Appendix A: Data Appendix — Parameter Calibrations for the 'Pseudo' Urban Indian Economy (Section 4)

In this appendix we detail the definitions, data sources, and estimation procedures used to calibrate the model parameters to represent a 'pseudo' urban Indian economy in Section 4 of the paper. The calibration draws on three data sources: (i) the Periodic Labour Force Survey (PLFS), January–December 2025 (NSSO, MoSPI, 2025) for earnings and labour force status; (ii) the NSS 75th Round Survey on Household Social Consumption: Education (Schedule 25.2), July 2017–June 2018, for education costs; and (iii) the Consumer Price Index sub-index for Education (Reserve Bank of India, 2025) for updating the education cost estimates to 2025 prices. The marriage market parameters are calibrated using market prices of consumer durables as of 2025. All parameter bounds are reported in Rs. lakhs at 2025 prices, over a 30-year generational horizon.

The calibrated parameter set consists of 600 communities, 100 drawn from each of six macro-regions. For each community c, both the H-caste and L-caste sub-communities are simulated, with a community-specific caste composition derived from survey-weighted PLFS household data as described in Section I below.

A. Classification of Cities

Tier-1 cities: million-plus metropolitan areas (Delhi NCR, Mumbai MMR, Bengaluru, Chennai, Hyderabad, Kolkata, Pune, Ahmedabad). **Tier-2 cities**: other Class-I urban centres where wages and prices are generally lower than metros. Cities are classified following the official listing in government notifications for House Rent Allowance (HRA) classification issued by the Ministry of Finance (last updated 2017).

B. Regional Classification

We aggregate the 36 States and Union Territories of India into six macro-regions, following standard NSS practice. The state-code assignments are: **North** ($st \in \{1, 2, 3, 4, 6, 7, 8, 37\}$: J&K, HP, Uttarakhand, Punjab, Haryana, Delhi, Rajasthan, Ladakh); **Central** ($st \in \{5, 9, 22, 23\}$: UP, Bihar,

MP, CG); **East** ($st \in \{10, 19, 20, 21\}$: WB, Jharkhand, Odisha, Bihar); **North-East** ($st \in \{11–18\}$: NE states and Sikkim); **South** ($st \in \{28, 29, 31, 32, 33, 34, 35, 36\}$: AP, Telangana, Goa, Karnataka, Kerala, TN, Puducherry, other southern UTs); **West** ($st \in \{24, 25, 27, 30\}$: Gujarat, Daman & Diu, Maharashtra, Dadra & NH).

C. Definition of Caste

We follow a broad definition of 'low' (L) caste encompassing all non-General households: **L-caste** = all households with social group code $sg \neq 9$ (i.e. Scheduled Caste, Scheduled Tribe, and Other Backward Classes; $sg \in \{1, 2, 3\}$). **H-caste** = General/upper-caste households ($sg = 9$). In the PLFS 2025 urban working-age sample used for calibration, all observations have $sg \in \{1, 2, 3, 9\}$, so this definition is equivalent to L-caste = SC/ST/OBC.

Note: Scheduled Tribes are included in the L-caste category because OBC reservations under the central quota structure are in play only after SC and ST reservations are filled. The proportion of L-caste households in the population, and its regional variation, is calibrated directly from PLFS data as described in Section I.

D. Definition of Education

'High' education corresponds to the completion of a graduate degree or above (general education level code $gedu\_lvl \in \{12, 13\}$ in PLFS). 'Low' education corresponds to all other educational attainment levels, including missing values. The cost of high education, parameter E, is calibrated from NSS 75th Round data and updated to 2025 prices as described in Section H.

E. PLFS 2025 Sample Definition

All earnings and labour force statistics are drawn from the Periodic Labour Force Survey (PLFS), January–December 2025 (Annual Report, NSSO, MoSPI, 2025). We use the first-visit person-level file (CPERV1.txt) and the first-visit household-level file (CHHV1.txt).

The estimation sample is restricted as follows: (i) urban sector (sector code $sec = 2$); (ii) working-age individuals, defined as age > 17 and age < 81 (i.e. 18–80 years inclusive); (iii) iden-

tified region (state code *st* mapped to one of the six regions in Section B). Earnings statistics are computed over all individuals in this sample with non-missing *earn_monthly_wcas* (monthly earnings inclusive of casual wage income, converted from weekly reference-period data where applicable). This definition includes individuals who report zero earnings; only those with missing earnings are excluded. Labour force status statistics draw on the Usual Principal Activity Status code (*pas*, Block 5.1).

The total urban working-age sample in PLFS 2025 comprises 364,826 individuals. All 364,826 have identifiable caste and region codes, so no additional exclusions arise from the caste or region criteria. Of these, 178,594 have non-missing *earn_monthly_wcas* (including 14,312 with zero earnings and one with negative earnings) and constitute the earner sample used for parameter calibration. Summary statistics for this sample are reported in Tables 4(a)–(d) of the paper.

F. Wages

We define the model's high wage w_H and low wage w_L from percentiles of the *earn_monthly_wcas* distribution, converted to generational (30-year) values. We assume that each working generation earns 30%, 50% and 100% of its annualised monthly wage in the first, second and third ten-year intervals of its working life respectively, so that the 30-year generational wage = $18 \times (12 \times$ monthly wage). All wage parameters are expressed in Rs. lakhs per generation.

National wage bounds (w_H, w_L)

The national high wage w_H is defined as the 85th–90th percentile of the monthly *earn_monthly_wcas* distribution across all urban working-age earners in PLFS 2025, giving a monthly range of (Rs. 35,000, Rs. 45,000). Converted to generational values: $w_H \in (75.6,\ 97.2)$ Rs. lakhs. The national low wage w_L is defined as the 15th–20th percentile, giving a monthly range of approximately (Rs. 5,214, Rs. 7,170) and a generational range of $w_L \in (11.3,\ 15.5)$ Rs. lakhs. These thresholds serve as the common reference points for deriving the probability parameters, as described in Section G.

Caste-specific wage bounds (w_H^i, w_L^i)

The caste-specific wages w_H^i and w_L^i are derived from the 85th–90th and 15th–20th percentiles of the *earn_monthly_wcas* distribution within each caste group, by region. For H-caste earners, the monthly 85th–90th percentile range is (Rs. 45,000, Rs. 55,000), giving generational $w_H^H \in$ (97.2, 118.8) Rs. lakhs; the 15th–20th percentile range is (Rs. 6,000, Rs. 8,500), giving $w_L^H \in$ (13.0, 18.3) Rs. lakhs. For L-caste earners, the corresponding monthly ranges are (Rs. 30,000, Rs. 40,000) for w_H^L — generational (64.8, 86.4) Rs. lakhs — and (Rs. 5,000, Rs. 6,952) for w_L^L — generational (10.8, 15.0) Rs. lakhs.

These caste-specific wages enter the model's budget constraint and durable accumulation equations. They capture the realised earnings levels for workers of each caste conditional on job type (high or low wage), reflecting the wage penalty embedded in Society W. The separation between the national wages used for probability derivation and the caste-specific wages used for the simulation is discussed in Section G.

G. Probabilities of Earning the High Wage

The conditional probabilities p_1, p_2, p_3, p_4 of earning the high wage are derived, for each caste group i and region c, from the earnings distribution of workers in that group using the formula:

$$p = \frac{X - w_L}{w_H - w_L}$$

where X is the expected earnings of workers of a specific productivity and education type within group $(i, \text{education}, \text{region})$, and w_H, w_L are the national wage bounds defined in Section F. Since productivity (α) is unobserved, we proxy the expected earnings of high-productivity workers by the 55th–60th percentile of the within-group *earn_monthly_wcas* distribution, and of low-productivity workers by the 40th–45th percentile. High-education workers ($gedu\_lvl \in \{12, 13\}$) and low-education workers are treated as separate groups. This yields, for each community c, the conditional probabilities p_1 (low-ed, low-prod), p_2 (low-ed, high-prod), p_3 (high-ed, low-prod), p_4 (high-ed, high-prod) for each caste.

The use of national — rather than caste-specific — wage bounds in this formula is deliberate. Caste-specific wages would mechanically yield higher conditional probabilities for L-caste workers relative to H-caste workers of the same type, because the L-caste within-group earnings distribution has a lower mean. Using national wage bounds ensures that $p_k^L < p_k^H$ for each productivity-education combination k, consistent with the model's premise that L-caste workers face barriers in accessing high-wage employment. The caste-specific wages w_H^i, w_L^i enter separately through the budget constraint, capturing the within-job wage penalty. This decomposition analytically separates the access channel (probabilities, Society P) from the wage channel (Society W), which is central to the paper's empirical strategy.

Unemployment correction

The conditional probabilities p_k derived above are defined over the sub-sample of current earners and therefore represent the probability of earning w_H *conditional on being employed*. To obtain unconditional probabilities consistent with the model — in which every working-age adult draws either w_H or w_L in each period — we scale each conditional probability by the employment rate within the labour force of the corresponding group:[1]

$$P_k^{ci} \;=\; p_k^{ci} \;\times\; e_{ij(k)}^{c}$$

where $e_{ij}^{c} = \Pr(\text{employed} \mid \text{in Labout Force})$ for caste i, education group $j(k)$, region c. The education mapping is: $j(k) =$ low-education for $k \in \{1, 2\}$ and high-education for $k \in \{3, 4\}$. Employment rates are computed as unweighted counts from PLFS 2025, for consistency with the unweighted percentile-based derivation of p_k.[2]

After applying the correction, the unconditional probabilities $P_1 < P_2 < P_3 < P_4$ are verified to hold in all 600 simulated communities for both caste groups. The ordering $P_k^L < P_k^H$ holds in 488 of 600 communities. Violations are concentrated entirely in Regions 3 (East) and 4 (North-East). In

[1] We define *in the labour force* as all individuals who are either employed ($pas \in \{11, 12, 21, 31, 41, 51\}$) or unemployed and actively seeking work ($pas = 81$). Persons who are out of the labour force ($pas \in \{91, 92, 93, 94, 95, 97\}$) are excluded from both the denominator and numerator of the employment rate.

[2] The correction factors are unweighted employment rates for consistency with the unweighted percentile-based derivation of the conditional probabilities p_k. Survey-weighted employment rates are found to be numerically close in all groups.

both regions, L-caste workers are predominantly SC/ST households who benefit from reservation-related employment policies and report earnings exceeding those of general-category workers at equivalent education and productivity levels. We retain these communities in the simulation sample. The empirical reversal of the discrimination direction in these regions is itself an interesting finding about regional heterogeneity in caste-based labour market outcomes in India.

H. Updating Education Cost Estimates to 2025 Prices

The cumulative cost-of-education estimates in Annexure 1 are drawn from the NSS 75th Round Survey on Household Social Consumption: Education (Schedule 25.2), conducted by the National Sample Survey Office (NSSO), MoSPI, over the period July 2017 to June 2018. Since these estimates pre-date our calibration target of 2025, we update them to 2025 price levels using the Consumer Price Index (CPI) sub-index for Education.

Data source. We use the monthly All India CPI Education sub-index (Combined, i.e. Rural and Urban together), series 'A.6.5) Education' (Final values), base year 2012=100, sourced from Table 19 of the Reserve Bank of India Bulletin (Reserve Bank of India, 2025). This series is compiled by NSO, MoSPI, and constitutes the official nationally representative measure of price changes in education-related goods and services — including tuition fees, stationery, books and uniforms — faced by Indian households.

Reference price level (denominator). The NSS 75th Round survey was fielded over the twelve-month period July 2017 to June 2018. We compute the simple average of the monthly Education CPI index values over these twelve months as the price-level base. The 12-month average is **139.24**.

Target price level (numerator). We compute the simple average of the monthly Education CPI index values over all months of 2025 for which Final values are available at the time of writing, viz. January through November 2025. This 11-month average is **193.21**.

Inflation multiplier. The price-level adjustment factor is:

$$\text{Multiplier} \;=\; 193.21 \,/\, 139.24 \;=\; \mathbf{1.39} \quad \text{(rounded to 2 decimal places).}$$

This implies that education costs rose by approximately 39% between the NSS 75th Round survey period and 2025. We apply this multiplier uniformly to the cumulative cost-of-education estimates in Annexure 1. The updated parameter range for E is therefore $(1.39,\ 4.25)$ Rs. lakhs — equivalently, the original bounds $(1.00,\ 3.06)$ Rs. lakhs at 2017–18 prices multiplied by 1.39.

Two caveats are warranted. First, the CPI Education sub-index covers a broad basket of education-related expenditures at all levels of schooling, and may not precisely track the cost trajectory of graduate education specifically. Second, the most recent dedicated NSS survey on Household Social Consumption: Education remains the 75th Round (2017–18). The NSS 80th Round Comprehensive Modular Survey on Education (CMS:E), released in August 2025, covers school education only and does not provide updated graduate-level cost estimates.[3]

Reference: Reserve Bank of India (2025). *Consumer Price Index (Base: 2012=100), All India — Rural, Urban, Combined.* Table 19, RBI Bulletin. Available at: https://data.rbi.org.in.

I. High-Status Household Durables and the Signal Function

The signal function $\Phi_S(B) = \Pr(\text{high-wage partner} \mid \text{durables stock } B)$, where $S = (\beta, \sigma^2)$, is parameterised by the social standard β — the level of durable ownership that is generally recognised as marking a household of high social standing — and σ, reflecting the degree of social 'scepticism' around this standard.

To calibrate β, we price a contemporary urban 'high-status' basket of durables at 2025 market prices. The basket follows the NFHS wealth-index style and comprises: car, 2 air conditioners, refrigerator, washing machine, smart television, laptops, smartphones (3 units), RO water purifier, inverter with battery backup, microwave oven, geysers (3), broadband router, decent furniture (basic 2BHK set), and LPG connection. Prices are differentiated by city tier. The Tier-1 basket totals approximately Rs. 24.2 lakhs and the Tier-2 basket approximately Rs. 12.8 lakhs (detailed itemisation in Annexure 2). Assuming the basket is renewed every 10 years (3 times over a 30-year generation), the 30-year signal threshold range is $\beta \in (38.4,\ 72.0)$ Rs. lakhs.

The scepticism parameter σ is calibrated to have a mean of 3, consistent with the difference

[3]The NSS 80th Round CMS:E (April–June 2025) reports an average per-student expenditure on school education of Rs. 23,470 in urban India — useful as a cross-check on the school-level components of our cumulative education cost, but not applicable to the graduate-level parameter E.

between the Tier-1 and Tier-2 basket values (approximately Rs. 11 lakhs), and is drawn from $\sigma \in (2.5,\ 3.5)$ across communities.

J. Probability of Low Productivity (q_L)

The probability of low productivity, $q_L = 0.85$, is set equal to the approximate share of the urban population residing in Tier-2 and Tier-3 cities within the urban total, following the interpretation that urban workers in smaller cities face less favourable matching in the high-wage segment of the labour market.

J. Generational discount factor (δ)

We calibrate the generational discount factor to 0.5 (in the range found in Manuelli, R.E. & Seshadri, A. (2014). Human Capital and the Wealth of Nations, *American Economic Review*, 104(9), 2736-62).

K. Drawing Weights and Community Simulation

For each of the 600 simulated communities, the steady-state distribution is computed and a sample of observations is drawn. The number of observations drawn from each community reflects its caste and regional composition of the national urban population, so that the pooled simulated dataset is representative of urban India.

For community c in region j, the drawing weights are:

$$\text{Observations from L-caste steady state:} \quad \text{round}\big(1000 \times \Pr(\text{L-caste, region } j)\big)$$

$$\text{Observations from H-caste steady state:} \quad \text{round}\big(1000 \times \Pr(\text{H-caste, region } j)\big)$$

where $\Pr(\text{caste } i, \text{region } j) = \Pr(\text{region } j) \times \Pr(\text{caste } i \mid \text{region } j)$ is the joint population share of caste-region cell (i, j). Both proportions are estimated as survey-weighted shares of urban households from PLFS 2025, using household multiplier *mult* as the survey weight. The grand total of simulated observations across all 600 communities is approximately 100,000.

L. The Stuck-at-the-Bottom Index

The Stuck-in-j Index for class (or generational income) j and caste i is defined as the proportion of period-$(t-1)$ households of caste i in class j who remain in class j in period t, evaluated at the steady-state distribution ($j = low$, $middle$, $high$). The index is computed directly from the steady-state distribution of states (or income histories) (I_t, I_{t-1}) in the simulated dataset, reported separately by caste group. No direct manipulation of the transition matrix is involved. The index is read from the simulated data rather than from the matrix itself. The Stuck-at-the-Bottom (S-B) Index is the Stuck-in-$\hat{j}$ index for the lowest class $\hat{j}$ in steady state equilibrium.

The S-B Index is a steady-state class-persistence probability. It is *not* an axiomatic mobility index in the tradition of Shorrocks (1978), and the paper does not prove decomposability or other axiomatic properties of the index. Its value lies in its interpretability — an S-B Index of 1 indicates complete intergenerational class persistence; an index of 0 indicates complete fluidity — and in the fact that it is reported separately by caste, which allows the excess burden of class persistence attributable to discrimination to be identified as the departure from the just-society benchmark, in which both castes have equal index values for each class.

Annexure 1: Calculation of the Cost of Education

Source: NSS KI (75/25.2) — Key Indicators of Household Social Consumption on Education in India, recalculated by MoSPI for report *Women and Men* (Table 3.17: Average expenditure (Rs.) per student in basic course in the current academic year, by type of course).

Level	Rural (Rs.)	Urban (Rs.)	Total (Rs.)	Years
Pre-primary	5,655	14,509	8,997	2
Primary	3,545	13,516	6,024	5
Upper Primary/Middle	3,953	15,337	6,886	3
Secondary	5,856	17,518	9,013	2
Higher Secondary	9,148	23,832	13,845	2
Total cumulative (pre-primary to higher secondary)	70,902	2,25,309	1,14,488	
General graduate	9,948 p.a.	26,934 p.a.	29,844 (3 yrs) Cumulative: Rs. 1,00,746 (India avg.) Rs. 3,06,111 (Delhi)	
Technical/Professional	50,307 p.a.	71,867 p.a.	2,01,228 (4 yrs) Cumulative: Rs. 3,15,716 (India avg.) Rs. 5,12,777 (Delhi)	

The cost of high education parameter E is calibrated to the **general graduate** track, representing cumulative costs from pre-primary through to completion of a three-year general graduate degree. At 2017–18 prices, $E \in$ (Rs. 1.01 lakhs, Rs. 3.06 lakhs), corresponding to Tier-1 (Delhi costs) and Tier-2 (all-India average) cities respectively. After the CPI Education inflation adjustment (Section H), the 2025-price bounds are $E \in$ (Rs. 1.40, Rs. 4.25) lakhs.

Annexure 2: Value of Durables in Wealth Index (2025 Market Prices)

Item	Tier 1 (Rs.)	Tier 2 (Rs.)	Notes
Car	12,00,000	6,84,000	Hyundai Creta (T1); Maruti Suzuki Dzire (T2)
Air Conditioners (2 units)	1,00,000	70,000	
Refrigerator	43,490	15,000	Double door Samsung (T1); single door Godrej (T2)
Washing Machine	32,000	15,000	Front load (T1); top load (T2)
Smart TV	32,000	10,000	
Laptops	1,50,000	50,000	2 units (T1); 1 unit (T2)
Smartphones (3 units)	1,20,000	45,000	Rs. 40,000 each (T1); Rs. 15,000 each (T2)
RO Water Purifier	20,000	15,000	
Inverter + Battery	60,000	40,000	Incl. maintenance
Microwave Oven	10,000	7,000	
Geysers (3 units)	21,000	12,000	
Broadband Setup (router + 1 yr)	20,000	5,000	
Decent Furniture (basic 2BHK)	6,00,000	3,00,000	
LPG Gas	14,400	10,800	
Total	**24,22,890**	**12,78,800**	≈ Rs. 24.2 lakhs (T1); ≈ Rs. 12.8 lakhs (T2)