India Residential Energy Survey (IRES) 2020

Design and data quality

Shalu Agrawal, Sunil Mani, Abhishek Jain, Karthik Ganesan, and Johannes Urpelainen

Technical Document | October 2020

Suggested citation: Agrawal, Shalu, Sunil Mani, Abhishek Jain, Karthik Ganesan and Johannes Urpelainen. 2020. *India Residential Energy Survey (IRES) 2020: Design and data quality.* New Delhi: Council on Energy, Environment and Water.

Peer reviewers: Dr Satish Kumar, President and Executive Director of the Alliance for an Energy Efficient Economy (AEEE); and Dr Michaël Aklin, Associate Professor, University of Pittsburgh.

Publication team: Alina Sen (CEEW) and Mihir Shah (CEEW).

Organisations: The Council on Energy, Environment and Water (ceew.in) is one of South Asia's leading not-for-profit policy research institutions. The Council uses data, integrated analysis, and strategic outreach to explain and change the use, reuse, and misuse of resources. It prides itself on the independence of its high-quality research, develops partnerships with public and private institutions and engages with the wider public. In 2020, CEEW has once again been featured across nine categories in the 2019 Global Go To Think Tank Index Report. It has also been consistently ranked among the world's top climate change think tanks. Follow us on Twitter @CEEWIndia for the latest updates.

Council on Energy, Environment and Water
Sanskrit Bhawan, A-10, Qutab Institutional Area
Aruna Asaf Ali Marg, New Delhi—110067, India

The authors

Shalu Agrawal | shalu.agrawal@ceew.in | @ShaluAgrawal12

A Programme Lead at CEEW, Shalu's research focuses on the changing residential energy landscape and exploring the role of institutional, technology, and policy reforms to enable access to affordable, reliable, and sustainable energy for all. She also supports The Council's work on power sector reforms and solar-powered irrigation. Shalu holds an MSc from University College London and a BTech in Electrical Engineering from IIT Roorkee.

Sunil Mani | sunil.mani@ceew.in | @sunilmani02

A Programme Associate at The Council, Sunil's work, for the last four years, has centred around collecting and using primary data to understand and improve the energy access situation at the household and community level in India. He holds a master's degree in economics from Shiv Nadar University, India.

Abhishek Jain | abhishek.jain@ceew.in | @ajainme

As a Fellow, Abhishek built and leads The Council's practices on energy access, rural livelihoods, and sustainable food systems. He is directing 'Powering Livelihoods', a $3 million initiative. He co-conceptualised and leads CEEW's flagship research on ACCESS (Access to Clean Cooking energy and Electricity—Survey of States). With more than nine years of experience, Abhishek has worked on multiple issues at the confluence of energy, economics, and environment. He is an alumnus of the University of Cambridge and IIT Roorkee.

Karthik Ganesan | karthik.ganesan@ceew.in | @KarthikGanesan6

Leading The Council's work on the power sector, Karthik has focused his research on the operational reform of discoms in India and t he competitiveness of power generation sources. He holds a master's degree in Public Policy from the Lee Kuan Yew School of Public Policy at the National University of Singapore. He also holds an BTech in Civil Engineering and an MTech in Infrastructure Engineering from IIT Madras.

Johannes Urpelainen | JohannesU@jhu.edu | @jurpelai

Johannes Urpelainen is the Prince Sultan bin Abdulaziz Professor of Energy, Resources and Environment at Johns Hopkins SAIS and the Founding Director of ISEP. He received his Ph.D. in Political Science from the University of Michigan in 2009 and spent the next eight years at Columbia University. He is the award-winning author of four books and over a hundred refereed articles on environmental politics, energy policy, and global governance.

Authors' contribution

Shalu Agrawal: Led the execution of the entire project including survey and questionnaire design, enumerator trainings, and data monitoring and cleaning.
Sunil Mani: Contributed to design, field-testing, and revision of the survey instrument, training of the enumerators, field supervision, data monitoring, cleaning, and analysis.
Karthik Ganesan: Co-conceptualised the project and gave inputs at all stages.
Abhishek Jain: Co-conceptualised the project and gave inputs at all stages.
Johannes Urpelainen: Co-conceptualised the project, Contributed to survey design.

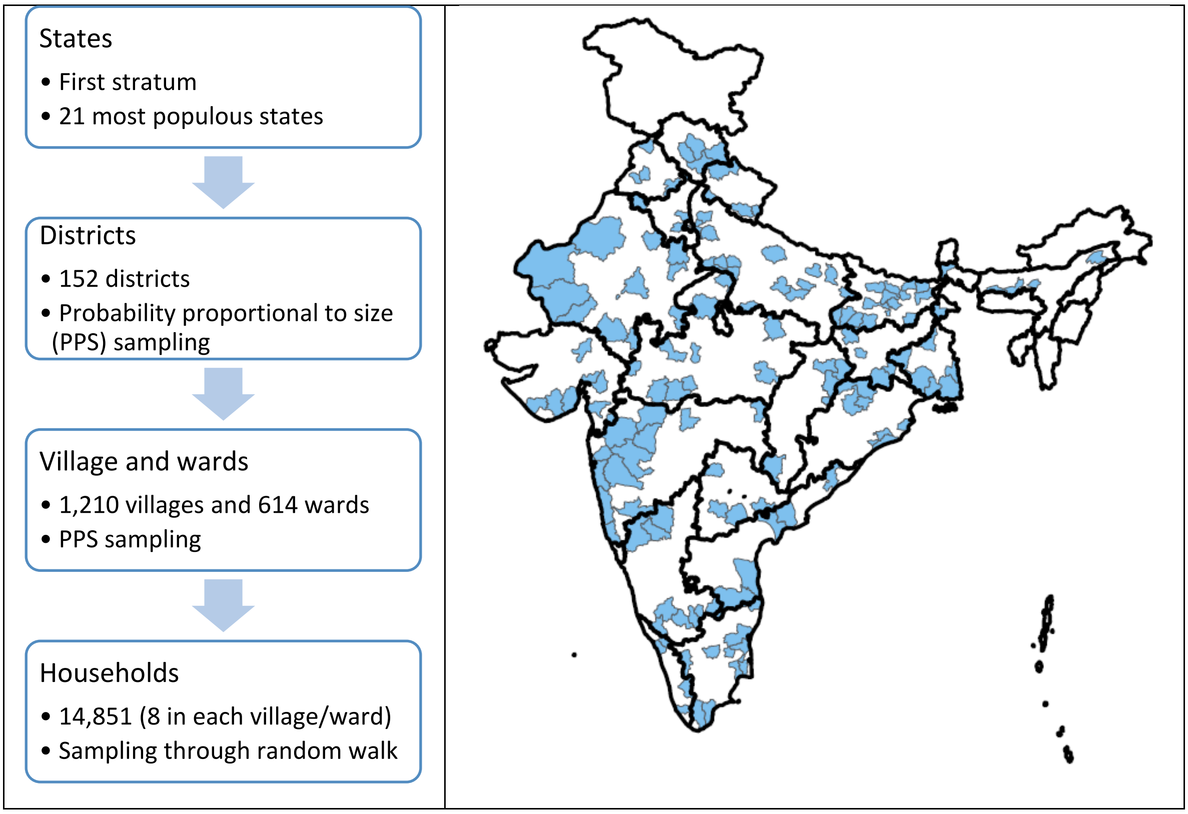
Contents

1. Sampling strategy

The India Residential Energy Survey (IRES) 2020, conducted by the Council on Energy, Environment and Water (CEEW), provides detailed information about the state of energy access and energy use patterns in Indian homes using a nationally representative sample. IRES 2020 covered 14,850 urban and rural households spread across 152 districts in the 21 largest states (by population) of India. The surveyed states together account for 97 per cent of the Indian population. Using the sampling strategy described below, IRES sample is representative of all Indian households.

The IRES used a stratified multistage probability sample design (Figure T1). Within each state, a select number of districts (*d*) was sampled randomly from *d*/2 number of strata. Within each of the sampled districts, we formed two basic strata: i) rural stratum comprising all rural areas in the district and (ii) urban stratum including all urban areas in the district. In each district, we sampled a total of 12 villages and urban wards from the urban and rural sampling frames, proportional to the urban and rural populations in the district. From each village/ward, we randomly surveyed eight households, which formed the ultimate stage units. An equal number of households (12 * 8 = 96) was sampled in each district. Of the sampled households, 34 per cent were urban and 66 per cent rural.

Figure T1: The IRES employed a stratified multi-stage sampling strategy, covering 152 districts in 21 Indian states



Source: Authors' analysis

1.1 Allocation of the total sample to states and sampling of districts

We assigned the number of districts to be sampled per state in proportion to the populations of the latter, while ensuring a minimum of 4 and a maximum of 16 districts in any state. We did this to avoid over-sampling in large states (such as Uttar Pradesh) and under-sampling in small states (such as Uttarakhand). The number of districts (*d*) allocated to every state was further rounded up to be a multiple of two.[1]

To ensure that the sampled districts had varied urbanisation rates and populations — which are correlated variables—we employed the following stratification strategy. We arranged all the districts within each state in descending order of population strength. From this frame, *d*/2 strata were formed such that each stratum had a more or less equal population. Then, from each stratum thus created, we sampled two districts based on probability proportional to size without replacement (PPSWOR).

1.2 Allocation of the sample within districts

We sampled 96 households in each district—8 households each from a total of 12 villages and urban wards. The number of villages and wards sampled across urban (*u*) and rural (*r*) stratum was decided based on the proportion of the urban–rural population in the district as per Census 2011. The allocation of villages/wards at the stratum level was also adjusted to a multiple of two, to ensure a minimum of two urban wards even in predominantly rural districts. Table T1 illustrates the sample distribution across states and rural/urban sector.

Table T1: Sample allocation across states and urban–rural sector

	State	State population	Total districts	No. of districts sampled	No. of villages sampled	No. of wards sampled	Household sample allocation	Households surveyed
1	Andhra Pradesh	49	13	5	40	20	480	498
2	Assam	31	27	4	38	10	384	399
3	Bihar	104	38	14	136	32	1,344	1,346
4	Chhattisgarh	26	18	4	40	8	384	386
5	Gujarat	60	26	8	56	40	768	798
6	Haryana	25	21	4	28	20	384	383
7	Himachal Pradesh	7	12	4	40	8	384	384
8	Jharkhand	33	24	4	38	10	384	386
9	Karnataka	61	30	8	60	36	768	788
10	Kerala	33	14	4	30	18	384	385
11	Madhya Pradesh	73	50	10	88	32	960	964
12	Maharashtra	112	35	14	88	80	1,344	1,409
13	National Capital Territory (NCT) of Delhi	17	9	4	0	48	384	395
14	Odisha	42	30	6	52	20	576	586
15	Punjab	28	20	4	30	18	384	397

[1] Telangana and Andhra Pradesh were part of a single, undivided state as per Census 2011. Hence, their combined sample of 10 districts was divided evenly—5 each.

16	Rajasthan	69	33	10	92	28	960	973
17	Tamil Nadu	72	32	10	70	50	960	969
18	Telangana	36	10	5	36	24	480	492
19	Uttar Pradesh	200	71	16	130	62	1,536	1,548
20	Uttarakhand	10	13	4	32	16	384	384
21	West Bengal	91	19	10	86	34	960	980
	Grand total	1,179	545	152	1,210	614	14,592	14,850

Source: Authors' analysis

Note: State populations are as per Census 2011.

1.3 Selection of villages and wards

For the rural sector: The list of villages in a district according to Census 2011 constituted the sampling frame. In each sampled district, we first arranged the villages in descending order according to the total number of households. From this frame, two sub-strata were formed so that each sub-stratum had a more or less equal population. From each sub-stratum, the required number of sample villages (r/2) was selected using probability proportional to size with replacement (PPSWR). Here, size is the total number of rural households in the village as per Census 2011 and r is the number of villages to be sampled from the district. We excluded villages with less than 50 households from the sampling frame to reduce the need for village replacements due to inadequate sample size.

For the urban sector: The list of urban wards in a district as per Census 2011 formed the sampling frame. In each sampled district, we arranged all the urban wards in descending order of the total number of households. Two sub-strata with more or less equal populations were formed from this frame. From each sub-stratum, the required number of wards (u/2) were selected using the PPSWR strategy. Here, size is the total number of rural households in the ward as per Census 2011 and u is the number of urban wards to be sampled from the district. We excluded wards with less than 50 households from the sampling frame.

Handling the outliers: In districts where a single ward accounted for more than 50 per cent of the total households in the district, it was difficult to divide the wards into two equal strata. We sampled such wards more than once, as shown in Table T2.[2]

Table T2: Wards sampled more than once

S.No.	Ward details	State	District	Wards sampled	Total households (Census 2011)
1	Chilla Saroda Bangar (CT) ward no.-0212	Delhi	East	2	15,516
2	DMC (U) (Part) ward no.-0087		Central	2	9,230
3	DMC (U) (Part) ward no.-0092		Central	2	10,733
4	DMC (U) (Part) ward no.-0149		Central	2	12,105
5	Gharoli (CT) ward no.-0216		East	2	14,380

[2] Such wards were considered as part of a separate stratum. Accordingly, the number of urban wards to be sampled were allocated across the urban sub-strata in proportion to the number of households.

6	Greater Mumbai (M Corp.) (Part) ward no.-0732	Maharashtra	Mumbai	3	108,293
7	Greater Mumbai (M Corp.) (Part) ward no.-2385		Mumbai Suburban	2	120,373
8	Navi Mumbai Panvel Raigarh (CT) ward no.-0001		Raigarh	2	46,920
9	Noida (CT) ward no.-0001	Uttar Pradesh	Gautam Buddha Nagar	6	153,474
10	Panchkula (M Cl) ward no.-0031	Haryana	Panchkula	2	2,621
11	Surat (M Corp.) ward no.-0042	Gujarat	Surat	2	67,515

Source: Authors' analysis

1.4 Selection of households

From each village and ward sampled for this study, we selected eight households for the survey using the random walk procedure. For this, we advised the survey team to choose a random geographic location (e.g., municipal school) in each village/ward and sample every i_{th} household, following the right-hand rule. The skip pattern, value of parameter i, was four in urban wards and villages with more than 400 households; a skip of two households was used in villages with less than 400 households. We determined the skip parameters based on two main factors: non-response rates and household density in urban and rural areas.

We instructed the enumerators to interview the head of each household. If the household head was not available, then the enumerators interviewed another adult member of the household who had adequate knowledge about the household situation and its key decisions. If members of the chosen household were not available or unwilling to participate, another household was selected following the prescribed skip pattern.

1.5 Substitution process

In exceptional circumstances where the agency was unable to survey a village/ward on the sample list, another unit from the same district, with a comparable household population (within +/– 15 per cent range) to the unit being replaced, was selected. The research team proposed replacements after confirming the reasons for the action. Overall, we had to replace 13 villages throughout the survey. The reasons for replacement included i) lack of permission from the village panchayat to conduct the survey, often due to communal tensions or protests related to the National Register for Citizens (NRC), ii) insurgency-related safety issues, iii) migration of the entire village, iv) the area being a high-security zone, or v) weather conditions limiting access to the village.

We replaced one district each in West Bengal and Assam due to protests against the Citizenship (Amendment) Act and NRC.[3] In Assam, Dima Hasao was replaced by Morigaon, while Murshidabad was replaced by Nadia district in West Bengal. Due to the NRC-related protests in West Bengal, only

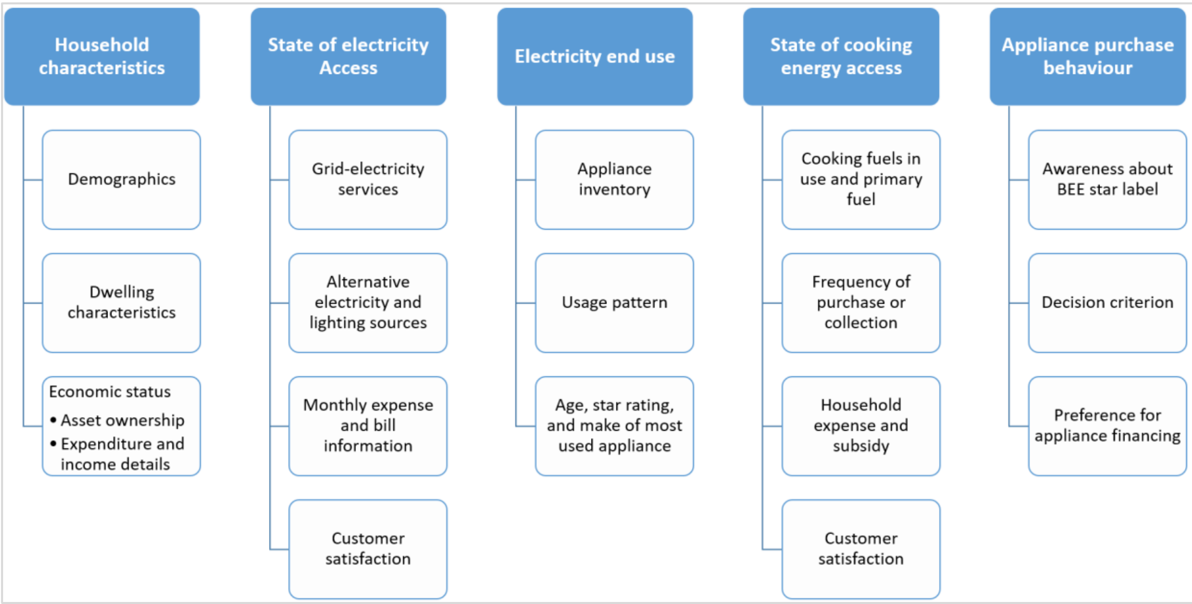
[3] In December 2019, the Government of India amended the Citizenship Act 1955 to give eligibility for Indian citizenship to illegal migrants who are Hindu, Sikh, Buddhist, Jain, Parsi, and Christian from Afghanistan, Bangladesh, and Pakistan, and who entered India on or before 31 December 2014. The act does not mention Muslims.

Nadia was safe enough to conduct surveys in the month of March. Therefore, Nadia was sampled more than once. Though this was a significant deviation from the original design, we did this to ensure an adequate sample size at the state level.

2. Questionnaire design

We designed the IRES questionnaire to capture socio-economic information about the households, state of electricity, cooking energy access, energy usage patterns, equipment characteristics for major end uses (cooking, lighting, space cooling and heating, water heating, entertainment, and other household needs), appliance purchase behaviour, and awareness about government schemes concerning energy-efficient appliances. We designed the questionnaire so that it could be completed in 30–45 minutes (depending upon the household's context). The median time per interview was 35 minutes. Figure T2 depicts the major parameters that the survey captured.

Figure T2: IRES 2020 questionnaire framework



Source: Authors' compilation

We developed the first draft of the questionnaire after reviewing existing survey instruments, borrowing elements from ACCESS 2018 by CEEW, the Residential Energy Consumption Survey (RECS) 2015 by the US Energy Information Administration (EIA), and the Residential Energy Consumption Survey 2019 by Prayas Energy Group. We revised our questionnaire significantly after inputs from peer reviewers from multiple organisations and stakeholder categories.

CEEW researchers piloted the second draft of the questionnaire in three districts—Gurgaon (Haryana), Kolkata (West Bengal), and Bengaluru (Karnataka). The final questionnaire incorporated inputs from the pilot studies; we translated it into 10 Indian languages—Assamese, Bangla, Gujarati, Hindi, Kannada, Malayalam, Marathi, Odiya, Tamil, and Telugu.

3. Data collection

The IRES was conducted between November 2019 and March 2020; the majority of the data collection occurred in December, January, and February. In this section, we describe the data collection process and reflect on the quality of data collected.

3.1 Data collection process

Professional interviewers from Market Xcel Data Matrix Private Limited conducted all the in-person interviews. The interviewers used handheld tablets to record the responses. We used the SurveyToGo application to gather data and ensure regular monitoring. The application allowed enumerators to conduct the interview in any of the ten vernacular languages or English and take pictures of electricity bills, if available. The digital versions of the questionnaires were thoroughly tested in mock interviews during training sessions and corrected before the survey rolled out.

The survey company employed a **team of 154 enumerators (one-third women), 40 supervisors and 20 regional managers** for the data collection. Survey training involved a session for supervisors in New Delhi followed by sessions for enumerators in nine locations across India. Each training session lasted for three days and involved classroom training, role-play exercises (on paper as well as digitally), and dry runs (mock surveys in the field). Enumerators who did not meet expectations were either retrained or dropped from the survey team.

Each enumerator was given a survey kit, which included a copy of questionnaire with detailed instructions, show cards to be used for certain questions, and the authorisation letter from CEEW. To gain the trust and cooperation of households and local leaders, enumerators also carried a copy of a letter of support for the survey from the Ministry of Power, Government of India. Enumerators contacted the sampled households at least twice to maximise the response rate.

3.1 Data quality and limitations

Survey data are vulnerable to multiple errors arising out of recall bias, enumerator bias, or measurement errors. The IRES data are no exception. We made the utmost effort to minimise these errors and ensure data quality using multiple strategies.

We built adequate checks, skips, and value limits (upper and lower bounds) into the data collection software to reduce incorrect, missing, or invalid responses. For some questions, show cards with pictures and response options were used to assist the respondents. Enumerators were trained to code responses that were phrased diversely and to avoid asking leading questions.

Throughout the data collection process, we carried out data quality checks on small data batches to identify various gaps, such as missing, incorrect, or inconsistent values, and deviations from expected trends or outliers. We reported all cases of incorrect responses to the survey agency for cross-verification or re-survey. Many observations were dropped and re-surveys conducted if the quality of the data was doubtful. CEEW researchers also visited multiple survey sites to observe the enumerators at work. This helped us prescribe timely, corrective measures for the interview process and to better understand the context of the responses.

Despite these efforts, we cannot completely reject the possibility of errors in the survey data. Potential users of the data sets must note the following possible sources of errors.

1. *Recall bias*

Questions about monthly household expenditure or monthly electricity expenditure are particularly vulnerable to recall bias. As these were sensitive issues, responses were difficult to cross-verify. We noticed certain inconsistencies in these two variables, so we suggest user prudence while using these data points.

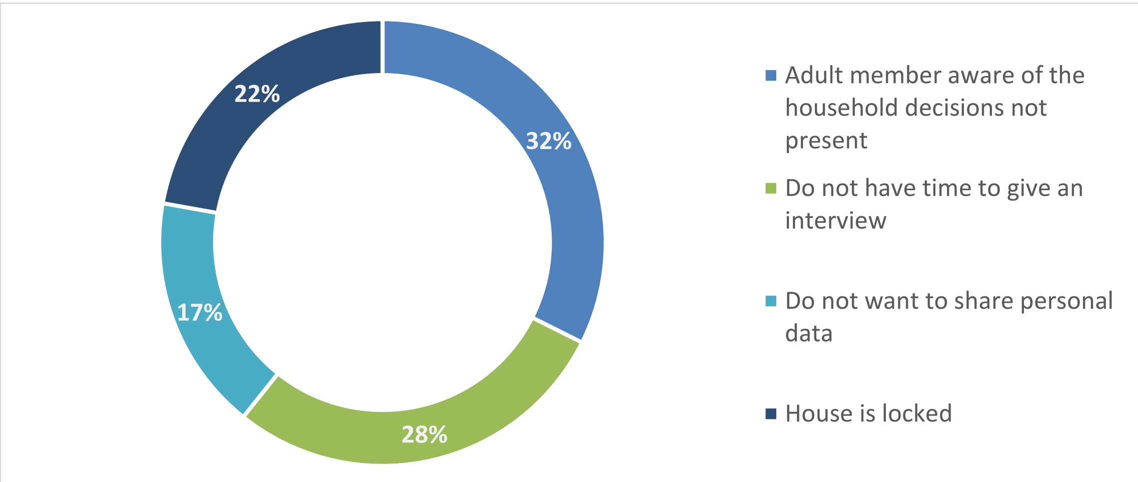
2. *Language-related errors*

The surveys were conducted in nine different languages. We attempted to minimise translation and interpretation errors through questionnaire reviews and pilots. However, given the use of multiple dialects in every state, some questions may not have been adequately administered for some households.

3. *Non-responses*

For the IRES, we observed an average non-response rate of 26 per cent; there was a higher non-response rate in urban areas (34 per cent) than rural areas (21 per cent).[4] Figure T3 records the key reasons for the non-responses. While the majority of non-responses occurred because the sampled house was locked or the required adult member was not at home, nearly 45 per cent of household members were unwilling to spend time on the interview or share their personal data.

Figure T3: Reasons behind non-responses recorded in the IRES survey

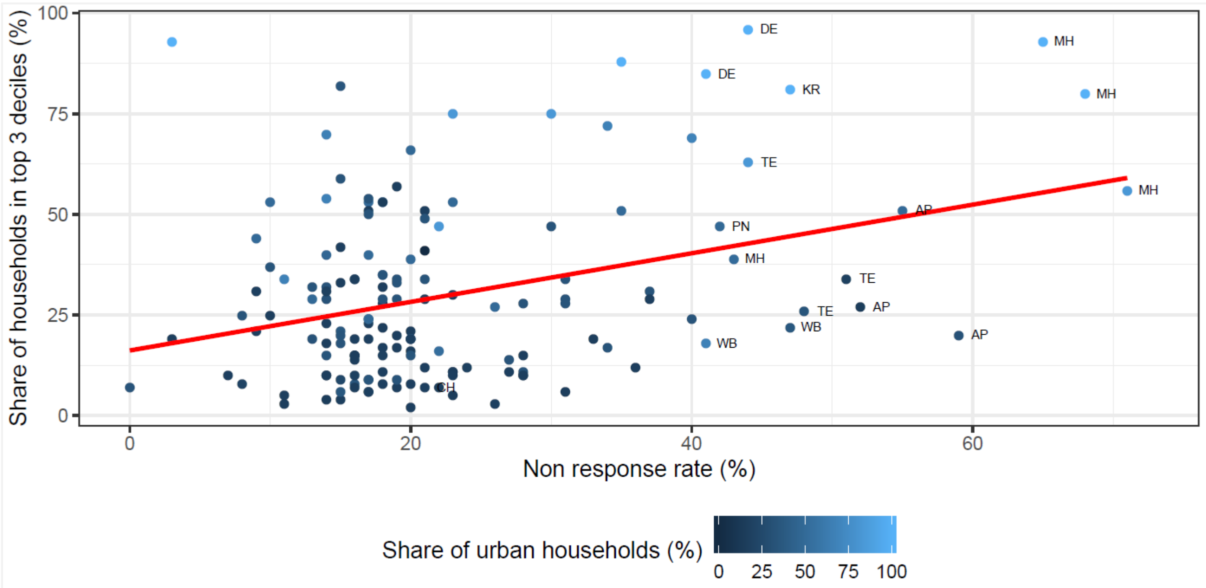


Source: Authors' analysis

[4] If for a targeted sample of *s* households, *r* is the number of houses that could not be interviewed, we estimate the non-response rate using the formula *r/(r+s)*. Thus, a non-response rate of 25 per cent for 100 surveyed households implies that the sample of 75 was achieved after non-response from 25 households.

Figure T4 shows that non-response rates were higher in districts with a higher share of wealthy households (with a 34 per cent correlation).[5] This suggests that wealthy or economically better-off households were more likely to refuse to participate in the survey. These trends indicate that the aggregate estimates for parameters that are strongly correlated with household wealth/income levels, such as owning an air conditioner, are likely understated, though the extent may vary across geographies. However, due to a lack of adequate information, we could not adjust our survey weights to account for such sampling biases.[6]

Figure T4: Non-response rate was higher in economically better-off areas and urbanised districts



Source: Authors' analysis

4. Survey weights

The IRES 2020 used a stratified multistage probability sample design. To produce population estimates, we employ design (base) weights for sample households at two levels—district and state (national). The design weight for each responding household is the number of households in the population that the household represents, estimated as the reciprocal of the probability of selecting that household for the IRES sample. For example, if the national-level weight for a household is 16,000, that household represents 16,000 households in India; if the state-level weight for that household is 13,000, it represents 13,000 households in the given state; and if the district-level weight for that household is 3,500, it represents 3,500 households in the given district. Due to a lack of adequate information, we did not make non-response and post-stratification adjustments to the survey weights. Box 1 illustrates the detailed procedure we used to calculate survey weights.

[5] We measured household wealth with the help of a wealth index. We created this wealth index using principal component analysis (PCA) on a select 12 indicators that together reveal the long-run economic status of a household. See section 7 for the detailed methodology.

[6] Such an adjustment requires some minimum information about key household indicators that are correlated with household wealth; these can be obtained despite non-participation of the concerned household. As we did not collect this information (a limitation of our study), we cannot adjust the survey weights for the bias due to non-responses.

Box 1: Estimating design weights
We estimated design (base) weights for each of the surveyed households to reflect the unequal probability of selection in line with the multi-stage sampling strategy discussed in Section 1. We first estimated the probability of sampling a given household (HH). Then, we used these probabilities to estimate the design weights, which are simply reciprocals of the probability values. Probability (P) of sampling k^{th} rural HH (p_r) = P of sampling i^{th} district from m^{th} strata * P of sampling j^{th} village from i^{th} district * P of sampling k^{th} HH from j^{th} village = p_d * p_v * p_h Where • p_d = No. of districts sampled from m^{th} strata * Total HHs in i^{th} district / Total HHs in m^{th} strata • p_v = No. of villages sampled from i^{th} district * Total HHs in j^{th} village / Total rural HHs in i^{th} district • p_h = No. of HHs sampled from j^{th} village / Total HHs in j^{th} village Similarly, we calculated the probability of sampling k^{th} urban HH (p_u) = P of sampling i^{th} district from m^{th} strata * P of sampling j^{th} ward from i^{th} district * P of sampling k^{th} HH from j^{th} ward = p_d * p_w * p_h We also corrected the design weights for under-/over-sampling of households by multiplying them with the ratio of planned to actually conducted surveys at the village/ward level.

5. Research ethics and confidentiality of information

The surveys were undertaken after conducting due diligence and obtaining Institutional Review Board (IRB) approval. In line with research ethics, enumerators communicated to every respondent the survey objectives, approximate time required, and nature of the questions. In each case, enumerators also got written or verbal consent, depending on the respondent's preference.

For verification and follow-up purposes, the survey agency holds and keeps confidential any information collected during the survey that might identify repondents or their households; for example, respondent details, address, and phone numbers. IRES 2020 data that CEEW has access to, and may publish, do not include identifying information.

6. Representativeness of the survey sample

In Table T3, we compare key parameters with the NSSO 68th Round (2011–12) at the national level to showcase the survey's representativeness. We observe that the household distribution by caste in the IRES data compares well with that in the NSSO 2011-12 data. However, there is some underrepresentation of households following Islam or Christianty, which may be because we did not specify any strata/soft-quota based on religion of the household.[7] We also find that the monthly per capita expenditure (MPCE) values in IRES data are lower than that in NSSO data. This may be on two counts: i) use of more robust methodology in NSSO and under-reporting of expenditure by households in IRES survey due to recall bias and ii) high non-response from higher income households in IRES. Around 15 per cent of households surveyed in the IRES did not reveal their monthly expenditure.

Table T3: Comparison of key parameters between IRES and NSSO 68th Round (household level)

Parameter	NSSO 68th Round 2011–12	IRES 2020
Distribution of households by caste (%)		
Scheduled Tribes	13.4	11.0
Scheduled Castes	15.4	20.5
Other Backward Classes	39.3	34.5
Other	31.9	34.1
Distribution of households by religion (%)		
Hinduism	75.8	87.5
Islam	12.9	8.1
Christianity	7	1.7
Sikhism	2	1.6
Jainism	0.3	0.2
Buddhism	1.1	0.9
Other	0.9	0.1
Average MPCE (INR)		
Rural	1,430	1,270
Urban	2,630	2,131
Average household size		
Rural	4.6	5.1
Urban	4.0	4.7

Source: Authors' analysis using NSSO 68th Round (2011–12) and IRES (2020) data

[7] Another reason for under-representation of these religions may be due to incorrect reporting by some households due to the then ongoing concerns related to NRC-CAA.

Tables T4–T6 provide details about the sample characteristics.

Table T4: Characteristics of IRES respondents

Characteristic	**Rural**	**Urban**	**Total**
Distribution of households by the age of the respondent (%)			
18-25 years	13	12	12
26-35 years	27	27	27
36-45 years	29	28	29
46-60 years	24	25	24
More than 60 years	7	8	8
Distribution of households by gender of the primary income earner (%)			
Male	74	62	71
Female	26	38	29

Source: Authors' analysis

Note: Some percentages may not total 100 per cent because of rounding up.

Table T5: Characteristics of the primary income earner from the IRES

Characteristic	Rural	Urban	Total
Mean age (years)	44	44	44
Distribution of households by gender of the primary income earner (%)			
Male	83	85	84
Female	17	15	16
Distribution of households by the education level attained by the primary income earner (%)			
Uneducated	24	13	21
Schooling (up to Class 8)	32	23	30
Schooling (between Classes 9 and 12)	35	43	37
College graduate/Diploma	9	21	12

Source: Authors' analysis

Note: Some percentages may not total 100 per cent because of rounding up.

Table T6: Other household-level characteristics from the IRES

Characteristic	**Rural**	**Urban**	**Total**
Distribution of households by the highest educational attainment of any household member (%)			
Uneducated	9	5	7
Primary schooling (up to Class 8)	18	10	16
Schooling (between Classes 8 and 12)	48	43	47
College graduate/Diploma	25	43	30

Distribution of households by the average age of all household members (%)			
Less than 18 years	29	24	27
18-40 years	41	43	42
41-60 years	24	27	25
More than 61 years	6	6	6
Distribution of households by type of house (%)			
Kachha house	28.3	3.9	21.1
Mixed (Semi-*pucca*) house	31.7	23.1	29.2
Pucca house	40.0	73.0	49.7
Distribution of household by monthly income (%)			
Up to INR 5,000	12.4	3.9	9.9
INR 5,001–10,000	41.1	21.6	35.4
INR 10,001–20,000	29.2	34.0	30.6
INR 20,000–30,000	8.4	18.5	11.4
INR 30,001–40,000	1.9	9.9	4.3
More than INR 40,000	1.2	6.0	2.6
Undisclosed	5.8	6.1	5.9

Source: Authors' analysis

Note: Some percentages may not total 100 per cent because of rounding up.

7. Household wealth index

For this study, we created a wealth (asset) index which can be used as a proxy for the long-run economic status of a household. We did so with the help of a principal component analysis (PCA) on a set of household wealth indicators, employing the method proposed by Filmer and Pritchett (2001). The benefit of using a PCA is that luxury assets with lower ownership are weighted more than the assets that are more commonly used. However, one fundamental limitation of using an asset index is that the weights on individual indicators are not theoretically grounded. Yet, such relative wealth indices are better indicators of households' economic well-being than stated values for household expenditure or income. This is mainly because expenditure and income levels can be subject to short-term fluctuations, and because of gaps in survey data due to intentional under-reporting, clustering, non-reporting, or recall bias.

7.1 Variables used to create the wealth index

We computed the wealth index using IRES survey data on 12 indicators, which range from basic amenities to items considered as luxury assets in the Indian context.[8] The assets we looked at are as follows:

- *Household characteristic*: type of house (*Pucca* or not)
- *Vehicle ownership*: two/three-wheelers and four-wheelers
- *Durable/non-durable equipment*:
    - *Social connectivity*: TV, computer (desktop/laptop)
    - *Space conditioning*: Fan, space conditioning (cooler/AC/room heater)
    - *Kitchen equipment*: Fridge, exclusive use of clean cooking fuels (Liquified petroleum gas (LPG)/ pressurised natural gas (PNG)/electricity)
    - *Other assets*: Electric iron, washing machine, water heater (geyser/immersion rod/solar/LPG geyser)

We did not include variables which have a small or negative correlation with monthly expenditure, such as house or bicycle ownership, or those with very high ownership, such as phone (95 per cent). We also excluded assets whose ownership has been significantly influenced by the presence or absence of government intervention (at the national or state level), such as toilets, grid-electricity connection, or access to a piped water supply. Further, we excluded items like water purifier and internet connection as their ownership rates are low and limited to urban areas. Given that the selected indicators are largely available across rural and urban areas, we computed a single wealth index for both the rural and the urban populations.

One limitation of this assessment is the lack of information on livestock ownership, which is common among rural households and an indicator of wealth. However, the use of livestock for both domestic and economic purposes makes it a rather challenging variable.

[8] In a recent study, Khosla (2018) used a similar approach to create an amenity index for households in Delhi based on a two-parameter item response model employing 15 indicators: two-wheeler (motorcycle/scooter), car, desktop and laptop computer, Wi-Fi internet connection, landline phone, air conditioner, water purifier, microwave, DVD player, smartphone, refrigerator, and washing machine.

7.2 Internal validity of the wealth index

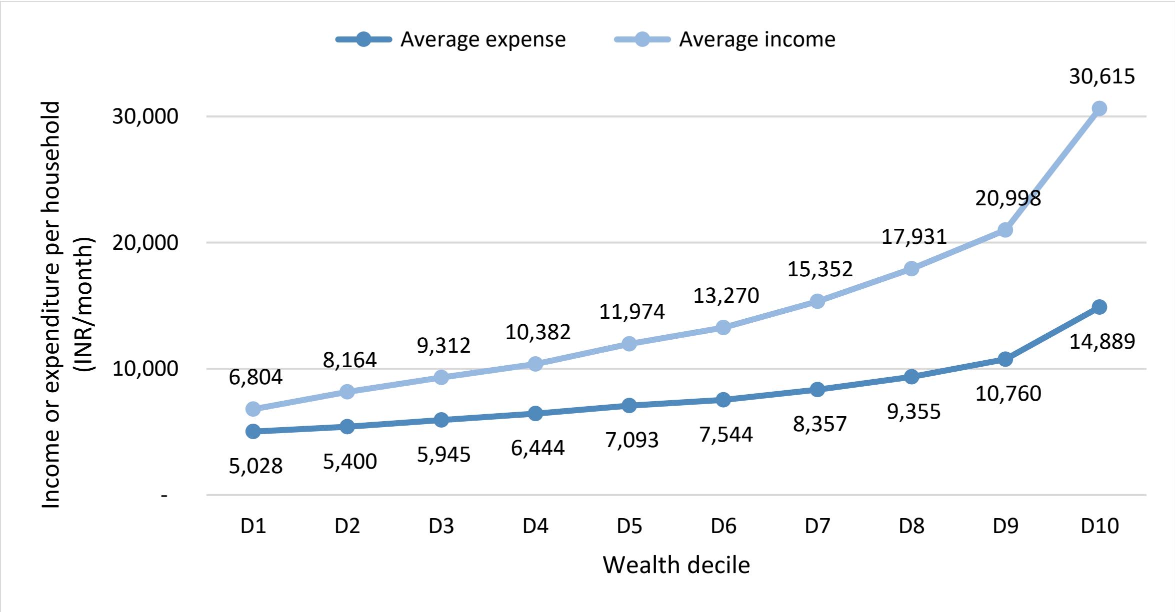
We assessed the internal validity of the index by looking at the characteristics of households classified into deciles based on the wealth index. Table T7 shows the share of India's population that has access to each asset. Evidently, the first PCA can discriminate well between rich and poor households. This is also clear from the increasing trends of average monthly expenditure and income across deciles (Figure T5).

Table T7: Asset ownership across wealth deciles (%)

Asset/ wealth indicators	D1 Poorest 10%	D2	D3	D4	D5	D6	D7	D8	D9	D10 Richest 10%
Fan	22	87	95	98	98	99	99	99	100	100
Television	6	3	63	80	81	88	93	97	98	99
Pucca house	5	26	14	39	56	52	56	77	88	93
Two-/three-wheelers	7	25	17	39	56	61	60	79	86	88
Exclusive clean cooking	10	10	16	36	50	48	61	79	89	94
Electric iron	1	1	6	6	11	30	33	43	62	83
Fridge	–	0	2	0	8	27	59	69	91	98
Space conditioning	–	1	2	1	6	11	17	20	40	73
Water heater	–	0	2	0	5	10	14	22	39	64
Four-wheelers	–	0	1	0	2	4	5	8	11	34
Computer	–	–	0	–	1	2	3	4	5	33
Washing machine	–	–	0	–	0	1	2	6	15	68
Average wealth index (mean score for the first PCA)	–2.62	–1.94	–1.47	–0.98	–0.49	–0.05	0.48	1.17	2.07	3.83

Source: Authors' analysis

Figure T5: Average income and expenditure levels of households across wealth deciles

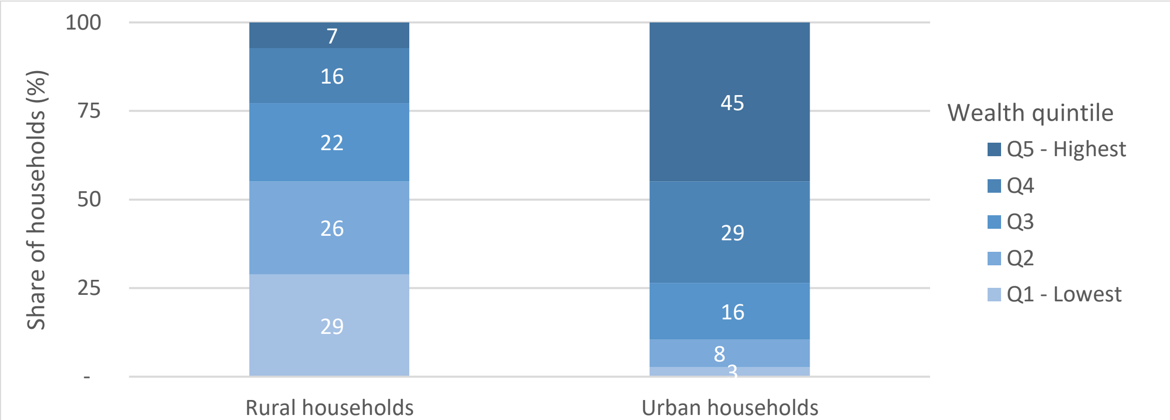


Source: Authors' analysis

7.3 Distribution of household wealth

We classified the households into five equal groups (quintiles) after ranking them in ascending order using the wealth index. We find that more than 75 per cent of the wealthiest households (top 20 per cent) are from urban India, while almost all households in the bottom quintile are from rural areas. Figure T6 shows the distribution of urban and rural households across wealth quintiles. This distribution is comparable to that observed in the National Family Health Survey 2015–16 (Balram Paswan et al. 2017). Around 60 per cent of Scheduled Tribe households and 50 per cent of Scheduled Caste households are in the bottom two wealth quintiles.

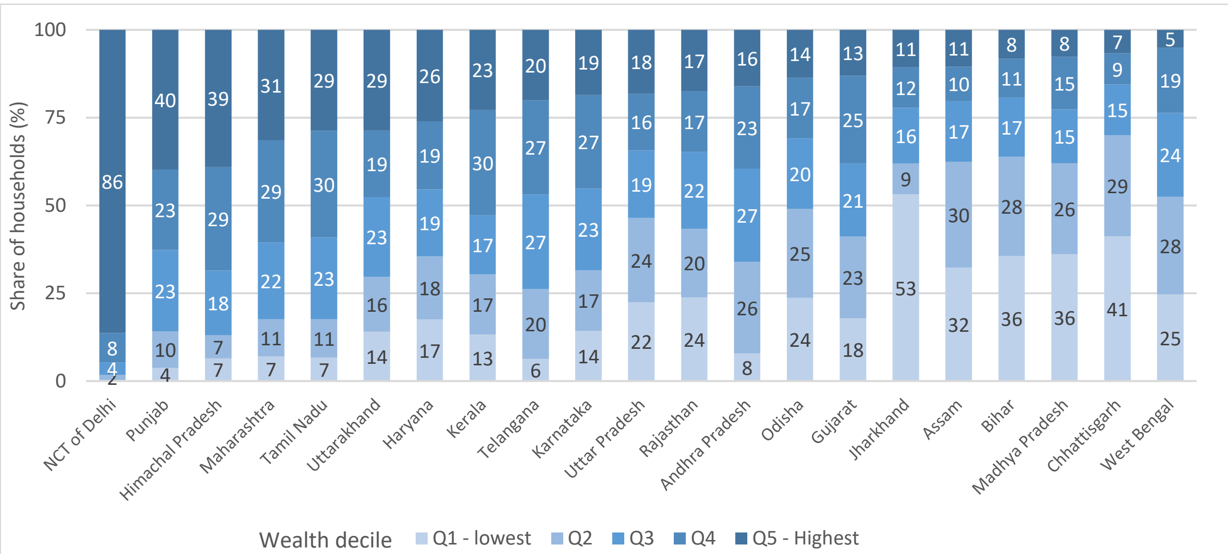
Figure T6: Three-fourths of urban households in India fall in the highest wealth quintile



Source: Authors' analysis

A state-level assessment reveals that Delhi (86 per cent), Punjab (40 per cent), and Himachal Pradesh (39 per cent) have the highest share of households in the wealthiest quintile. In contrast, Jharkhand (53 per cent), Chhattisgarh (41 per cent), Madhya Pradesh (36 per cent), and Bihar (36 per cent) have the highest share of households in the poorest quintile (Figure T7).

Figure T7: Household wealth distribution varies significantly across states



Source: Authors' analysis

References

Balram Paswan, S.K. Singh, Hemkhothang Lhungdim, Chander Shekhar, Fred Arnold, Sunita Kishor, Abhishek Singh, et al. 2017. "National Family Health Survey (NFHS-4) 2015-16 India." *International Institute for Population Sciences (IIPS) and ICF*. https://doi.org/kwm120 [pii]10.1093/aje/kwm120.

Filmer, Deon, and Lant H Pritchett. 2001. "Estimating Wealth Effects without Expenditure Data-or Tears: An Application to Educational Enrollments in States of India." *Demography* 38 (1): 115–32.

Khosla, Radhika. 2018. "Plugging In - Energy Demand in Indian Residences." *Kleinman Center for Energy Policy*. Pennsylvania.

COUNCIL ON ENERGY, ENVIRONMENT AND WATER (CEEW)

Sanskrit Bhawan, A-10, Aruna Asaf Ali Marg
Qutab Institutional Area
New Delhi 110 067, India
T: +91 11 4073 3373

info@ceew.in | ceew.in | @CEEWIndia