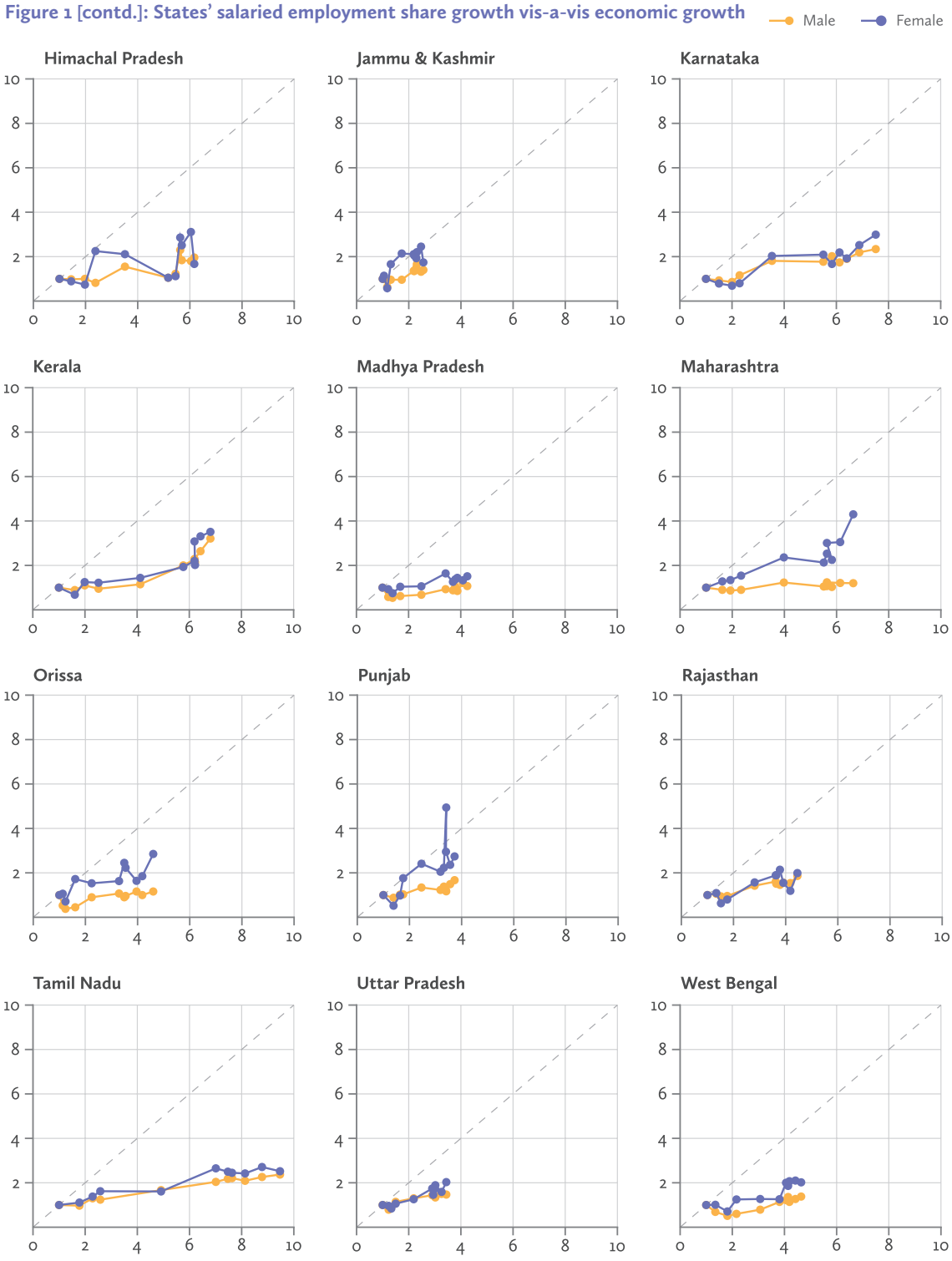


Figure 1 [contd.]: States' salaried employment share growth vis-a-vis economic growth



State Panel Regression

In the state panel regression, the independent variables are log per capita NSDP, interaction of log per capita NSDP and state controls, with state fixed effects.The regression model captures the effect of growth on the proportion of salaried youth for a given state i at time point t. The cross-state regression framework allows us to estimate the semi-elasticities for structural change for each state .The standard errors are clustered at the state level.

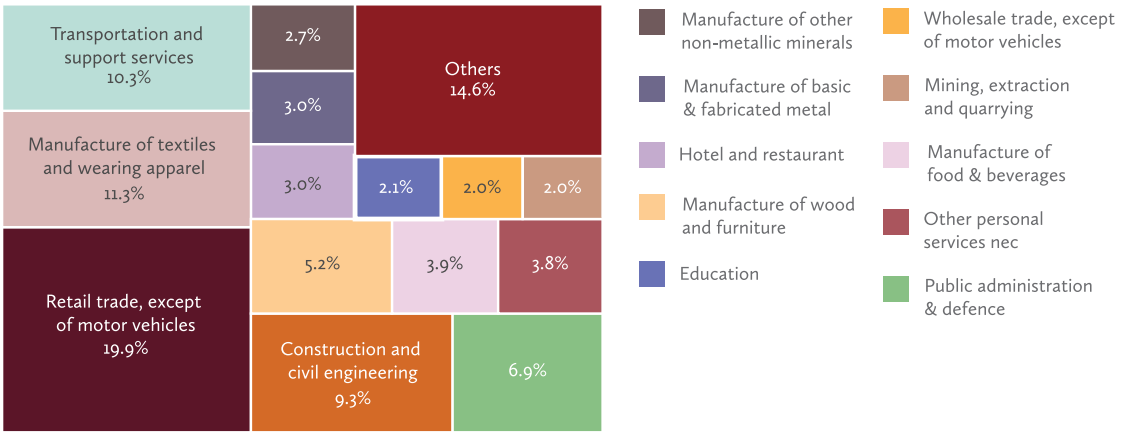
$$prop\_salaried\_youth_{it} = \alpha + \beta_1 \, log\_percapita\_nsdp_{it} + \beta_2 \, state_{it} + \beta_3 log\_percapita\_nsdp_{it} * state_i + \varepsilon_{it}$$

Here $\beta_1 + \beta_3$ gives the semi elasticity. Multiplying this by 100 gives us an easy interpretation of how doubling of growth (100 percent change in per capita NSDP) affects a salaried youth share in percentage point terms.

Chapter 3

Figure 2: Industrial composition of non-graduate men

a: 1983



b: 2023

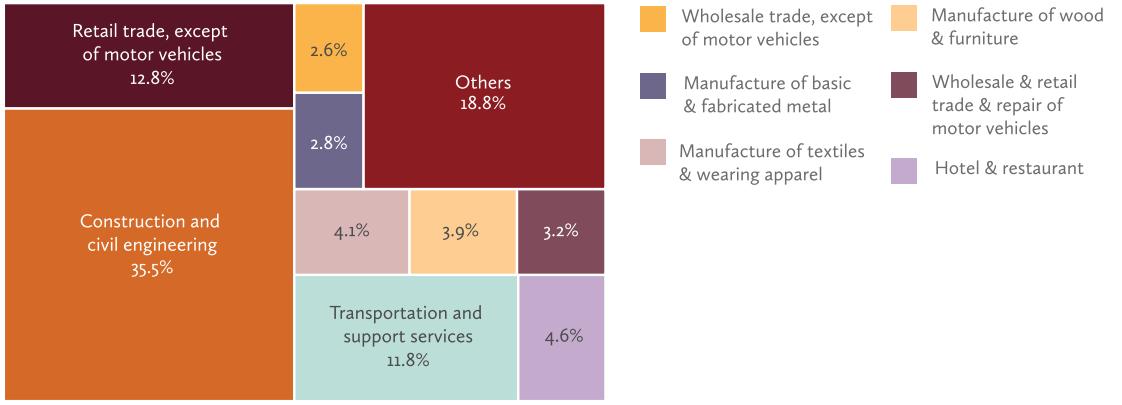
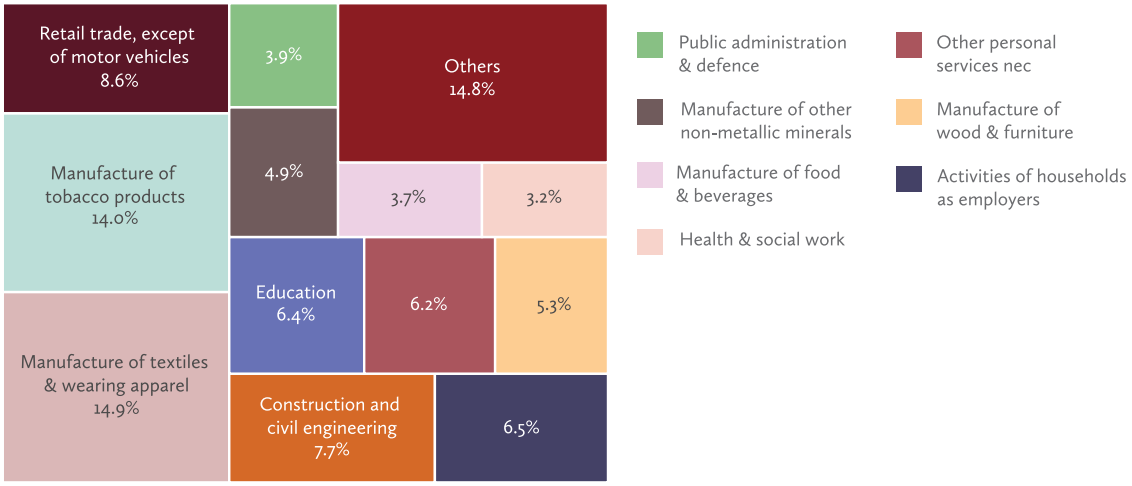


Figure 3: Industrial composition of non-graduate women

a: 1983



b: 2023

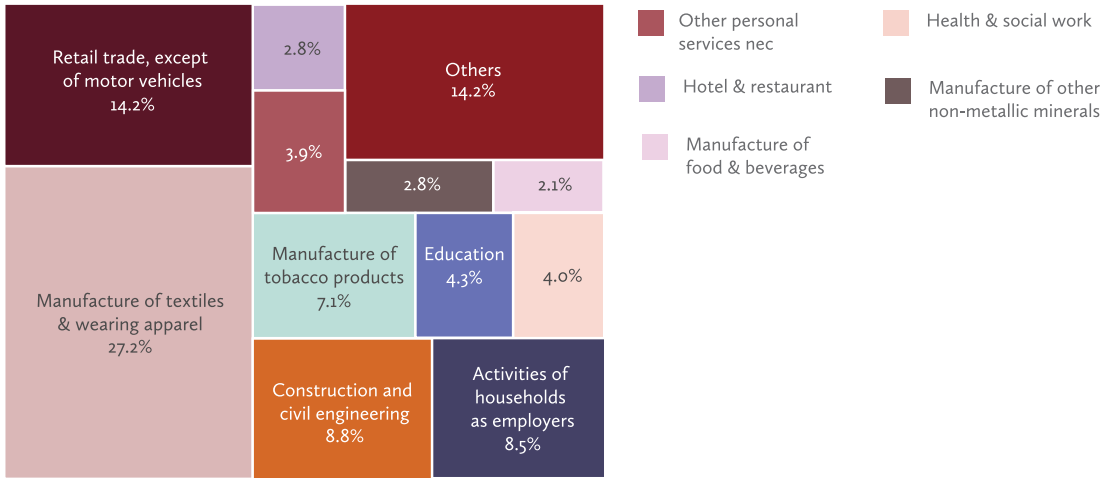
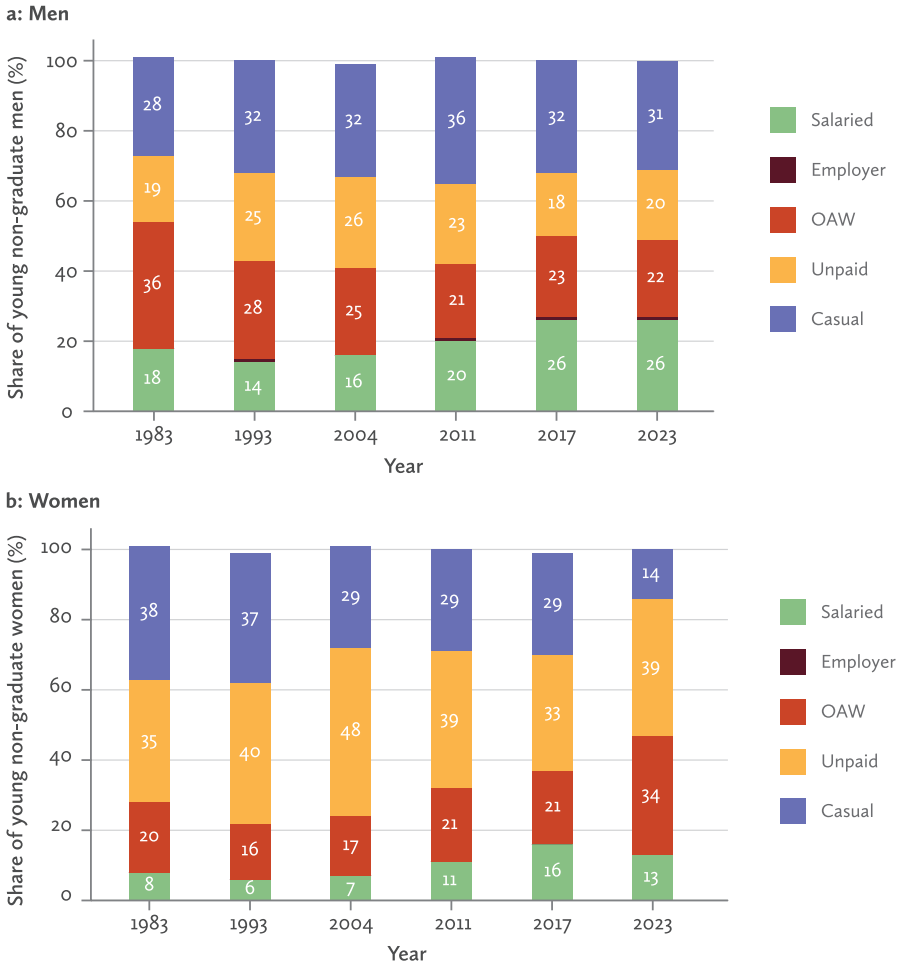


Figure 4: Employment type for non-graduate men and women



Chapter 4

Estimating increase in teacher numbers by district

For estimating if the increase in the number of teachers has been commensurate with enrolment by districts, we use the following regression model:

$$\text{total\_teachers}_{it} = \beta^d_0 + \beta^d_1 t\text{students}_{it} + \text{State Fixed Effects} + \varepsilon^d_{it}$$

Here i refers to the college and t, the year. $total\_teachers_{it}$ refers to the number of teachers in institute i in the year t. $students_{it}$ refers to the number of students enrolled in institute i in the year t. The number of students is scaled to units of 100, i.e. a 1 unit increase in $students_{it}$ indicates an increase in 100 students. The beta coefficient, represents the baseline increment in the number of teachers for an increase in 100 students. To estimate the heterogeneous effects of beta across the districts, run the regression in samples restricted to colleges in a given district. We do this for each of the 640 districts and report the respective beta-coefficients in Figure 10. We also include state fixed effects and a residual term.

Estimating increase in teacher numbers by district: panel analysis

We estimate the increase in the number of teachers commensurate to increase in enrolment by each year, we use the following regression model:

$$\text{total\_teachers}_{it} = \beta_0 + \beta_1 students_{it} + \sum_t \gamma_t Y_t + \sum_t \delta_t (students_{it} \times Y_t) + \alpha_i + \varepsilon_{it}$$

Here *i* refers to the college and *t*, the year. $total\_teachers_{it}$ refers to the number of teachers in institute *i* in the year *t*. $students_{it}$ refers to the number of students enrolled in institute *i* in the year *t*. The number of students is scaled to units of 100, i.e. a 1 unit increase in $students_{it}$ indicates an increase in 100 students. The beta coefficient, represents the increment in the number of teachers for an increase in 100 students. We interact the student enrolment with year dummies, to estimate the heterogeneous effects of beta for across the years. The gamma coefficient captures year specific effects. δ, the interaction coefficient between students and year, accounts for the year-specific effects on teacher numbers to vary by enrollment. We also include college fixed effects and a residual term. Therefore, in this estimation, we account for college-specific and time-specific variations. The coefficient of interest is ($\beta + \delta$) which gives us the average increase in teachers within colleges, for each year.

Chapter 6

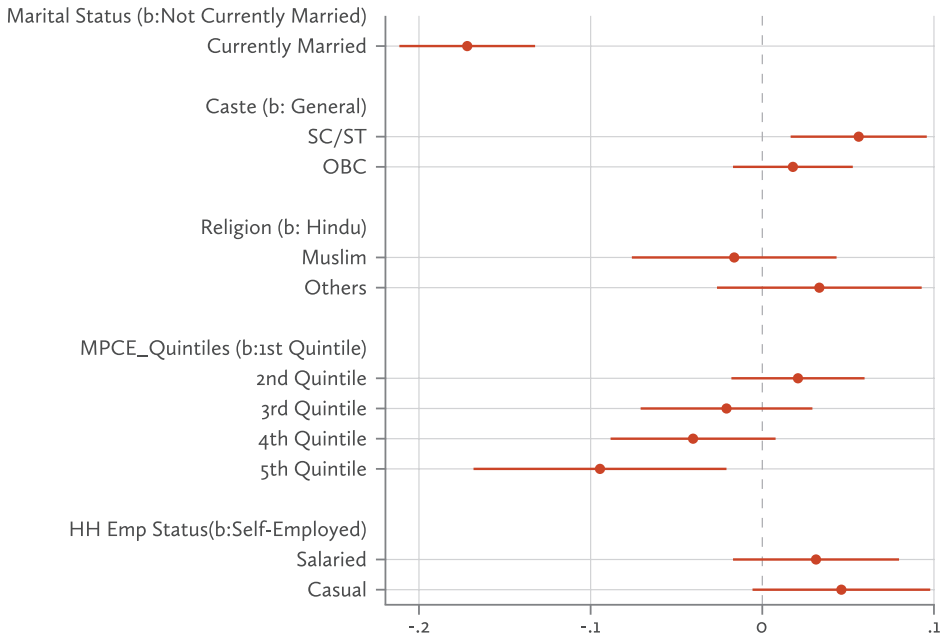
Correlates of unemployment

We estimate a linear regression model to estimate correlates of unemployment for men using PLFS 2023-24. The dependent variable in all the regressions is a binary variable taking the value 1 if the man is unemployed and 0 otherwise.

We control for individual characteristics, namely age, marital status (1 if married, 0 otherwise). We include controls for household attributes including religion (Hindu, Muslim, Others with Hindu as the base), caste (SC/ST or not), technical education (1 if received any technical education and 0 otherwise) and employment type of household head (Self employed, Salaried and Casual worker with Self employed as the base). We accounted for the sector (urban or rural) differences in income by first generating quintiles of monthly per capita expenditure (MPCE) separately for the urban and rural populations. We then combined these into a single categorical variable—MPCE Quintile—where each individual was assigned a quintile according to their respective region. We include state dummies to control for state-level differences.

$$unemp_i = \alpha + \gamma_1 \text{age}_i + \gamma_2 \text{married}_i + \gamma_3 \text{caste}_i + \gamma_4 \text{religion}_i + \gamma_5 \text{urban}_i + \gamma_6 \text{MPCE\_Quintile}_i + \gamma_7 \text{head\_of\_household\_employment}_i + \gamma_8 \text{state}_i + \varepsilon_i$$

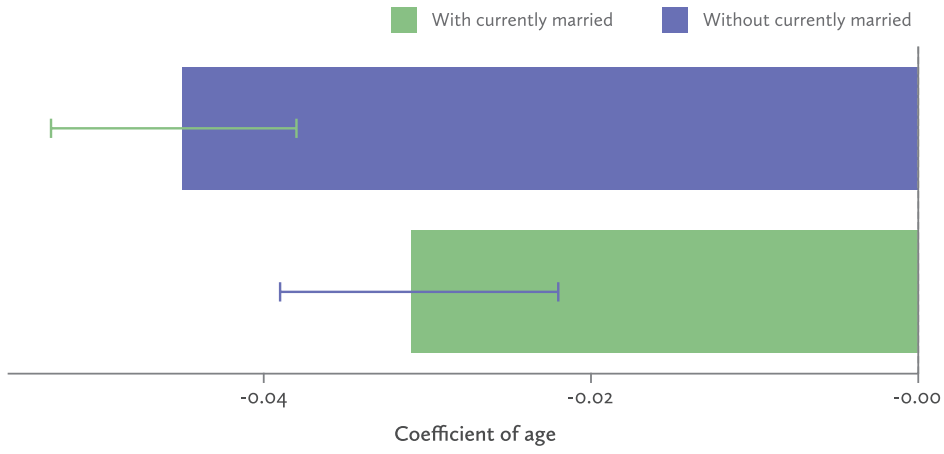
Figure 5: Association between marital status, caste, household income and household head employment type on likelihood of youth unemployment



Sources and Notes: PLFS 2023-24. The graph shows coefficients from a regression of young male graduate unemployment on individual and household attributes.

Using pooled PLFS rounds (2021-22, 2022-23, 2023-24) , we estimate a similar regression model of unemployment - both including and excluding the marital status variable in the specification. This approach allows us to control for the effect of marriage and better isolate the true impact of age on unemployment.

Figure 6: The inverse relation between age and likelihood of unemployment is muted when we control for marriage



CMIE-CPHS School to Work Transition Regression analysis

We now present the results of a linear probability model where the dependent variable is a binary one that takes the value 1 if an individual found employment within a year of declaring themselves to be looking for work, and zero otherwise. The regressors are the individual's age, religion, caste, education, father's education, rural/urban location and state of residence. The education variable is also binary (12th pass or degree holder). We include a dummy to capture the post-Covid scenario. We present the results of two models - one where the dependent variable captures any type of employment and one where it only takes into account white collar permanent salaried employment.

In line with the results presented earlier, individuals from relatively disadvantaged backgrounds (Muslims and marginalised castes) are more likely to find employment compared to their relatively privileged counterparts but if we restrict the analysis to white collar employment this is no longer the case. As expected again from results shown earlier, a college graduate is less likely to find employment in a one year period, but significantly more likely to find white-collar permanent salaried work as compared to a school graduate. We also find that sons of relatively more educated fathers (education exceeding 12th standard) are less likely to be employed within one year but significantly more likely to find white collar work, if they do find employment. This suggests that with better social capital as reflected in the education levels of fathers, individuals are more likely to wait longer for better jobs.

$$\textit{Found employment in 1 year}_i = \alpha + \gamma_1 \text{sector}_i + \gamma_2 \text{age}_i + \gamma_3 \text{religion}_i + \gamma_4 \text{education}_i + \gamma_5 \text{head\_of\_household\_education}_i + \gamma_6 \text{time\_period}_i + \gamma_7 \text{state}_i + \varepsilon_i$$

Table 2: The likelihood of finding employment within a year is higher as age increases

	Coefficient	Robust std. err.	t	P>t	[95% conf.	interval]
Sector (b: Rural)						
Urban	**-0.061**	**0.015**	**-3.950**	**0.001**	**-0.092**	**-0.029**
Transition age (b : 17)						
18	0.017	0.060	0.280	0.781	-0.107	0.141
19	-0.059	0.050	-1.180	0.247	-0.161	0.043
20	0.029	0.047	0.610	0.546	-0.068	0.126
21	0.037	0.069	0.540	0.597	-0.105	0.179
22	**0.113**	**0.049**	**2.330**	**0.028**	**0.013**	**0.213**
23	**0.135**	**0.064**	**2.100**	**0.045**	**0.003**	**0.267**
24	**0.189**	**0.072**	**2.650**	**0.013**	**0.043**	**0.336**
Religion (b: Hindu)						
Muslim	**0.105**	**0.030**	**3.480**	**0.002**	**0.043**	**0.168**
Other	-0.002	0.085	-0.020	0.984	-0.175	0.172
Caste (b: General)						
OBC	0.040	0.026	1.540	0.136	-0.013	0.094
SC/ST	**0.050**	**0.025**	**1.990**	**0.057**	**-0.002**	**0.102**
Education (b : Non-grad)						
Grad/PG	**-0.057**	**0.025**	**-2.250**	**0.033**	**-0.110**	**-0.005**
Father education (b: Upto 5th)						
More than 5th, up to 10th	-0.020	0.022	-0.900	0.377	-0.066	0.026
More than 10th, up to 12th	**-0.097**	**0.027**	**-3.640**	**0.001**	**-0.151**	**-0.042**
More than 12th	**-0.182**	**0.035**	**-5.170**	**0.000**	**-0.254**	**-0.110**
Time period (b: pre covid)						
postcovid	**0.120**	**0.037**	**3.210**	**0.003**	**0.043**	**0.196**

Azim Premji University

Burugunte Village, Survey No 66, Bikkanahalli Main Rd,
Sarjapura, Bengaluru, Bengaluru Karnataka – 562125

080-2441 4000
www.azimpremjiuniversity.edu.in

Facebook: /azimpremjiuniversity **Instagram:** @azimpremjiuniv **Twitter:** @azimpremjiuniv