Policy Research Working Paper 9994

Poverty in India Has Declined over the Last Decade But Not As Much As Previously Thought

Sutirtha Sinha Roy
Roy van der Weide

WORLD BANK GROUP
Poverty and Equity Global Practice &
Development Research Group
April 2022

Policy Research Working Paper 9994

Abstract

The last expenditure survey released by India's National Sample Survey organization dates back to 2011, which is when India last released official estimates of poverty and inequality. This paper sheds light on how poverty and inequality have evolved since 2011 using a new household panel survey, the Consumer Pyramids Household Survey conducted by a private data company. The results show that: (1) extreme poverty is 12.3 percentage points lower in 2019 than in 2011, with greater poverty reductions in rural areas; (2) urban poverty rose by 2 percentage points in 2016 (coinciding with the demonetization event) and rural poverty reduction stalled by 2019 (coinciding with a slowdown in the economy); (3) poverty is estimated to be considerably higher than earlier projections based on consumption growth observed in national accounts; and (4) consumption inequality in India has moderated since 2011.

This paper is a product of the Poverty and Equity Global Practice and the Development Research Group, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at ssinharoy@worldbank.org and rvanderweide@worldbank.org.

Poverty in India Has Declined over the Last Decade But Not As Much As Previously Thought*

Sutirtha Sinha Roy and Roy van der Weide†

Originally published in the Policy Research Working Paper Series on *April 2022*.
This version is updated on *May 2024*.
To obtain the originally published version, please email prwp@worldbank.org.

Keywords: poverty, inequality, India

JEL Classification: I32

*The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. The authors gratefully acknowledge financial support from the UK Government through the Data and Evidence for Tackling Extreme Poverty (DEEP) Research Program. Excellent research support was provided by Ruchi Avtar, Khushboo Chaudhary and Serene Vaid. The authors are most grateful to Peter Lanjouw for organizing a seminar to solicit valuable feedback from Gaurav Datt, Chris Elbers, Maitreesh Ghatak, Himanshu, Abhiroop Mukhopadhyay, Rinku Murgai, and Martin Ravallion, which has greatly benefited the paper. The authors are equally grateful to our colleagues Junaid Kamal Ahmad, Zoubida Allaoua, Surjit Bhalla, Andrew Dabalen, Indermit Gill, Kristen Himelein, Johannes Hoogeveen, Dean Mitchell Jolliffe, Aart Kraay, Nandini Krishnan, Christoph Lakner, Ambar Narayan, Odyssia Sophie Si Jia Ng, Pedro Olinto, Berk Ozler, Carmen Reinhart, Bob Rijkers, Paul Andres Corral Rodas, Carolina Sanchez-Paramo, Nayantara Sarma, Hans Timmer, Tara Vishwanath, Nobuo Yoshida, and members of the World Banks Global Poverty Monitoring Group and Office of the Chief Economist for South Asia at the World Bank for their very helpful comments and suggestions.

†ssinharoy@worldbank.org; rvanderweide@worldbank.org

1 Introduction

Household consumption expenditure surveys conducted by the National Sample Survey (NSS) organization are the main source of poverty and inequality statistics in India. These surveys support the development of major data-driven policies in India and are used as inputs in the estimation of GDP and India's consumer price index (CPI).[1] The latest NSS expenditure survey that is publicly available for India is from 2011. As the Indian economy has undergone significant changes since then, the release of the 2017-18 round of the survey had been eagerly anticipated. Unfortunately, it was ultimately decided to withhold the unit level survey data and its main results.[2] Using leaked estimates of the empirical distribution function of household consumption, Subramanian (2019) shows that poverty increased in rural India between 2011 and 2017 and that consumption inequality moderated (both in rural and urban areas). The rise in rural poverty neither sits well with consumption trends reported in national accounts data nor with proxy indicators of household welfare derived from official and non-official sources (including labor force surveys, surveys on agricultural household incomes, national family and health surveys of DHS, nighttime lights, etc.).

In the absence of an official consumption survey, several studies have attempted to fill the gap in poverty and inequality data by exploiting alternative data sources. Newhouse and Vyas (2019) and Edochie et al. (2022) impute household consumption into different choices of non-expenditure surveys, namely the Survey of Expenditure on Services and Durables (conducted in 2014-15) and the Survey on Social Consumption on Health (conducted in 2017-18). Chen, et al. (2018) and Felman, et al. (2019) predict growth in mean household consumption based on national accounts data.[3] Bhalla, Bhasin and Virmani (2022) build predictions using night-time lights and changes in state gross domestic product data. Desai (2020) estimates poverty using consumption data obtained from a sub-round of the India Human Development Survey conducted in 2017[4,5]. All studies report a reduction in headcount poverty in India in the years

[1]Given that approximately 18 percent of the worlds population lives in India, its poverty and inequality numbers are also crucial for any efforts to track global poverty, see e.g. Chen and Ravallion (2010).

[2]The government raised concerns about the quality of the NSS-2017 household expenditure data according to the following press release: https://pib.gov.in/Pressreleaseshare.aspx?PRID=1591792.

[3]The relationship between poverty reduction and growth in India has been studied earlier in Datt and Ravallion (2011). See also the cross-country study by Ravallion (2012) on the intricate relationship between poverty and growth.

[4]Desai (2020) is limited to only three states in India, namely Uttarakhand, Bihar and Rajasthan.

[5]More recently, Gupta, Malani and Woda (2021b) use consumption data from the Consumer Pyramid Household Survey to directly estimate poverty for 2019. However, the paper makes no attempt to make newly obtained estimates of poverty comparable to estimates for 2011, preventing assessments of how poverty evolved after 2011

following 2011 - contradicting the headline estimates of the leaked 2017 NSS survey. These apparently contradictory results, combined with restrictions on the release of the NSS-2017 consumption survey, has given rise to a new Great Indian Poverty Debate, a sequel to the debate from the 1990s (Deaton and Kozel, 2005; Kijima and Lanjouw, 2005).

The private sector has recently stepped in by fielding its own household consumption survey called the Consumer Pyramid Household Survey (CPHS). The CPHS may be preferred to alternative data sources used to date for several reasons (but remains second-best to the NSS household consumption expenditure survey for poverty measurement). First, it collects detailed expenditure information on about 115 items, offering household consumption data for the first time since the NSS-2011. Second, the CPHS contains a panel of approximately 174,000 households that covers 28 states representing over 95% of India's population. Third, it is conducted continuously at four-month intervals since its launch in January 2014. This opens the possibility of tracking poverty and inequality at a frequency higher than what has been traditionally feasible based on NSO's quinquennial consumption expenditure surveys. The CPHS is already being used in empirical research. Chanda and Cook (2020) and Chodorow-Reich et al. (2020) use it to estimate the impacts of the demonetization policy, Deshpande (2020) and Gupta et al. (2021a, 2021b) have used the survey to quantify the impact of Covid induced lockdowns on labor market indicators, and Ghatak et al. (2020) employ the CPHS to study rates of consumption and savings in low-income households in India.

Despite these advantages, the CPHS also has its limitations. The CPHS adopts a measure of consumption that is not readily comparable to that of the NSS, stemming from differences in survey instruments. Furthermore, scholars have questioned the representativeness of the survey compared to NSS surveys, due to differences in sample design and geographical coverage (for instance Somanchi and Dreze 2021 and Somanchi, 2021). Both of these differences will have important impacts on poverty estimates for India (e.g., Deaton, 2003).

The objective of this paper is two-fold. First, we conduct a comprehensive examination of potential biases in the CPHS survey and propose adjustments to the survey weights that transform the CPHS into a nationally representative dataset. The outcome of this work will hopefully serve as a public good for anyone looking to use the CPHS for their empirical research. Second, we use the reweighted CPHS to construct NSS-compatible measures of poverty and inequality for the years 2015 to 2019. The challenge in this second objective is similar to that of Tarozzi (2007) which seeks to establish comparability in welfare aggregates across rounds of NSS' consumption expenditure surveys that adopt different recall periods.[6]

We consider two approaches to imputing NSS-compatible consumption into the CPHS. Our preferred approach identifies the relationship between CPHS- and NSS-consumption, and then use this relationship to convert observed CPHS consumption into NSS-type consumption (within the CPHS survey). As a robustness check, we also impute NSS-type consumption on the basis of non-expenditure predictors of consumption that are shared between the CPHS and NSS (i.e. demographics, education, employment, dwelling characteristics, and asset ownership). Both approaches yield qualitatively similar results. We validate our estimates of the levels and trends in poverty and inequality by means of an inclusive set of corroborative evidence that brings in every available source of official and non-official data that could help rationalize the trends in mean consumption, poverty and inequality in India over the last decade.

Our findings are as follows. First, the poverty headcount rate in India is estimated to have declined by 12.3 percentage points since 2011.[7] Our preferred estimates suggest that the poverty head-count rate is 10.2 percent in 2019, down from 22.5 percent in 2011. Second, reductions in rural areas are more pronounced than in urban areas. Rural and urban poverty dropped by 14.7 and 7.9 percentage points during 2011-2019. Third, urban poverty rose by 2 percentage point in 2016 (coinciding with the demonetization event) and rural poverty rose by 10 basis points in 2019 (coinciding with a slowdown in the economy). Fourth, we observe a slight moderation in consumption inequality since 2011, but by a margin smaller than what is reported in the unreleased NSS-2017 survey.[8] Finally, the extent of poverty reduction during 2015-2019 is estimated to be notably lower than earlier projections based on growth in private final consumption expenditure reported in national account statistics. Our analysis stops just before the

[6]Similar methods have been applied to estimate consistent poverty measures when recent household survey data is unavailable and older estimates are considered outdated (Douidich, et al., 2016); to report poverty rates at finer levels of spatial disaggregation (Elbers, et al., 2003); and, to validate official estimates of poverty when comparability of data across surveys is compromised due to changes in instruments (Tarozzi, 2007).

[7]This suggests an extension of the steady progress observed in India over the last two decades, see e.g. Gravel and Mukhopadhyay (2010). However, Dreze and Sen (2012) note that this progress does not extend to all indicators as growth in select nutrition and health indicators, for example, have been more muted. Similarly, Ravallion (2016) notes that despite high growth and a fall in headcount rates in developing countries, the minimum levels of living for the global poor has not moved by much over the past three decades. Castello-Climent and Mukhopadhyay (2013) and Castello-Climent et al. (2018) show that in growth in India is sensitive to changes in tertiary education levels -- suggesting that changes in higher levels of education can impact poverty through the growth channel.

[8]The observed reductions in inequality and poverty are accompanied by major expansion of social security programs in India (e.g. school meals, child care services, employment guarantee, food subsidies, and social security pensions) in the past (see e.g. Dreze and Khera (2017)); and, an expansion of household access to bank accounts, cooking gas, access to toilets, electricity, housing, etc in recent periods, see e.g. Subramanian and Felman (2022).

lockdown measures were imposed due to Covid-19 and therefore cannot speak to changes in poverty headcounts in the aftermath of the pandemic.

The rest of the paper proceeds as follows. Section 2 provides a detailed overview of the known differences between the survey instrument and sample design of CPHS and NSS and sets up both datasets to achieve closest possible comparability based on this knowledge. Section 3 examines the results of the reweighting exercise while Section 4 introduces our two approaches to estimating NSS-consistent measures of consumption. Section 5 reports headline poverty and inequality estimates and reports the results from robustness checks. Section 6 corroborates our findings using a range of independent data sources. We conclude in Section 7.

2 Data

2.1 Consumer Pyramid Household Survey (CPHS)

The CPHS is a stratified multi-stage survey with towns and villages from the 2011 population census as its primary sampling units (PSU) and households as its ultimate sample unit (USU). CPHS' first stage stratum is a spatial unit called Homogeneous Region (HR), which is a set of contiguous districts with similar agroclimatic conditions, urbanization levels, female literacy rates and number of households. The latest round of CPHS consists of 102 HRs spread over 28 states and 514 districts in India (out of total of 36 states and 718 districts in India), with each HR further divided into rural and urban sub-strata. The latest round of CPHS' rural sample comprises 63,430 households selected randomly from 3,965 villages and 110,975 households from 7,920 urban census enumeration blocks (CEBs).

The CPHS' consumption module contains monthly household expenses for about 115 unique items. A quarter of these relate to food, while others include expenditures on clothing, footwear, cosmetics, toiletries, appliances, restaurants, utilities, transport, communication, education, health, monthly loan repayments and other miscellaneous items. CPHS interviews households three-times a year, at four-month intervals referred to as waves. Households report item-wise consumption for each of these four months. Household interviews are scheduled such that survey estimates are nationally representative for each month of the CPHS wave. In addition to consumption expenditures, CPHS collects data on demographic information, incomes, employment status of members, asset ownership and consumer sentiments of the household. The CPHS does not conduct a listing exercise. Instead, it uses household and population growth projections from Registrar General and Census Commissioner of India to calculate household and population level sampling weights.

The CPHS' sample has evolved over time with household dropping out of the original panel and new replacement households being added. A notable number of households were deleted and added to the CPHS panel during the first five waves of data collection (Figure 1). For that reason, we begin our analysis of CPHS data from 2015-16.[9] There are large net additions to the rural panel during the third wave of 2017. The number of sampled districts increased from 422 to 503 between the second and the third wave of 2017. The newly added districts are concentrated in the comparatively poor and rural areas of the country (with a 2011 mean household consumption per capita that is 18 percent lower when compared to the districts that were already part of the sample).

Response rates in the CPHS vary between 80.6 and 87.6 percent over the 2014 to 2019 period. The highest non-response rates are observed during the pandemic-induced lockdown of 2020. The fraction of households from the first wave of 2014 that remained in the panel until December 2019 is 16.9 percent.[10] On average, the probability that a household will survive the panel is halved after about 7 waves of data collection. Further information on the sample is available on the CPHS' official website.

2.2 NSS surveys and other data sources

We use a range of secondary data sources to correct for biases in the CPHS and to validate our estimates of poverty and inequality for the 2015 to 2019 period.

NSS consumption surveys: The 68$^{\text{th}}$ round of NSS conducted between July 2011 and June 2012, is the latest official source of consumption data publicly available for India. The survey reports consumption expenditure values with a 30-day recall period[11] and consists of a sample of over 100,000 households spread across all Indian states. Survey estimates are representative at the district level. The poverty headcount rate at the \$1.90 poverty line is 22.49 percent and the Gini coefficient is 35.71 using consumption per capita based on uniform recall period. We also use select moments derived from the leaked cumulative distribution function that is estimated from the 2017 NSS consumption expenditure survey round for robustness checks.

Other official surveys: Despite there being no contemporaneous NSS and CPHS expenditure surveys, there are three official non-expenditure surveys that allow us to

[9]Vyas (2020) offers a detailed account of the execution challenges by the survey team until the first wave of 2015, especially related to inclusion of excess CEBs in the urban sample.

[10]CMIE makes an attempt to revisit households that are locked on the same day or sometimes the next day in villages. In urban areas, repeated re-visits are conducted spread over several days. If households are consistently locked or unoccupied over three waves, they are dropped from the panel.

[11]We continue the existing practice of measuring poverty and inequality from older NSS rounds based on the uniform recall period (URP)

Figure 1: Percentage of samples added and deleted over survey waves.
Notes: Based on Vyas (2020).

observe changes in socioeconomic variables since 2011. These are: (i) periodic labor force surveys (PLFS) of 2017-18, 2018-19 and 2019-20; (ii) the situation assessment of agricultural households (SAAH) of 2013 and 2019; and, (iii) the all-India Debt and Investment Surveys (AIDIS) of 2013 and 2019. The PLFS provides estimates of wage growth for casual and salaried wage workers, while AIDIS surveys track the evolution of physical and financial assets ownership overtime. The SAAH surveys allow us to study income inequality across agricultural (and predominantly rural) households. Following Himanshu (2019), we use these surveys to construct updated estimates of consumption, earnings, income and asset inequality.

The PLFS furthermore contains a single self-reported expenditure variable referred to as "usual household consumption expenditure", which may serve as a proxy for the respondent's monthly consumption. Mehrotra and Parida (2021) have used this "usual consumption expenditure" variable to document a large increase in headcount poverty in 2019-20. In Appendix 5, we examine this welfare aggregate and detect the presence of significant bunching of consumption around multiples of Rs. 1000 - consistent with theory of satisficing documented in Krosnick (2018). Our simulations suggest that these rounding off errors can have a considerable impact on estimates of poverty and inequality.

We also use the National Health and Family Surveys (NFHS) to obtain estimates of changes in consumer durable assets and access to public services, such as electricity, water and toilet on household premises. We follow Somanchi (2021) and use the publicly released state-level aggregates of 14 states from the NFHS' 2019 round to validate our reweighting strategy (see section 3).

Finally, we use changes in real rural wages reported in Kundu (2019) to validate estimated changes in the consumption distribution for rural India observed after 2011.

Non-official surveys: We rely on two private survey data sources to further our understanding of household consumption since 2014. The first is the India Human Development Survey (IHDS) subsample round, comprising of a sample of 4,828 households from three states of Rajasthan, Bihar, Uttarakhand and fielded during February to July 2017. The first two rounds of IHDS are nationally representative household panels with waves conducted in 2004 and 2011. Households interviewed in the third subsample round of 2017 are part of IHDS' original panel (Desai, 2020). Consumption aggregates from IHDS are based on a basket of 52 items. Average national consumption growth between 2004 and 2011 based on IHDS is 3.8 percent compared to compared to 3.5 percent growth reported in NSS. Historically, the mean consumption growth from the two surveys have closely tracked each other.

We also use publicly reported quarterly growth estimates of fast-moving consumer goods (FMCG) from Nielsen to track consumption trends. These estimates are based on Nielsen's extensive network tracking sales, stock and prices of FMCG goods across brick-and-mortar shops and online channels in rural as well as urban centers.

National accounts and remote sensing data. We use growth in private final consumption expenditure (PFCE) per capita based on national accounts and night-time lights data from 2014 to 2020 from Beyer, et al. (2021) to validate our main results. Nighttime light data are aggregated to the district level and measured in Nanowatts/cm2/steradian.

2.3 Differences between CPHS and NSS consumption surveys

In this section, we systematically document differences between CPHS and NSS consumption surveys that hamper direct comparisons of consumption levels between the two surveys.

Sampling differences. First, the rural and urban substrata in the two surveys constitute different geographical units. The rural FSUs in the NSS' 2011-12 survey were drawn based on 2001 population census village boundaries, whereas the rural FSUs in the CPHS are based on the 2011 round of the census. The number of statutory towns in India has grown by 6 percent between 2001 and 2011 census rounds (ORGI, 2011) as villages evolved into towns, resulting in a divergence in the urban-rural classification between the two surveys. From a poverty measurement perspective this could matter because growth of smaller towns has an impact on rural poverty (Gibson et al., 2017).

Second, larger villages and towns are more likely to be selected in the NSS, whereas differently sized villages have an equal probability of being sampled into the CPHS. More specifically, the NSS draws FSU locations based on population size. In comparison, the CPHS selects rural villages from the rural strata using simple random sampling; for urban areas, CPHS stratifies cities into four groups based on their population and then draws urban FSUs using simple random sampling. Within the FSUs from the CPHS, households have unequal sampling probabilities as households on the main street may have a higher likelihood of selection into the sample relative to other households (see Pais and Rawal, 2021; Dreze and Somanchi, 2021 for details).

Third, the NSS-2011 survey implemented a second stage stratification process, selecting a greater fraction of households in state-regions that had a higher proportion of non-agricultural occupations in rural areas and urban households with mean per capita consumption expenditure between the 1st and 6th decile based on the NSS' 2009-10 expenditure survey. The CPHS in contrast, randomly selects households in rural and urban areas without second-stage stratification, with higher urban draws compared to rural. Despite comparatively larger urban samples, the absence of a second stage stratification in the CPHS means that representation of households from both ends of the income distribution is left to chance. In the NSS, representation of urban households from the 1st to 6th deciles of the distribution is embedded into the sampling design.

Fourth, the CPHS defines households as the physical unit where a group of individual members reside; whereas the NSS defines household as a group of individuals who normally live together and share a common kitchen. The CPHS' definition implies homeless people or families living in construction sites are excluded in the survey. This choice could potentially further contribute to under-coverage of the poorest households in the CPHS.

Fifth, unlike the NSS, the CPHS does not conduct a listing exercise. Instead, it uses projections of household and population growth from India's census organization to construct sampling weights. The NSS does conduct a listing exercise at the start of every round and uses this frame to estimate household level weights. Population weights in the NSS are calculated as the product of the household's sampling weight and its household size; in CPHS population weights are based on the population projections and not the number of household members observed in the survey.

Differences in instruments. Sixth, the NSO uses a more detailed consumption module comprising of over 345 items, compared to 114 unique items captured in the CPHS. Expenditures on household appliances, personal transport equipment, other durables are notably not covered in the CPHS consumption survey. Both surveys contain information on household asset ownership. Additionally, the NSS' expenditure based on uniform recall period captures household consumption over the past thirty days, whereas the CPHS collects consumption based on the past four calendar months. Differences in recall periods across surveys can have large impacts on estimates of poverty (Deaton, 2003; Deaton and Dreze, 2002; Tarozzi, 2007).

Seventh, the CPHS household consumption aggregate includes expenditures on insurance premiums and loan repayments, which are excluded in NSS' consumption expenditure aggregate.

2.4 Addressing differences in instrument design

In this section we document the necessary adjustments we applied to the CPHS datasets in order to address the differences in instrument design between the two surveys. First, we pool the CPHS interviews conducted during the second and third wave of a calendar year and the first wave of the following year to match (as closely as we can) the NSS-2011 reference period of July 2011 to June 2012. The second wave of CPHS starts in May and the first wave ends in April, with households reporting consumption for the past calendar month. Accordingly, CPHS consumption reference period will correspond to April (the month prior to May, when interviews begin) through March of the following year(the month prior to April, the last month of interview). The 2019-20 round of the CPHS consumption overlaps with the first week of the covid induced lockdowns (as the lockdowns in India were imposed on March 24$^{\text{th}}$, 2020), and as such may provide limited evidence on how household consumption, poverty and inequality were impacted at the start of the lockdowns [12]

Second, we exclude districts that are covered by the NSS consumption survey but not by the CPHS to obtain geographical consistency in our analysis. The excluded non-overlapping districts represent about 4.8 percent of the country's population in 2011. Third, in an effort to approximate the NSS' 30-day uniform recall period, we retain item-wise household expenditures for the month preceding the CPHS survey and ignore values that are reported with a lag of two to four months. Fourth, we construct a harmonized basket of items across the two surveys. Expenditures on loan repayments, insurance premiums and household's private transfers to emigrated members are discarded from the CPHS -- while expenditures on durables, household appliances, etc.

[12] All references CPHS consumption years in this paper refer to the financial year starting April to March. That is, CPHS 2015 refers to the corresponding months in 2015-16.

are discarded from the NSS consumption survey. On average, the harmonized basket of goods accounts for about 96 percent of per capita consumption expenditure in the NSS-2011. Fifth, we standardize CPHS' custom industry codes by constructing a concordance with the national industrial classification (NIC, 2008). Sixth, we discard the longitudinal properties of the CPHS by randomly selecting one wave out of a possible three waves in a year.[13]

We adjust individual level sampling weights for non-response using an adjustment factor provided in the CPHS. This non-response adjusted weight, by design, adds-up to the Census' population projections for a given year. We choose not to rely on these individual weights as due to the passage of time -- the last available census is now a decade old -- population projections are likely to become imperfect. One of these imperfections stems from faster than expected fall in fertility rates in 2019 reported in the recent National Family and Health Survey round of 2019-21[14]. Instead, we reconstruct individual level survey weights by multiplying household level weights (provided in the CPHS survey) and the household size (observed in the household roster) for each round.[15] This approach allocates the same sampling weight to each household member and relies on the population distribution observed in the survey rather than the Census' estimated population distribution.[16] Henceforward, we refer to these reconstructed weights as reported CPHS weights and implement a reweighting procedure (that produced adjusted weights) to achieve national representativeness.

2.5 Addressing differences in sampling design

Comparisons of selected statistics obtained with the CPHS with those obtained with several nationally representative surveys identify key biases that raises concern about measurement of poverty and inequality using CPHS data with reported weights. For this reason, we undertake a systematic reweighting exercise with the objective to transform the CPHS into a nationally representative survey (and thereby correct for these biases). Following recent literature (Wittenberg, 2009; Tack and Ubilava, 2013), we adopt the

[13]Not all households are interviewed in all three waves in a year, due to households being unavailable, locked at the time of survey or other reasons. For households that are visited more than once a year, we choose one visit at random.

[14]If the fertility rate falls to below replacement level, it signals that the population is stabilizing.https://indianexpress.com/article/india/fertility-rate-falls-to-below-replacement-level-signals-population-is-stabilising-7639986/

[15]The individual level weights that are bundled in CPHS survey dataset are based on population projections from the Census. As these projections can become dated overtime, we observe the household size captured in CPHS' survey roster and calculate individual weights as the product of CPHS' reported household weight and its size.

[16]Note that the non-response adjusted household weights are still based on census' household level projections.

max-entropy approach advocated by Jaynes (1957).

The reweighting procedure consists of two steps. First, we use assets, demographic and education variables observed in the NFHS-2015 (as well as the CPHS) to reweigh all CPHS rounds from 2015 to 2019[17]. Second, we use demographic, education and labor market indicators observed in the PLFS rounds of 2017, 2018 and 2019 to further adjust the sampling weights in each round of the CPHS[18]. The second reweighting step allows us to account for changes in socio-economic indicators over time.

For the selection of target variables (on which to reweigh), we prioritize non-expenditure indicators that exhibit comparatively large biases in the CPHS relative to the benchmark surveys that are assumed to be nationally representative. An example of such a target variable is the share of undereducated adults (comprising of illiterate and below primary levels of education). We deliberately do not include all indicators that are shared between the CPHS, PLFS and NFHS in the set of target variables. This facilitates convergence of the max-entropy procedure (Zhang and Yoshida, 2022), and more importantly, sets aside a set of indicators that can be used to validate the reweighting exercise.

The adjusted sampling weights are obtained by matching the weighted means of the target variables between the CPHS and the benchmark representative surveys at the state-rural or urban levels (max-entropy minimizes distances between the weighted means obtained in the CPHS and the benchmark surveys). Following existing practices (e.g. Chen et al., 2018; Haziza and Beaumont, 2017; Kolenikov, 2014), the adjusted individual level weights obtained are winsorized at the 0.25^{th} and 99.75^{th} percentile level. We achieve national level representation by multiplying the resulting normalized weights with the rural and urban population populations of each state. The population estimates are obtained from the NFHS-2015 for 2015 and 2016 rounds; and from the PLFS 2017 to 2019 for the remaining periods. Finally, the household level weights are reconstructed by dividing the adjusted individual level weights by the household size observed in the survey.

[17]We use the following set of target indicators for reweighting at the first step: dummy variables for ownership of air conditioners, cars, computers, refrigerators, television sets, two-wheelers, washing machine; dummies for household sizes 1 and 2, sizes 3 and 5; dummy variables for hindu, muslim, scheduled caste, schedule tribe, other backward classes households; total number of members less than 10 years old, over 60 years old; and, total members with below primary level of education, primary level and secondary level of education.

[18]We use the following set of target indicators for reweighting at the second step: dummy variables for female headed household; scheduled caste, scheduled tribe and other backward classes households; dummy variables for household sizes 1 to 5; total members working in casual, salaried and self-employed jobs; total number of members less than 10 years old, over 60 years old; and, total members with below primary level of education, primary level of education and secondary level of education.

3 Comparing CPHS to benchmark surveys

Our starting point is a CPHS dataset containing one observation per household per year, where consumption is reported with a one-month recall and individual level sampling weights reflect the observed population distribution. Nominal consumption expenditures in both the CPHS and NSS surveys are deflated to 2011-12-rupee prices using monthly CPI-IW and CPI-AL price indices for urban and rural observations, respectively. We also adjust for spatial price differences using 2011 PPP exchange rates from the International Comparison Program following Atamanov, et al. (2020).

3.1 Non-expenditure variables

Demographic characteristics: According to Somanchi (2021), the share of children under the age of 10 in CPHS-2019 is 8.9 percentage points lower than the official sample registration survey (SRS) of 2018. This under-coverage is balanced by shares of people aged 40 to 65 years being 11.9 percentage points higher in CPHS-2019 than SRS 2018. CPHS also reports a higher share of households with 2 to 5 members but undercounts households with either a single member or those with more than 6 members. Finally, the CPHS is seen to over-represent Hindu households compared to the benchmark surveys such as NFHS-4.

Figure 2 compares trends in key demographic indicators using the NSS-2011 consumption expenditure survey, the NSS-2014 survey on services and durable goods consumption and the PLFS surveys of 2017 through 2019 as the nationally representative benchmark surveys. The figure shows both the magnitude of the biases observed in the CPHS and the extent to which these biases are corrected by means of reweighting the CPHS. The distribution of household size and its trend estimated using the CPHS now closely match the estimates observed in the nationally representative NSS-surveys. The over-representation of Hindu households is also accounted for. The population shares for other religions similarly match with those observed in the NSS surveys. Biases observed in the composition of scheduled caste, scheduled tribes (and other classes), share of female headed units and households with extended family members living in the same house are also largely resolved through reweighting.

The one demographic variable for which a bias persists is the share of members aged between 0 and 18 years for which a gap of up to 5 percentage points between the CPHS and the NSS-surveys is observed.

Asset ownership and access to services: Somanchi (2021) also documents that the shares of households with access to electricity, water, toilet and ownership of a television and refrigerator are notably higher in the CPHS -- 2015 and 2019 compared to the

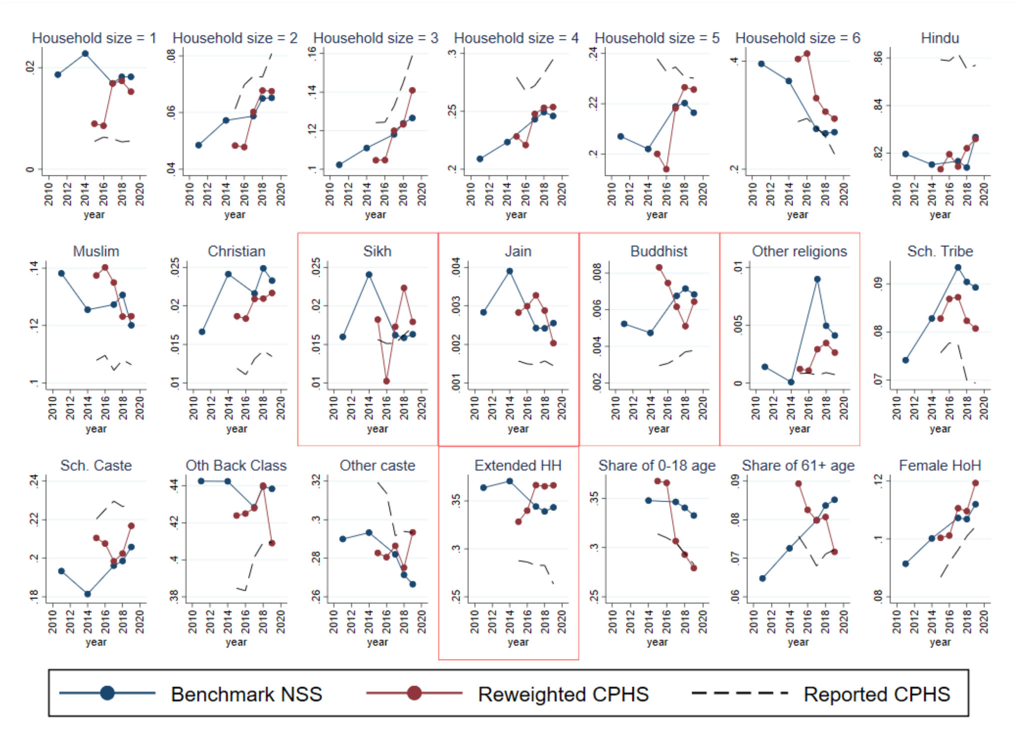


Figure 2: Key demographic indicators from benchmark NSS surveys and CPHS.
Notes: Reweighted CPHS series is based on maxentropy adjusted sampling weights; reported CPHS is based on individual level weights reported in the survey. The figure denotes the share of population for each indicator. The graphs highlighted in red indicate variables that were not included in the set of target variables used for reweighting. Gaps in almost all indicators are closed after reweighting, except for share of individuals between 0 to 18 years of age.

NFHS from the same years. Our analysis finds that ownership of washing machines, two-wheelers and pucca-roof and walls are similarly inflated in the CPHS. Households owning air-conditioning units and computers, however, are under-represented in the CPHS with gaps becoming more pronounced by 2019. These assets tend to be owned by the richest households of the population -- suggesting potential under-representation of richer households (in addition to missing the poorest households).

Asset ownership based on the reweighted CPHS closely matches ownership levels observed in NFHS 2015, closing the gap observed in reported CPHS data (Panel (a), Figure 3). Notable bias corrections are also observed for other indicators such as the share of households with pucca wall and roof (which are not included in the set of target variables for reweighting). The share of electrified households is also seen to match between the CPHS and benchmark survey. Access to water and toilet within premises however are found to be over-represented in the CPHS, also after reweighting with NFHS as benchmark. A candidate reason for this discrepancy is the difference in instrument design (these indicators are not in the set of targeting variables). In the NFHS, access to water and toilet within the household premises are collected through a detailed list of options, eliciting specific types of water sources and toilet waste disposal technologies available to the household. The CPHS in contrast, collects this information through binary yes or no questions without distinguishing between sources or disposal methods.

Comparison of CPHS and NFHS in 2019 (restricted to 14 states where asset ownership and public service access data is presently available) serves as a validation, as the reweighting for this year does not include asset ownership or access to services as target variables (these indicators are not available in the PLFS-2019). The results in Panel (b) of Figure 3 confirms the bias correction that is achieved for these non-target variables.

The largest gap in asset ownership between the CPHS and NFHS 2019 is for households owning television sets (10 percentage point) and air conditioning units (6 percentage points). The reweighting procedure does however reduce the bias by a significant margin: without reweighting, households owning TV sets would be 24 percentage points higher in the CPHS.

Education levels: Undereducated people are severely under-represented in the CPHS with only 2 percent of the 2018 adult population (ages 15 to 49 years) having not received a formal education. By comparison, the periodic labor force survey (PLFS) from the same year estimates that the share of adults without formal education is 17 percent. By 2019, adults without formal education are virtually eliminated from the CPHS sample, while the PLFS-2019 continues to estimate this share of the population at approximately 17 percent. Somanchi (2021) similarly observes that female illiteracy is estimated with a significant bias in the CPHS (in selected states the mean values from the CPHS-2019 are as much as 45 percentage points lower than what is observed in the NFHS-5).

Figure 4 compares adult education levels (ages 15 to 49) in CPHS and PLFS for 2017 to 2019. The share of adult education attainment at the state level observed in the CPHS is plotted against the shares observed in the benchmark PLFS survey. Estimates above (below) the diagonal indicate states where education shares are estimated to be higher (lower) in the CPHS relative to the PLFS. Panel (a) of Figure 4 shows population shares of adults with below primary level education (which includes those with non-formal education as well non-literates). Panels (b) to (d) compare state level shares of primary, secondary and higher educated adults, while panel (e) plots the share of adults with graduate, certificate or post-graduate levels of education.

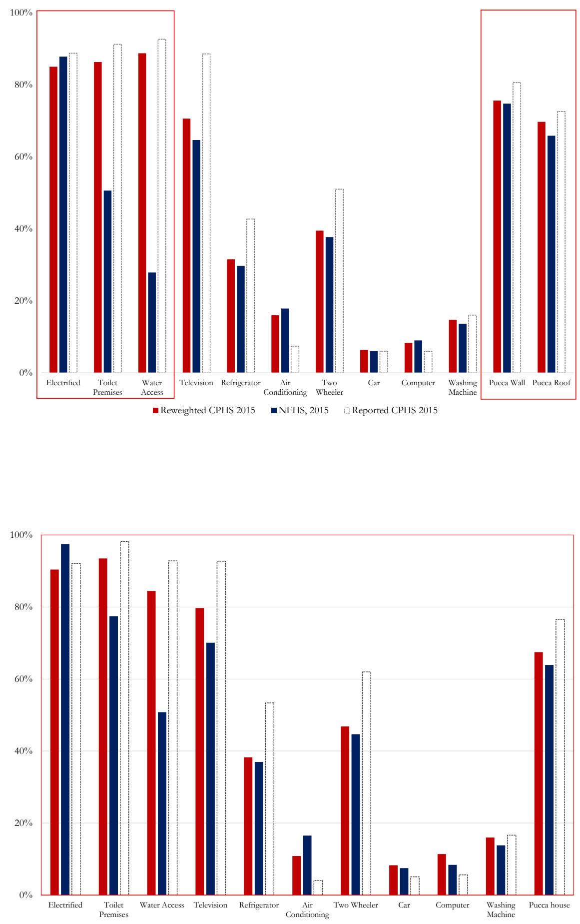


Figure 3: Access to services and asset ownership: NFHS and CPHS 2015 (panel (a); top), NFHS and CPHS 2019 (panel (b); bottom)
Notes: Figure shows asset ownership shares and access to public services. Electrified households in CPHS are defined as those that pay non-zero amounts towards electricity; in NFHS these include households possessing an electrical connection. Toilet in premises in NFHS includes all households that do not have a toilet facility or conduct open defecation. Water in premises in NFHS includes those that have piped water in dwelling unit or use improved water sources. Pucca houses are those that have both pucca walls and pucca roofs. NFHS 2019 all-India estimates are produced by multiplying state-level ownership shares with estimated number of households reported in state-level fact sheets by DHS. Graphs highlighted with a red box denote indicators that were not included in the set of target variables for reweighting. All indicators in 2019 belong to this group.

Overall, reweighting has helped close the biases for these education variables that are observed in the CPHS when using the reported weights. Discrepancies in education levels are most notable in states where illiteracy (or below primary level education) among adults is high. Reweighting is seen to be more successful in correcting biases in 2017 and 2018 than in 2019. But even in 2019, reweighting comes a long way in reducing the bias in states with high shares of illiterate or non-formal education. The estimates for higher education levels are largely scattered along the diagonal, confirming the successful bias correction. Figure 1 in Appendix 1.1 shows that the large bias in female illiteracy using reported CPHS data as documented in Somanchi (2021) is largely resolved after reweighting.

The NSS survey on education consumption conducted in 2017-18 provides another opportunity to compare education statistics derived from the (reweighted) CPHS against. As this survey is not used in the reweighting procedure, this comparison helps provide external validity of the adjustments made to the sampling weights. Panels (a), (b) and (c) from Figure 5 show the results for all adults, males and females above the age of 15, respectively. Reassuringly, all education level shares obtained using the adjusted CPHS sampling weights are within 1 percentage points from the benchmark survey. This denotes a notable improvement compared to the estimated obtained using the reported CPHS weights.

Labor force indicators: Abraham and Srivastava (2019) observe a 3.2 percentage point gap in labor force participation rates among males between the CPHS-2017 and the PLFS from the same year. Labor force participation rate for females in the CPHS are about half that of what is estimated by the PLFS. Basole, et al.(2021) finds that the average real incomes in the CPHS of 2018 are about 30 percent higher when compared to the PLFS from the same year[19]. Despite the higher average incomes, wage inequality is lower in the CPHS relative to the PLFS: estimates of the Gini coefficient of income inequality for the two surveys are 0.42 and 0.44, respectively (excluding zero wage earners). Our analysis furthermore finds that the share of casual wages workers is higher in the CPHS than in the PLFS.

Figure 6 shows log monthly salaries and log daily wages for both the CPHS and benchmark surveys (these indicators are not included in the set of target variables used in reweighting). Reweighting closes the gap in monthly salaries and daily wages that is observed when using reported CPHS weights. The bias correction is larger for rural than for urban wage incomes. Unlike Basole, et al.(2021), we exclude income from self-employment in our analysis as determining profits from work requires detailed enumeration of cost and revenue parameters of an enterprise -- which are not recorded in either survey.

Reweighting is also seen to account for the gap in wage inequality between the CPHS

[19]Basole, et al. (2021) include earnings from self-employed work in their analysis.

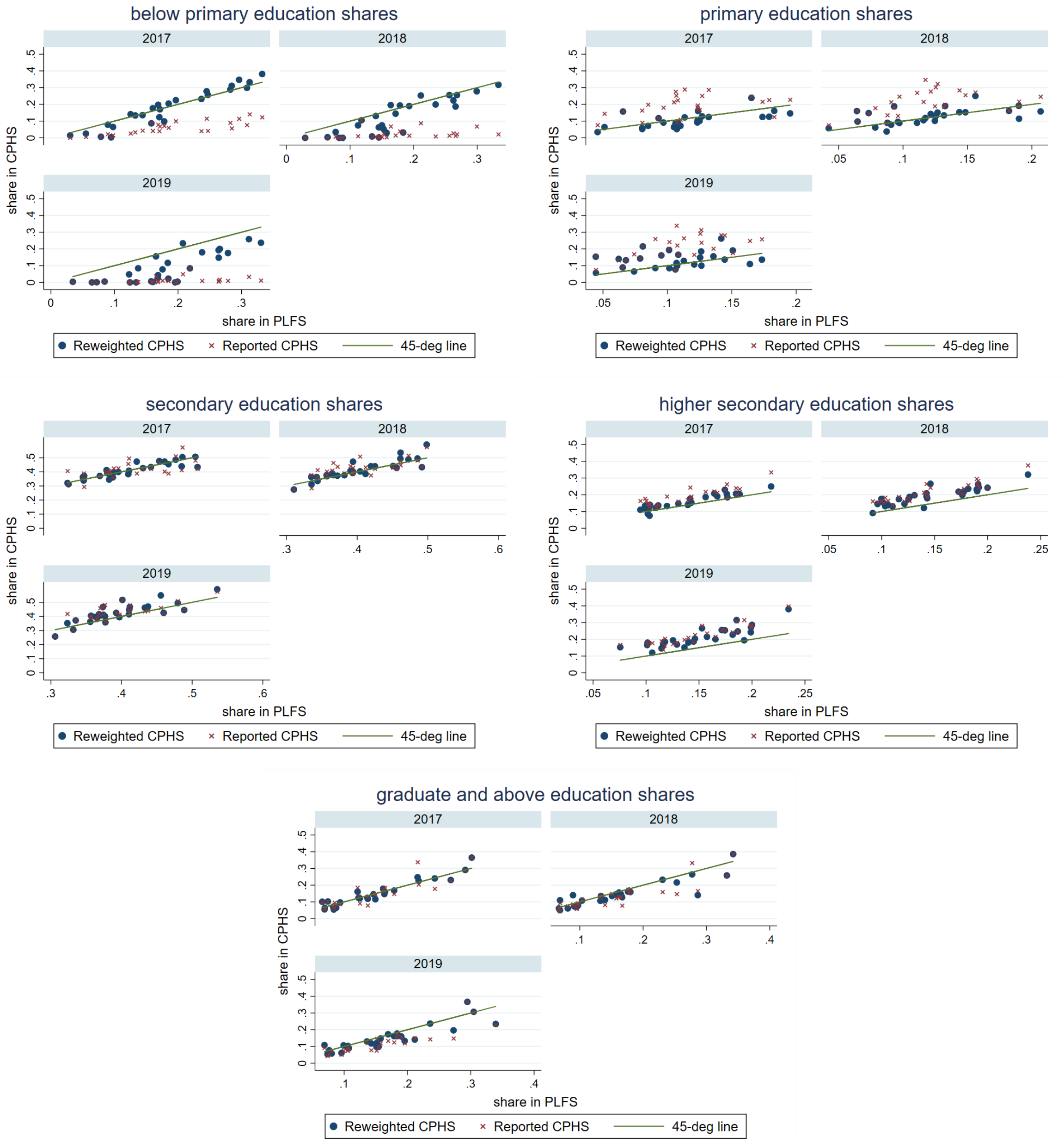


Figure 4: State level educational attainment in PLFS, Reported CPHS and Reweighted CPHS: Below primary education shares (panel (a); top-left), Primary education shares (panel (b); top-right), Secondary education shares (panel (c); middle-left), Higher secondary education shares (panel (d); middle-right), Graduate and above education shares (panel (e); bottom)
Notes: Scatter points denote education attainment shares at the state level from reported and reweighted CPHS in the vertical axis and PLFS in the horizontal axis. PLFS data includes only the first visit to each household. Sample includes adults ages 15-49 in both surveys. Estimates are constructed using individual level weights from both surveys.

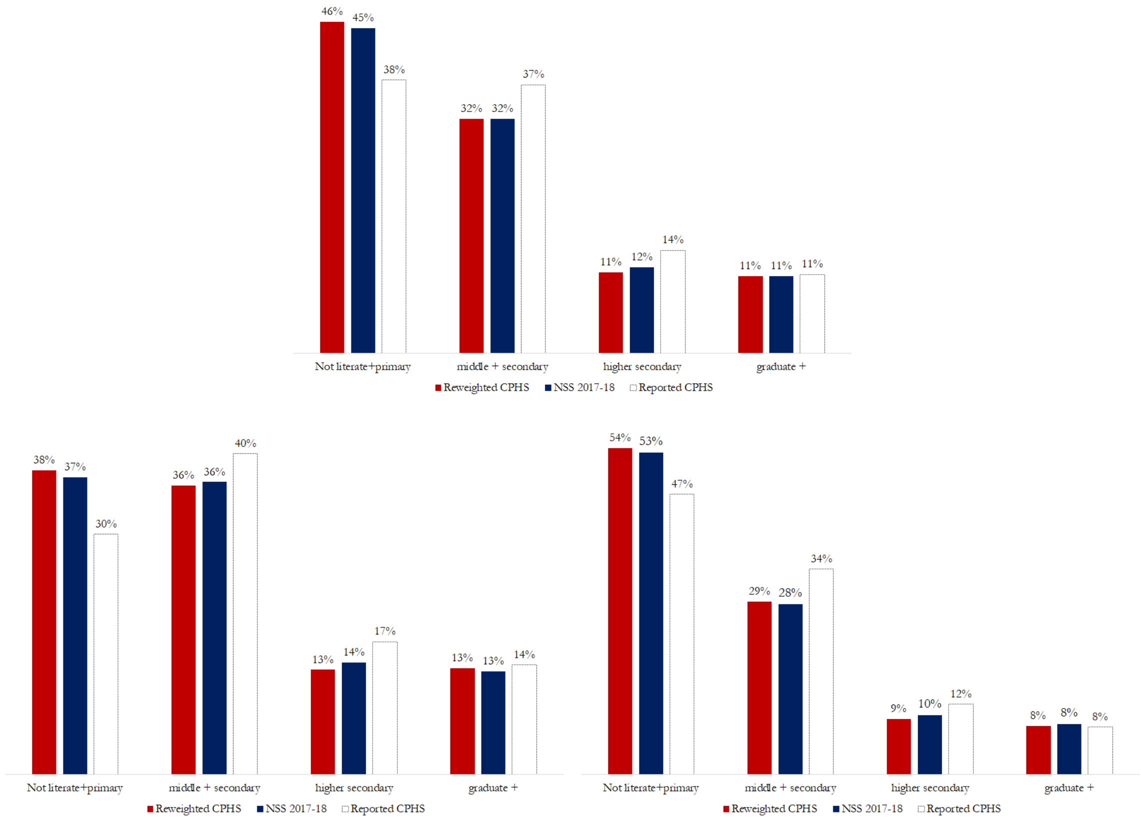


Figure 5: Comparison of education levels with NSS 75th round survey on education consumption (2017-18): All adults (panel (a); top), Male adults (panel (b); bottom-left), Female adults (panel (c); bottom-right)
Notes: Sample includes individuals over the age of 15. Individual level sampling weights used to produce weighted estimates in both surveys.

and PLFS (Figure 7). The Gini coefficient for salaried incomes (Panel (a)) obtained using the adjusted CPHS weights closely approximates the PLFS values for 2017 and 2018 . Despite a three-basis point inequality difference between the two surveys in 2019, reweighting corrects the divergent trend in earnings inequality for that year. Casual wage inequality (Panel (b)) is about four-basis points higher in the CPHS compared to the PLFS for all years. The reweighted series nonetheless helps align the annual trends in casual wage inequality between the CPHS and the PLFS. Gaps in casual wages inequality (after reweighting) are higher in rural areas. Figure 2 in Appendix 1.2 suggests that the gap in casual wage inequality is largely due to differences at lower deciles of daily wage income, especially in 2019. The deciles of salaried incomes for the reweighted CPHS and PLFS are seen to be close to each other.

Figure 3 and Figure 5 in appendices 1.3 and 1.5 compare estimates of other labor market indicators such as labor force participation rates (LFPR), worker population

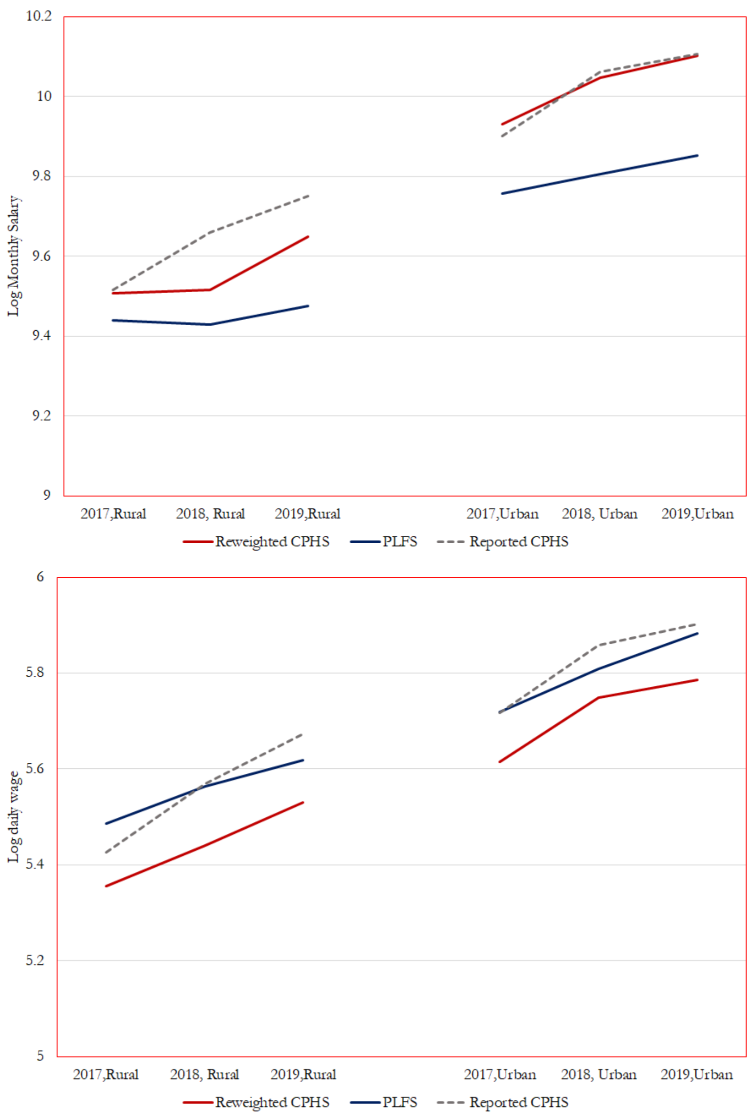


Figure 6: Comparison of average monthly salaries (panel (a); top) and daily wages (panel (b); bottom) across CPHS and PLFS
Notes: Monthly salaries and daily wages are in log nominal terms. Sample in both surveys include households with non-zero salaries and wages. Salaries and wages from PLFS are based on all visits made to the household. The red outline shows that indicators of wage income were not included in the set of targeting variables used for reweighting.

rates (WPR), and workforce composition.[20] For all of these indicators, reweighting largely resolves the biases that are observed with reported weights. This is expected as these indicators are included in the set of target variables. The bias observed for female LFPR (Figure 4 of Appendix 1.4) is partially accounted for.

[20]LFPR and WPR are not included in the set of target variables for reweighting

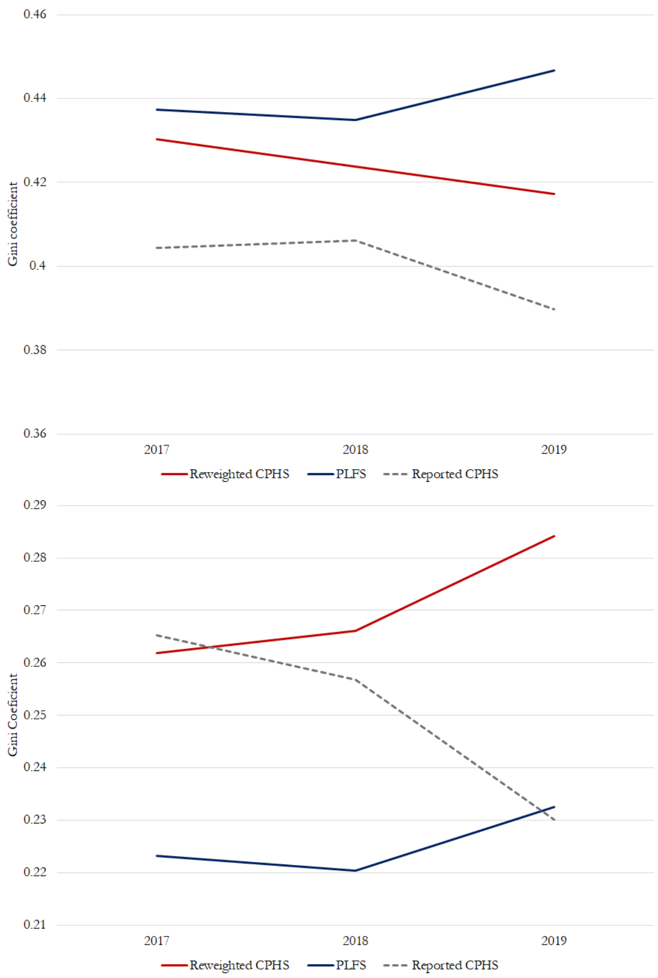


Figure 7: Inequality in monthly salaries and daily casual wages after reweighting: Salaried Workers (panel (a), top); Casual wage workers (panel (b), bottom)
Notes: Monthly salaries and daily wages are in nominal terms. Sample in both surveys include households with non-zero salaries and wages. Salaries and wages from PLFS are based on all visits made to the household. The red outline denotes that these variables were not included in the set of targeting variables used for reweighting.

3.2 Expenditure

Mean nominal consumption per capita obtained using reported CPHS weights is approximately 33 to 35 percent of private final consumption expenditure (PFCE) per capita from official national accounts (NAS). Similar fraction of consumption from survey to NAS (S-NA) is observed for the unreleased 2017 consumption expenditure survey. In comparison, S-NA share of the NSS-2011 consumption round was 41 percent (based on URP consumption aggregate). Nominal per capita consumption growth in the CPHS is higher than growth in nominal per capita PFCE reported in 2017, 2018 and 2019

(Table 1). The reverse is observed in 2016-17. The absence of a clear pattern could partly stem from the fact that data from national accounts are themselves a source of contention (see e.g. Subramanian, 2019 and Goyal and Kumar, 2020 for details).

In Figure 8, the variance of log consumption per capita in the CPHS is lower than the variance observed in the NSS-2011 (on average 0.267 in the CPHS compared to 0.368 in the NSS-2011). The gap in consumption inequality is larger in urban areas. The Gini coefficient of inequality obtained using reported CPHS weights would rank urban India at par with Sweden, the 25$^{\text{th}}$ most equitable country in the world. By comparison, the NSS-2011 would rank urban India around the 60th most unequal country in the world.

The third moment of the log consumption per capita distribution is also markedly lower in the CPHS when compared to the NSS-2011. Figure 9 compares the third moment between the two surveys for urban and rural separately [21]. The gap in the third moment is larger in urban India, and larger than the gaps observed for the second moment (Figure 8). The second and third moment of log per capita consumption in CPHS are on average about 27 and 70 percent lower than the respective moments from the NSS-2011.

Year	Mean per capita consumption expenditure (MPCE, nominal)	Private final consumption expenditure per capita (PFCE, nominal)	Growth in survey nominal MPCE	Growth in nominal PFCE per capita
2015-16	2193	6334		
2016-17	2315	7026	5.6%	10.9%
2017-18	2558	7638	10.5%	8.7%
2018-19	2846	8457	11.3%	10.7%
2019-20	3143	9179	10.4%	8.5%

Table 1: Comparison of levels and trends in nominal consumption per capita in CPHS and National Account Statistics (NAS).
Notes: Per capita consumption estimates are in nominal terms. Private final consumption expenditure is based on Statement 1.12 of national accounts statistics (NAS). The population estimates are also from NAS. Consumption per capita in CPHS is approximately 32 to 34 percent of PFCE per capita from NAS across years.

Comparing expenditure and non-expenditure statistics derived from the CPHS to

[21] Defined as $E[(x - E(x))^3]$ where x is the log consumption per capita

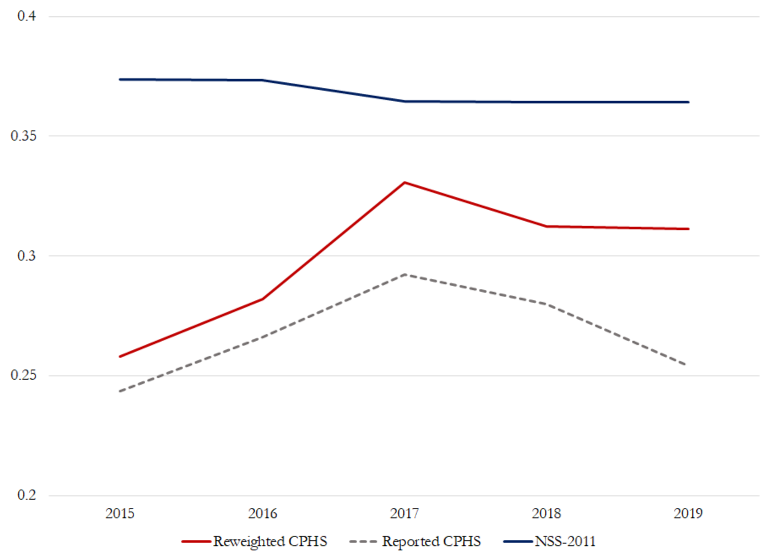


Figure 8: Variance of log consumption per capita
Notes: Consumption per capita is deflated using CPI-AL and IW for rural and urban areas. Sample includes districts that are common between CPHS and NSS-2011. The set of districts in CPHS have slightly evolved overtime. This causes a change in the geographic composition of samples overtime, resulting in small changes in the variance of log consumption in NSS-2011 overtime. All estimates are weighted by individual level sampling weights.

those obtained from nationally representative benchmark surveys confirms that: (1) the CPHS arguably under-represents the poorest as well as the richest households in the population; and (2) the under-coverage of the poor and the rich is more pronounced in urban areas, despite a larger sample of urban households in the CPHS compared to other nationally representative surveys. Pais and Rawal (2021) surmise that the absence of a sampling frame and biased selection of households within primary sampling units of CPHS could be a source of these discrepancies.

Comparing log consumption per capita using reweighted CPHS and NSS-2011, we obtain the following stylized facts:

Variance of log consumption per capita in the CPHS is lower than the variance in the NSS; reweighting helps reduce this gap but does not fully close it. The variance of log consumption per capita obtained using reported CPHS weights is 27 percent lower than the variance of log consumption from the NSS-2011 (Figure 8). This gap in variance is reduced to 19 percent after reweighting, which is consistent with the corrections we observed for education and asset ownership etc. Despite this improvement, a 19 percent gap represents a considerable discrepancy between the two surveys. Furthermore, the gap is larger in urban areas (log consumption variance in urban and rural using adjusted CPHS weights is 23 and 4 percent lower than what is observed in the NSS-2011).

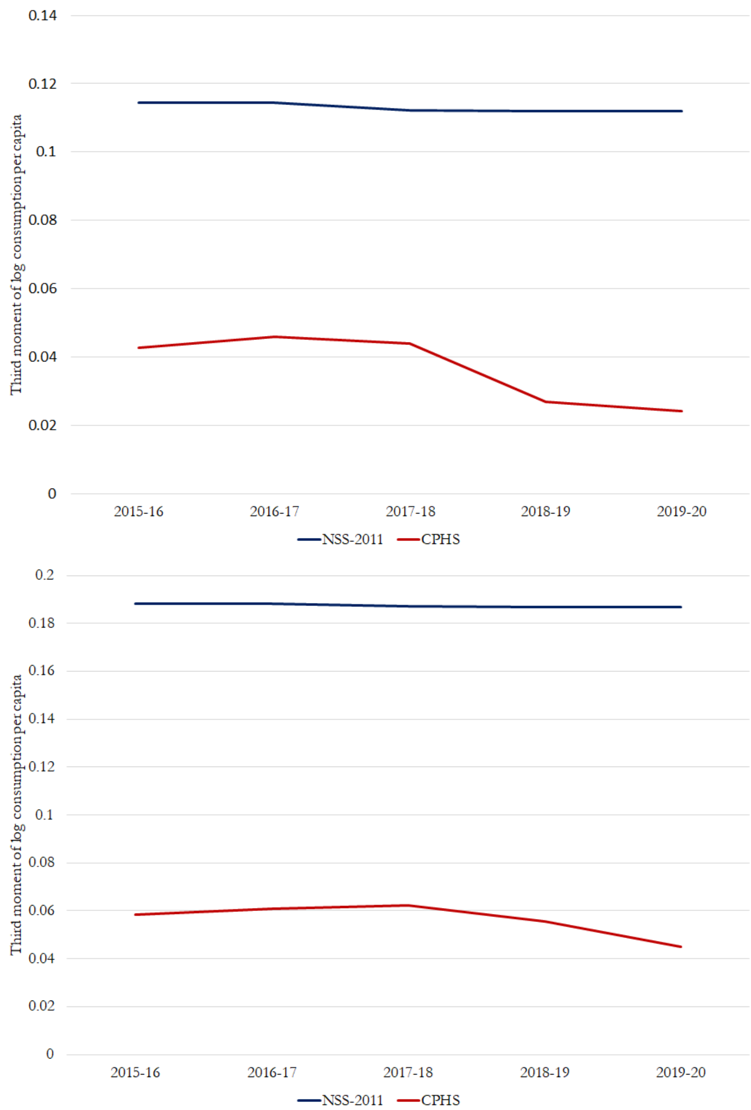


Figure 9: Third moment of log consumption per capita using reported CPHS: Rural (panel (a); top) and Urban (panel (b); bottom)
Notes: Estimates are constructed using reported people weights in CPHS and NSS. The third moment of log consumption per capita is much lower in reported CPHS than NSS-2011. The gaps in the third moments are much bigger than the second moment and are larger for urban than rural areas.

The third moment of the log consumption (per capita) distribution in the CPHS too is lower than the third moment observed in the NSS; and reweighting does little to close this gap. The third moment of log consumption per capita obtained using adjusted CPHS weights is 63 percent lower than the third moment from the NSS-2011 (Figure 10). Figure 11 shows that the third moment in the

CPHS is closer to zero than any other NSS consumption expenditure survey conducted over the past 35 years. The distribution of log consumption per capita from the CPHS is notably closer to a normal distribution while the consumption in NSS is observed to be closer to a non-normal distribution. The gap in the third moment between the two surveys is found to be larger than the gap that is observed for the variance. For both moments, the gaps are most notable for urban India.

The third moment of log consumption observed in the NSS is remarkably stable over time (most notably after 2004). Figure 11 shows that this is true for both urban and rural areas. The stability of the third moment across years is observed despite difference in recall periods used in various NSS survey rounds over the years. A similarly stable pattern is also observed for the fourth moment of log consumption per capita (not reported here).

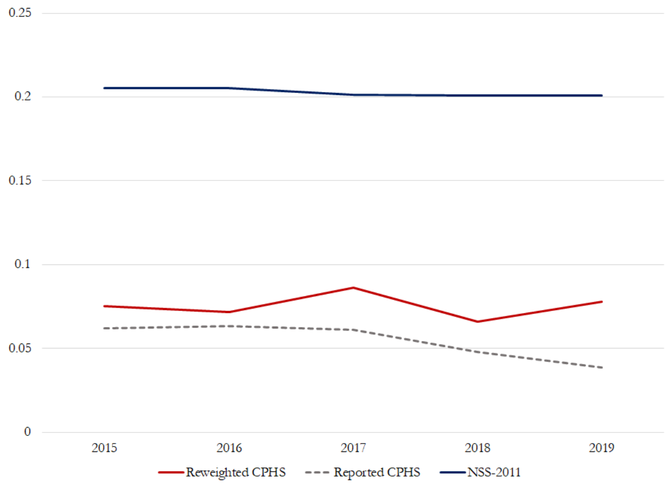


Figure 10: Third moment of log consumption per capita
Notes: Consumption per capita is deflated using CPI-AL and IW for rural and urban areas. Sample includes districts that are common between CPHS and NSS-2011. The set of districts in CPHS have slightly evolved overtime. This causes a change in the geographic composition of samples overtime, resulting in small changes in the variance of log consumption in NSS-2011 overtime. All estimates are weighted by individual level sampling weights.

Figures 8 and 10 show that there is a significant increase in the second and third moment of log CPHS-consumption in 2017 that is not ironed out by re-weighting. This spike stands out relative to the year-on-year fluctuations observed after 2017, which are notably smaller. It follows that the increase in CPHS-consumption dispersion in 2017 coincides with an approximately 20 percent expansion of the sampled districts in the third wave of 2017 (Figure 12). The newly added districts are disproportionately from poorer rural areas of India. Consequently, the standard deviation of log consumption

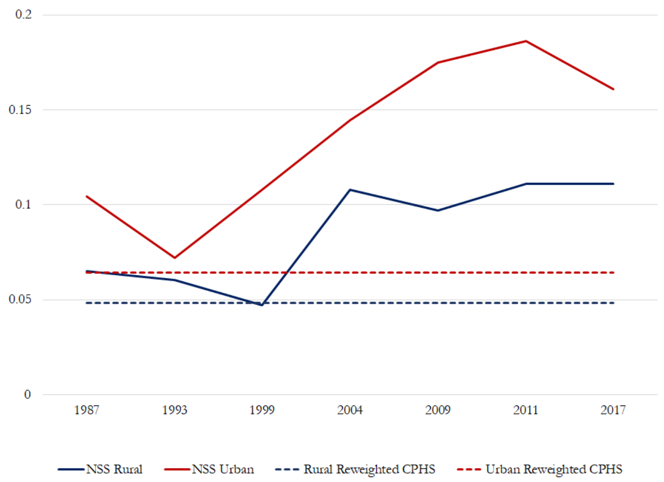


Figure 11: Third moment of log consumption per capita based on reweighted CPHS and 35 years of NSS consumption expenditure survey rounds
Notes: The third moment is calculated using real consumption per capita deflated using CPI-AL for rural and CPI-IW for urban samples. Urban deflators for years prior to 2001 are based on Povcalnet's India deflators provided at http://iresearch.worldbank.org/PovcalNet/Docs/CountryDocs/IND.htm#. The third moment of consumption for 2017 is derived from fractiles of state rural and urban consumption reported in the leaked survey report of NSS-2017.

per capita increased from 0.525 before the 2017-wave 3 to 0.560 after the expansion, while the third moment increased from 0.069 to 0.082. The implications of these changes for poverty and inequality estimation are reviewed in Section 4.2 and Appendix 3.3.

4 Two approaches to measuring poverty and inequality using the CPHS

4.1 Approach 1

Model

Approach 1 imputes NSS-type household consumption into the CPHS using predictors of household consumption that are available in both surveys. Let y_i measure NSS consumption expenditure for household i and let z_i be a vector of household characteristics (shared between the NSS and CPHS) that will serve as predictors of NSS-consumption. Assume that the relationship between log NSS-consumption and the household's characteristics (which will also be referred to as the consumption model) satisfies:

$$\log y_i = c + \beta z_i + u_i, \tag{1}$$

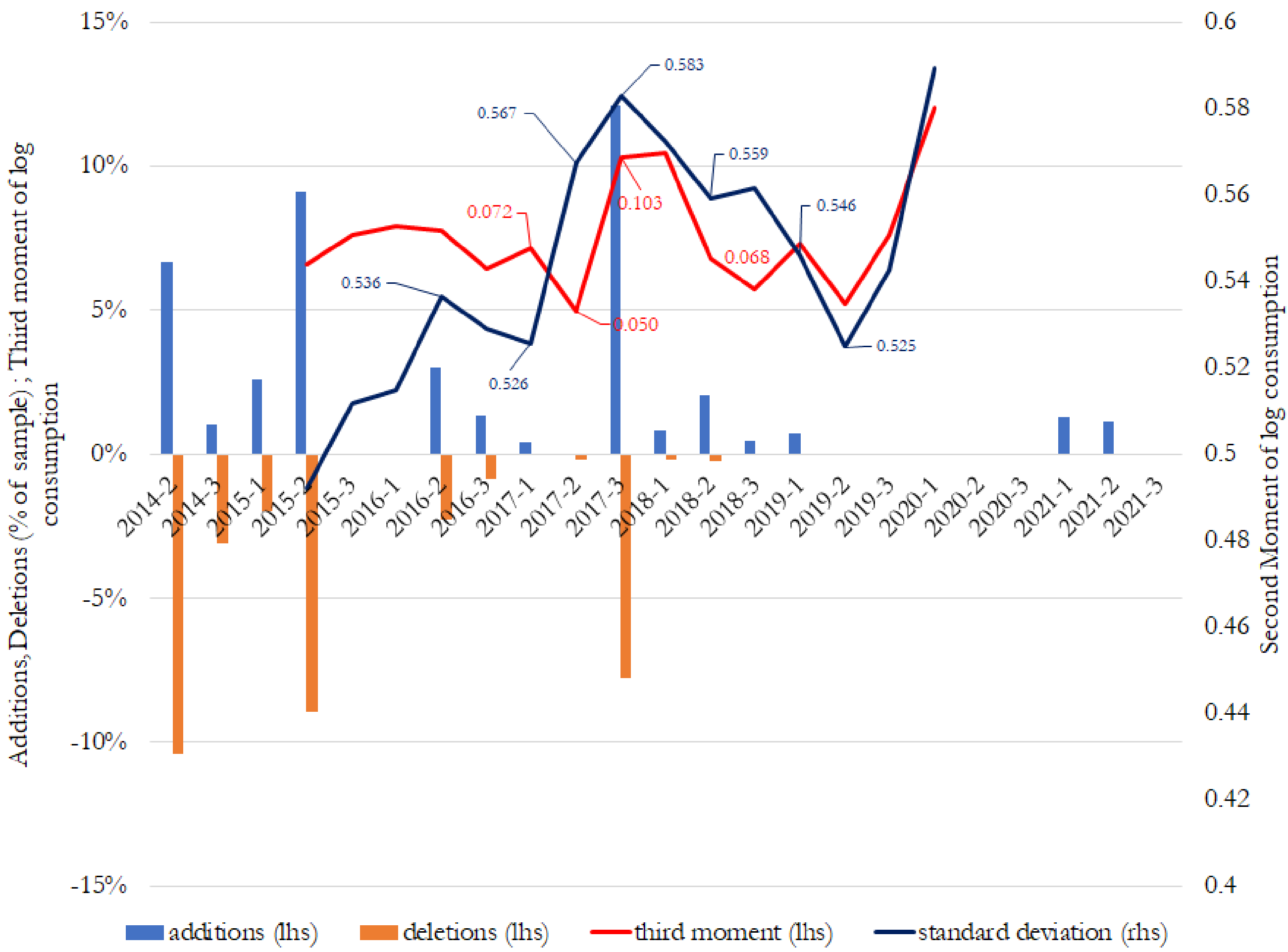


Figure 12: Net sample additions and the second and third moment of log consumption per capita by wave

Notes: The wave-wise moments of log consumption per capita are constructed using wave-level consumption vectors and the adjusted weights for the whole year. For instance, the moments for second and third wave of 2015 and the first wave of 2016 in the figure are calculated using the adjusted weights for 2015-16, as outlined in section 2.5. Weights for other waves are similarly based on adjusted weights of respective years. Note that the standard deviation is plotted using the secondary vertical axis.

where u_i is an independent identically distributed error term with mean zero. No further assumptions are made about the distribution of u_i.

The candidate set of predictors that are available in both the CPHS and NSS include household demographics, education, employment, asset ownership variables and consumption dummies. The latter dummy variables are derived from observed expenditures on selected categories, such as: (i) Clothing, footwear, accessories; (ii) Books, newspapers, stationery, tuition, hobbies; (iii) Furniture and fixtures; and, (iv) Cooking and household appliances. The dummy for a given category equals 1 if the household spent a non-zero amount on items from that category, and 0 otherwise. The items represent goods that are more likely to be dropped from (included in) a household's consumption basket when the household is subjected to negative (positive) income shocks, thereby improving the model's ability to capture temporal changes in economic conditions. Figure 6 in Appendix 2.1 examines the evolution of premium good consumption in CPHS overtime.

Implementation

The consumption model is estimated using data from the NSS and then applied to impute NSS-type consumption into the CPHS. Success of this approach is contingent on: (a) model stability (i.e., the model estimated in 2011 continues to apply in the years for which the CPHS is available), (b) sufficient predictive power of the model (i.e., the predictors are sufficiently correlated with household consumption), and that (c) the predictors are consistently measured between the two surveys. The analysis presented in section 3 confirms that the levels and trends in demographics, education and asset ownership observed in the (reweighted) CPHS are consistent with those observed in the nationally representative benchmark surveys. [22]

The regression model, estimated separately for urban and rural India, is shown in Table 2 (the coefficients related to principal industry of occupation is suppressed for formatting purposes). The urban model fits the data better when compared to the rural model, which is consistent with consumption models estimated to data from other countries (e.g. Douidich, et al., 2016). Overall, families with higher share of dependents (members below the ages of 18 and above the age of 61) are associated with lower consumption per capita, while households with more educated members and greater ownership of assets are associated with higher per capita consumption.

	(1)	(1)
Dependent variable: Log consumption per capita	Rural	Urban
1-member household	0.74***	0.99***
	(38.10)	(56.03)
2-member household	0.53***	0.64***
	(46.90)	(47.17)

[22]Figure 7 of the Appendix 2.2 shows the share of principal industry codes of households are also consistent across NSS-2011 and CPHS. Share of households with agriculture as the principal industry code are excluded in the graph for ease of representation: 39.1 percent of households in NSS-2011 and 33.1 percent of households (averaged across years) in CPHS belong to this category. In NSS-2011, principal industry code refers to the industry from which the households obtained their maximum income. In CPHS, we construct this variable based on the industry code of the household head. Households with missing principal industry code (due to head of household being unemployed or no member of the household being active in the labor market) are set to zero.

3-member household	0.37***	0.45***
	(45.73)	(46.13)
4-member household	0.24***	0.29***
	(38.64)	(38.89)
5-member household	0.12***	0.14***
	(22.63)	(20.35)
Multigeneration family	-0.00	0.01
	(-0.94)	(1.29)
Extended family	0.03***	0.06***
	(3.90)	(7.59)
Share of 0 to 18 years old members in family	-0.18***	-0.22***
	(-21.21)	(-20.27)
Share of 61+ years old members in family	-0.04*	-0.03
	(-2.34)	(-1.50)
Female headed households	-0.04***	-0.04***
	(-6.12)	(-5.24)
Log (age of household head)	0.03***	-0.02
	(3.34)	(-1.71)
Any member with higher than middle to high school level of education	0.02***	0.03**
	(3.41)	(2.98)
Share of members with middle to high school level of education	0.12***	0.13***
	(12.49)	(10.95)
Any member with diploma to post graduate level of education	0.05***	0.05***
	(7.87)	(8.10)
Muslim household	0.03***	-0.02*
	(5.18)	(-2.29)
Christian household	0.09***	0.03*
	(5.54)	(2.05)
Sikh household	0.14***	0.03
	(9.41)	(1.50)
Jain household	0.07	-0.01
	(1.03)	(-0.26)
Buddhist household	-0.03	0.04
	(-1.07)	(1.75)
Zoroastrian and other religions	-0.07	0.07
	(-1.40)	(1.00)

Scheduled Castes	0.09***	0.01
	(12.09)	(0.63)
Other Backward Classes	0.16***	0.05***
	(23.47)	(3.77)
Other castes	0.19***	0.12***
	(24.93)	(8.82)
Electrified household	0.11***	0.15***
	(21.47)	(10.98)
Rented household	0.22***	0.25***
	(14.25)	(28.61)
Television owning household	0.17***	0.16***
	(35.56)	(19.77)
Air conditioner owning household	0.08***	0.05***
	(9.55)	(8.00)
Washing machine owning household	0.08***	0.17***
	(6.33)	(23.99)
Refrigerator owning household	0.24***	0.23***
	(32.50)	(37.41)
Car owning household	0.15***	0.30***
	(12.11)	(32.03)
Computer owning household	0.23***	0.25***
	(14.37)	(32.63)
Household owns the homestead	-0.00	0.00
	(-0.33)	(0.19)
Inverter owning household	0.13***	0.05***
	(9.67)	(5.57)
Dummy for Clothing, footwear, accessories	0.20***	0.14***
	(48.27)	(29.96)
Dummy for Books, newspapers, stationery, tuition, hobbies	0.08***	0.13***
	(20.68)	(23.83)
Dummy for Furniture and fixtures	0.24***	0.24***
	(29.90)	(20.71)
Dummy for Cooking and household appliances	0.13***	0.12***
	(14.78)	(15.01)
Constant	6.17***	6.46***
	(171.10)	(134.87)

Observations	41,915	31,923
R-squared	0.4674	0.6314

Table 2: Regression coefficients from the imputation model.
Notes: Standard errors in parentheses. * $p< 0.05$, ** $p < 0.01$, *** $p < 0.001$. Regressions are weighted by person level weights from respective surveys. Coefficients of harmonized industry codes are suppressed to keep the results tractable. The regression coefficients reported are based on a set of districts common between NSS-2011 and CPHS' 2015. As CPHS expanded to a few more districts in the following years, the set districts common to the two surveys expanded slightly resulting in slightly different regression coefficients across years.

The error term from the regression model is accounted for when imputing NSS-type consumption into the CPHS. Given the non-normality observed in the NSS, we follow Elbers, et al. (2003) by drawing the errors from the empirical residuals with equal probability (to preserve the empirical distribution for the errors observed in the NSS). Errors terms for households in the CPHS are standardized using the mean and standard deviation, multiplied by the root mean square error term and added to the predictions of the imputation model into CPHS.

Figure 13 compares the mean, variance, and third moments of the imputed (log) NSS-type consumption into the CPHS to the moments of observed (log) consumption from both the NSS and the CPHS. The means of imputed NSS-type consumption and observed CPHS consumption are nearly identical in rural areas. In urban areas, NSS-type consumption is on average approximately ten percentage points higher when compared to observed CPHS consumption. This suggests that the CPHS under-estimates consumption in urban India (consistent with observations made in Dhingra and Ghatak, 2021).

The variance of the imputed NSS-type consumption is seen to match the variance of observed NSS-2011 consumption in both rural and urban India, i.e. the use of imputed consumption and adjusted CPHS weights fully closes the gap in variance between the two surveys. Unfortunately, this does not extend to higher moments. While the use of imputed NSS-type consumption in the CPHS helps reduce the gap in third moments (compared to observed log consumption in the NSS), the remaining gap is economically significant and will bias estimates of poverty and inequality if not addressed. This motivates our second approach which is outlined next.

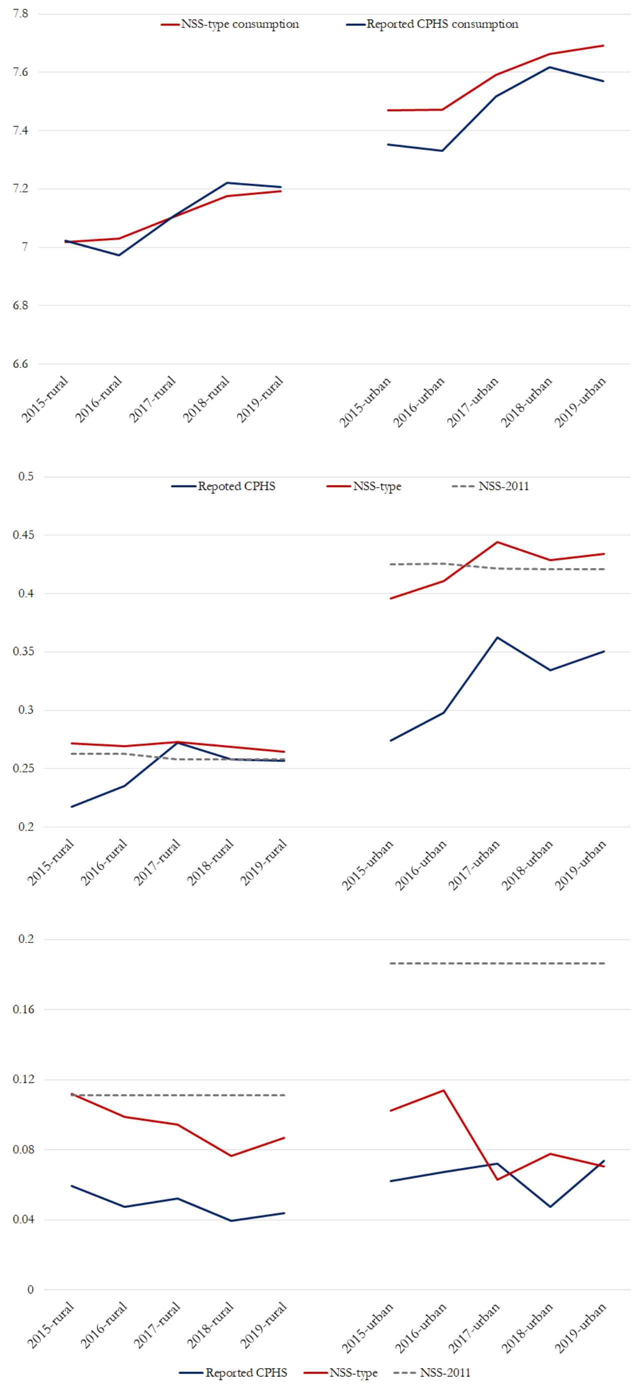


Figure 13: Three moments of log consumption per capita: Mean (panel (a); top), Variance (panel (b); middle), Third moment (panel (c); bottom)
Notes: NSS-type consumption is obtained using non-expenditure variables in CPHS and the regression coefficients reported in Table 2. All estimates are based on reweighted individual level weights. Consumption is in real terms deflated using CPI-AL and IW for rural and urban areas. All three moments are calculated using log real consumption per capita.

4.2 Approach 2

Model

Approach 2 uses a single predictor to impute NSS-type consumption into the CPHS, namely observed CPHS consumption, a variable that is arguably highly predictive of NSS-type consumption, but which is entirely ignored in approach 1. In other words, in this approach we will convert the observed CPHS consumption into NSS-type consumption. Let CPHS-consumption expenditure for household i be denoted by x_i. Section 3.2 establishes the following stylized facts: (a) The variance of NSS log consumption is higher than the variance of CPHS log consumption. The re-calibration of the survey weights has reduced this gap in the second moment, but some gap still remains, and (b) CPHS log-consumption is near normally distributed, while NSS log consumption shows a more marked deviation from normality. Specifically, the third moment of NSS log consumption is approximately twice the third moment of CPHS log consumption. (A similar ordering applies to the fourth moment.)

To accommodates the above-mentioned stylized facts, consider a model where CPHS log consumption is described as a linear combination of NSS log consumption and a normally distributed error term:

$$\log x_i = a + b \log y_i + \sigma \varepsilon_i, \tag{2}$$

where ε_i is an independent identically distributed error term with mean zero and unit variance. In practice we do not observe y_i and x_i for the same household i given that the two measures of consumption come from different cross-sectional surveys with their own samples of households that cannot be linked. Accordingly, the model that describes the relationship between the two cannot be estimated using standard regression analysis (which is the reason why observed CPHS consumption was excluded as a predictor in approach 1). Instead, the parameters a, b, and σ will be estimated using method of moments.

A minimum of three moment conditions will be required. The first three moments of the log consumption distribution are natural candidates. The mean and variance of both sides of eq. (2) solve:

$$\mu_x = a + b\mu_y \tag{3}$$

$$\sigma_x^2 = b^2\sigma_y^2 + \sigma^2, \tag{4}$$

where μ_q and σ_q^2 evaluate the mean and variance of the variable q, respectively. At this point we have two moment conditions and three unknown parameters, meaning that a third moment condition is required to obtain identification. For the third moment, we obtain:

$$
\begin{aligned}
(\log x_i - \mu_x)^3 \quad = \quad & b^2 (\log y_i - \mu_y)^2 [b (\log y_i - \mu_y) + \sigma\varepsilon_i] \\
& + \sigma^2 \varepsilon_i^2 [b (\log y_i - \mu_y) + \sigma\varepsilon_i] \\
& + 2b\sigma\varepsilon_i (\log y_i - \mu_y) [b (\log y_i - \mu_y) + \sigma\varepsilon_i] .
\end{aligned}
$$

The first two moments do not require any assumption about the distributional form of ε. Identification through the third moment, however, rests on the non-normality of the log consumption distributions.

Assumption 1 *Assume that* ε_i *is normally distributed, and that* $\log x_i$ *and* $\log y_i$ *are non-normally distributed.*

Under Assumption 1, we have $E[\varepsilon_i^3] = 0$, while $E[(\log y_i - \mu_y)^3]$ and $E[(\log x_i - \mu_x)^3]$ are presumably non-zero. It is furthermore assumed that ε_i is uncorrelated with $\log x_i$. This similarly opens the door for identification. It follows that:

$$
E\left[(\log x_i - \mu_x)^3\right] = b^3 E\left[(\log y_i - \mu_y)^3\right], \tag{5}
$$

since $E[(\log y_i - \mu_y)] = E[\varepsilon_i^3] = 0$. This yields the following estimator for b:

$$
b^3 = \frac{E[(\log x_i - \mu_x)^3]}{E[(\log y_i - \mu_y)^3]}. \tag{6}
$$

Note that identification fails when log incomes are normally distributed, in which case $E[(\log y_i - \mu_y)^3] = E[(\log x_i - \mu_x)^3] = 0$. Given the estimate for b, estimates of a and σ^2 can be obtained by solving equations (3) and (4):

$$
\begin{aligned}
a &= \mu_x - b\mu_y \\
\sigma^2 &= \sigma_x^2 - b^2 \sigma_y^2.
\end{aligned}
$$

It will be convenient to re-arrange the model as follows:

$$
\frac{\log x_i - a}{b} = \log \tilde{x}_i = \log y_i + \left(\frac{\sigma}{b}\right) \varepsilon_i. \tag{7}
$$

Given estimates for a, b, and σ, we can treat $\log \tilde{x}_i$ as observed data.

The next challenge is to extract a drawing for $\log y_i$ given an observed value for $\log \tilde{x}_i$. To this end, we assume that the distribution for $\log y_i$ can be described by a normal mixture distribution. Let the cumulative distribution function for NSS log consumption be denoted by F_y.

Assumption 2 *F_y can be represented by a normal mixture distribution of the form:*

$$F_y = \sum_j \pi_j F_j, \tag{8}$$

where F_j are normal distribution functions with mean m_j and variance s_j^2, and where π_j are non-negative mixing probabilities that sum up to 1.

Under Assumption 2, the distribution for $\log \tilde{x}_i$ denoted by G_x can also be represented by a normal mixture distribution. It follows that:

$$G_x = \sum_j \pi_j G_j, \tag{9}$$

where G_j are normal distribution functions with mean m_j and variance $\nu_j = s_j^2 + \sigma^2/b^2$.

Since $\log \tilde{x}_i$ is observed, the normal mixture distribution G_x can readily be estimated (see for example the FMM package in Stata). This gives us estimates for π_j, m_j and ν_j. Note that this also identifies two-thirds of the parameters of F_y (as the parameters π_j and m_j are shared between F_y and G_x). To fully identify F_y, we also need estimates for s_j^2, which can be obtained by combining estimates for ν_j with the estimates for σ^2 and b, as: $s_j^2 = \nu_j - \sigma^2/b^2$ (provided that $\nu_j > \sigma^2/b^2$; if this condition is violated, we could reduce the number of components by one until all mixture components satisfy this condition).

At this point we have an estimate of the unconditional distribution F_y for NSS log consumption $\log y_i$. What we really want is an estimate of the distribution for $\log y_i$ conditional on the observation of CPHS log consumption $\log \tilde{x}_i$ for household i. Let us denote this conditional distribution by $F_{y|x}$. It follows that $F_{y|x}$ is also a normal mixture distribution (see e.g. Elbers and van der Weide, 2014), i.e. $F_{y|x}$ satisfies $F_{y|x} = \sum_j \alpha_j F_{j|x}$, where $F_{j|x}$ are normal distribution functions with mean $m_{j|x}$ and variance $s_{j|x}^2$. Lemma 2 from Elbers and van der Weide (2014) shows that the parameters that define $F_{y|x}$ can be derived from the parameters of the normal mixture F_y and the estimate for $\tilde{\sigma}^2 = \sigma^2/b^2$:

$$\begin{aligned}
m_{j|x_i} &= (1-\gamma_j)m_j + \gamma_j \log \tilde{x}_i \\
s_{j|x_i}^2 &= \left(\frac{1}{s_j^2} + \frac{1}{\tilde{\sigma}^2} \right)^{-1} \\
\alpha_j &= \tilde{\alpha}_j / \sum_j \tilde{\alpha}_j,
\end{aligned}$$

with:

$$\begin{aligned}\gamma_j &= \frac{s_j^2}{s_j^2+\tilde{\sigma}^2}\\ \tilde{\alpha}_j &= \pi_j\varphi\left(\log\tilde{x}_i; m_j, s_j^2+\tilde{\sigma}^2\right),\end{aligned}$$

where $\varphi(x;m,v)$ is a normal density function with mean m and variance v evaluated at the value x. Note that when the variance of the error term tends to zero (i.e. $\tilde{\sigma}^2 \to 0$), the conditional mean $E[\log y_i|\log\tilde{x}_i]$ will tend to $\log\tilde{x}_i$ while the conditional variance will tend to zero, as they should.

A practical way to proceed is to draw an observation of NSS log consumption from the conditional distribution $F_{y|x}$ for each household, and evaluate the welfare measures of interest. We draw 50 observations of NSS-type log consumption for each household in the CPHS sample, and then compute the aggregate welfare indicator (i.e. poverty and inequality) for each $k=1,\ldots,50$. The mean and standard deviation evaluated over the K realizations will serve as the point estimate and standard error of the welfare indicator.

Alternatively, when measuring head-count poverty for example, one could evaluate for each household the probability that their NSS log consumption is below the poverty line conditional on the observation of their CPHS log consumption value -- and then compute the mean value of these probabilities across all households in the sample. Let the poverty line for log consumption be denoted by z. The probability that household i is poor equals:

$$H_i = \sum_j \alpha_j\Phi\left(\frac{z-m_{j|x_i}}{s_{j|x_i}}\right), \tag{10}$$

where Φ is the standard normal distribution function. Head-count poverty can then be estimated by:

$$H = \sum_i w_i H_i, \tag{11}$$

where w_i denote survey weights that are assumed to sum up to 1.

Implementation

The assumed model (see eq. 2) contains three parameters: a, b, and σ^2. As described above, a minimum of three moments (for both the NSS and CPHS log consumption data) are required to estimate all three of these parameters. All three moments of the CPHS log consumption distribution can readily be estimated using the observed CPHS consumption data. Estimation of the moments from the NSS consumption distributed is complicated by the fact that there is no NSS survey for the same moment in time for which we have CPHS. We have established however that the third moment of NSS consumption is remarkably stable over time, allowing us to use the third moment estimated to observed NSS consumption from the NSS-2011. For the second moment, we consider two options, namely estimate it using (a) observed NSS consumption data from the NSS-2011, and (b) imputed NSS-type consumption in the CPHS (which we established does reasonably well in matching the second moment from the observed NSS log consumption data). The first moment (mean log consumption), which is the least stable moment over time, is obtained from the imputed NSS-type consumption data. The resulting estimates of the three parameters a, b, and σ^2 for the different years are shown in Figure 14.

The next step is to estimate the parameters of the unconditional distribution of NSS log consumption, which is assumed to follow a Normal Mixture distribution. Normal mixtures (NM) are very flexible. Two or three components are generally sufficient to closely fit any empirical distribution function underlying household consumption data.[23] In our case, it offers two practical advantages. First, it follows that the distribution of NSS log consumption conditional on CPHS log consumption too follows a NM distribution. Second, the parameters of the NMs associated with both the unconditional and conditional distribution of NSS log consumption can readily be derived from the parameters of the NM estimated to CPHS log consumption combined with the parameters governing the relationship between CPHS and NSS consumption (i.e. a, b, and σ^2).

We start by fitting a NM with three components for the unconditional NSS log consumption distribution. When the estimated variance of one or more of the components is negative, the number of components is reduced by one, until all components are estimated to have positive variance. See assumption 2 for details on the positive variance constraint (and why positive variance is not necessarily guaranteed). Negative variance estimates are only obtained for urban samples during 2018.

Once we have an estimate of the conditional distribution, we obtain 50 random draws of NSS consumption for each household in the CPHS sample (conditional on each household's CPHS consumption value). For each of the 50 realizations of NSS consumption data, we evaluate the corresponding poverty headcount rates and selected measures of inequality. The point estimates of poverty and inequality are obtained by averaging over the 50 different realizations.

When a new NSS household consumption survey becomes available, both NSS-consumption and CPHS-consumption can be observed for the same year (albeit in different surveys with their own sample of households). Accordingly, one could estimate

[23] To illustrate, we report the empirical goodness of fit for the mixed normal distributions for the years 2015 and 2019 in Figure 8 of Appendix 3.1

all three moments of NSS log consumption using the observed data and adopt our method of moments estimator to obtain estimates of a, b, and σ^2 for that year -- and subsequently assume that all three parameters remain constant over time until the next NSS household consumption survey becomes available (which is when the first three moments derived from observed consumption data can be updated). Alternatively, one could continue to adopt the version of Approach 2 we are currently using, namely estimate moments that are found to be comparatively stable over time from observed household (log) consumption data and estimate moments that are found to be less stable from up-to-date imputed consumption data. The latter (and currently adopted) approach may be preferred when the CPHS sample is subjected to notable changes that may significant introduce changes in moments that are not accounted for by re-weighting. See Appendix 3.3 for a further discussion on the changes made to the CPHS sample (most notably during the third wave of 2017) and its implication for our method of estimation.

On the choice between Approaches 1 and 2, it should be noted that the two approaches rely on their own set of assumptions. The validity of these assumptions will be context-specific and may vary over time. Approach 1 assumes that the relationship between NSS-consumption and household characteristics such as demographics, education, and employment is stable over time, while Approach 2 assumes that the relationship between NSS-consumption and CPHS-consumption is stable over time. Where possible one should implement both approaches (thereby considering different assumptions) and inspect robustness. Appendix 3.2 compares the relative ranking of households based on their observed CPHS consumption and imputed consumption based on approach 1 of section 4.1 and approach 2 of section 4.2.

5 Results

5.1 Main estimates of poverty and inequality

Both approaches yield qualitatively similar levels and trends in headcount poverty estimated at the $1.90 line: poverty is about 12.3 percentage points lower in 2019 than 2011 (see Figure 15). Estimates of poverty obtained using observed CPHS consumption data are seen to be up to 3.5 percentage points higher when compared to estimates obtained using NSS-compatible measures of consumption. By the same token, our estimates of poverty are notably higher than previous estimates obtained by the World Bank's Povcalnet database and other scholars, see e.g. Edochie, et al. (2022); Newhouse and Vyas (2019) and Gupta, Malani and Woda (2021b). Estimates from World Bank's

Povcalnet are included in Figure 15 for comparison. The projections in Povcalnet are extrapolated using the consumption distribution of NSS-2011 and applying the growth in private final consumption expenditure observed in national accounts. The method therefore assumes that inequality has remained unchanged since the NSS-2011[24]. We compare our approach to Newhouse and Vyas (2019) and Edochie, et al.(2022) in Section 5.2 and reflect on the potential reasons for why their estimates are lower. Gupta, Malani and Woda (2021b) use the raw CPHS data to construct headcounts for 2019 and the post-pandemic period; our reservations with this approach are documented in Section 3.

The rate of poverty reduction between 2004 and 2011 is estimated at approximately 2.5 percentage points per year. After 2011 poverty reduction has slowed down. By our estimates, poverty has declined by an average of 1.3 percentage points per year between 2011 and 2018. It should be noted that at lower levels of poverty, it would take increasingly larger rates of consumption growth and/or reductions in inequality to sustain the high rates of poverty reduction (e.g. Bourguignon, 2003).

Figure 12 in Appendix 4.1 dis-aggregates the trends in poverty by rural and urban. Three observations stand out. First, rural poverty in 2019 is 14.7 percentage points lower than in 2011 while urban poverty reduced by 7.3 points over the same period. This is consistent with a continuation of the rural-urban poverty convergence observed over the past six decades (see Datt, Ravallion and Murgai, 2019).[25] Second, urban India experienced a churn in poverty trends around 2016. Urban poverty rose by 2 percentage points in that year followed by a rapid rise in consumption that drove poverty down by 3.2 percentage points in the following year. Third, the fastest poverty reduction occurred in the years 2017 and 2018. Thereafter, the rate of poverty reduction stalled considerably.

Headcount poverty rates at the international \$3.2 and \$5.5 poverty lines are shown in Figure 13 of Appendix 4.2. A similar reduction in poverty is observed for both lines. The average rate of poverty reduction at \$3.2 and \$5.5 was 2.1 and 0.8 percentage points per year between 2004 and 2011. By comparison, all years since 2015 clock an average rate of 1.2 and 0.6 percentage points poverty reduction per year relative 2011. The \$5.5 line also shows poverty rising between 2018 and 2019. This dynamic is detected in the consumption data but not by changes in demographic and asset levels. The rise is mainly on account of urban households where headcount rates rose by 2.5 percentage

[24]Povcalnet projections can allow for some changes to the distribution. For instance, the 2014.5 estimate employs a pass-through rate of 0.559 for urban and 0.733 for rural areas; see box 1.3 in World Bank (2018) and box 1.2 in World Bank (2020) for details. However, the distribution within rural and urban areas is assumed to be unchanged.

[25]Note that rural poverty reduction in the decade(s) prior to 2004 was more modest and heterogeneous, see e.g. Lanjouw and Murgai (2009) and Himanshu et al. (2013).

in that year.

Let us also inspect time-trends in inequality. Figure 16 shows our estimates of the Gini coefficient for the years under consideration. Both approaches are found to produce qualitatively similar results.[26] We observe a slight moderation in consumption inequality in India since 2011. This could in part be attributed to the fact that top-income households are under-represented in household surveys (whether NSS or CPHS). Consequently, consumption inequality estimated from household survey data capture distributional changes for households that are in the bottom 95 percent, say, of the distribution. To the extent that the income or consumption growth since 2011 is largely concentrated in the top end of the distribution (Chancel and Piketty, 2019), our household survey-based estimates of consumption inequality will be downward biased.

Figure 14 in Appendix 4.3 reveals that the moderation of inequality has been larger in rural than urban areas. Since 2015, changes in rural inequality have been less pronounced than urban areas. Urban inequality dropped in 2018 which coincides with the year in which the rate of poverty reduction was its highest. Figure 15 in Appendix 4.4 shows that other measures, namely, poverty gap and mean-log deviation yield trends in poverty and inequality dynamics that are consistent with the main results.

Finally, in Figure 17, we connect our estimates of poverty and inequality for India over the last decade with estimates dating back to 1993. It can be seen that our estimates of headcount poverty preserve the long-term trend of poverty reduction that is observed in India over this period. By the same token, our estimates suggest that the current poverty rate is higher than the forecasts based on pass-through adjusted consumption growth from national accounts (under the assumption of distribution neutrality). For consumption inequality we observe a trend reversal around 2011 (see Figure 18). Inequality is estimated to have steadily increased between 1993 and 2011. By our estimates inequality has started to moderate after 2011.

5.2 Robustness analysis

Our preferred specification in approach 2 assumes a linear relationship between observed CPHS consumption and NSS consumption. We allow for heterogeneity (i.e. different relationships) between urban and rural India. It is possible however, that there are additional heterogeneities that should be accounted for. For instance, Gibson and Kim (2007) observe that the measurement errors in household consumption are systematically correlated with household size. Similarly, Beegle, et al. (2012) find that

[26]The inequality based on reported CPHS consumption range between 0.2965 and 0.3213 across years (not included in the figure) -- considerably lower than the estimates of inequality obtained using NSS-type consumption measures.

in addition to household size, the number of adults in the household, the education level of the household head and asset ownership levels can induce systematic differences between different measures of household consumption.

To test whether any potentially important heterogeneities are overlooked by our preferred specification, we allow the linear relationship between CPHS and NSS consumption to vary by these household characteristics. We consider six binary household level indicators: households with more than three adults, households with at least one member with a high level of education, household head with over primary levels of education, households with agriculture as the primary industry, Hindu households, and households that belong to schedule caste, schedule tribe or other backward classes. Each of these will be combined with the rural-urban indicator, such that four different linear relationship are estimated for each of these six cases.

Figure 19 plots the headcount poverty rate at the $1.90 line for each of the six specifications -- each accounting for a different choice of heterogeneity (labeled as the "heterogeneous" series). The "homogenous" series refers to our main specification that only accounts for heterogeneity between rural and urban India. All six specifications, each accounting for a different form of heterogeneity, produces similar levels and trends in headcount poverty than the estimates obtained with our preferred specification. The one outlier is the headcount estimate obtained for 2018 that accounts for heterogeneity in household head literacy.

We can further check the robustness of our imputation model of approach 1 by estimating poverty in 2004 and comparing it to the actual estimates for the year. This "back casting" exercise generates poverty figures for 2004 based on the estimated coefficients in Table 2 and imputing consumption for 2004 based on the NSS consumption round for the year. The back casted estimates can also help compare our approach to those of Newhouse and Vyas (2019) and Edochie, et al. (2022). As all three papers use the same training and validation dataset (NSS-2011 and 2004 respectively), these comparisons can reveal the accuracy of prediction across papers.

Figure 14: Parameters for method of matching moments: a (panel (a); top), b (panel (b); middle), σ^2 (panel (c); bottom)

Notes: $b = (\text{third moment}_{\text{cphs}}/\text{third moment}_{\text{nss}})^{1/3}$. *Parameter* b_{2011} *and* b_{2017} *are based on the third moments of log consumption from NSS-2011 and NSS-2017 respectively.* $a = \mu_{\text{cphs}} - b * \mu_{\text{nss}}$. $a_{t(=2011 \text{ or } 2017)}$ *is calculated using* b_t *and the mean of imputed log consumption from approach 1.* $s^2 = \sigma^2 = \sigma^2_{\text{cphs}} - b^2\sigma^2_{\text{nss}}$. *Parameter* $s^2_{t(=2011 \text{ or } 2017)}$ uses the variance of log consumption *from imputed NSS-type consumption and the corresponding* b_t. *All consumption values are in real terms and deflated using CPI-AL and CPI-IW for rural and urban samples.*

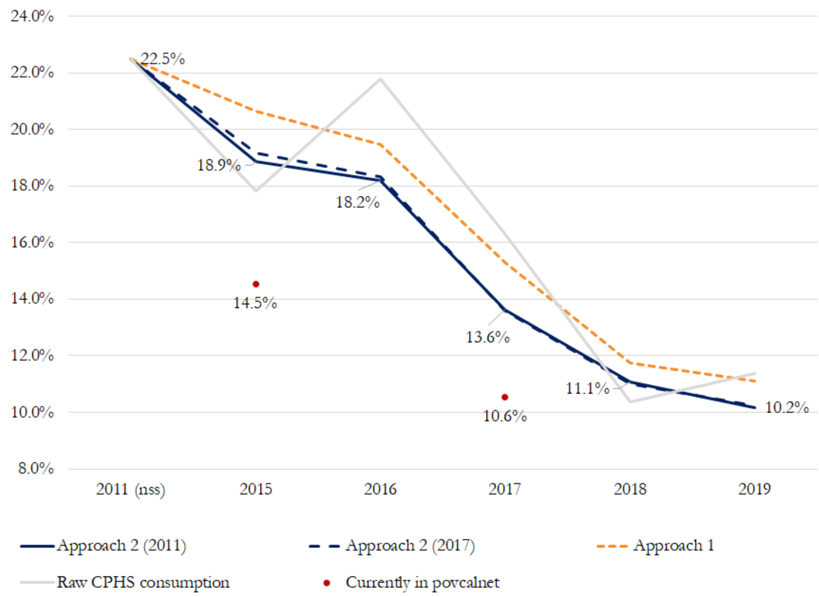


Figure 15: Headcount poverty estimates at the $1.90 line
Notes: Refer to section 4.1 and 4.2 for details on Approach 1 and Approach 2 respectively. Estimates currently in Povcalnet are based on the line-up method: growth in real HFCE from national accounts statistics is multiplied by a pass-through rate and applied to NSS-2011 consumption distribution. The Povcalnet estimates denoted in the figure are for the corresponding calendar years. The equivalent estimate for the financial years are: 15.8 percent for 2015-16 and 9.8 percent for 2017-18.

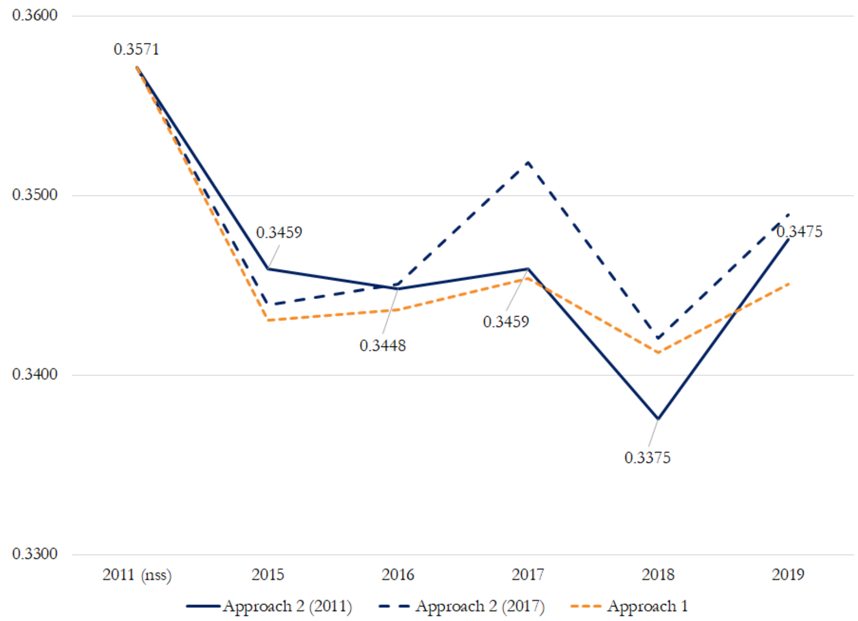


Figure 16: Gini measure of inequality
Notes: Refer to section 4.1 and 4.2 for details on Approach 1 and Approach 2 respectively. Gini measure of inequality is calculated using PPP adjusted household consumption. PPP exchange rate of 13.173 and 16.017 , updated as of May 2020, are used for rural and urban areas. distribution.

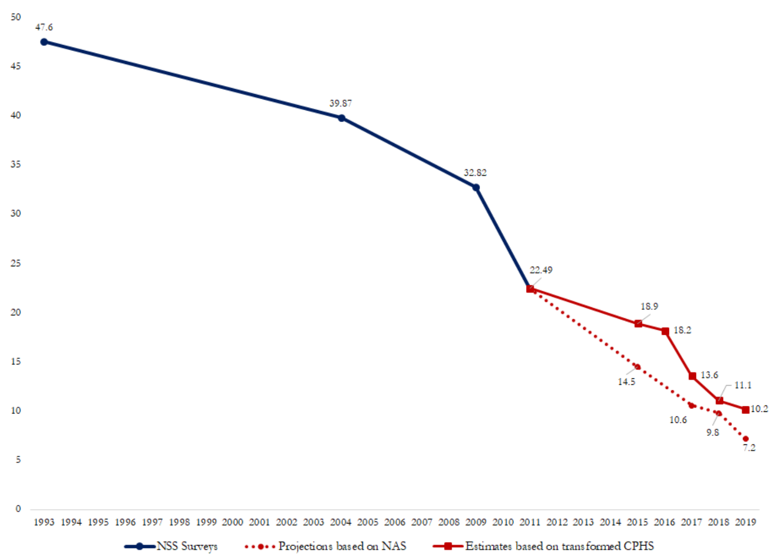


Figure 17: Poverty Headcount at $1.90 line
Notes: "NSS survey" denotes estimates based on NSS survey rounds; "Projections based on NAS" pass-through adjusted consumption growth from national accounts; and, "Estimates based on transformed CPHS" are based on Approach 2 (2011) of this paper.

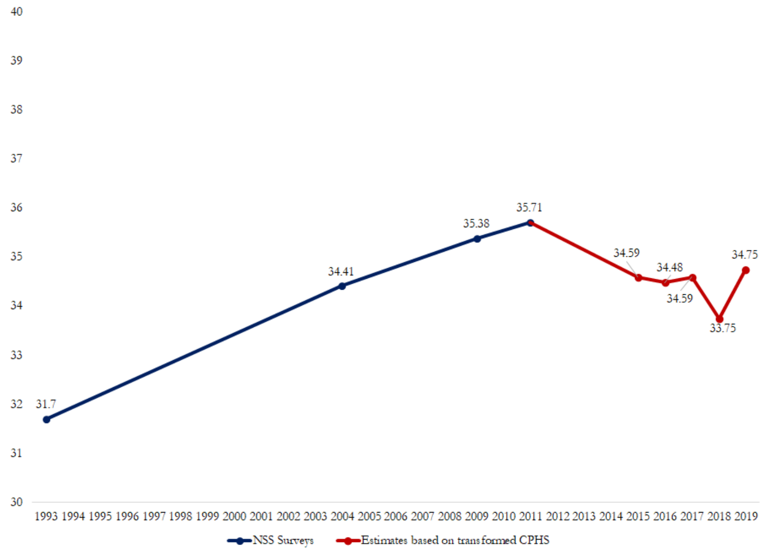


Figure 18: Inequality based on Gini measure
Notes: "NSS survey" denotes estimates based on NSS survey rounds; "Projections based on NAS" pass-through adjusted consumption growth from national accounts; and, "Estimates based on transformed CPHS" are based on Approach 2 (2011) of this paper.

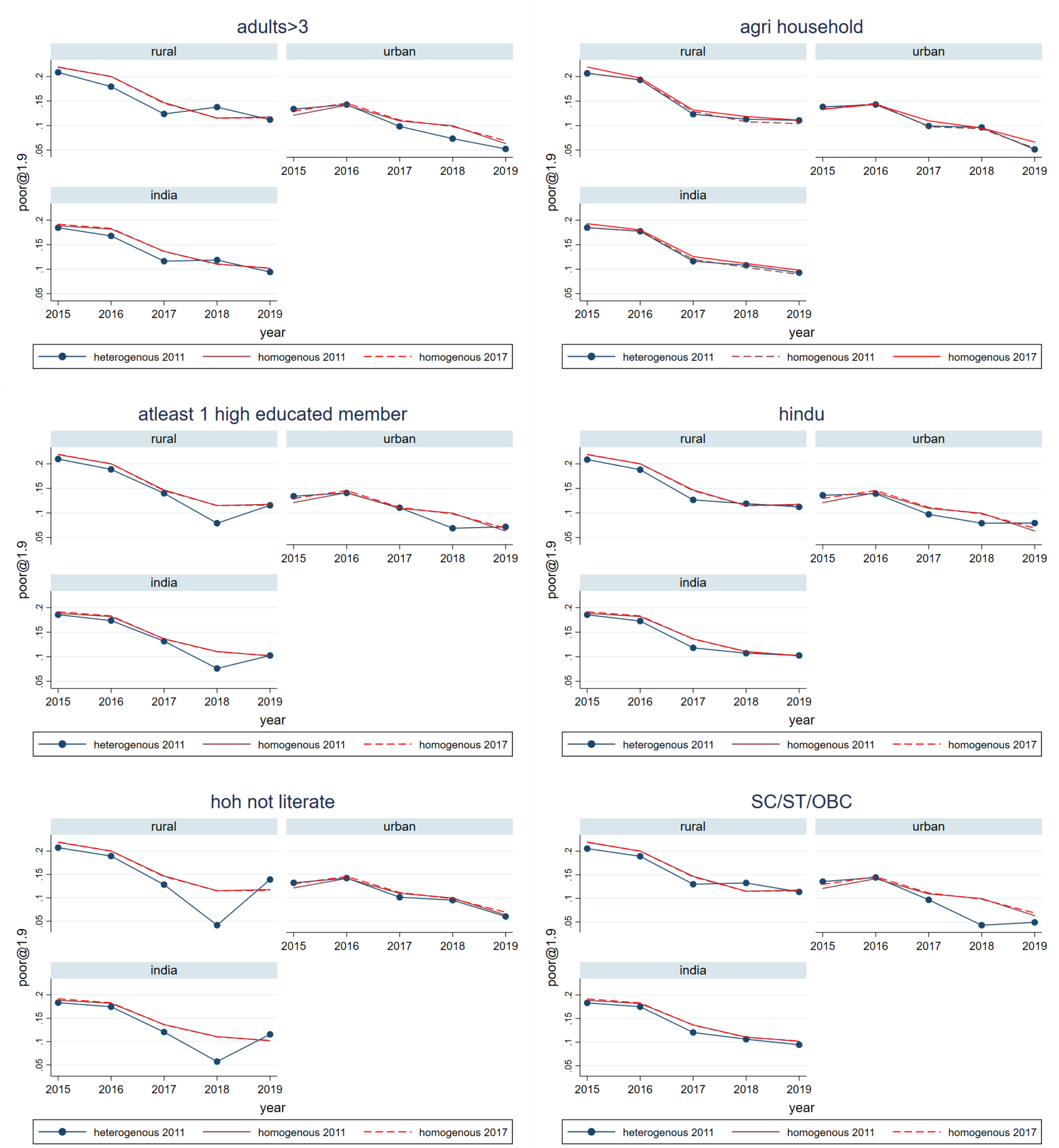


Figure 19: Headcount poverty rates after stratifying the rural and urban samples by household-level indicators: more than 3 adult members (panel (a); top-left), agricultural household (panel (b); top-right), at least 1 highly educated member (panel (c); middle-left), hindu household (panel (d); middle-right), non-literate head of household (panel (e); bottom-left), scheduled caste,tribe or other backward classes (panel (f); bottom-right)

Notes: The “homogenous” series denote headcounts based on a relationship fitted using only the rural and urban moments of the data. The moments are estimated using both NSS-2011 or NSS-2017. The “heterogeneous” series depicts a relationship fitted by further stratifying the rural and urban samples by on the six household-level indicators shown in the title of the graph.

In Figure 20, we plot the gap between back casted poverty projections and the actual poverty rate for 2004 across studies. Estimates closer to the horizontal axis show that the predicted poverty rates were close to the actual rate observed in 2004. The graph shows that approach 1 of our study predicts 2004 poverty rate to be 3.4 percentage points lower than the actual headcount across India and 3.2 percentage points lower rate for urban samples.[27] In comparison, estimates from Newhouse and Vyas (2019) are 2.2 percentage point apart from the actual national rate but the differences for urban samples are 9.2 percentage points higher. Deviations from the actual poverty rate in Edochie, et al. (2022) are in the same direction as our estimates but the magnitude is considerably higher in their study across all samples. Overall, these out-of-sample predictions for NSS-2004 suggest that our approach yield estimates that are closer to the actual headcount rate across rural, urban and all-India samples.

We believe that the inability to model changes in household asset ownership overtime could have led the earlier papers to overestimate poverty reduction in 2015 and 2017 and produce incompatible back casted estimates of poverty for 2004 (asset indicators were unavailable in the surveys used in the two papers). Our analysis in Section 6 using PLFS shows that asset indicators are important predictors of household consumption; failing to capture these indicators leads to divergent poverty estimates even within the same survey.

6 Corroborative evidence

Our estimates of poverty are at odds with findings from the leaked NSS-2017 survey which shows a rise in poverty between 2011 and 2017. Both sources point to a moderation of inequality since 2011, but the magnitude of changes to inequality are significantly higher in the NSS-2017 relative to our estimates. In this section, we corroborate our main findings using a range of independent data sources.

6.1 Headcount poverty has declined after 2011 with larger reductions in rural areas

Estimated consumption levels sit well with private final consumption expenditure (PFCE) reported in national accounts. A number of earlier studies have shown that there are systematic differences in consumption growth reported in national accounts statistics (NAS) and household surveys (see e.g. Ravallion, 2003; Datt

[27]Mean consumption per capita in the 2004 survey is 83.88 PPP dollars. The mean imputed 2004 consumption is 82.684 (1.4 percent lower than the survey mean).

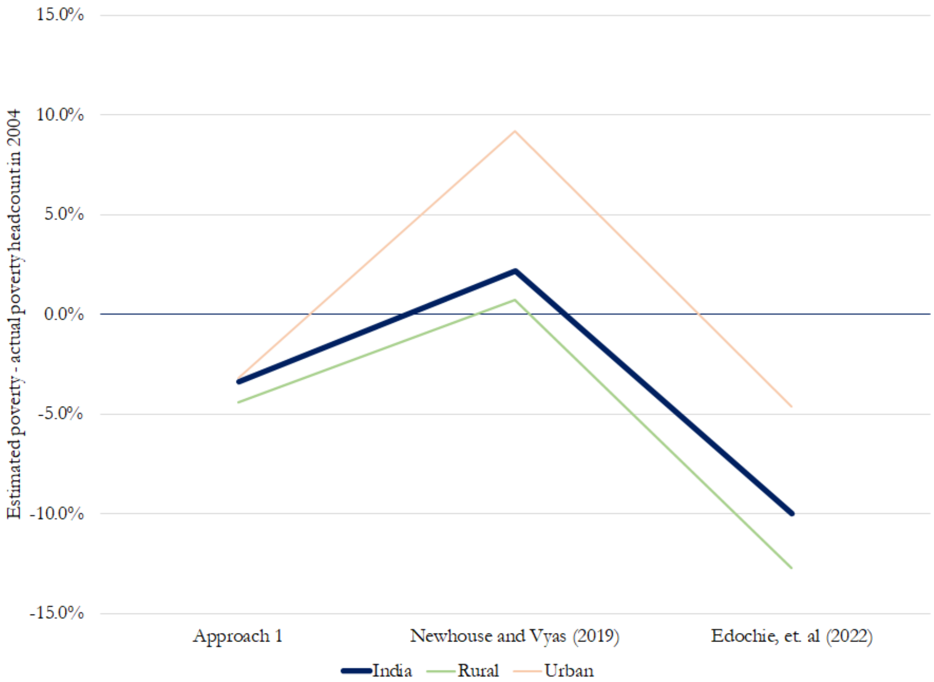


Figure 20: Backward predictions of poverty headcount at $1.90 for 2004 based on previous attempts and the two approaches
Notes: Horizontal axis depicts the gap between backward predictions of poverty and the actual poverty rate in 2004. The gap for Newhouse and Vyas (2019) is calculated using the PPP exchange rate of 14.975, all others are based on PPP exchange rate of 15.28 updated as of May 2020. Back casting estimates from previous papers are based on their respective preferred specifications. The imputation model used in approach 1 is the same as in section 4.1 except for the dummy variable for inverter ownership, NSS-2004 did not collect data on ownership of this asset

and Ravallion, 2002; Deaton, 2005 and Pinkovskiy and Sala-i-Martin, 2016). These differences are due to methodological differences as well as differences in the scope of consumptions covered by the two sources. For instance, PFCE in NAS includes financial intermediation services indirectly measured (FISIM), an indicator quantifying the value of financial intermediation in the country. FISIM is unlikely to be directly related to household consumption levels. Consequently, growth in PFCE from NAS is discounted by a factor known as the pass-through rate, to facilitate comparisons with consumption growth reported in household surveys. Edochie, et al. (2022) estimates the pass-through rate to be 0.67 for India.

Figure 21 shows that mean nominal consumption per capita from the NSS-2011 is Rs. 1652. Applying the discounted PFCE growth rate to this value, the 2015 consumption is estimated to be Rs. 2193. Average consumption per capita from our approach is approximately 3 percent lower (see Subramanian, 2019 for a potential explanation).

In 2016, the mean consumption from our approach is 4 percent lower than the PFCE derived measure. This was the year of demonetization of currency notes. Several observers, including the Chief Economist to the Government of India (CEA, 2017), have noted that the event may have resulted in a short-term economic shock to informal sector households. Since consumption in national accounts are based on the formal sector of the economy, observers predict that the growth in PFCE in 2016 has overlooked shocks to the informal sector. This could rationalize the 4 percent gap between the survey measure of consumption from our approach and the prediction based on PFCE. By 2017, the gap in nominal PFCE per capita between the two sources is almost eliminated. In 2018, our estimate of consumption is about 4 percent higher than the predicted value based on NAS and by 2019, the survey-based measure of consumption are about 8 percent higher than PFCE. The gaps in later years are plausibly due to higher pass-through rates.

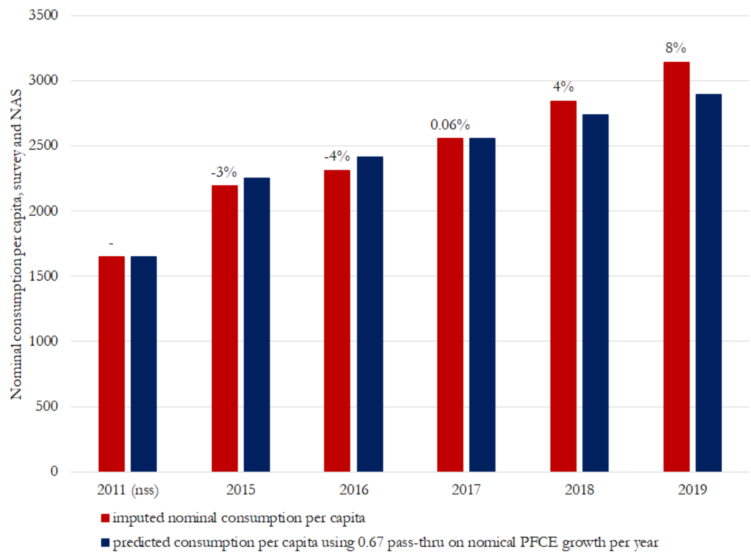


Figure 21: Mean consumption per capita from NAS and imputed NSS into CPHS
Notes: Consumption values are in nominal terms. The NAS estimate is calculated by discounting growth in nominal PFCE by 67% and applying it to the mean survey consumption observed in NSS-2011. The mean NSS-consumption of 2011 is derived by restricting the sample to the states that are covered in CPHS. The labels in the graph indicate the percent difference in per capita consumption from the NSS-type series and PFCE from NAS.

The growth in per capita PFCE suggests improvements in the standards of living in India since 2011. All else equal, this would predict a decline in poverty since 2011. This observation is confirmed independently by Felman, et al. (2019).

The third round of IHDS, conducted between February to July 2017,

provides further confirmation that poverty in India is lower in 2017 than in 2011. Consumption trends in past rounds of the IHDS and NSS surveys have tracked each other closely -- both surveys were conducted in 2004 and 2011 and predicted comparable drops in extreme poverty over this period. A limitation of IHDS-3 is that it is limited to the states of Bihar, Rajasthan and Uttarakhand. For this validation exercise therefore, we restrict the CPHS sample to these three states.

The IHDS captures consumption using the mixed recall period whereas the CPHS consumption used in our analysis corresponds more closely to the uniform recall period. Furthermore, IHDS-3 consumption values reported in Desai (2020) are in constant 2017 values and deflated using the *monthly* CPI-AL and CPI-IW series. The consumption values in our analysis are in constant 2011 terms deflated using *yearly* CPI-AL and CPI-IW series. For these reasons, we will be comparing changes in real consumption across the two sources (rather than comparing levels).

Real consumption grew at an annualized rate of 2.7 percentage points between the IHDS 2011-12 and 2017. The average annualized consumption growth over the same period in our analysis (approach 2) is 1.5 percent.[28] Real consumption growth in the IHDS-3's rural and urban samples are 3.8 and -0.7 percent per year. By comparison, consumption growth in rural and urban in our analysis is 1.7 percent and 0.6 percent, respectively. Both surveys therefore point to faster growth in rural areas than urban areas. The differences in consumption recall and deflators used in the two surveys could account for the difference in magnitudes of the observed growth rates.

Correlates of consumption, such as durable asset ownership, are similar across the two surveys. Thirty-two percent of households in the IHDS-3 states own motorcycles and cars and 21 percent possess air coolers and air conditioners. In the reweighted CPHS, ownership shares of these two assets are 34 and 22 percent, respectively. Growth in monetary and non-monetary indicators in the IHDS-3 therefore are consistent with the observation that poverty in 2017 is lower than in 2011.

Another assessment of poverty since 2011 can be made by comparing rural headcounts to rural wages produced by India's Labor Bureau. Monthly wages for agricultural and non-agricultural occupations are available since 1998. We take a weighted average of wages across occupations to construct a composite monthly rural wages series. The series is then deflated using monthly CPI-AL series and collapsed at the yearly level by taking a simple average across months.

Figure 22 correlates the growth in average annual wages for rural workers with year-on-year changes in rural poverty headcounts from our analysis (approach 2). As real

[28] The average real consumption in NSS-2011 for the three states is 1259.01 (constant 2011 rupees). For rural and urban areas, the mean consumption in NSS-2011 is 1141.57 and 1885.60 respectively.

rural wage growth is approximately 0.9 percent in 2016, poverty reduction occurs slowly, falling by 1.9 percentage points in the two consecutive years. In 2017, wage growth sharply accelerates as rural poverty fell by 5.3 percentage points. The moderation of wage growth to about 1.7 percent in 2018, slowed the rate of rural poverty reduction down to 3.2 percentage points that year. In 2019, rural wages fall below 2018 levels resulting in a 0.2 percentage point rise in poverty. The rate of rural poverty reduction observed in our analysis therefore sits well with the trends in real rural wage growth: the two series have a correlation of -0.94 across years.

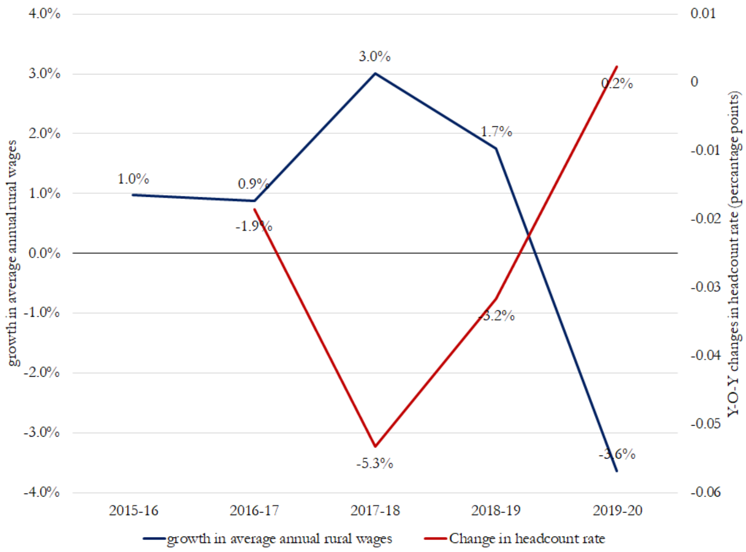


Figure 22: Relationship between real rural wage growth and rate of rural poverty reduction
Notes: Monthly wages for agricultural and non-agricultural occupations are from Labour Bureau of the government of India. A composite rural wage series is constructed by constructing a weighted average of agricultural and non-agricultural occupations using 59.32% and 40.68% as weights respectively. Wages are then deflated using the monthly CPI-AL series and collapsed at the yearly level (reference period: March to April of consecutive years). Rural headcount rates are based on approach 2 (2011).

Finally, poverty reduction since 2011 can be validated using periodic labor force surveys (PLFS). The first round of the PLFS was conducted in the same year as the unreleased NSS 2017 consumption survey. An alternative poverty rate for 2017 can therefore be derived by imputing consumption into the PLFS instead of the CPHS (using approach 1). Table 3 compares average consumption based on imputations into the PLFS (denoted by "PLFS-NSS"[29]) and based on imputations

[29] The variables used in imputation include all non-expenditure variables that are common to PLFS and NSS-2011, namely: dummy variables for household sizes 1 to 5; multigeneration family; extended

into the CPHS (denoted by "CPHS-NSS"[30]). Mean consumption per capita from the PLFS 2017 is estimated at Rs. 2385, which is approximately 7 percent higher than the NSS-2011 on an annualized basis. Note that these predictions rely only on changes in non-expenditure variables -- meaning that the growth of non-monetary predictors of consumption, as captured by the nationally representative official survey, must have been positive since 2011. This is further evidence that poverty in 2017 is lower than in 2011.

	2017		2018		2019	
	PLFS	**CPHS**	**PLFS**	**CPHS**	**PLFS**	**CPHS**
PLFS-NSS	2385	-	2525	-	2712	-
CPHS-NSS	-	2557	-	2843	-	3139
CPHS-NSS-PLFS	2404	2443	2548	2539	2758	2803

Table 3: Mean consumption per capita based on different imputation models and surveys.
Notes: Mean consumption values are deflated using CPI-AL and CPI-IW in rural and urban areas. The PLFS and NSS-2011 samples excludes states which are not included in CPHS. "PLFS-NSS" denotes consumption per capita based on an imputation model that uses variables that are common to PLFS and NSS-2011 (see footnote 28); "CPHS-NSS" denotes a model using a set of variables that are common between CPHS and NSS-2011 (see footnote 29); and, "CPHS-NSS-PLFS" denotes the model using variables common across all three surveys (see footnote 30).

Nevertheless, the first two rows in Table 3 underscore potential differences between the imputed consumptions into the CPHS and PLFS: Consumption imputed into the CPHS is about 7 to 16 percent higher than the PLFS. It should be noted, however, that the two consumption estimates are not a strict like-to-like comparison: the consumption imputed into the CPHS is based on demographic as well as asset variables, whereas imputations into the PLFS are based only on slower-moving demographic indicators (asset variables are unavailable in PLFS). To construct comparable vectors of imputed consumption across the surveys, we select a set of demographic indicators that are

family; share of 0 to 18 years old members in family; share of 61+ years old members in family; female headed households; log (age of household head); any member with higher than middle to high school level of education; share of members with middle to high school level of education; any member with diploma to post graduate level of education; dummy variables for Muslim; Christian; Sikh; Jain; Buddhist; Zoroastrian and other religions; scheduled castes; other backward classes; other castes; principal industry code of the household; household type; any regular salaried member in the household; household size and an interaction between the two variables. For urban sample we also include a dummy for cities that had over a million population in the 2011 census.

[30]The list of variables used in imputation are the same as in Table 2 of the main text

available in all three surveys (NSS, PLFS and CPHS) and re-estimate the model. The resulting consumption values, labeled as "CPHS-NSS-PLFS"[31] in Table 3, are about 0-2 percent apart across the years. Similarly, Figure 23 shows that the corresponding poverty rates at the $1.90 line are approximately 1.3 to 2.4 percentage points apart. This reasonably close correspondence adds further support to the robustness of our results. The analysis also underscores the importance of accounting for asset ownership in the household consumption models.

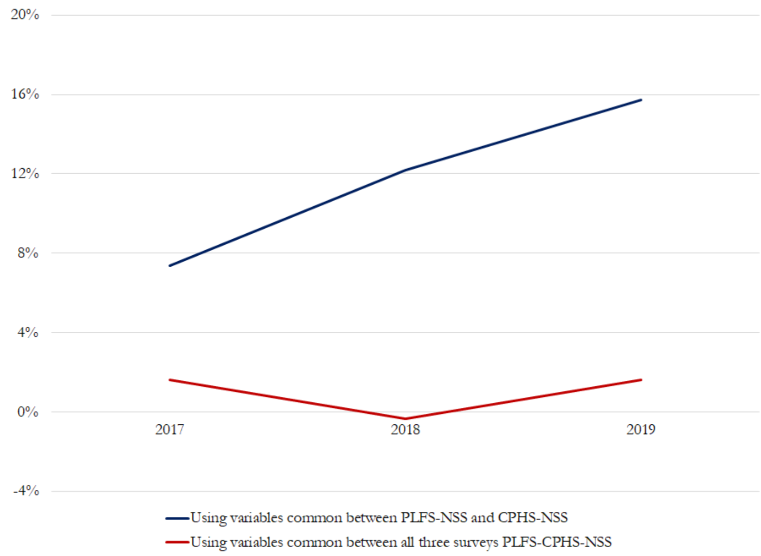


Figure 23: Differences in poverty headcounts using consumption imputed into CPHS and PLFS
Notes: Headcount poverty rates are based on consumption imputed into CPHS and PLFS using a common set of indicator variables (corresponding to "CPHS-NSS-PLFS" in Table 3). Mean consumption values are deflated using CPI-AL and CPI-IW in rural and urban areas. The PLFS and NSS-2011 samples excludes states which are not included in CPHS.

6.2 In the years following 2015, poverty reduction rates are highest in 2017-2018 and moderated in 2019

Faster growth in casual wages since 2011 supports observed reductions in extreme poverty. Historically, casual and salaried wage growth have been correlated with changes in poverty and inequality estimates. In 2011, for instance, only 8 percent of households below the $1.90 line had at least one member in the household with

[31]For PLFS, the list of indicators is the same as footnote 26, except household type; any regular salaried member in the household; household size and their interaction; and, the dummy for cities that had over a million population in the 2011 census. For CPHS, this includes all the variables in Table 2, except the asset variables.

regular salaried wages. In contrast, 50 percent of households at the top decile of the consumption distribution had a regular salaried wage earner. Observing the growth in casual wages may therefore provide useful indications about changes in poverty.

Figure 24 shows that the annualized growth in real casual wages between 1993-2004 and 2004-2011 was 1.8 and 6.8 percent, respectively (data obtained from ILO, 2018). The slower growth in casual wages during the first period translates to a poverty headcount reduction of 0.7 percentage points per year while the rapid wage growth in later period coincides with a brisk poverty reduction rate of 2.5 percentage points per year. More recently, casual wage grew at an annualized rate of 4.1 percent between 2011-2017 as poverty fell by 1.5 percentage points over the period. Casual wage growth is highest in 2017-2018, coinciding with a poverty reduction rate of 2.8 percentage points. In 2018-2019, casual wage growth turned negative. The poverty reduction rate slowed down to -0.8 percentage points during this time. The trajectory of casual wage growth therefore supports the observation that poverty in 2017 is lower than in 2011 and that the highest poverty reduction rates are observed in the years 2017 and 2018 followed by lower rates of poverty reduction. (Overall, casual wage growth and percentage point reduction in poverty headcount rates over 26 years have a correlation of -0.93.)

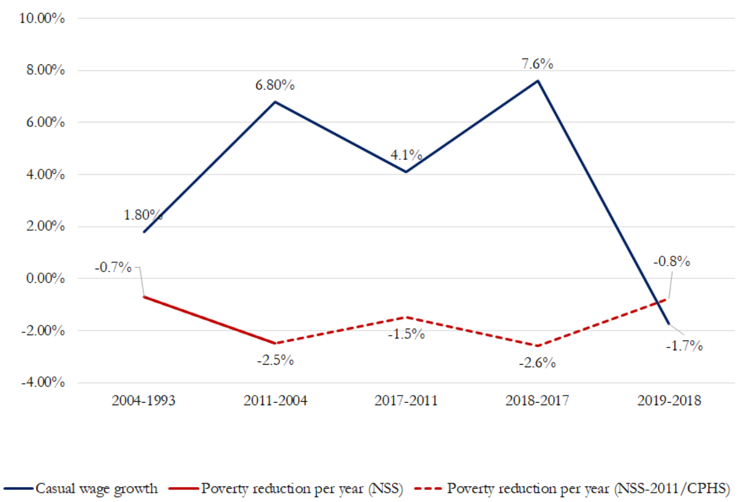


Figure 24: Growth in casual wages is historically correlated with reduction in poverty
Notes: Casual wage growth estimates for 1993, 2004 and 2011 are based on (ILO,2018). Wage growth for 2017, 2018 and 2019 are based on periodic labor force surveys. Wages in both sources are deflated using CPI-AL and IW.

A similar pattern emerges when we inspect yearly growth in night-time lights and sale of fast-moving goods in surveys conducted by Nielsen. Night-time lights data is obtained from Beyer, Jain and Sinha (2021). The authors obtained raw night-time lights data from VIIRS-DNB Cloud Free Monthly Composites (version 1) and corrected the raw data for outlier observations (averaging cells overtime and clustering areas based on the intensity of night-time lights). These corrections follow the approach advocated by Elvidge, et al. (2017). Values of night-time lights are reported in nanowatts per square kilometer. We collapse the monthly nighttime-lights aggregates from Beyer, Jain and Sinha (2021) to yearly levels before evaluating growth rates.

Nielsen's surveys track sales of consumer goods through retail store level surveys, covering a network of mom-and-pop stores as well as modern retail stores in 52 cities and 2700 villages across India. The instrument collects quantities, prices and sale values of both branded and non-branded items. We use estimates of quarterly growth in store-level sale values from publicly available sources[32]. The quarterly growth values are aggregated at the yearly level by taking simple averages, see Figure 25.

Both night-time lights and Nielsen's store-level surveys indicate welfare indicators peaked in 2017 and 2018. This period coincides with rapid rate of poverty reduction in our analysis. The sources also suggest a slowdown in 2016 and 2019 which further supports our finding that the rate of poverty reduction peaked in 2017-2018 and moderated in 2019.

6.3 A rise in urban poverty in 2016 followed by a rapid rise in consumption in 2017

Consumption growth trends from the IHDS-3 can help validate a break in poverty trends around 2016. The break in poverty reduction around 2016 coincided with a rise in urban poverty in that year. Household consumption strongly rebounded thereafter. Households interviewed by the IHDS in February to April 2017 reported a negligible rise in consumption since 2011-12. In contrast, household consumption for interviews conducted between May to July 2017 is 5 percent higher than 2011 on an annualized basis. Consumption of the first cohort of households was plausibly affected by the demonetization of currency notes in November 2016 followed by rapid growth in consumption as the economy was remonetized. We observe similar trends in our analysis albeit with smaller magnitudes. Consumption growth for the first cohort of

[32]List of all sources: http://bsmedia.business-standard.com/_media/bs/img/article/2016-08/09/full/1470687448-3888.jpg, https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/india-FMCG-growth-snapshot-q3-2018.pdf, https://images.assettype.com/afaqs/2020-01/200d87dc-162d-41ae-8fde-299faec4927f/Q4_2019_FMCG_Final_Deck.pdf. Quantity growth for 4^{th} quarter of 2016 was not available online. 2015-16 references the period starting the third quarter of CY2015 to the second quarter CY2016.

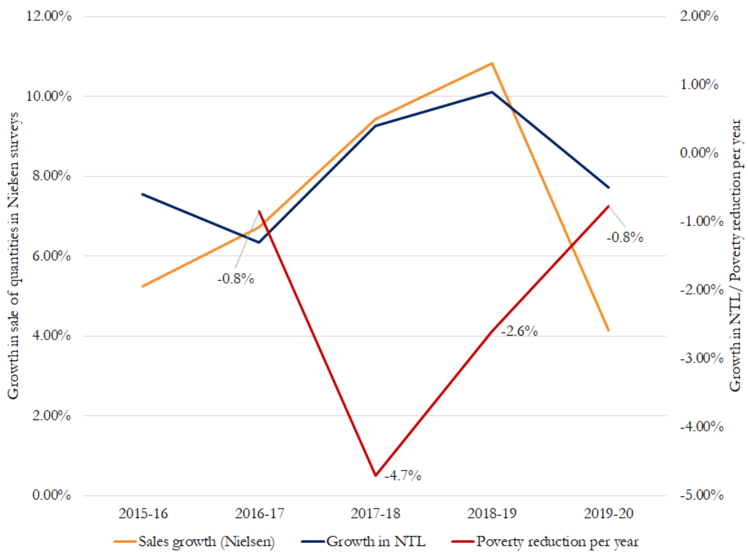


Figure 25: Growth in night-time lights and sales of fast-moving consumer goods in Nielsen surveys
Notes: Nighttime-lights data is obtained from Beyer, Jain and Sinha (2021). The values are reported in nanowatts per square kilometer and averaged across months to construct a yearly aggregate. Nielsen data is from retail-store level surveys. Refer to footnote 29 for reference to publicly accessible data sources.

households was 0.5 percent annualized since 2011, while consumption of the second cohort grew at 1.9 percent per year.

Chodrow-Reich et al. (2020) show that demonetization shocks had dissipated by mid-2017 despite having a large impact in the short-term. The authors estimate a 14-log point difference in nighttime lights before demonetization and immediately after the event. Using an estimate of 0.3 for the GDP-nighttime-lights elasticity, the authors predict short-term GDP changes to be approximately 4.2 log points. But by the spring of 2017, GDP rebounded significantly and reached levels observed in the pre-demonetization period -- suggesting that the monetary shocks had dissipated as all areas were remonetized. The authors support their night-time analysis using a range of administrative data on ATM cash withdrawals, deposit and credit data from banks and a composite indicator for economic activity. Changes in almost all indicators support a churn in economic activity at the end of 2016 followed by sharp rebounds by early-to-mid 2017. Our main findings for the same time period are consistent with the empirical observations from this literature.

6.4 No rise in consumption inequality since 2011, but indications of a rise in 2019

The unreleased NSS-2017 shows a moderation in inequality but the magnitude of the reduction is comparatively large. Based on leaked NSS-2017 results, Subramanian (2019) estimates rural and urban consumption inequality to have reduced by 0.0291 and 0.0387 Gini points since 2011 (based on modified mixed reference period in both NSS rounds). The direction of changes to inequality between NSS-2011 and NSS-2017 agrees with our findings. Our results differ, however, on the magnitude of the inequality reduction. Based on our estimates, average inequality reduction since 2015 in rural- and urban-India are 0.0007 and 0.007 Gini points (using the uniform recall periods of NSS-2011).

In Figure 26, we put the inequality estimates in a global context. Data on inequality is obtained from World Development Indicators. Countries that report at least one estimate of the Gini coefficient between 2009-2013 (two years before and after NSS-2011) and 2015-2019 (two years before and after NSS-2017) are included. We average the Gini coefficients for each of the two time-periods and evaluate the difference in mean values to observe how much inequality has changed between the two points in time across countries. The MMRP-2011 level of Gini and the change in inequality based on NSS-2017 data is highlighted in blue; whereas the URP-2011 level of Gini and the change in inequality from our analysis is highlighted in red. It follows that there are only a handful of countries that report inequality reductions that are comparable to what is reported between the NSS-2011 and NSS-2017. By comparison, the rate of reduction based on URP-2011 and our analysis is found to sit well with global trends.

Quintile consumption growth estimates in IHDS-3 show higher consumption growth in the bottom parts of the distribution. Figure 27 compares quintile consumption growth rates from the IHDS-3 to our estimates. Average consumption growth in the bottom quintile of the distribution is higher than the growth rates observed for households at the top end of the distribution in both sources. These patterns are consistent with the observed moderation in consumption inequality. Desai (2020) finds that the Gini measure of inequality has fallen by 0.023 points between 2011-12 and 2017. Over the same period, inequality based on our estimates fell by 0.07 Gini points.

NSS' All-India Debt and Investment Surveys (AIDIS) show that wealth inequality too has fallen. Using past rounds of NSS' All-India Debt and Investment Surveys (AIDIS), Himanshu (2019) shows that gross wealth inequality increased by 0.01 and 0.08 Gini points between 1991-2002 and 2002-2012. The direction of changes

Figure 26: Inequality reduction between 2009-2013 and 2015-2019 across the world

Notes: Cross-country Gini measures of inequality are obtained from the World Development Indicators. Observations restricted to countries reporting an inequality estimate in 2009-2013 and 2015-2019. The horizontal axis shows the average inequality of a country in the baseline period (2009-2013); the vertical axis shows changes in inequality across periods. Changes in MMRP level of inequality is based on MMRP based urban inequality measures from NSS-2011 and NSS-2017. Change in URP-2011 is based on URP measure of urban inequality in NSS-2011 and the average urban inequality for 2015-2019 using approach 2 (2011). Country codes represent: MDA - Moldova, ARE United Arab Emirates, MKD North Macedonia, MDV Maldives, NGA Nigeria, GMB The Gambia, SLV El Salvador, HND Honduras, BWA Botswana.

in wealth inequality have therefore tracked changes in consumption inequality from NSS-surveys for over two decades. Figure 28 shows that wealth inequality in the 2018 round of the AIDIS survey has moderated relative to levels observed in 2012. Following historical patterns, this finding further supports a fall in consumption inequality since 2011.

Inequality in wages offers complementary evidence on inequality moderating in recent periods. Himanshu (2019) uses labor force surveys to examine changes in wage inequality. Changes in wage and consumption inequality have not always moved in the same direction. For instance, Himanshu (2019) finds that both wage and consumption inequality rose markedly between 1993-94 and 2004-05. But by 2011-12, wage inequality had moderated while consumption inequality continued to rise. The analysis suggests that a sharp increase in real wages for casual workers

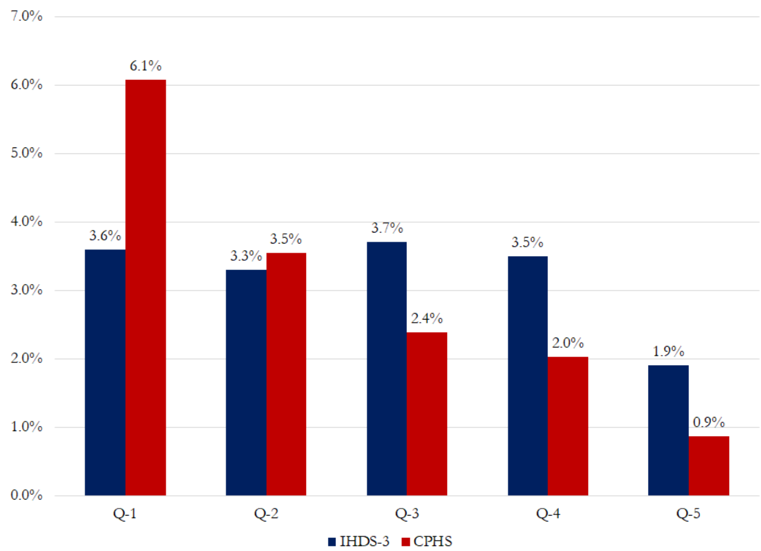


Figure 27: Mean consumption growth across consumption quintiles in IHDS-3 and CPHS
Notes: Consumption is deflated using CPI-AL and IW in both surveys. IHDS uses monthly deflators; CPHS deflated using annual values. Sample of CPHS restricted to states of Bihar, Rajasthan and Uttarakhand -- states where IHDS-3 was conducted. Sample includes households reporting consumption for the period February 2017 to July 2017 in both surveys.

between 2004-05 and 2011-12 relative to other workers may have contributed to the moderation in wage inequality during this period.

We extend the analysis on changes in wage inequality using recent rounds of the periodic labor force data in Figure 29. The results show a fall in wage inequality after 2011 with a larger moderation in urban areas. The year-to-year trend in the figure also suggests that wage inequality attained a minimum in 2018 followed by an increase in 2019. The overall trends in rural and urban wage inequality, as well as the year-on-year changes, are well aligned with our estimates of consumption inequality.

We next examine whether the fall in wage inequality is induced by a disproportionate growth in wages for casual workers relative to salaried earners. As noted earlier, only 8 percent of households from the bottom decile of the consumption distribution in 2011 have a member working in a regular salaried job. By comparison, 50 percent of households from the top decile have at least one salaried member. A higher wage growth of casual workers would therefore indicate a growth in the bottom part of the welfare distribution and a moderation in inequality. Figure 30 confirms that this is indeed the case. Real wage growth for casual wage workers is positive between 2011 and 2017 while wage growth for salaried workers has been negative. The differences in wage growth between the two types of workers is highest in 2017-2018, which is

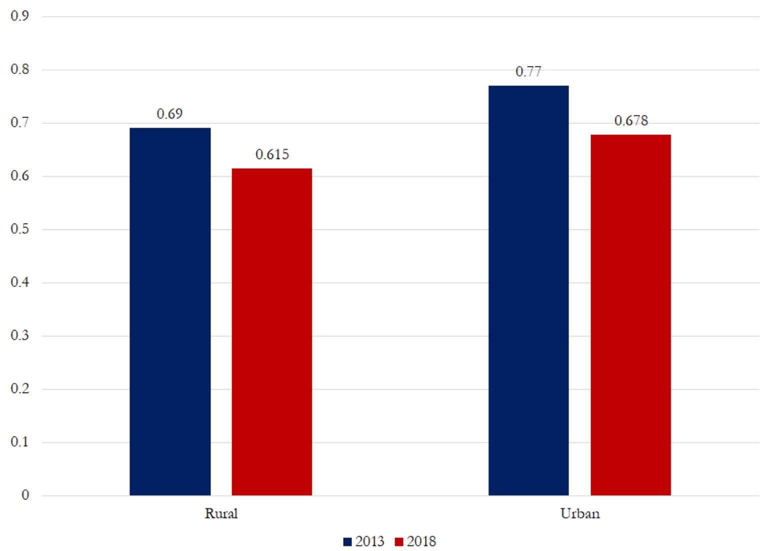


Figure 28: Changes in gross wealth inequality from AIDIS surveys of 2013 and 2018
Notes: Gini estimates of wealth inequality for 2013 are based on Sarma, Saha and Jayakumar (2017); estimates for 2018 are based on NSS' report accompanying survey data (statement 3.26, page 66). Estimates are based on gross wealth ownership and exclude values of durable assets owned by the household. Wealth values include both physical as well financial assets.

consistent with the observation that inequality bottomed-out in that year. As wage growth for casual workers fell in 2019, wage inequality levels rose back up.

Farmers with small landholding sizes have experienced higher income growth. Incomes from the NSS' situation assessment of agricultural household (SAS) surveys provide another opportunity to examine distributional changes in rural incomes. Using earlier rounds of this data, Himanshu (2019) reports a drop in the Gini coefficient of inequality for farm earnings from 0.63 to 0.58 between 2002 and 2012. His analysis suggests that the reduction in inequality can be attributed (at least in part) to NSS' definition of farmers that excludes agricultural workers with incomes below Rs. 3,000 from its sample.

Figure 31 examines the changes in agricultural incomes between SAS survey rounds of 2013 and 2019 by the size of landholding (the NSS' definition for farmers did not change during the two rounds). Real incomes for farmers with the smallest landholdings have grown by 10 percent in annualized terms between the two survey rounds compared to a 2 percent growth for farmers with the largest landholding. Rural households owning smaller pieces of land are more likely to be poorer than others. For example, 30 percent of households with consumption per capita below the $1.90 line in NSS-2011 possess less than 0.01 hectare of land. In contrast, only 4 percent of poor households possess more

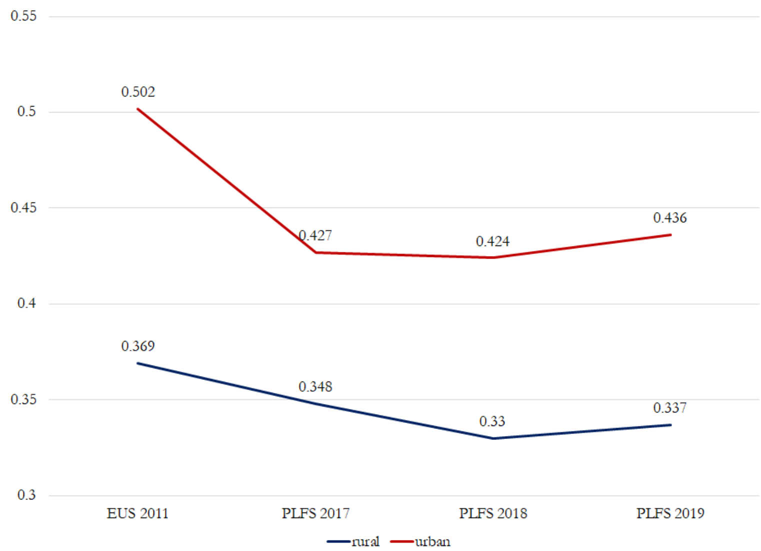


Figure 29: Changes in Gini measure of inequality over time
Notes: Wages of casual and salaried workers are included in the sample; wages of self-employed workers (~50% of the labor force) excluded due the absence of detailed profit or less statement. Sample includes workers reporting non-zero levels of wages. Wages are deflated using CPI-AL and IW and adjusted for rural and urban specific PPPs to account for cost-of-living differences in the areas.

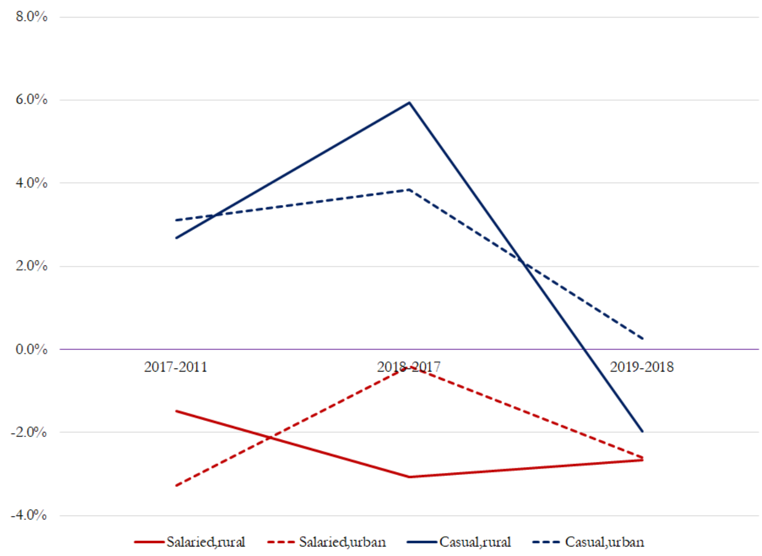


Figure 30: Real casual wages grew while salaried wages fell between 2011 and 2017
Notes: Wages of casual and salaried workers are included in the sample; wages of self-employed workers (~50% of the labor force) excluded due the absence of detailed profit or less statement. Wages are deflated using CPI-AL and IW. Sample includes workers reporting non-zero levels of wages.

than 10 hectares of land. Growth in incomes of the smallest landholders in rural areas (which constitute a larger share of the poor populations) therefore provides further evidence of a moderation in rural income inequality.

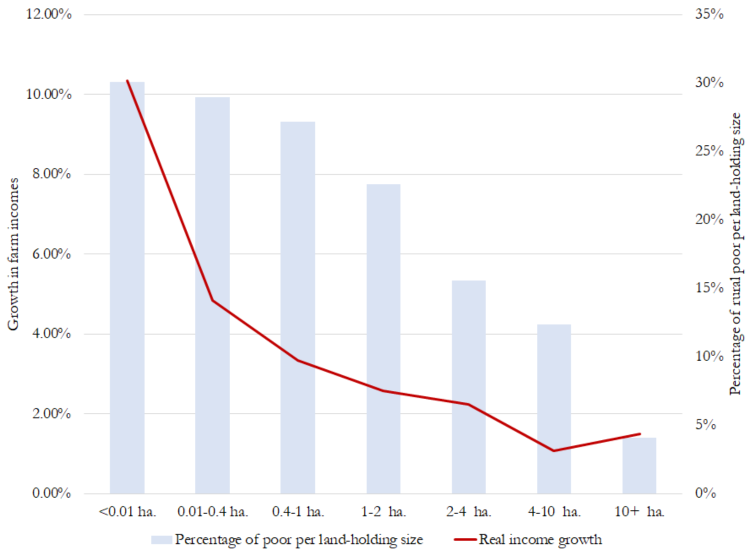


Figure 31: Growth in real incomes of agricultural households between 2013 and 2019
Notes: Rural incomes include income from wages, net receipt from crop production, net receipt from farming of animals and net receipt from non-farm business. Income from leasing of out of land is excluded from total incomes of 2019 to make consistent comparisons with the 2013 round, where this data was not collected. Data obtained from survey reports of SAS-2013 (statement 12) and SAS-2019 (statement 5.1A). Income values are deflated using the CPI-AL series. Share of poor by land-holding size is calculated by restricting the data to states where CPHS was conducted.

7 Conclusion

India has not released a new household consumption survey since the NSS from 2011. By extension, the country has not released any official estimates of poverty and inequality for over a decade now. Given the significance of these numbers, numerous scholars have made attempts to obtain estimates of how poverty and inequality may have evolved in India after 2011 using a variety of alternative (both official and non- official) data sources, see e.g Newhouse and Vyas (2019), Edochi et al. (2022), Desai (2020), Mehrotra and Parida (2021). The apparent disagreement between these estimates has given rise to a new poverty debate in India, a sequel to the Great India Poverty Debate from the 1990s (see e.g. Deaton and Kozel, 2005).

A new household consumption survey was introduced in 2014, the Consumer Pyramid Household Survey (CPHS), collected by the private data collection company called the Center for Monitoring Indian Economy (CMIE). This is the first time since the NSS-2011 there is household consumption expenditure data to work with, opening new doors for the measurement of poverty and inequality in India. There are two limitations of the CPHS however that have to be addressed. The first is that the survey in its current form is not nationally representative (see e.g. the biases documented in Somanchi, 2021). The second is that it uses its own measure of consumption expenditure that is not readily comparable to the NSS measure of consumption.

This paper makes a comprehensive effort to address both of the above-mentioned concerns. We implement a rigorous reweighting exercise using multiple nationally representative benchmark surveys to obtain adjusted sampling weights that make the CPHS nationally representative. The adjusted weights will be put in the public domain and hopefully serve as a public good to anyone looking to use the CPHS. We address the second concern by estimating the relationship between CPHS- and NSS-consumption and using this to impute NSS-type consumption directly into the CPHS. This allows us to compare our estimates of poverty to the official estimates for 2011, and by extension evaluate how poverty and inequality have evolved over the last decade.

We find that extreme poverty in India has declined by 12.3 percentage points between 2011 and 2019 but at a rate that is significantly lower than observed over the 2004-2011 period. Poverty reduction rates in rural areas are higher than in urban areas. We detect two incidences of rising poverty in our period of analysis: urban poverty rose by 2 percentage points in 2016 during the demonetization event and fell sharply thereafter; and, rural poverty rose by 10 basis points in 2019 likely due to a growth slowdown. Our estimates of poverty for recent periods are more conservative than earlier projections based on consumption growth in national accounts and other survey data. Finally, we do not find evidence of rising consumption inequality in our analysis. Our findings are supported by a comprehensive set of independent data sources.

The approach we developed to convert CPHS consumption into NSS consumption could be used to monitor poverty between the NSS years, thereby increasing the frequency of India's poverty estimates. The approach may also find use outside of India. The first-best approach of course is to work with actual up-to-date household consumption expenditure data. Any imputation-based estimates of poverty and inequality are inferior to survey-direct estimates that are obtained from observed household consumption data. Imputation methods are necessitated when real up-to-date household consumption data are not available. When the imputation methods considered in our study are used to estimate poverty and inequality for the years in between NSS rounds, the precision of these estimates is increased when the gaps in time that need to be bridged are reduced (i.e. when the frequency of NSS surveys is increased) -- as the assumptions underlying the imputation methods come under increasing pressure when the most recent household consumption survey becomes increasingly outdated.

References

Abraham, Rosa and Shrivastava, Anand (2019). How comparable are indias labour market surveys. Report. Centre for Sustainable Employment.

Atamanov, Aziz, Lakner, Christoph, Mahler, Daniel Gerszon, Tetteh Baah, Samuel Kofi and Yang, Judy (2020). The effect of new ppp estimates on global poverty. Global Poverty Monitoring Technical Note 12. The World Bank.

Bank, World (2018). Piecing together the poverty puzzle. Poverty and Shared Prosperity Report: 2018. World Bank.

Bank, World (2020). Reversal of fortunes. Poverty and Shared Prosperity Report: 2020. World Bank.

Basole, Amit, Abraham, Rosa, Lahoti, Rahul, Kesar, Surbhi, Jha, Mrinalini, Nath, Paaritosh, Kapoor, Radhicka, Mandela, Nelson, Shrivastava, Anand, Dasgupta, Zico, Gupta, Gaurav and Narayanan, Rajendran (2021). State of working india 2021: one year of covid-19. Report. Centre for Sustainable Employment, Azim Premji University.

Beegle, Kathleen, De Weerdt, Joachim, Friedman, Jed and Gibson, John (2012). Methods of household consumption measurement through surveys: Experimental results from tanzania. *Journal of Development Economics*, **98,** number 1, 3–18.

Beyer, Robert CM, Franco-Bedoya, Sebastian and Galdo, Virgilio (2021). Examining the economic impact of covid-19 in india through daily electricity consumption and nighttime light intensity. *World Development*, **140,** 1–13.

Bhalla, Surjit, Bhasin, Karan and Virmani, Arvind (2022). Poverty, inequality, and growth in india: 2011-2018. mimeo. IMF.

Bourguignon, Franois (2003). The growth elasticity of poverty reduction: explaining heterogeneity across countries and time periods. *Inequality and growth: Theory and policy implications*, **1,** number 1.

Castello-Climent, Amparo, Chaudhary, Latika and Mukhopadhyay, Abhiroop (2018). Higher education and prosperity: From catholic missionaries to limonosity in india. *Economic Journal*, **128,** 3039–3075.

Castello-Climent, Amparo and Mukhopadhyay, Abhiroop (2013). Mass education or a minority well educated elite in the process of growth: The case of india. *Journal of Development Economics*, **105,** 303–320.

CEA, Economic Survey 2016-17 (2017). Economic outlook and policy challenges. Technical Report Volume I. Ministry of Finance.

Chancel, Lucas and Piketty, Thomas (2019). Indian income inequality, 19222015: From british raj to billionaire raj? *Review of Income and Wealth*, **65,** S33–S62.

Chanda, Areendam and Cook, C. Justin (2020). Was india's demonetization redistributive? insights from satellites and surveys. *Insights from Satellites and Surveys.*

Chen, Shaohua, Jolliffe, Dean Mitchell, Lakner, Christoph, Lee, Kihoon, Mahler, Daniel Gerszon, Mungai, Rose, Nguyen, Minh Cong, Prydz, Espen Beer, Sangraula, Prem, Sharma, Dhiraj, Yang, Judy and Zhao, Qinghua (2018). Povcalnet update: Whats new. Global Poverty Monitoring Technical Note 2. The World Bank.

Chen, Shaohua and Ravallion, Martin (2010). The developing world is poorer than we thought, but no less successful in the fight against poverty. *Quarterly Journal of Economics*, **125,** number 4, 1577–1625.

Chodorow-Reich, Gabriel, Gopinath, Gita, Mishra, Prachi and Narayanan, Abhinav (2020). Cash and the economy: Evidence from indias demonetization. *Quarterly Journal of Economics*, **135,** number 1, 57–103.

Datt, Gaurav and Ravallion, Martin (2002). Is indias economic growth leaving the poor behind. *Journal of Economic Perspectives*, **16,** number 3, 89–108.

Datt, Gaurav and Ravallion, Martin (2011). Has india's economic growth become more pro-poor in the wake of economic reforms? *World Bank Economic Review*, **25,** number 2, 157–189.

Datt, Gaurav, Ravallion, Martin and Murgai, Rinku (2019). Poverty and growth in india over six decades. *American Journal of Agricultural Economics*, **102,** number 1, 4–27.

Deaton, Angus (2003). *Regional poverty estimates for India, 1999-2000*, Research Program in Development Studies edn. Princeton University.

Deaton, Angus (2005). Measuring poverty in a growing world (or measuring growth in a poor world). *Review of Economics and Statistics*, **87,** number 1, 1–19.

Deaton, Angus and Dreze, Jean (2002). Poverty and inequality in india: A re-examination. *Economic and Political Weekly* 3729–3748.

Deaton, Angus and Kozel, Valerie (2005). Data and dogma: The great indian poverty debate. *World Bank Research Observer*, **20,** number 2, 177–199.

Desai, Sonalde, Banerji, Manjistha, Barik, Debasis, Tiwari, Dinesh and Sharma, Om Prakash (2020). A glass half full: Changes in standards of living since 2012. India Human Development Survey Data Brief 2020-01. IHDS.

Deshpande, Ashwini (2020). The covid-19 pandemic and gendered division of paid and unpaid work: Evidence from india. Discussion Paper No. 13815. IZA.

Dhingra, S. and Ghatak, M. (2021). How has covid-19 affected indias economy? *Economics Observatory*, **30**.

Douidich, Mohamed, Ezzrari, Abdeljaouad, Van der Weide, Roy and Verme, Paolo (2016). Estimating quarterly poverty rates using labor force surveys: A primer. *World Bank Economic Review*, **30,** number 3, 475–500.

Dreze, J. and Somanchi, A. (2021). View: New barometer of indias economy fails to reflect deprivations of poor households. *The Economic Times* J, **une 21**.

Dreze, Jean and Khera, Reetika (2017). Recent social security initiatives in india. *World Development*, **98,** 555–572.

Dreze, Jean and Sen, Amartya (2012). Putting growth in its place. *Economy Perspective.*

Edochie, Ifeanyi Nzegwu, Freije-Rodriguez, Samuel, Lakner, Christoph, Herrera, Laura Moreno, Newhouse, David, Roy, Sutirtha Sinha and Yonzan, Nishant (2022). What do we know about poverty in india in 2017/18? Policy Research Working Paper 9931. The World Bank.

Elbers, C., Lanjouw, J. and Lanjouw, P. (2003). Micro–level estimation of poverty and inequality. *Econometrica*, **71,** number 1, 355–364.

Elvidge, Christopher D, Baugh, Kimberly, Zhizhin, Mikhail, Hsu, Feng Chi and Ghosh, Tilottama (2017). Viirs night-time lights. *International Journal of Remote Sensing*, **38,** number 21, 5860–5879.

Felman, Josh, Sandefur, Justin, Subramanian, Arvind and Duggan, Julian (2019). Is indias consumption really falling? Blog. Center for Global Development.

Ghatak, Maitreesh, Kotwal, Ashok and Ramaswami, Bharat (2020). What would make indias growth sustainable. Blog. The India Forum.

Gibson, John, Datt, Gaurav, Murgai, Rinku and Ravallion, Martin (2017). For indias rural poor, growing towns matter more than growing cities. *World Development*, **98,** 413–429.

Gibson, John and Kim, Bonggeun (2007). Measurement error in recall surveys and the relationship between household size and food demand. *American Journal of Agricultural Economics*, **89,** 473–489.

Gideon, Michael, Helppie-McFall, Brooke and Hsu, Joanne W. (2017). Heaping at round numbers on financial questions: The role of satisficing. *Survey research methods*, **11, No. 2**.

Goyal, Ashima and Kumar, Abhishek (2020). Indian growth is not overestimated: Mr. subramanian you got it wrong. *Macroeconomics and Finance in Emerging Market Economies*, **13.1,** 29–52.

Gravel, Nicolas and Mukhopadhyay, Abhiroop (2010). Is india better off today than 15 years ago? a robust multidimensional answer. *Journal of Economic Inequality*, **8,** 173–195.

Gupta, Arpit, Malani, Anup and Woda, Bartek (2021a). Explaining the income and consumption effects of covid in india. NBER Working Papers w28935. National Bureau of Economic Research.

Gupta, Arpit, Malani, Anup and Woda, Bartosz (2021b). Inequality in india declined during covid. NBER Working Papers w29597. National Bureau of Economic Research.

Haziza, David and Beaumont, Jean-Franois (2017). Construction of weights in surveys: A review. *Statistical Science*, **32.2,** 206–226.

Himanshu (2019). Inequality in india: A review of levels and trends. WIDER Working Paper 2019/42. Helsinki: UNU-WIDER.

Himanshu, Lanjouw, Peter, Murgai, Rinku and Stern, Nicholas (2013). Nonfarm diversification, poverty, economic mobility, and income inequality: A case study in village india. *Agricultural Economics*, **44,** 461–473.

ILO (2018). India wage report: Wage policies for decent work and inclusive growth. Technical Report. International Labor Organization.

Jaynes, E.T. (1957). Information theory and statistical mechanics. *Physical review*, **106.4,** number 620.

Kijima, Yoko and Lanjouw, Peter (2005). Economic diversification and poverty in rural india. *Indian Journal of Labour Economics*, **48,** number 2.

Kolenikov, Stanislav (2014). Calibrating survey data using iterative proportional fitting (raking). *The Stata Journal*, **14,** 22–59.

Krosnick, Jon A. (2018). *Questionnaire design*, The Palgrave handbook of survey research edn. Palgrave Macmillan.

Kundu, Sujata (2019). Rural wage dynamics in india: What role does inflation play. RBI Occasional Paper 40. Reserve Bank of India.

Lanjouw, Peter and Murgai, Rinku (2009). Poverty decline, agricultural wages, and nonfarm employment in rural india: 1983-2004. *Agricultural Economics*, **40,** 243–263.

Mehrotra, Santosh and Parida, Jajati Keshari (2021). Poverty in india is on the rise again. Blog. The Hindu.

Newhouse, David Locke and Vyas, Pallavi (2019). Estimating poverty in india without expenditure data: A survey-to-survey imputation approach. Policy Research Working Paper 8878. The World Bank.

ORGI (2011). Provisional population totals: Urban agglomerations and cities. Technical Report. Registrar General and Census Commissioner of India.

Pais, Jesim and Rawal, Vikas (2021). Cmies consumer pyramids household surveys: An assessment. Blog September 3. The India Forum.

Pinkovskiy, Maxim and Sala-i Martin, Xavier (2016). Lights, camera income! illuminating the national accounts-household surveys debate. *Quarterly Journal of Economics*, **131,** 579–631.

Ravallion, Martin (2003). The debate on globalization, poverty and inequality: why measurement matters. *International affairs*, **79,** 739–753.

Ravallion, Martin (2012). Why don't we see poverty convergence? *American Economic Review*, **102,** 504–523.

Ravallion, Martin (2016). Are the world's poorest being left behind? *Journal of Economic Growth*, **21,** 139–164.

Somanchi, Anmol (2021). Missing the poor, big time: A critical assessment of the consumer pyramids household survey. Web 11 Aug. SocArXiv.

Subramanian, Arvind (2019). India's gdp mis-estimation: Likelihood, magnitudes, mechanisms, and implications. CID Faculty Working Paper 354. Center for International Development, Harvard University.

Subramanian, Arvind and Felman, Josh (2022). Indias stalled rise: How the state has stifled growth. Technical Report January/February. Foreign Affairs.

Subramanian, S. (2019). Letting the data speak: Consumption spending, rural distress, urban slow-down, and overall stagnation. Blog 11. The Hindu Centre for Politics and Public Policy.

Tack, Jesse B. and Ubilava, David (2013). The effect of el nio southern oscillation on us corn production and downside risk. *Climatic change*, **121.4,** 689–700.

Tarozzi, Alessandro (2007). Calculating comparable statistics from incomparable surveys, with an application to poverty in india. *Journal of Business and Economic Statistics*, **25,** 314–336.

Vyas, Mahesh (2020). Impact of lockdown on labour in india. *The Indian Journal of Labour Economics*, **63,** 73–77.

Wittenberg, Martin (2009). Weights: Report on nids wave 1. NIDS Technical Paper 2. National Income Dynamics Study.

Zhang, Kexin and Yoshida, Nobuo (2022). How to correct for sampling bias in poverty projections using phone surveys. mimeo. The World Bank.

Appendix 1 Reweighting Results

1.1 Adult female education shares

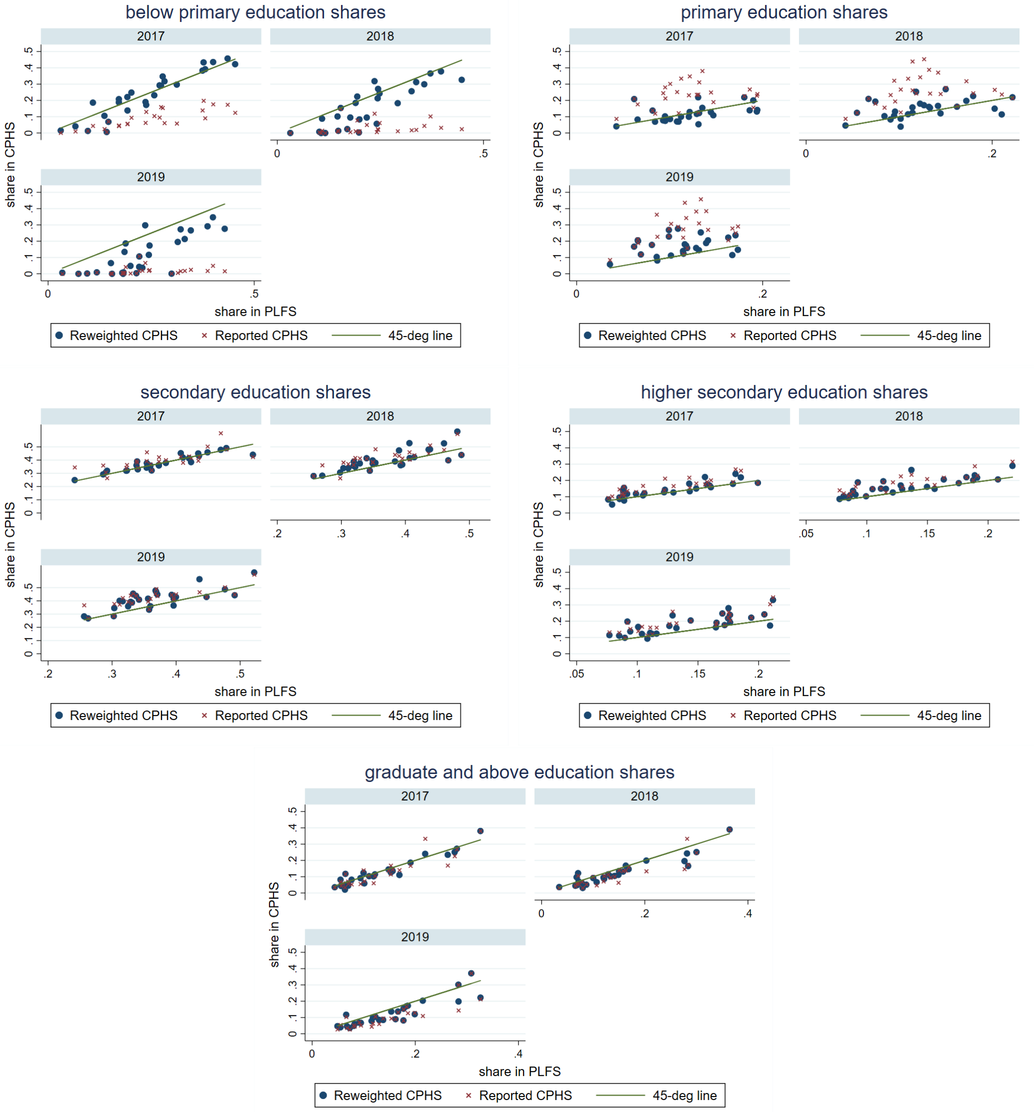
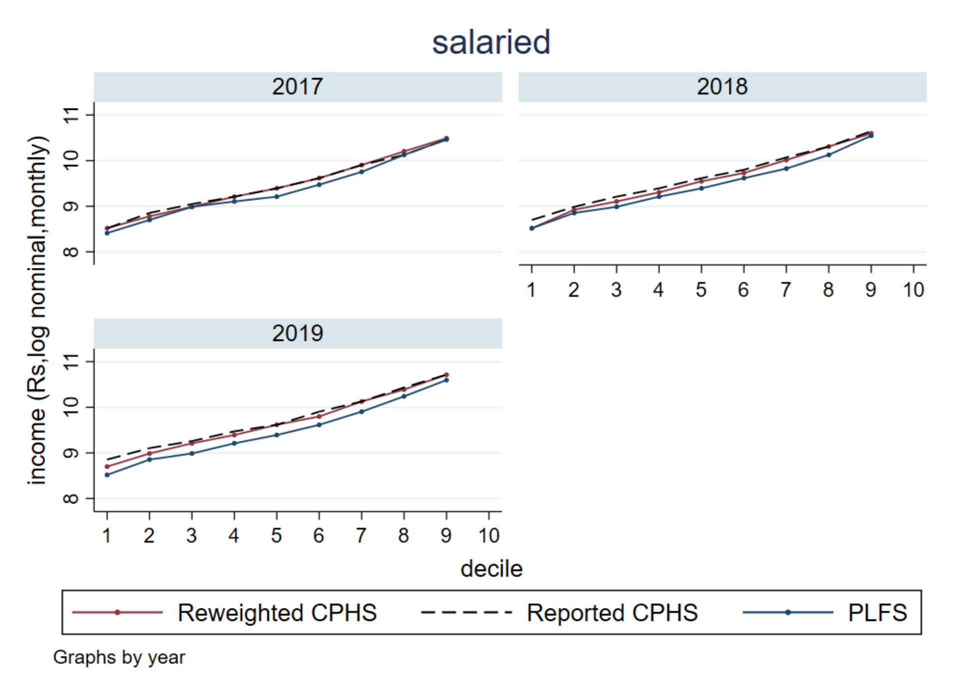


Figure 1: State level educational attainment in PLFS, Reported CPHS and Reweighted CPHS: Below primary education shares (panel (a); top-left), Primary education shares (panel (b); top-right), Secondary education shares (panel (c); middle-left), Higher secondary education shares (panel (d); middle-right), Graduate and above education shares (panel (e); bottom)
Notes: Scatter points denote education attainment shares at the state level from reported and reweighted CPHS in the vertical axis and PLFS in the horizontal axis. PLFS data includes only the first visit to each household. Sample includes adult females ages 15-49 in both surveys. Estimates are constructed using individual level weights from both surveys.

1.2 Distribution of monthly salary and daily wage incomes



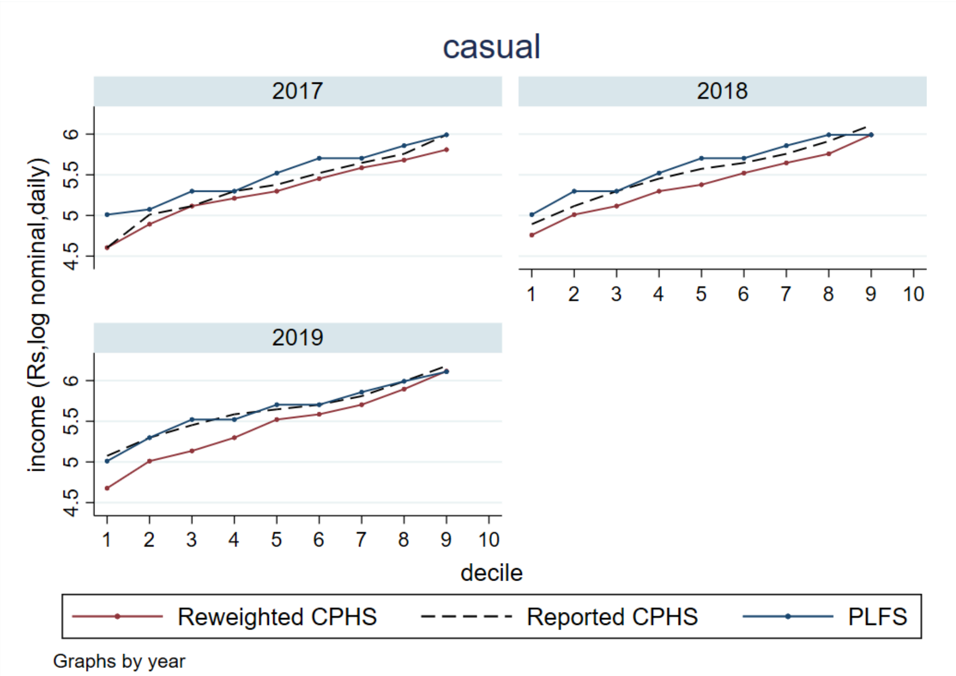
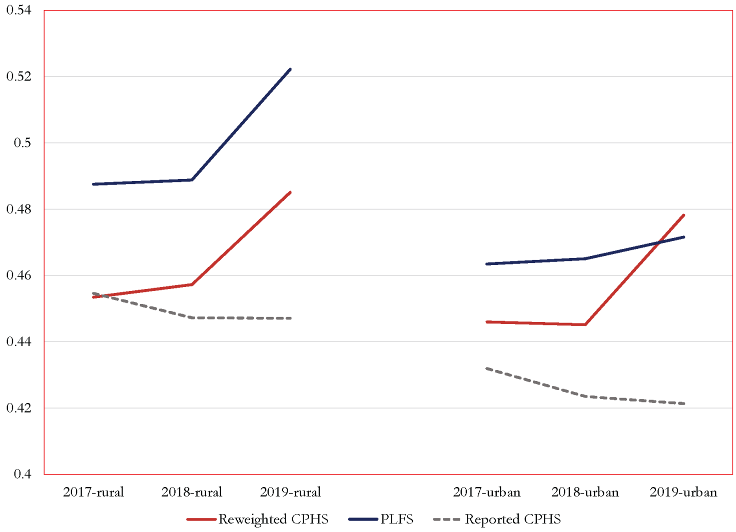


Figure 2: Deciles of monthly salaries and daily casual income: Monthly salaried incomes (panel (a); top), Daily casual wages (panel (b); bottom)
Notes: Monthly salaries and daily wages are in nominal terms. Sample in both surveys include households with non-zero salaries and wages. Salaries and wages from PLFS are based on all visits made to the household

1.3 Labor force participation rate and Worker population ratio



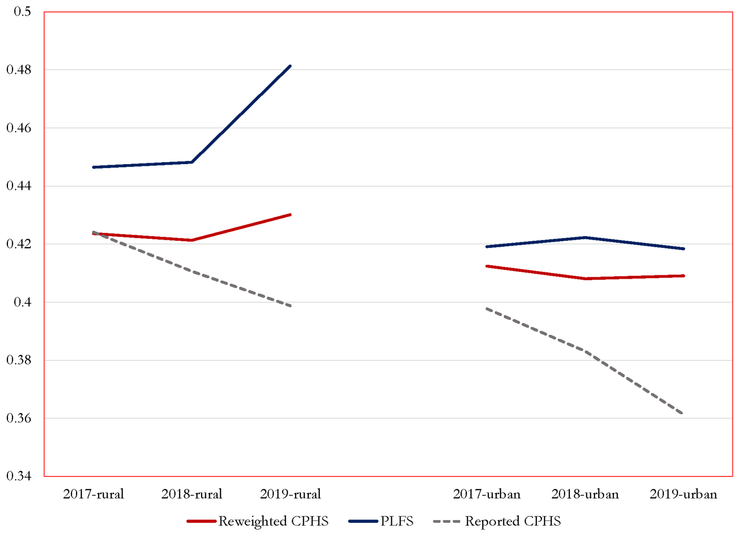


Figure 3: Key Labor Market Indicators: Labor Force Participation Rate (panel (a); top), Worker Population Ratio (panel (b); bottom)
Notes: Labor force participation rate and worker population ratio from PLFS is based on data from multiple visits. The red outline shows that the two indicators were not included in the set of targeting variables used for reweighting.

1.4 Female Labor force participation rate

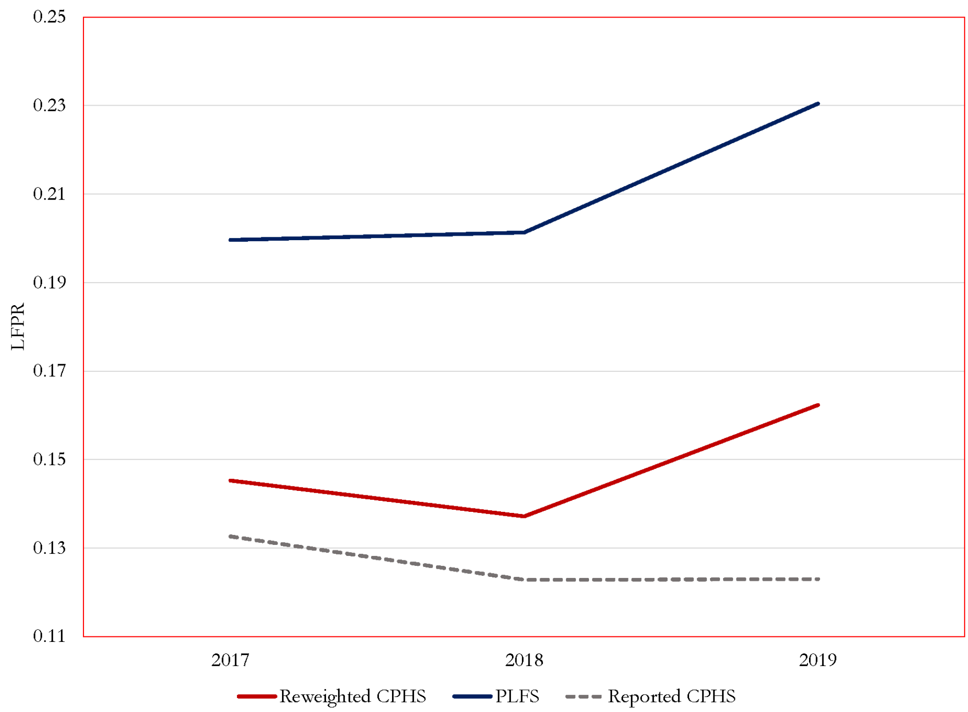


Figure 4: Key Labor Market Indicators: Female Labor Force Participation Rate
Notes: Labor force participation rate from PLFS is based on data from multiple visits. The red outline shows that female labor force participation rate was not included in the set of targeting variables used for reweighting.

1.5 Composition of workforce

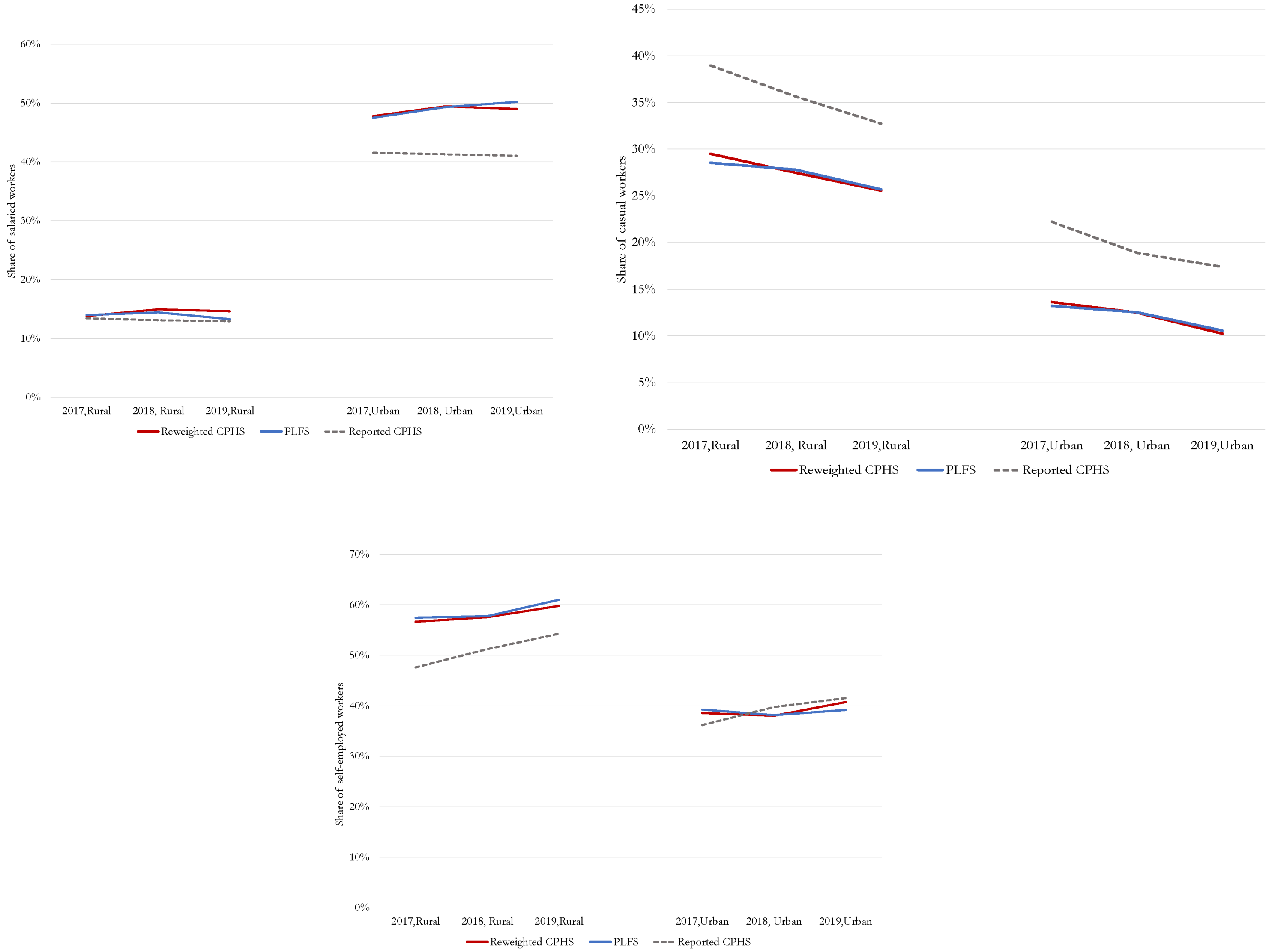
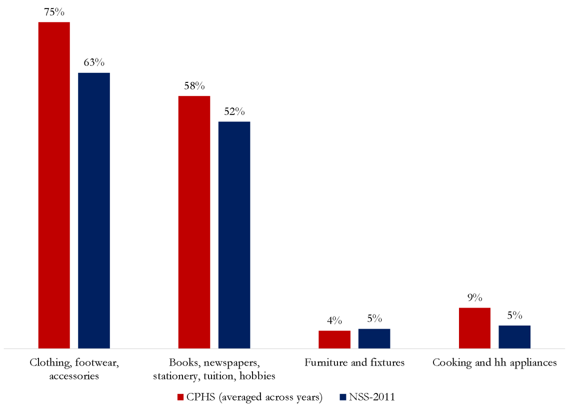


Figure 5: Composition of workforce across PLFS and CPHS: Share of Salaried Workers (panel (a); top-left), Share of casual wage workers (panel (b); top-right) and Share of self-employed workers (panel (c); bottom)
Notes: Salaried workers in CPHS include those that have either temporary or permanent employment arrangement. Share of workers from PLFS are based on data from multiple visits. The variable is included in the set of target variables used for reweighting.

Appendix 2 Implementing Approach 1

2.1 Examining dummy variables of consumption



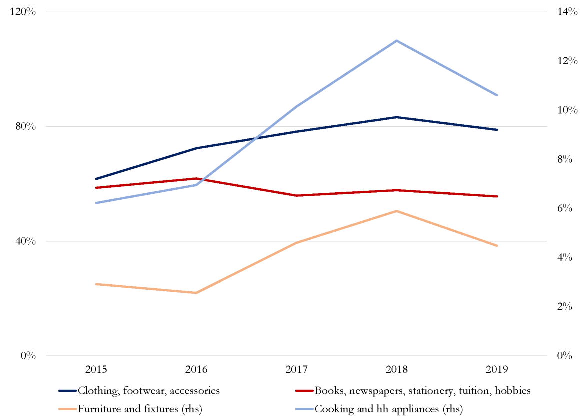


Figure 6: Share of household consuming premium goods and evolution over-time in CPHS: Share of households consuming items in CPHS and NSS-2011 (panel (a); top), Changes in the share of households consuming items (panel (b); bottom)
Notes: Figures indicate share of households that consume non-zero amounts of each item. The estimates are based on household level weights. CPHS estimates are based on reweighted sampling weights. Estimates from CPHS in Panel (a) are based on average household shares across 2015-2019 rounds. Panel (b) uses dual-axis: furniture and fixtures; and, cooking and household appliances use the vertical axis on the right-hand side. We define "Premium goods" as those that are likely to be dropped from a households consumption basket in the face of an adverse economic shock.

2.2 Examining principal industry code of the household

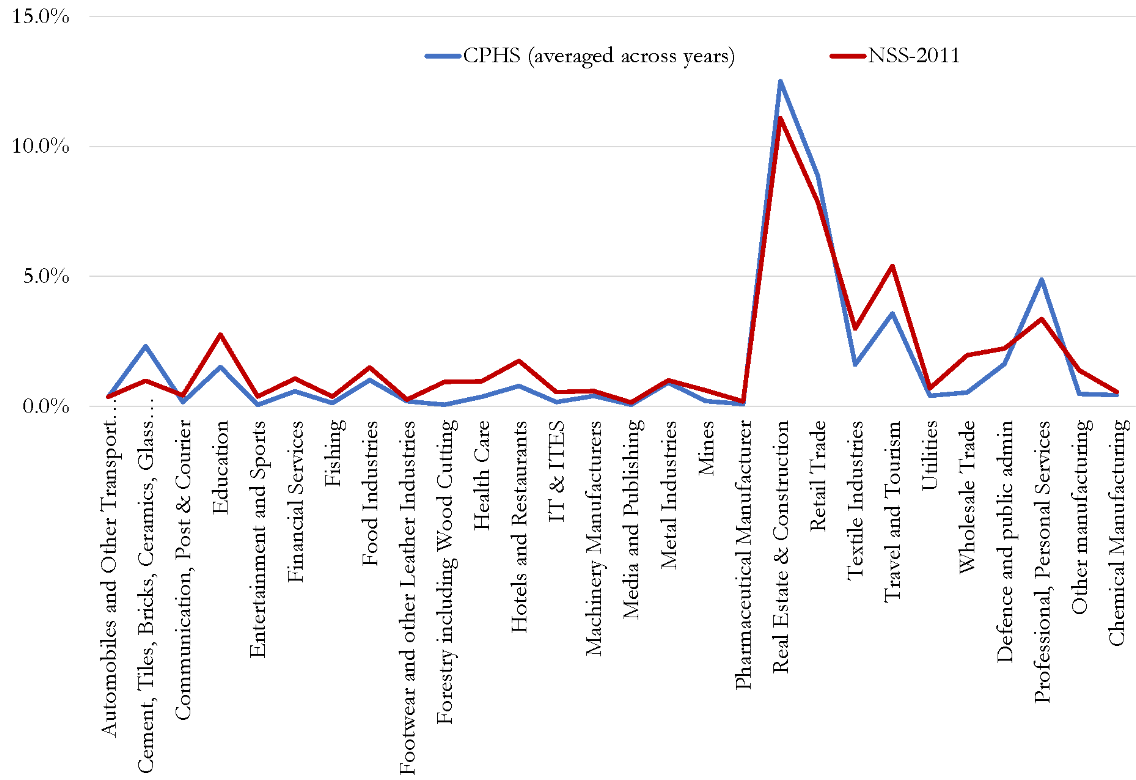


Figure 7: Comparison of principal industry code of occupation of households in CPHS and NSS-2011
Notes: Figures indicate the principal industry of occupation for a household. In NSS-2011, this indicator is defined in terms of the NIC-2008 industry classification and references the industry code of the member with the maximum level of earnings in the household; in CPHS, we define this variable as the industry code of the household head. We standardize the custom industry codes used in CPHS using a cross-walk. The horizontal axis depicts the standardized industry codes from this cross-walk. Reported estimates are based on household level weights; CPHS estimates are based on reweighted sampling weights.Shares of households with agriculture as the principal industry is omitted in the graph. These are 39.1 percent of households in NSS-2011 and 33.1 percent (averaged across years) in CPHS.

Appendix 3 Implementing Approach 2

3.1 Goodness of Fit of the mixed normal distribution

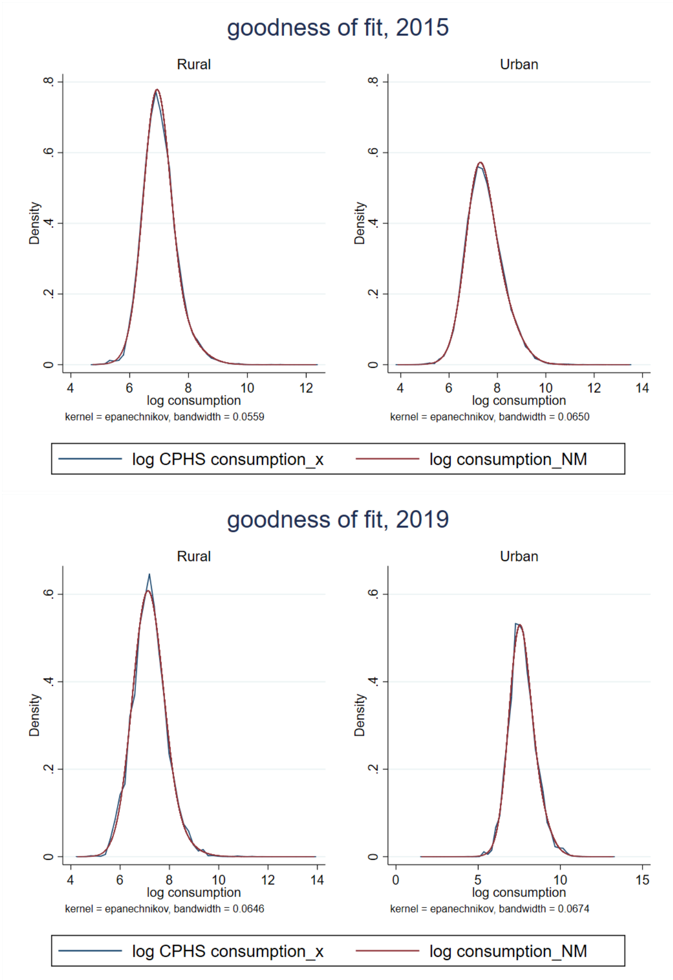


Figure 8: Examining the goodness of fit for mixed normal distributions: CPHS 2015-16 (panel (a); top), CPHS 2019-20 (panel (b); bottom)
Notes: Log CPHS "consumption-x" denotes the transformed log consumption from CPHS using equation 7 ($(logx_i - a)/b$). Log "consumption-NM" denotes the fitted consumption from a mixed normal distribution with two components. Consumption is in real terms and graphs are weighted using individual level weights.

3.2 Ranking households based on the three estimates of consumption

Figure 9 evaluates how sensitive the relative position of households in the consumption distribution is with respect to the choice of consumption measure. Quintile ranks are assigned to households based on their observed CPHS consumption and NSS-type consumptions from each year. We then compute the share of households that switch quintile rank when switching consumption measure. Panel (a) of the Figure shows that 27 and 23 percent of households in the 1st quintile of consumption from approach 1, originally belonged to quintiles 2 and 3 of the reported CPHS distribution; 26 percent of the households retained their first quintile rank before and after the transformation. In contrast, 66 percent of households ranked in the richest quintile retailed their ranking before and after the transformation of approach 1. This suggests that approach 1 trims the mass of households at the middle of the distribution and shifts the distribution leftwards, leaving the richest part of the distribution relatively intact. Panels (b) of the Figure shows that approach 2 has a smaller impact: As high as 90 percent of households in the 1st and 5th quintile preserve their ranking after transformation. The transformation impacts households in the 3rd quintile the most: approximately 60 percent of the households in the 3rd quintile of transformed consumption preserved their quintile rank based on reported consumption and the remaining are allocated either the 2nd or the 4th quintile rank.

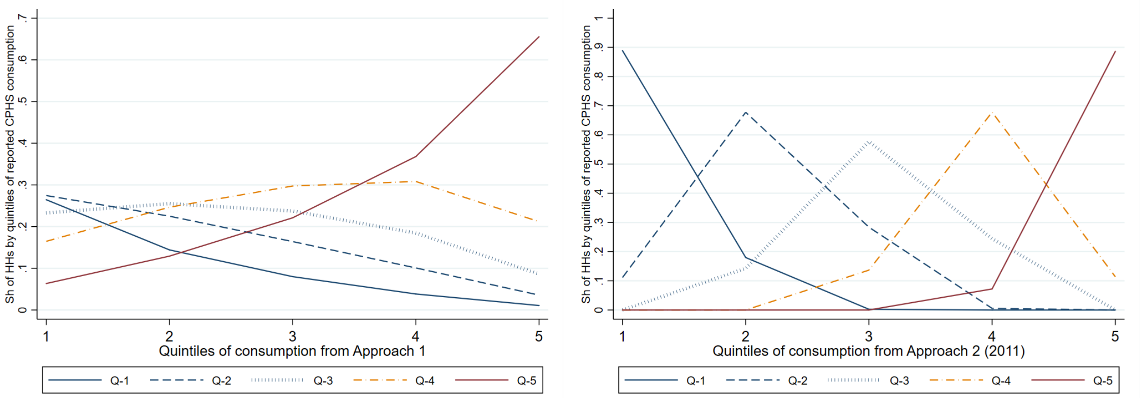


Figure 9: Changes in the relative ranking of households after transformations: Approach 1 (panel (a); top), Approach 2 (panel (b); bottom)
Notes: The figure compares the relative rank of a household before and after the two transformations. The quintile rank in the legend denotes the rank of the household in the CPHS reported consumption (prior to transformations). Results for approach 2 are based on matching higher moments to NSS-2011.

3.3 Implication of rural sample expansion

In this sensitivity analysis we consider a variation of Approach 2 that assumes that the relationship between CPHS-consumption and NSS-consumption (the parameters a, b, and σ^2) are constant over time, such that all year-on-year changes in poverty and inequality are due to variation in the observed CPHS-consumption distribution. We estimate the time-invariant parameters by first averaging our estimates of b (which does not depend on the values of a and σ^2) across all years. Next, we estimate a and σ^2 conditional on the resulting estimate of b, and then average the estimates of both a and σ^2 over time.

Figure 10 compares the resulting poverty and inequality trends to our preferred estimates. All poverty estimates are largely in agreement with each other for the years after 2016-17. The variation on Approach 2 (where the parameters a, b, and σ^2 are held constant) produces a nearly identical estimate of poverty for 2019 when compared to our preferred approach (Approach 2 where estimates of a, b, and σ^2 are adjusted over time). The headcount poverty estimates for 2015 and 2016, however, are significantly different. Poverty under our preferred approach (original Approach 2) shows a continued decline between 2011, 2015 and 2016. The variation on Approach 2 (denoted by "moments averaged (2015-2019)") shows a drastic reduction in poverty between 2011 and 2015, followed by a sharp increase in 2016. Inequality too shows an abrupt decline in 2015-2016, followed by a steep increase in 2017, when estimated using the variation on Approach 2 ("moments averaged (2015-2019)") and then settles at a comparatively level higher than our preferred approach in 2019-20.

Table 4 shows that our preferred approach ("Approach 2 (2011)") detects a rise in urban poverty in 2016 but no rise in rural poverty. The variation on Approach 2 ("moments averaged (2015-2019)") picks up an increase in both urban and rural poverty during this year. Is there corroborative evidence that would either confirm or reject an increase in rural poverty in 2016 (and a reduction in the year prior)? In figure 11 below, we plot real rural wages (covering agricultural and low-skilled non-agricultural occupations) between January 2015 and December 2017 and highlight the mean rural wage for the periods corresponding to 2015-16, 2016-17 and 2017-18. A 6-percentage point higher rural poverty estimated by the variation on Approach 2 ("moments averaged (2015-2019)") over our preferred estimate for 2016 would be consistent with a moderation in real rural wages during this time. No such decline in rural wages is observed between 2015-16 and 2016-17.

	Rural		
	Moments averaged (2015-2019)	**Approach 2 (2011)**	**Difference**

	Rural		
2015-16	17.5%	21.9%	4.4%
2016-17	26.4%	20.0%	−6.4%
2017-18	17.2%	14.7%	−2.5%
2018-19	10.1%	11.5%	1.4%
2019-20	11.1%	11.7%	0.6%
	Urban		
	Moments averaged (2015-2019)	**Approach 2 (2011)**	**Difference**
2015-16	8.8%	12.1%	3.3%
2016-17	17.1%	14.1%	−2.9%
2017-18	11.3%	10.9%	−0.3%
2018-19	7.3%	10.0%	2.7%
2019-20	6.9%	6.3%	−0.5%

Table 4: Estimates of poverty headcount at 1.90 line based on two variants of Approach 2
Notes: The series 'moments averaged (2015-2019)' indicates poverty and inequality estimate based on approach 2 using time-invariant a and b and σ parameters.

Consistent with the rural wage data, Nielsen store level surveys conducted between April and June of 2016 show that rural consumption growth (year-on-year) is positive in almost all products and higher than in urban areas.[33] Yearly rural consumption growth in April-June 2016 (corresponding to 2016-17 reference period in our sample) is 2.5 percentage points higher for FMCG products, 3.8 percentage points higher for food products, 1.2 percentage points higher for non-food products; and, 0.4 percentage points higher for over-the-counter sale of medicines than yearly consumption growth in urban areas.

In summary, the corroborative evidence that is available for the years 2015 through 2016 do not sit well with the increase in rural poverty during that time period as estimated by the variation on Approach 2 considered here (where a, b, and σ^2 are held constant), lending greater confidence to the estimates obtained by the version of Approach 2 where a, b, and σ^2 are adjusted over time.

The year-on-year changes in poverty and inequality obtained when holding a, b, and σ^2 constant may in large part stem from the expansion of the CPHS survey sample in 2017 wave 3, where over 80 new districts were added to the sample (the number of

[33]refer to https://www.business-standard.com/article/companies/fmcg-sales-growth-slows-to-3-2-in-apr-jun-116080900004_1.html

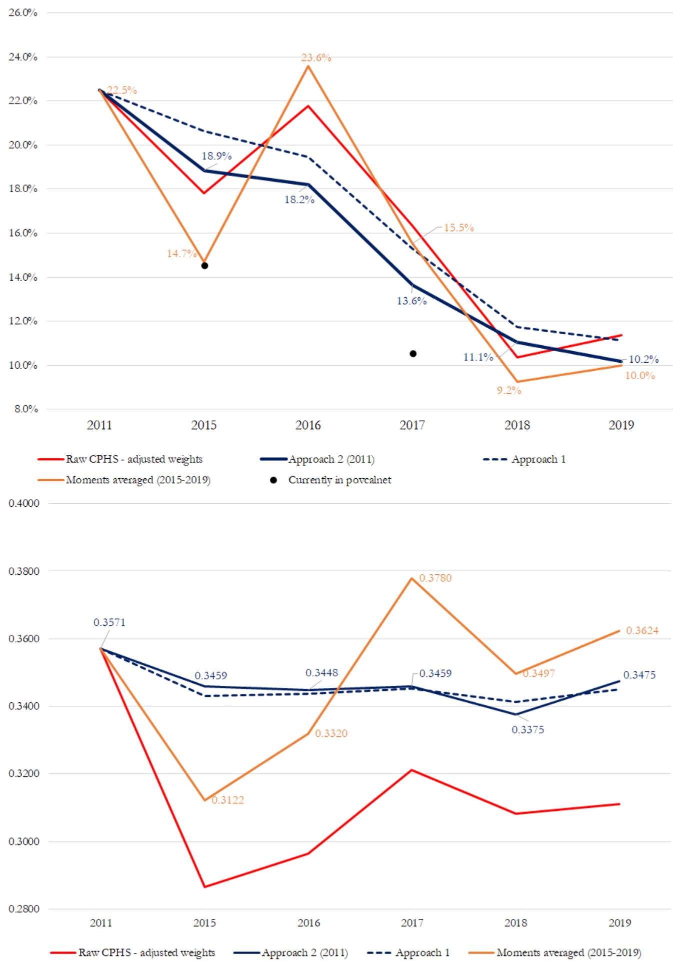


Figure 10: Changes in estimates of poverty (panel (a); top) and inequality (panel (b); bottom) based on different approaches
Notes: The figure compares the year-on-year changes in poverty and inequality based on different approaches. The series 'moments averaged (2015-2019)' indicates poverty and inequality estimate based on approach 2 using time-invariant a and b and σ parameters.

districts increased from 422 to 523). The bulk of the newly introduced districts during this change are from poorer rural locations in the country. This resulted in a significant increase in the dispersion of household consumption (and a similarly significant increase in the third moment) as seen in Figure 12. While these changes also introduced a shift in the first moment of household consumption, this is largely accounted for by re-weighting (i.e. by using the adjusted sampling weights). The re-weighting does, however, not resolve the abrupt changes to the second and third moments of the log consumption distribution. The corresponding fluctuations in the higher moments line-up with the comparatively large fluctuations in inequality and poverty that are observed prior to 2017 when using observed CPHS consumption data (or without adjusting the estimates

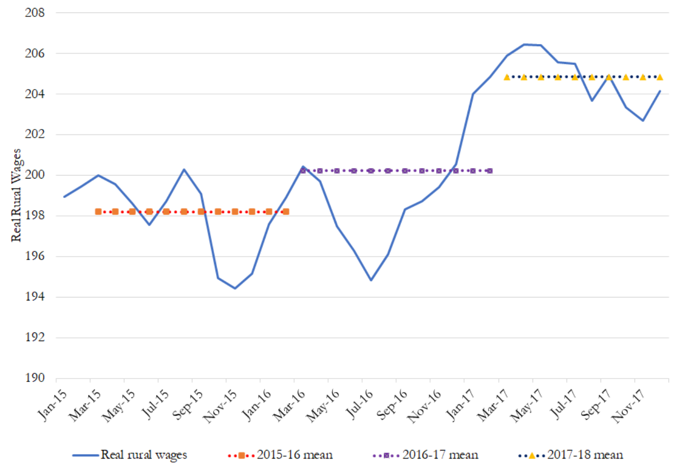


Figure 11: Real rural wages 2015-2017
Notes: Monthly wages for agricultural and non-agricultural occupations are from Labour Bureau of the government of India. A composite rural wage series is constructed by constructing a weighted average of agricultural and non-agricultural occupations using 59.32% and 40.68% as weights respectively. Wages are then deflated using the monthly CPI-AL series and collapsed at the yearly level. The mean rural wage for the years corresponding to 2015-16, 2016-17 and 2017-18 are highlighted (reference period: March to April of consecutive years).

of a, b, and σ^2 over time). The survey sample appears to have stabilized after 2017 - yielding estimates that are stable across the two variants of Approach 2.

As the change to the survey sample in 2017 disproportionately affected the rural sector, i.e. the sample expansion at this time was mainly for rural areas, the divergence in poverty and inequality for the years prior to 2017 should be largely concentrated in rural India. Table 4 confirms that this is indeed the case: differences in urban headcounts are more muted when compared to rural prior to 2017.

Appendix 4 Additional Estimates of poverty and inequality

4.1 Rural and urban poverty headcount at the 1.90 line

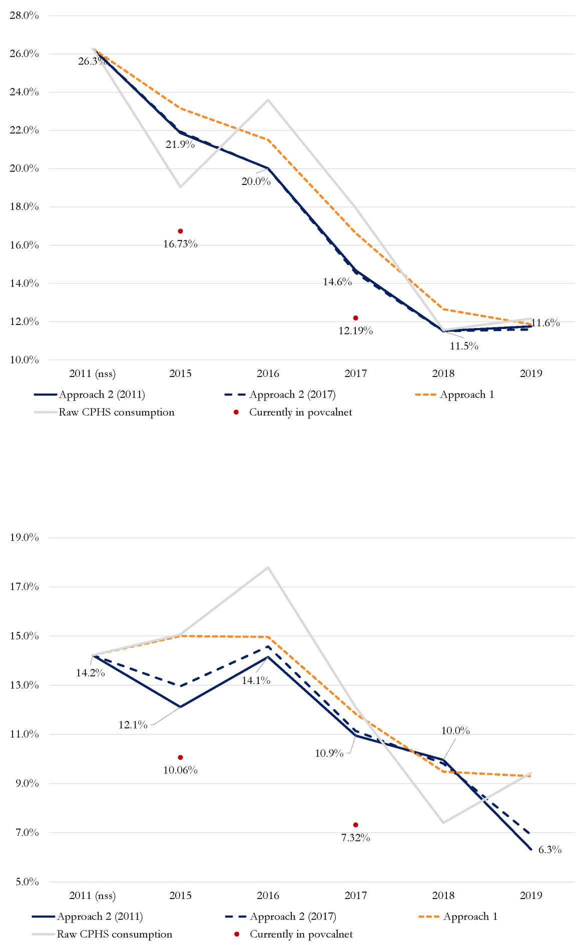


Figure 12: Headcount poverty rate since 2015 at the international 1.90 poverty line: Rural (panel (a); top), Urban (panel (b); bottom)
Notes: Refer to section 4.1 and 4.2 for details on Approach 1 and Approach 2 respectively. Estimates from Povcalnet are based on the line-up method: growth in real HFCE from national accounts statistics is multiplied by a pass-through rate and applied to the NSS-2011 consumption distribution. The Povcalnet estimates denoted in the figure are for the corresponding calendar years. The equivalent estimate for the financial years in rural are: 18.2 percent for 2015-16 and 11.3 percent for 2017-18; and in urban: 6.8 percent for 2015-16 and 9.3 for 2017-18.

4.2 Poverty headcount at the 3.30 and 5.50 lines

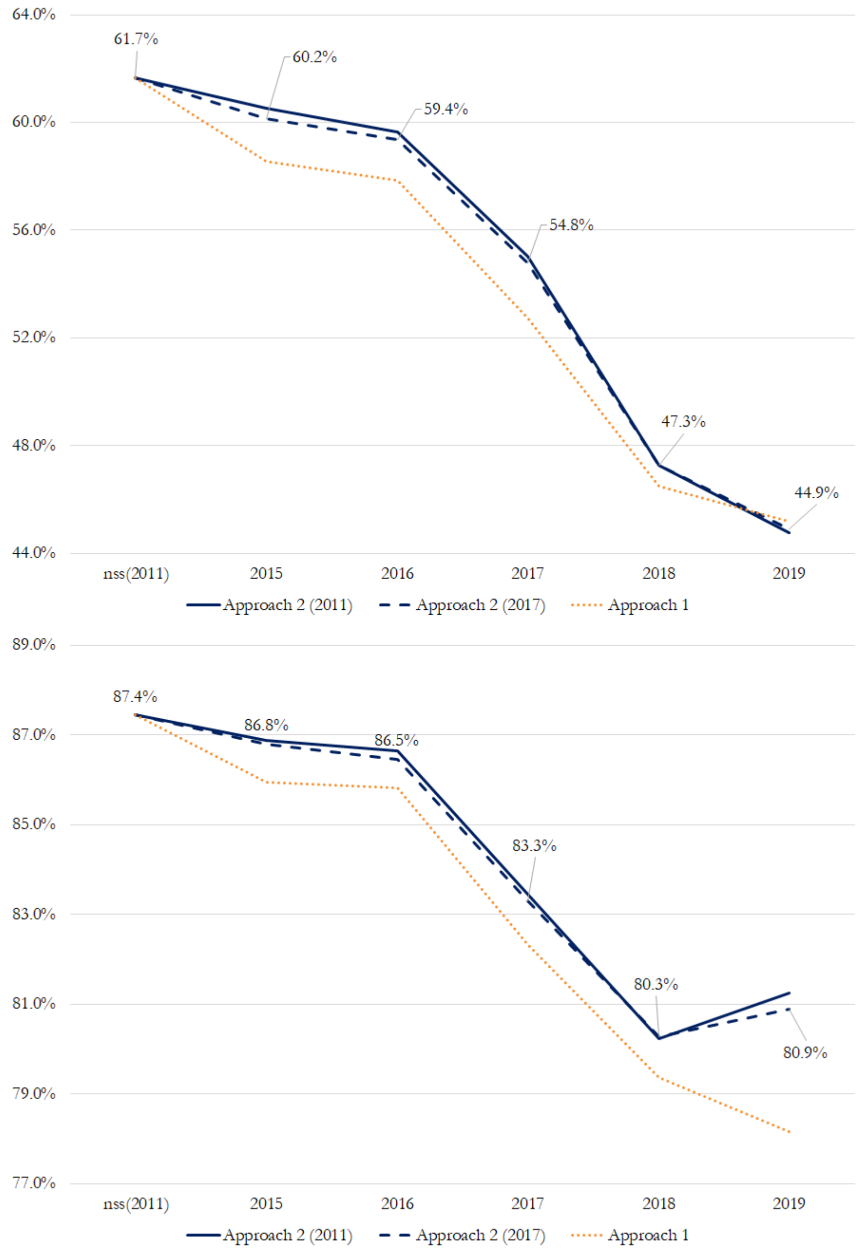


Figure 13: Headcount poverty rate since 2015 at: 3.30 line (panel (a); top), 5.50 line (panel (b); bottom)
Notes: Refer to section 4.1 and 4.2 for details on Approach 1 and Approach 2 respectively.

4.3 Rural and urban inequality

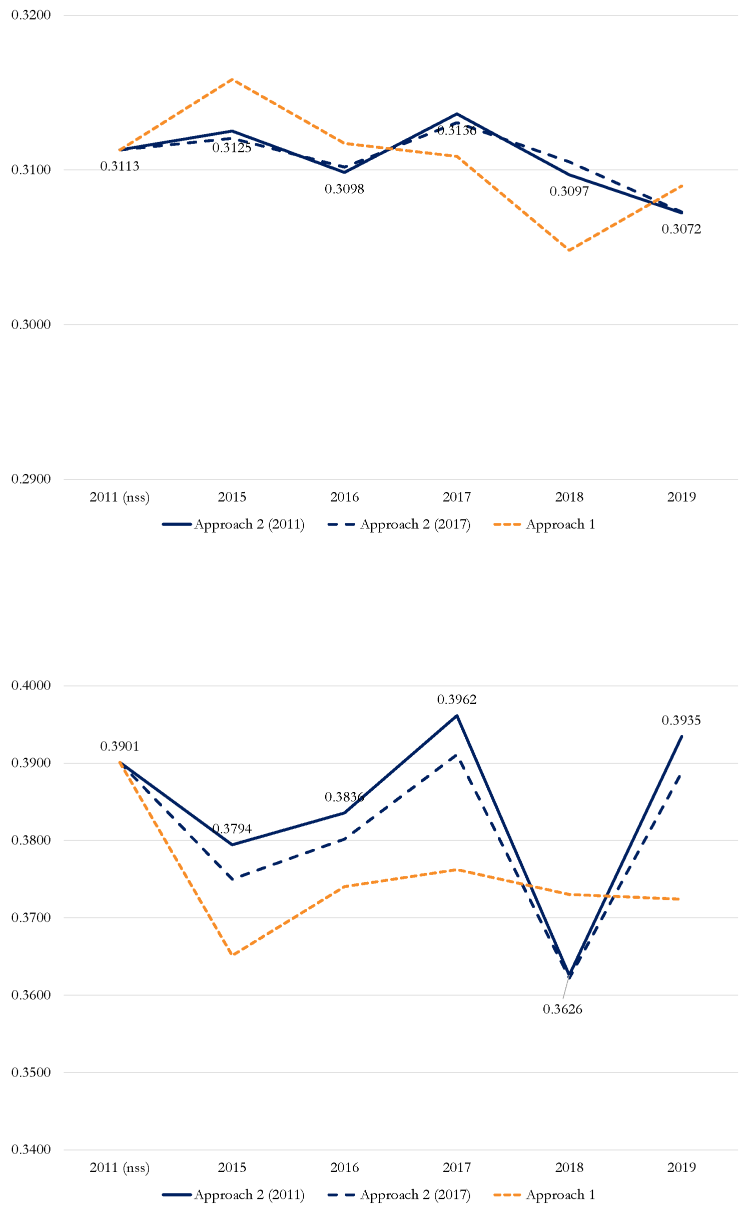


Figure 14: Gini measures of inequality: Rural (panel (a); top), Urban (panel (b); bottom)
Notes: Refer to section 4.1 and 4.2 for details on Approach 1 and Approach 2 respectively.Gini measure of inequality is calculated using PPP adjusted household consumption updated as of May 2020. PPP exchange rate of 13.173453 and 16.017724 are used for rural and urban areas

4.4 Poverty gap and Mean Log Deviation

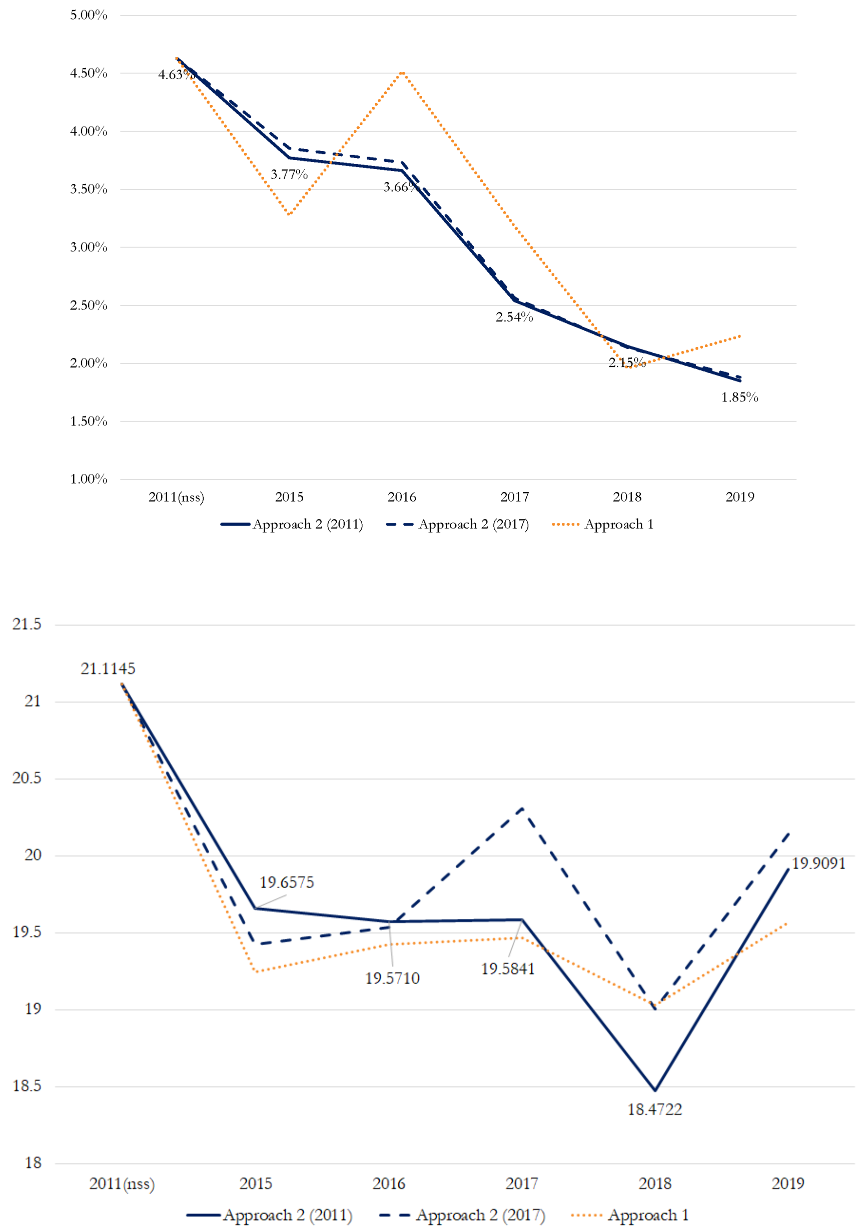


Figure 15: Poverty Gap (panel (a); top) and Mean Log Deviation (panel (b); bottom)
Notes: section 4.1 and 4.2 for details on Approach 1 and Approach 2 respectively.Following updates MLD is calculated using PPP adjusted household consumption updated as of May 2020. PPP exchange rate of 13.173453 and 16.017724 are used for rural and urban areas.

Supplementary Appendix 1: Discussion on CES 2022-23 fact sheet released by the NSO in February 2024 and other studies estimating poverty in India

1 Appendix: Discussion on other related estimates of poverty and inequality in India

After a gap of 12 years, India's NSO released a factsheet with select indicators of household consumption per capita for 2022-23. A broad pattern of agreement appears to be emerging amongst scholars in the debate that has since followed. There is a consensus that poverty declined significantly between 2011 and 2022. As the underlying microdata are yet to be released, various scholars (applying slightly different approaches to the aggregate data in the factsheet and various poverty lines) have estimated the headcount poverty rate in 2022-23 to be between 2-8% (down from 22.5% vs. 12.4% in 2011 using URP vs. MMRP). At the lower end of this range, sits Bhalla and Bhasin (2024) with an estimated headcount of 2%, then Subramanian (2024) with 3% and Ghatak and Kumar (2024) with two estimated values of 5% and 8% using two different approaches. Finally, using CPHS data from 2022-23, we estimate the extreme poverty headcount rate to be about 4.6% based on MMRP and 8.6% using URP.[1] In summary, our MMRP estimates for 2022-23 based on CPHS data are in agreement with Subramanian (2024), Ghatak and Kumar (2024) and Bhalla and Bhasin (2024) estimates based on the official factsheet data.

Deshpande (2024) and Himanshu (2024) have taken an exception to comparing consumption and poverty numbers from CES' 2022 and 2011 rounds. These studies have highlighted incompatibilities arising from (1) differences in the way the 2011 and 2022 surveys were executed (multiple household visits in the case of 2022-23 in contrast to a single visit in 2011-12) (2) differences in the number of consumption items captured between the two questionnaires, and (3) the use of imputed values for items consumed in the 2022-23 survey that were received free of cost through public programs. The approaches outlined in our paper are designed to address concern (2) – related to the incomparability of the 2022-23 and 2011-12 survey instruments. To recap, we use household consumption data from the Consumer Pyramid Household Survey (CPHS), conducted by the private agency CMIE, and impute consumption values compatible with NSS 2011-12. By construction, these estimates can be compared to consumption indicators from the NSS 2011-12, and follows earlier efforts by Tarrozzi (2007) to address similar breaks in comparability during the 1999-2000 round of NSS consumption survey.[2]

[1]The extreme absolute poverty line in our analysis is set at 1.90 line (at older 2017 PPP exchange rates, which approximately equals the 2.15 at new 2017 PPP).

[2]One technical point: the methods in our paper are designed to yield poverty estimates based on the uniform recall period. In contrast, the NSS 2022-23 factsheet is based on the Modified Mixed Recall Period (MMRP) consumption. Generally, for the same basket of items, MMRP is observed to offer

The analysis presented in the main text of this paper reports poverty estimates for 2015 to 2019. In the subsection 1.1 below, we extend our analysis to 2022-23. The resulting (imputed) consumption distribution closely approximates the consumption distribution reported in the NSS 2022-23 factsheet at lower deciles – most relevant for poverty measurement. Our estimates for top deciles are generally lower than those reported in the NSS 2022-23 factsheet. We hypothesize that this gap is due to differences in survey instruments between 2011 and 2022. The latter includes consumption across an expanded set of items – we conjecture that these are more likely to be consumed by richer households than the poorer ones, causing consumption at top deciles in 2022-23 survey to exceed its 2011-12 compatible counterpart. Although we are unable to conclusively test this hypothesis – due to the unavailability of microdata – we expect poverty estimates to be less sensitive to the differences in consumption at the top end of the distribution.

Based on our review of the recent literature, scholars disagree not on the extent of poverty reduction but rather on how we got to approximately 5 to 8% poverty rate in 2022-23. Our analysis suggests that poverty declined gradually over the past 10-15 years, with temporary disruptions in poverty reduction during demonetization in 2016 (based on our analysis presented in the main text) and a slowdown during the COVID pandemic period (based on section 1.1 of the appendix). In contrast, Ghatak and Kumar (2024) contend that poverty remained flat between 2011 and 2019, unaffected by demonetization, remonetization and other economic developments that occurred during this time, and then steeply dropped from 18-25% percent in 2019 to 5-8% percent in 2022-23. Additionally, Bhalla et al. (2022) estimate a steep decline in poverty in earlier years, with poverty falling below 10 percent by 2017, followed by a more gradual decline in subsequent years.

1.1 Comparing our imputed consumption values to those reported in the CES 2022-23 factsheet

In February 2024, the NSS released estimates of mean per capita consumption and selected fractiles of per capita consumption based on the 2022-23 round of CES. This notification filled a gap in official household consumption statistics in India dating back almost 11 years.The CES 2022-23 collected consumption using the Modified Mixed Recall Period (MMRP) method. Unlike the 2011 round, the 2022-23 CES did not collect consumption based on the Uniform Recall Method (URP). Our poverty estimates for the

a higher consumption value than URP, resulting in lower poverty headcount rates. We employed a back-of-the-envelope methodology to convert our URP-based imputed consumption expenditure values in CPHS to their MMRP equivalent.

period 2015 to 2019, however, are based on the Uniform Recall Period (URP) method. Differences in questionnaire design between CES-2011 and 2022-23, an updated and expanded list of enumerated items, and changes in survey protocol involving multiple household visits adopted in the latest round, compared to the earlier one, have raised further doubts about the compatibility of consumption data across the 2011 and 2022 rounds.[3]

To compare our consumption and poverty estimates to the latest official figures, we employ a back-of-the-envelope approach to transform URP-based expenditure values predicted by our methodology to their MMRP equivalent. The relationship between MMRP- and URP-denominated consumption values is derived from the 2011 round containing both consumption measures. More specifically, we start by observing the relationship between the percentile values of log consumption in URP and MMRP terms within the 2011 CES, as depicted in Figure 1.[4] Under the assumption that the relationship between the two consumption measures has remained stable over time, we apply approaches 1 and 2 to the CPHS 2022-23 data to obtain compatible URP expenditures for 2022-23.[5] Finally, we transform the URP consumption into its MMRP equivalent, using the log-linear relationship in Figure 1.

Figure 2 shows that the predicted rural and urban consumption (in nominal and MMRP terms) using our approaches and the CPHS 2022-23 data is about 5 and 10 percent lower than the official CES 2022-23 estimates in the factsheet. In Figure 3, the official consumption is higher mainly on account of richer urban households. In rural areas, the difference between predicted and official consumption at 5^{th}, 10^{th}, and 20^{th} percentiles are respectively -7%, 2% and -2%. Similarly, for urban areas, the gap between predicted and official statistics at the same percentiles is 2%, 11%, and 8%.[6] It is reassuring to observe the smallest gaps in reported and predicted consumption values among rural households at the lower end of the distribution, as these house-

[3]For a select set of articles reviewing these incompatibilities, see Deshpande (2024) and Himanshu (2024

[4]NSS-2011 does not report URP and MMRP consumption values for the same household. As a result, we collapse the distinct samples of URP and MMRP households in NSS-2011 to their percentile values. These percentiles are then used to transform URP to its MMRP equivalent value.

[5]We use NFHS 2019-2021 and PLFS 2022-23 first to adjust sampling weights in CPHS 2022-23. The reweighting algorithm uses a subset of target variables specified in the paper's main text. This selection enhances the convergence of the re-weighting algorithm; weights are winsorized as in the main text. Approach 1 is estimated using Stata's MI command and follows the model specified in the main text. Approach 2 consumption is then derived by matching the third moment to NSS-2011 (or NSS-2017) and the first and second moments from consumption derived using approach 1 – as in the main text.

[6]An expanded item set included in the 2022-23 round could account for higher consumption values among urban households in richer fractiles. We cannot further examine this hypothesis because of the unavailability of unit-level data from CES 2022-23

holds are the main contributors to the national poverty headcount.[7] Finally, based on imputed MMRP consumption values, which by construction are fully compatible with NSS 2011-12, we estimate headcount poverty rate in 2022-23 at \$1.90 to be 4.6% at the national level – comparable to the 5% estimate of Subramanian(2024) and Ghatak and Kumar(2024). The equivalent URP poverty estimates for 2022-23 for national, urban and rural India are respectively 8.6%, 9.7% and 5.5%.

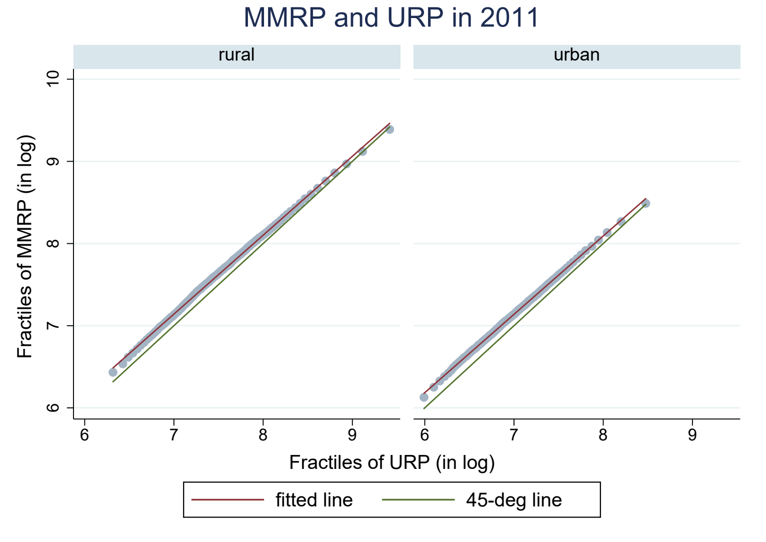


Figure 1: Relationship between percentiles of log consumption in MMRP and URP terms in NSS-2011
Notes: the horizontal axis shows percentiles of log consumption in URP terms. The vertical axis shows percentiles of log consumption in MMRP terms. The fitted line is used to transform URP-denominated consumption in CPHS 2022-23 into its MMRP equivalent, assuming that this line of fit remains unchanged overtime.

1.2 Discussion of Ghatak and Kumar(2024)

Ghatak and Kumar (2024) have argued that the lack of progress on poverty between 2011 and 2019 sits well with the empirical observation that India's share of agriculture in GDP and employment has remained steady since 2004: "Most poor countries are characterized by low rates of urbanization and a relatively high weight of agriculture in domestic output (or GDP) and employment ... The decline of the share of agriculture seemed to stop by 2004, settling at around 15-17% of GDP over the next two decades until now ... Both history and theory predict that poverty will decrease as the share

[7] Given the MMRP poverty rate of 12.4% in 2011-12, consumption at the bottom 2 to 3 deciles are most relevant for poverty measurement

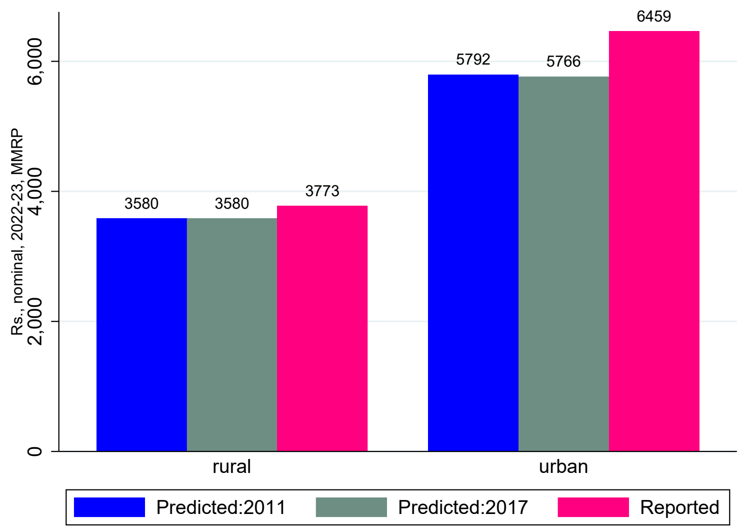


Figure 2: Predicted versus official – mean per capita consumption (nominal, MMRP)
Notes: the "reported" estimates are based on the 2022-23 factsheet released by the NSS in February 2024. The predicted 2011 and 2017 consumption values denote mean consumption per capita (in nominal, MMRP) from CPHS 2022-23 using our approach 2, with the third moment of consumption obtained from NSS-2011 and the unreleased 2017 NSS survey, respectively

of agriculture declines. It is simply not common for economies to transition from low to middle-income status without shedding the weight of agriculture in total output".

To rationalize the steep decline in poverty from around 18-25% in 2019 down to 5-8% in 2022, the authors hypothesize that the COVID pandemic introduced "a systematic shift in the level of spending on welfare programmes that clearly targeted the consumption of the poor", that resulted in a "distribution-led improvement in the poverty headcount". Put differently, "the government was forced by an unprecedented situation (albeit, not of its own making) to become an 'accidental welfarist."' The authors acknowledge that this rationalization may appear counter-intuitive at first glance: "it seems strange to argue that a pandemic that will be remembered for its brutal impact on the livelihoods of the poor instead caused improvements in their living standards. But the pandemic was also instrumental in forcing the state machinery into action and open its coffers to emergency welfare measures".

Their article reviews trends in the agriculture share of GDP up to 2022, while trends in the agriculture share of employment are reported only up to 2019. To obtain agricultural employment shares for recent years, we reproduce estimates from the "India Employment Report 2024" by the International Labor Organization (ILO) in Table 1 [8]

[8] the report is accessible at https://t.ly/hb8h_

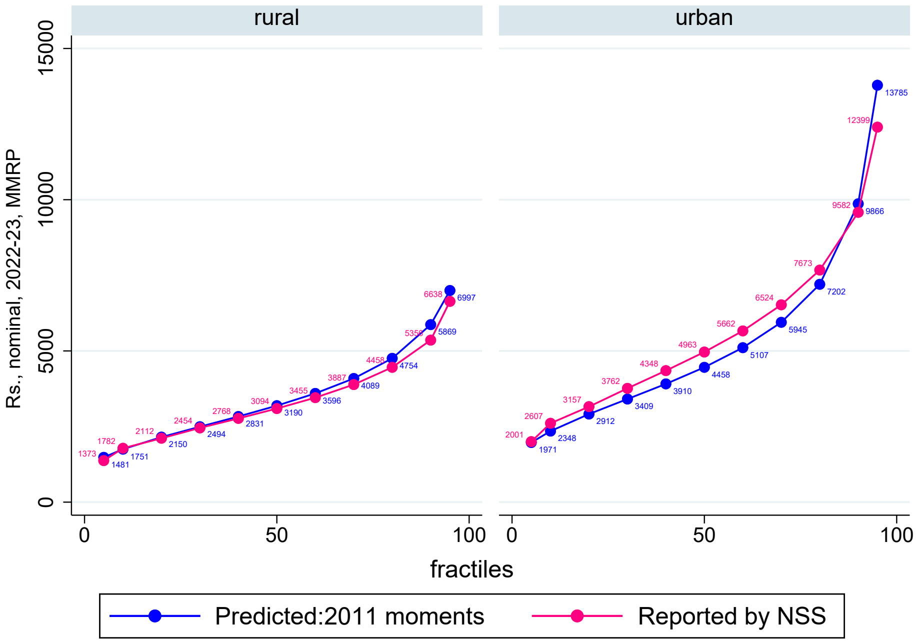


Figure 3: Predicted versus official – mean per capita consumption (nominal, MMRP)
Notes: the "reported by NSS" estimates are official statistics reproduced from the 2022-23 factsheet. The factsheet contains deciles of MMRP-denominated consumption and values at the 5th and 95th percentile. The "Predicted: 2011 moments" indicates MMRP equivalent consumption predicted using approach 2, with the third moment of consumption matched to NSS-2011. The 2017 variant of approach 2 yields values that are close to the 2011 series and excluded from the figure.

ILO's estimates show that the agricultural share of employment declined by 0.4% and 2.6% from 2000 to 2012 and from 2012 to 2019, respectively. The fall in agriculture was associated with a 9.2% rise in construction employment during the first decadal period. This rise in construction employment activity contributed to improved wages among unskilled workers relative to skilled labor within rural areas and improved relative wages for rural male workers compared to urban (Jacoby and Dasgupta, 2018). Datt, Ravallion and Murgai (2016) identified this construction boom as one of the plausible drivers of sharp poverty reduction in India between 2000 to 2011. Agricultural employment shares continued to decline after 2012 as the share of services sector employment expanded (Table 1). However, the onset of the Covid-19 pandemic introduced a break in sectoral employment trends, with the share of agricultural employment rising by 9% between 2019 and 2022.

The assumed relationship between structural transformation and poverty reduction

Table 1: Compound rate of growth of population, labor force, workforce and employment across sectors

	2000 to 2012	2012 to 2019	2019 to 2022
Population (15+)	2.39	2.07	1.15
Labor force (15+)	1.54	0.56	4.62
Workforce (15+)	1.55	0.01	5.29
Agriculture	-0.39	-2.55	8.93
Manufacturing	2.89	-0.33	3.00
Construction	9.15	2.18	6.37
Services	-0.67	10.8	1.09

in Ghatak and Kumar (2024) would predict greater poverty reduction before 2019 than in the following years since agriculture employment shares declined between 2012 and 2019 and rose until 2022. This prediction would be consistent with our analysis: a period of sharp poverty decline between 2000-2011 (based on official CES surveys), poverty reduction continuing but at a reduced pace until 2019 (based on our analysis of CPHS data in the main text), and finally, a gradual reduction from approximately 11% in 2019 to 8.5% in 2022 due to Covid induced shocks (based on section 1.1 above). Moreover, the trends in agricultural employment would be at odds with Ghatak and Kumar's assertion of a steep decline in poverty between 2019 to 2022 (from 18-25% down to 8%) – corresponding to a period with 9 % rise in agricultural employment, and little to no changes in poverty between 2011 and 2019 – despite a 2.6 % drop in agricultural employment during these years.

Ghatak and Kumar's diagnosis of a flat poverty trend between 2011 and 2019, followed by sharp poverty reduction after 2019, is also at odds with other economic indicators. First, while the exact magnitude of economic growth in India may be subject to debate, few would argue against positive growth during the 2011-2019 period. To obtain a flat trend in poverty between 2011 and 2019 under those conditions would imply a zero poverty-growth elasticity (which would require extreme assumptions). Second, a steep decline in poverty between 2019 and 2022 (from 18-25% down to 8%) would presumably involve a significant rise in real wages of casual workers (engaged in jobs other than public works) in the years following 2019 as poorer households are significantly more likely to work in casual wage jobs. Data on the level of casual wages plotted in Figure 4 (with percentages denoting annualized growth since 2011) does not reveal such patterns.

Finally, Ghatak and Kumar (2024) conjecture that the decline in poverty between 2020 and 2022 could partly stem from redistributive interventions implemented following the Covid-19 pandemic. A decline from 18-25% down to 5-8% represents, however,

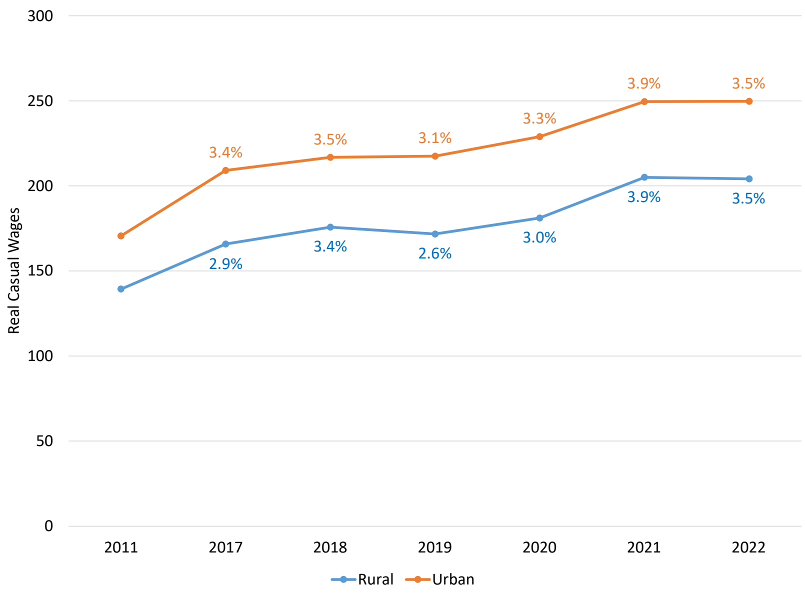


Figure 4: Casual Wages: Levels and Growth relative to 2011
Notes: Figure reports real wages of casual workers (in jobs other than public works). Wages are deflated using CPI-AL and IW for rural and urban samples, respectively. The percentages denote annualized growth in real wages relative to 2011.

a rather large change in poverty to be explained by a temporary introduction of redistributive policies alone. The Factsheet reports aggregate consumption data with and without government transfers and shows that the effect of transfers on poverty reduction is only marginally positive. Subramanian (2024) shows the differences in mean consumption with and without such transfers are only 2.31 and 0.96 percent in rural and urban areas. The small differences in consumption and their marginal impact on poverty are not surprising, considering transfers under India's emergency covid response programs were, on average, adequate enough to cover just half of the per capita grain and pulses requirements of the poorest household and large enough to cover only 5 percent of their per capita expenditure (Bhattacharya and Sinha Roy, 2021).

1.3 Discussion of Bhalla et al. (2022)

Bhalla et al. (2022; henceforward referred to as BBV) uses state-level GDP growth data to project mean household consumption growth and estimate headcount poverty between 2011 and 2020, assuming that state inequality levels have remained unchanged over this time period (as the state GDP data does not offer any information on distributional changes). For another comparative summary of BBV versus our analysis, we refer the interested reader to a recent blog post by Justin Sandefur (2022).

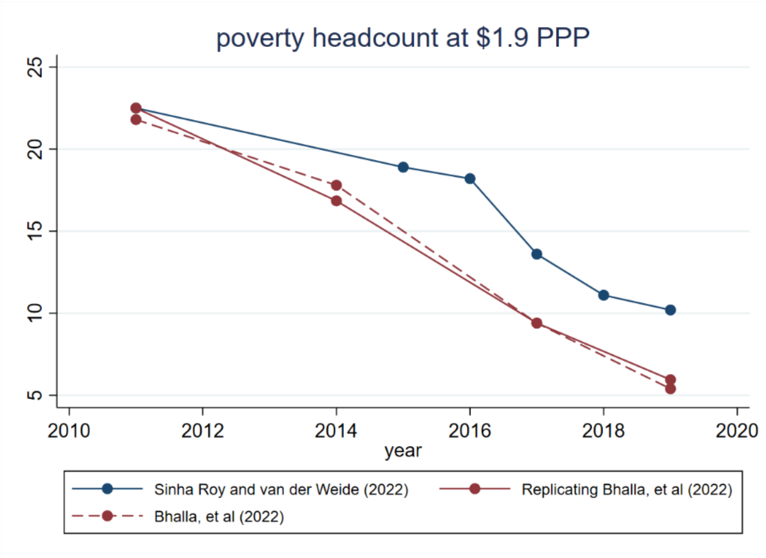


Figure 5: Replicating headcount estimates of Bhalla, et al (2022)
Notes: Our replication of Bhalla et al. (2022) starts with the empirical nominal consumption distribution denominated in URP terms from the 2011 consumption expenditure survey. Mean household consumption at the state level for the years following 2012 is estimated using growth rates in per capita gross state domestic product (GSDP), assuming distribution neutrality within the state and a pass-through value of 1. State level population projections are based on projections from national accounts. Consumptions for the years following 2013 are deflated using state (rural and urban) CPI series with 2012 as the base year. State-level CPI indices for January to December 2011 are calculated using the inflation observed between 2011 and 2012 in the CPI series which has 2010 as its base year. Consumption for all years is deflated to January-December 2011 rupee prices, converted to PPP values using the PPP exchange rate of 13.173 and 16.018 for rural and urban India as per Atamanov et al. (2020), before estimating headcounts at the $1.90 PPP line.

Figure 5 shows the poverty estimates from RR and BBV side by side, where the solid red line represents our re-production of BBV following the approach outlined in their paper. The two studies paint a markedly different picture of how headcount poverty, the population share whose consumption lies below the poverty line of $1.90 a day (in 2011 PPP), has evolved over the last decade. While both studies estimate a decline in poverty, they disagree on the magnitude of extreme poverty in India. BBV estimate that in 2019 about 5 percent of India's population lives in extreme poverty. That is half the extreme poverty rate estimated by RR. It should be expected that different data and methods yield different results. Yet, the magnitude of the difference in estimated poverty rates warrants a reflection on the possible sources of the discrepancy.

1.3.1 GDP growth does not pass through to consumption growth one-for-one

A candidate explanation stems from the fact that poverty projections based on national accounts are sensitive to the value of the pass-through rate that governs the relationship between mean household consumption growth and GDP growth. BBV observe that household consumption and national consumption from national accounts are growing at approximately the same rate between 2004 and 2011. Based on this empirical observation, BBV estimate the pass-through rate to be 1, which is somewhat higher than estimates documented in the existing literature based on cross-country regressions. It should be noted, however, that different choices of GDP growth series will generally have different pass-through rates (i.e., have a different relationship with mean household consumption growth). Rather than assuming that national GDP growth and state GDP per capita growth share the same pass-through rate (as BBV does), one would ideally work with an estimate for the series that is used to project mean household consumption growth levels, which in the case of BBV is the growth in nominal state gross domestic product.

When we estimate the pass-through rate by regressing mean household consumption growth on state GDP growth for the 2004 – 2011 period, we obtain an estimate of 0.916. While this is very close to the unit pass-through assumed by BBV and within the range of historical pass-through estimates they identified, this modest difference makes a meaningful difference for estimates of poverty. Our data also allows us to account for heterogeneity in pass-through by computing separate pass-through rates for each state (simply by dividing mean household consumption growth by state GDP growth for each state). The resulting poverty trends for both choices are shown in Figure 6, where the only change we have made is adjusting the value of the pass-through rate (all else is kept the same as in BBV).

It follows that the divergence in poverty rates between RR and BBV can largely be accounted for by adopting estimates of the pass-through rate fitted to the state GDP series. Using the uniform (i.e., average) pass-through of 0.916 (from the cross-state regression) approximately halves the gap in poverty rates between RR and BBV. Accounting for between-state differences in pass-through comes close to closing the gap completely. One difference between the two estimated poverty trends that continues to stand out is the break in the trend around 2016, corresponding to the demonetization event in India, which can be observed in RR but not in the estimates by BBV.

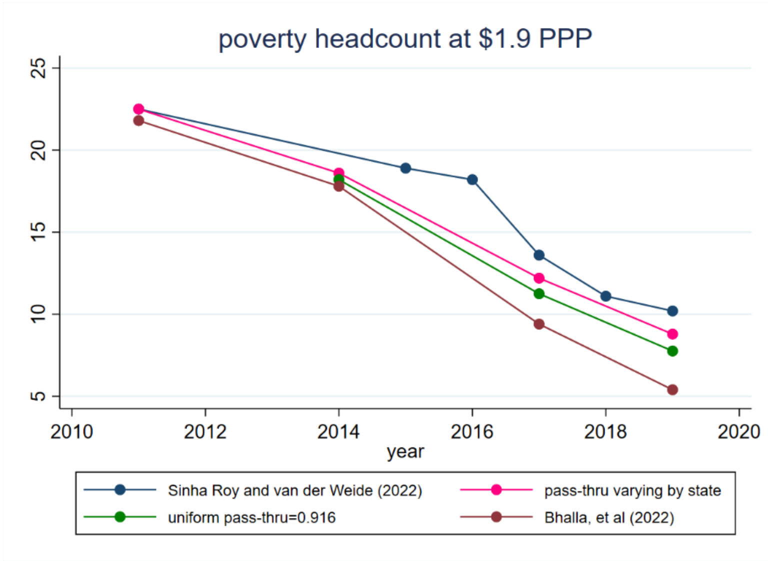


Figure 6: Estimates of poverty headcount based on State GDP specific pass-through rates
Notes: Uniform pass-through of 0.916 is estimated by regressing growth in state-level nominal consumption per capita, using the 2011 and 2004 round of CES, on growth in nominal state gross domestic product per capita over the same period. Following earlier literature, such as Lakner, et al (2022) and Mahler, et al (2021), we omit the intercept from the pass-through regression. The state-specific pass-throughs are estimated by the ratio of nominal survey consumption per capita growth and the nominal state GDP per capita for 2004 and 2011.

1.3.2 Incorporating food subsidies into the poverty calculation

Another innovation adopted by BBV is that they value household expenditures on subsidized rice and wheat at market prices (rather than at the subsidized prices for which households purchased the rice and wheat). This is by no means an unreasonable choice. It should be noted, however, that this approach implies abstracting from quality differences between goods. To illustrate this, we regress the unit price of four choices of goods (rice, wheat, milk, and salt) against log household expenditure from the 2011 consumption expenditure survey of the NSS, see Figure 7. All expenditures concern market goods (i.e., the purchase of subsidized rice and wheat are excluded). The slopes of the fitted lines arguably capture differences in quality and location. Salt is included as a control as it represents the most homogenous good among the four with little to no variation in quality, such that the slope coefficient captures variation in location, not quality. Subtracting this slope for salt from the slopes observed for rice, wheat, and milk indicates that a notable degree of the variation in prices can be attributed to quality-differences, richer households on average purchase higher quality versions of the same good at higher prices. Deaton (1988) proposes an approach to account for these quality differences.

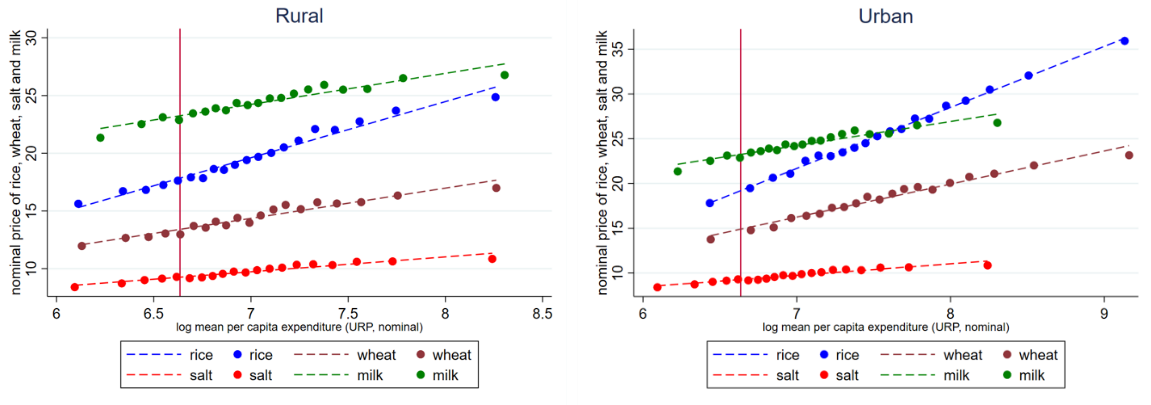


Figure 7: Quality differences induce differences in unit prices for the same item

Notes: Unit prices are calculated as the ratio of household expenditure on items and quantities consumed in the 2011 round of CES. The unit price for rice and wheat depicted in the figure does not include household cereal consumption from the public distribution system. The scatter plots in the figure are averages calculated at 20 bin intervals, while the fitted line is based on observed data; the figure is constructed using Stata's binscatter function.

Abstracting from quality differences for the purpose of poverty measurement is certainly a defensible choice. This choice has several important implications, however, that warrant further study. First, virtually every consumption good can be purchased at different quality levels and consequently exhibit variation in unit prices. Hence, a consistent application of this approach would require applying the price adjustments across a wider set of goods (not just subsidized rice and wheat). Second, what price should one value the expenditures at? The average market price assumed by BBV is generally significantly higher than the average price households below or around the poverty line purchase the goods for in rural areas (see Figure 8). In urban areas, the assumed uniform prices are notably closer to (or just below) the nominal unit price paid by poor households. Third, the same price used to adjust household expenditure values should ideally also be used to adjust the value of the national poverty line. For example, the amount of rice included in the basic needs basket underlies the poverty line should be evaluated at the same price. Using subsidized prices to value the poverty line but higher market prices to value household expenditures (as BBV currently does) will lead to an underestimation of poverty. It should be noted that BBV adopt the international poverty line of $1.9 a day and hence do not have to evaluate the price of the basic needs consumption basket for India. By the same token, however, the international poverty line is derived from cross-country data on poverty lines that reflect the cost of acquiring the basic needs basket. Conceptually, this means that if one were to change the prices used to evaluate household expenditures, one would have to use the same prices to evaluate the cost of the basic needs basket to verify whether households can afford the basic needs consumption basket.

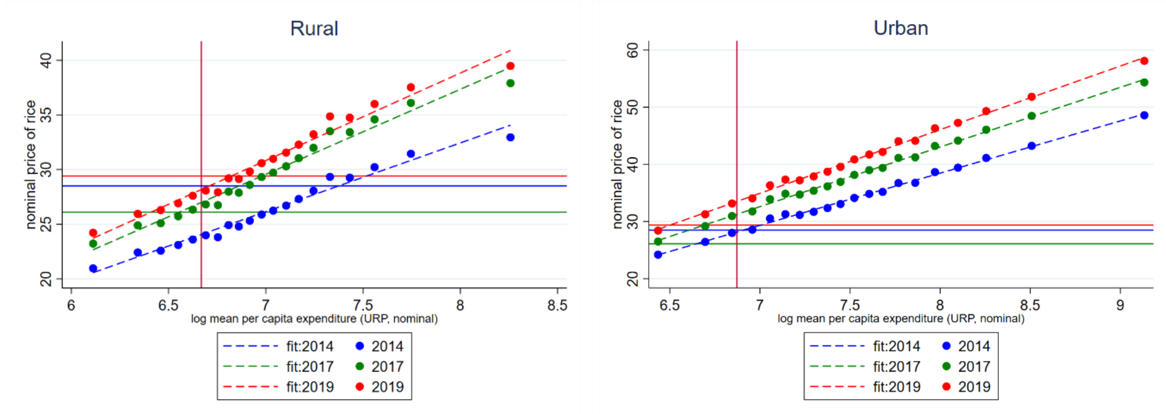


Figure 8: Selection of a uniform market price can impact estimates of poverty due to quality differentials
Notes: Unit prices for 2014, 2017 and 2019 are calculated using the observed unit price of rice in CES 2011 round and the food price inflation reported in the CPI state (rural and urban) series – the BBV's preferred choice of deflator series. The horizontal lines in different colors represent the price at which BBV values PDS consumption of rice. The vertical red line is the \$1.90 poverty line – poor households are on the left of this line. The figure shows that in rural areas, the unit price of rice implied by state-level CPI food indices is lower than the price of rice assumed by BBV across years. This discrepancy will overestimate poverty reduction in rural areas but underestimate poverty reduction in urban. Since discrepancies are markedly larger in rural samples and the rural population is about 67 percent of the overall population, the national poverty estimates at BBV's assumed price levels likely overestimate average consumption and underestimate national poverty headcounts. As in Figure 7, the scatter plots indicate mean values across 20 bins using Stata's binscatter function.

1.4 Discussion of other studies estimating poverty based on "usual consumption expenditure" aggregate

In the absence of official consumption expenditure surveys since 2011, selected researchers have used an alternative consumption variable called the "usual household consumption expenditure" to examine changes in average consumption and estimate poverty. This variable first appeared in NSS' 72nd round surveys, conducted in 2014-

15, and more recently in periodic labor force surveys of 2017 to 2019. Mehrotra and Parida (2021) use this consumption variable to show that the headcount poverty in India rose from 25.7 to 30.5 percent between 2011-12 and 2019-20 (based on the Tendulkar Committee's poverty lines). Similarly, Himanshu uses the same variable to show that there was a decline in rural and urban consumption of 4.4 and 4.8 percent per annum, respectively since 2015-16.[9]

The usual consumption expenditure is a single expenditure variable in NSS surveys. It is constructed by the enumerator by first establishing the usual expenditure for household purposes in a month, then determining purchase values of all household durables in the past year and dividing it by 12, and finally, imputing the approximate usual consumption from wages in-kind, home-grown stock and free collection of goods based on her own assessment of market prices for these products. The survey instrument does not require the enumerator to input the values of each component separately -– instead, the enumerators aggregate the components and enter lumpsum into the instrument.

It is hypothesized that the aggregation of components by enumerators, as well as the demands on respondent attentiveness needed to classify expenditures across components correctly, increases the scope for measurement error(s). With that in mind, respondents (or enumerators) can be expected to round off consumption values – consistent with theories of satisficing (Krosnick, 2018). Gideon et al. (2017) show that rounding off is a common coping strategy respondents adopt when they encounter difficult information retrieval questions in a survey. The extent to which these rounding-off errors can impact poverty estimates is an empirical question.

Let us examine the extent of bunching in the usual consumption expenditure around round numbers. Figure 9 plots the densities of household expenditures from NSS 2014-15 (Schedule 1.5) and NSS 2014-15 (Schedule 21.1) in multiples of Rs. 1000. The horizontal axis shows the remainder value when usual consumption is divided by 1000 (that is, the modulus function). Values clustered around 0 indicate usual consumption expenditure values that are exact multiples of Rs. 1000; those clustered around 500 depict expenditures that is 500 more than a multiple of 1000, and so on.[10] The figure suggests significant heaping of consumption: 60 percent of households in both surveys rounded off consumption to the nearest Rs. 1000 value and an additional 15 percent of households rounded off their welfare aggregate to the nearest Rs.500. In comparison,

[9] https://www.livemint.com/opinion/columns/opinion-what-happened-to-poverty-during-the-first-term-of-modi-1565886742501.html

[10] We choose the NSS 2014-15 round to conduct this assessment because it is the first full year for which the usual consumption expenditure welfare aggregate was captured by NSS. It is also the survey closest to the NSS 2011 survey

incidences of consumption being rounded off in NSS-2011 and CPHS-2015 is limited: households are almost equally likely to report any consumption estimate in multiples of 1 to 1000.

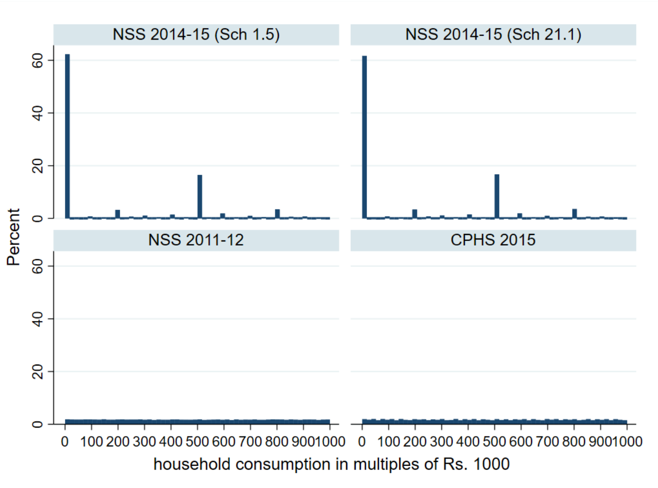


Figure 9: Fraction of households by reported levels of consumption
Notes: the horizontal axis is the modulus of reported household consumption with respect to 1000. For instance, the value 0 indicates that the consumption reported in the survey is in multiples of Rs. 1000. The value 1 indicates a usual consumption value of Rs. 1 more than a multiple of Rs. 1000, and so on. Fractions are unweighted; consumption is in nominal terms and at the household level in all surveys.

These rounding-off errors will introduce errors in estimates of consumption poverty and inequality. To quantify the potential impact of rounding-off errors, we conduct two simulations. In the first simulation exercise (designed to replicate the heaped distribution of consumption observed in NSS 2014-15 (Sch. 1.15) of Figure 9 in the NSS-2011), the consumption values recorded in the NSS-2011 are rounded down to the nearest Rs. 1000 values. 62.2 percent of households in the NSS 2014-15 report consumption in multiples of 1000. The same proportion of households in the NSS-2011 are chosen randomly with their reported consumption values rounded down to the closest Rs.1000 multiple. The second simulation exercise rounds up the actual consumption in NSS-2011 to the nearest Rs. 1000, similarly replicating the heaped distribution observed in the NSS 2014-15.[11] Table 2 below shows the extent of rounding off bias in headcount and inequality through these simulations. In cases where consumption is rounded down, the headcount rate at the 1.90 line is 5.8 percentage points higher than the actual estimate

[11] In practice, errors in reporting and rounding up or down of consumption is likely a function of household characteristics: richer households may find it more difficult to aggregate consumption from diverse sources mentally. Conversely, the enumerator could make mistakes in attributing the correct market prices for self-produced consumption

at the all-India level. When consumption is rounded upwards, headcount rates are 11.2 percentage points lower. Similarly, inequality is 0.015 Gini points higher and 0.021 Gini points lower in cases of downward and upward rounding-off consumption, respectively.

	Poverty headcount rate at 1.9 international line		
	Observed consumption	**Rounding down**	**Rounding up**
Rural	26.3%	32.1%	15.1%
Urban	14.2%	15.5%	8.3%
India	22.8%	27.4%	13.2%
	Gini measure of inequality		
	Observed consumption	**Rounding down**	**Rounding up**
Rural	0.3113	0.3279	0.2923
Urban	0.3901	0.3996	0.3743
India	0.3540	0.3692	0.3335

Table 2: Sensitivity of poverty and inequality estimates to rounding errors. Notes: Estimates due to rounding errors are constructed by simulating the heaped distribution of usual consumption expenditure variable in NSS 2014-15 rounds into the 2011-12 consumption survey. The estimates for rural and urban India in the table are the same as Povcalnet. However, there is a small difference in the all-India figures due to differences in rural and urban population shares assumed in Povcalnet.

References

Atamanov, Aziz, Lakner, Christoph, Mahler, Daniel Gerszon, Tetteh Baah, Samuel Kofi and Yang, Judy (2020). The effect of new ppp estimates on global poverty. Global Poverty Monitoring Technical Note 12. The World Bank.

Bhalla, Surjit and Bhasin, Karan (2024). India eliminates extreme poverty. Blog Post https://www.brookings.edu/articles/india-eliminates-extreme-poverty/. Brookings.

Bhalla, Surjit, Bhasin, Karan and Virmani, Arvind (2022). Poverty, inequality, and growth in india: 2011-2018. mimeo. IMF.

Bhattacharya, S. and Sinha Roy, S. (2021). Intent to implementation. mimeo. The World Bank.

Datt, Ravallion M. G. and Murgai, R. (2016). Growth, urbanization and poverty reduction in india. *National Bureau of Economic Research* N, **o. w21983**.

Deaton, A. (1988). Quality, quantity, and spatial variation of price. *The American economic review* 418–430.

Deshapande, A. (2024). What consumption expenditure survey leaves unanswered. Blog https://indianexpress.com/article/opinion/columns/what-consumption-expenditure-survey-leaves-unanswered-9187003/. Indian Express.

Ghatak, M. and Kumar, R. (2024). Poverty in india over the last decade. Technical Report https://www.theindiaforum.in/economy/poverty-india-over-last-decade. India Forum.

Gideon, Michael, Helppie-McFall, Brooke and Hsu, Joanne W. (2017). Heaping at round numbers on financial questions: The role of satisficing. *Survey research methods*, **11, No. 2**.

Himanshu (2024). We need a bridge consumption survey to accurately analyse indian poverty. Blog https://www.livemint.com/opinion/online-views/we-need-a-bridge-consumption-survey-to-accurately-analyse-indian-pover-11709815759161.html. Livemint.

Jacoby, H. G. and Dasgupta, B. (2018). Changing wage structure in india in the post-reform era: 1993–2011. *IZA Journal of Development and Migration*, **8,** 1–26.

Krosnick, Jon A. (2018). *Questionnaire design*, The Palgrave handbook of survey research edn. Palgrave Macmillan.

Lakner, Mahler D. G. Negre M. C. and Prydz, E. B. (2022). How much does reducing inequality matter for global poverty?. *The Journal of Economic Inequality*, **20-3,** 1–26.

Mahler, D. G., Castaneda Aguilar, A. and Newhouse, D. (2021). Nowcasting global poverty. Policy Research Working Paper 9860. World Bank.

Mehrotra, Santosh and Parida, Jajati Keshari (2021). Poverty in india is on the rise again. Blog. The Hindu.

Sandefur, J. (2022). The great indian poverty debate, 2.0. Blog Post. Center for Global Development.

Subramanian, S. (2024). The household consumption expenditure survey 2022-23: What the 'fact sheet' suggests about welfare, inequality, and poverty in india. Technical Report https://www.theindiaforum.in/economy/household-consumption-expenditure-survey-2022-23. The India Forum.

Tarozzi, A. (2007). Calculating comparable statistics from incomparable surveys, with an application to poverty in india. *Journal of business and economic statistics*, **25(3),** 314–336.

Supplementary Appendix 2: Discussion on Dreze and Somanchi

Annex: Response to Dreze and Somanchi

May 2024

1 Introduction

The National Sample Survey Office (NSSO) in India released its last household consumption expenditure survey in 2011, which served as the primary source of data for tracking household standards of living, including official estimates of poverty and inequality in India. A private sector company by the name of CMIE has recently stepped in to fill this gap by introducing a new source of household consumption expenditure survey data called the Consumer Pyramid Household Survey (CPHS). It did not take long for the CPHS to be adopted by scholars to address various outstanding empirical questions.[1] The present study (henceforward SR) employs the CPHS to produce updated estimates of poverty and inequality in India for the years 2015 to 2019.

Despite several advantages[2], SR underscore several limitations of the CPHS. Most prominently, key statistics related to demographics, education, asset ownership, etc. deviate significantly from statistics obtained using benchmark surveys such as the PLFS, NFHS, and non-expenditure surveys conducted by the NSS. This suggests that the sample of households is not representative of the full population. Earlier scholarly work has independently reached the same conclusion (Dreze and Somanchi, 2021; Somanchi,

[1]This includes efforts to estimate the impacts of India's demonetization policy (Chanda and Cook, 2020; Chodorow-Reich et al., 2020), to estimate the impact of Covid induced lockdowns on economic activities (Beyer et al., 2021) and labor market indicators (Deshpande, 2020; Gupta et al., 2021a, 2021b; Agarwal, 2021); to track India's social protection response to Covid (Bhattacharya et al., 2021); investigate rates of consumption and savings in low-income households in India (Ghatak et al., 2020) and to examine women's work under social norms in the marriage market (Afridi et al., 2023).

[2]Several features of the CPHS make it an attractive alternative for poverty measurement in the absence of NSS rounds: (1) availability of consumption aggregates for 115 items, (2) a large sample size of 174,000 households that covers 28 states representing over 95% of India's population, and (3) uninterrupted rounds of survey data collection during 2015-2019 which allows for high-frequency monitoring of household standards of living. It should be noted that SR consider NSS' consumption expenditure rounds to be the main dataset for poverty measurement and propose using the CPHS to fill gaps in poverty statistics only for years in between NSS' consumption expenditure survey.

2021; Pais and Rawal, 2021; Srija and Singh, 2021). In Somanchi (2021) and Dreze and Somanchi (2021), the authors argue that the CPHS undercovers remote locations and over-samples households located on the main street of each location/village/town, which they term the "main street" bias of the CPHS. Main street bias is therefore proposed to be a candidate reason for observed non-representativeness in the CPHS. To mitigate such concerns, SR adopt a reweighting approach (called maxentropy) in an effort to restore representativeness in the CPHS. Sections 2.5 and 3 of SR provides a detailed discussion of the reweighting procedure and a comprehensive evaluation that compares reweighted CPHS indicators against those derived from benchmark sources.

A recent article by Dreze and Somanchi (2023; DS-2023 hereafter) has taken a critical view of the reweighting approach. The study tests the effectiveness of the approach by means of a simulation exercise mainly using the government's Periodic Labor Force Surveys (PLFS). The premise is that the CPHS undercovers poor households. DS's simulation experiment systematically drops poor households from the PLFS to simulate a CPHS-styled survey that underestimates poverty. In their baseline scenario, households are randomly dropped if their observed consumption falls below the 40th percentile of the distribution; two other scenarios vary the quintile threshold and the number of observations excluded. The authors then adopt the reweighting approach and evaluate whether the adjusted sampling weights are able to reproduce the estimates of poverty and mean consumption observed in the nationally representative version of the survey (prior to dropping poor households). Their study concludes that while reweighting may be effective at resolving biases in statistics on several non-expenditure variables, considerable biases remain in expenditure and poverty statistics.

DS's simulation study makes a valuable contribution to advancing the poverty measurement debate in India. The study highlights an important limitation of re-weighting; when surveys exclude households based on their consumption expenditure level, reweighting is of limited use in addressing the resulting bias in estimates of poverty. It is also important to note however that the inability to correct for biases that stem from excluding households based on their actual consumption levels may be of lesser practical relevance. As the CMIE does not observe a household's consumption expenditure level prior to data collection (there would be no need to conduct the consumption survey if it did), it is unlikely that the consumption based scenarios simulated in DS-2023 characterize the biases observed in the CPHS.[3]

The "main street" bias conjectured in Dreze and Somanchi (2021; DS-2021 hereafter) arguably describes a more plausible scenario, where households are excluded based on their location or accessibility (or other heuristics that CMIE can observe prior to conducting the survey). Indeed, households located well off the main street and/or in remote areas may be excluded as collecting data for these households is presumably more expensive than others. Alternatively, CMIE's field operational protocols could cause enumerators to exhaust cluster-level sampling quotas at the main street itself before moving on to more remote families. Other reasons not related to accessibility are also plausible. For example, selected households such as casual workers sampled for interviews may not have the time to respond to surveys because their job may not allow taking breaks, or households from minority groups may not feel inclined to participate in survey interviews.

The objective of this annex is to probe "main street" bias as the candidate channel through which CPHS experiences loss of representativeness – and evaluate whether reweighting is more successful in restoring representativeness under this scenario.[4] Our findings are three-fold. First, we reproduce the empirical results obtained by DS-2023 using a different dataset, confirming that reweighting is of limited use when households are excluded from surveys based on their consumption expenditure levels. Second, reweighting is found to be significantly more effective when surveys over-sample accessible households or households residing on the "main street"; more than three-quarters of a 5 percentage point bias in poverty estimates is resolved by means of reweighting. Third, we provide empirical evidence that corroborates the "main street" bias argument put forward in DS-2021; we estimate that access to the main road is about 3-4 percentage points higher in the CPHS when compared to nationally representative surveys such as the NSS' 76$^{\text{th}}$ round and the PLFS.

[3]Take for example a case where a CMIE enumerator systematically ignores poorest households. This study will show that as long as the enumerator relies on observed correlates of household poverty (caste, location on main street, building materials, etc.), re-weighting can partially restore representativeness in a way that significantly reduces bias in estimate of poverty. In the implausible case where the enumerator can precisely observe actual consumption levels of the household before conducting the survey, the re-weighting procedure will be less successful in reducing bias in poverty measurement.

[4]Note that "main street" bias in this case refers to bias stemming from the exclusion of households based on non-consumption attributes such as access to the main street, remoteness more generally, in combination with other characteristics such as caste, literacy and other indicators. All of these additional attributes can limit a household access to poverty and by extension, reduce consumption. This study identifies the source of biases in CPHS by distinguishing between two cases: (1) where households are excluded precisely based on their consumption and (2) where households are excluded due to other non-expenditure related variables that are correlated with main street bias as well as consumption.

2 Excluding poor households from sample

2.1 Notation and assumptions

Let us assume that log household consumption per capita y_{ah} satisfies:

$$y_{ah} = \beta_a^T x_{ah} + \sigma_a \varepsilon_{ah}, \tag{1}$$

where where the subscripts a and h indicate area and household, respectively, x_{ah} is a vector with covariates (including the intercept), and ε_{ah} is a zero expectation error term with unit variance. The model parameters β_a and σ_a are allowed to vary across urban/rural and state-urban/rural, respectively.

The premise is that the survey undercovers poor households. Broadly speaking, there are two scenarios through which poor households end up being excluded from the survey sample: (1) households are excluded when their log consumption per capita is below a given threshold (this is the scenario put forward by DS-2023), and (2) households are excluded based on a combination of selected covariates (think of variables capturing remoteness, literacy, caste etc. as in DS-2021). The exclusion conditions are modeled by $y_{ah} < \tau$ or $\gamma^T x_{ah} < \tau$.

Reweighing will be used in an effort to correct for this. The effectiveness of reweighting will depend on the criteria for excluding households from the sampling-frame. Note that in the first scenario, the error term ε_{ah} co-determines whether households are included in the sampling-frame, or not. As the error term is not observed (i.e., as household consumption is not observed) and cannot be used as a target variable to reweigh on, reweighting is expected to be less successful when poor households have been undercovered based on the first criteria. The exclusion of households described in the second scenario similarly results in an undercoverage of poor households to the extent that the selected covariates are sufficiently strong predictors of household (log) consumption – and that these covariates can be used as target variables to reweigh on.

Let the population mean of log consumption per capita for area a be denoted by μ_a, let S represent the truncated sample of households, and let the expectations operator $E_a[.]$ evaluate the mean value across households in area a (observed in the truncated sample) using adjusted weights. The adjusted weights are assumed to reproduce the true population mean of x_{ah}. Let us formalize this in the following assumption.

Assumption 1 *After reweighting, the truncated sample mean of $\beta^T x_{ah}$ matches the population mean:*

$$E_a[\beta^T x_{ah} | x_{ah} \in S] = \mu_a. \tag{2}$$

For ease of exposition, it will also be convenient to assume normally distributed errors.

Assumption 2 *The error term ε_{ah} has mean zero, is normally distributed, and is orthogonal to x_{ah}.*

As sampling error associated with the (unbiased) estimators of the parameters β_a and σ_a are of secondary importance, we will simply work with the true parameters and ignore these errors.

2.2 Estimating mean (log) consumption

Let us first focus on estimating mean log consumption. In the case of scenario 2 (excluding households based on observable attributes), reweighting is seen to yield unbiased estimates of mean log consumption:

$$\begin{aligned} E_a[y_{ah} | x_{ah} \in S] &= E_a[\beta_a^T x_{ah} + \sigma_a \varepsilon_{ah} | x_{ah} \in S] && (3)\\ &= E_a[\beta_a^T x_{ah} | x_{ah} \in S] && (4)\\ &= \mu_a. && (5)\end{aligned}$$

The last step is due to Assumption 1.

DS-2023 demonstrates by means of a simulation experiment that estimates of mean log consumption will not be unbiased in the case of scenario 1, where household are excluded based on their level of consumption expenditure. Under Assumptions 1 and 2, an analytic expression for the bias term can be obtained. Let the area mean and variance of y_{ah} conditional on household attributes x_{ah} be denoted by $\mu_{ah} = \beta_a^T x_{ah}$ and σ_a^2, respectively. Expected log consumption for a given household conditional on their log consumption expenditure exceeding the threshold τ (i.e., corresponding to the sample where poor households are excluded) is seen to solve:

$$E_{ah}[y_{ah} | y_{ah} > \tau] = \mu_{ah} + \sigma_a \left(\frac{\phi((\tau - \mu_{ah})/\sigma_a)}{1 - \Phi((\tau - \mu_{ah})/\sigma_a)} \right), \tag{6}$$

where $\phi(z)$ and $\Phi(z)$ are the standard normal probability density function and cumulative distribution function, respectively. Taking expectations across households from area a, we obtain:

$$E_{ah}[y_{ah}|y_{ah}>\tau] = E_a[\mu_{ah}] + \sigma_a E_a\left[\frac{\phi((\tau-\mu_{ah})/\sigma_a)}{1-\Phi((\tau-\mu_{ah})/\sigma_a)}\right] \quad (7)$$

$$= \mu_a + \sigma_a E_a\left[\frac{\phi((\tau-\mu_{ah})/\sigma_a)}{1-\Phi((\tau-\mu_{ah})/\sigma_a)}\right]. \quad (8)$$

The second term on the right-had-side is the bias term, which increases with σ_a and is reduced to zero when $\sigma_a \to 0$.

2.3 Estimating head-count poverty

When the sample excludes poor households (with log consumption per capita below τ; scenario 2), as simulated by DS-2023, the truncated cumulative distribution function for y_{ah} conditional on household attributes x_{ah} (under the assumption of normally distributed errors; Assumption 2) will take the form:

$$F(y|y>\tau, x_{ah}) = \frac{\Phi\left(\frac{y-\mu_{ah}}{\sigma_a}\right) - \Phi\left(\frac{\tau-\mu_{ah}}{\sigma_a}\right)}{1-\Phi\left(\frac{\tau-\mu_{ah}}{\sigma_a}\right)}. \quad (9)$$

Accordingly, the probability that a household with attributes x_{ah} lives below the (log) poverty line z equals:

$$Prob[y_{ah}<z|y_{ah}>\tau, x_{ah}] = \frac{\Phi\left(\frac{z-\mu_{ah}}{\sigma_a}\right) - \Phi\left(\frac{\tau-\mu_{ah}}{\sigma_a}\right)}{1-\Phi\left(\frac{\tau-\mu_{ah}}{\sigma_a}\right)}. \quad (10)$$

It can be seen that evaluating the mean value across households from area a will generally not reproduce the population mean (i.e., the true headcount poverty rate). The reason for this is that the mean value of $\Phi((z-\beta_a^T x_{ah})/\sigma_a)$ involves higher moments of $\beta_a^T x_{ah}$. While reweighting is assumed to fix the first moment, it is not guaranteed to fix higher moments. Hence, Assumption 1 cannot rule out a bias in this case.

What separates poverty estimation from the estimation of mean log consumption is that poverty estimates will also be biased in the case where households are excluded based on covariates, even after reweighting.[5]

It follows that there are two sources of bias in estimates of poverty and mean (log)

[5] To see this, observe that the probability that a household with attributes x_{ah} lives below the poverty line is a non-linear function of x_{ah}. Hence, the bias emerges for the same reason, namely that reweighing need not address biases in higher moments of the distribution of x_{ah}.

consumption: (1) bias in moments of $\beta^T x$ (distribution of household covariates), and (2) selection on the household idiosyncratic error term. Reweighing is assumed to fix the first moment of $\beta^T x$, but not necessarily higher moments. This first source of bias matters only when estimating welfare measures that involve non-linear functions of household log consumption, such as the poverty head-count rate (the non-linearity makes that mean poverty will be sensitive to higher moments of the distribution of $\beta^T x$). Estimates of mean (log) consumption, on the other hand, is only sensitive to the first moment of the distribution of $\beta^T x$. The second source of bias, selection on the household error term, applies to all choices of welfare measures (both poverty and mean log consumption), but only when households are excluded based on their level of consumption expenditure (which is co-determined by the household error term). In summary, when households are excluded based on consumption, both sources of errors apply (which bias both estimates of poverty and mean log consumption). When households are excluded based on covariates such as remoteness, only the first source of error applies (leaving estimates of mean log consumption unbiased). How the two sources of bias compare in terms of magnitudes is ultimately an empirical question.

3 Diagnosing which scenario is more plausible

3.1 A simple diagnostic

The analytic expressions for the bias in mean log consumption derived in Section 2.2 could be employed to test whether poor households are excluded from the sample based on their consumption expenditure levels or their non-consumption attributes. Let b_a equal the difference between the expected value of y_{ah} (log consumption per capita) and the expected value of $\beta_a^T x_{ah}$ (predicted log consumption based on household covariates) for area a. Note that the latter is denoted by μ_a. Under Assumptions 1 and 2, b_a equals zero when households are excluded based on non-consumption attributes, but takes on the following non-zero value when households are excluded based on their consumption levels (see eq. (8)):

$$b_a = E_a \left[\sigma_a \frac{\phi((\tau - \beta_a^T x_{ah})/\sigma_a)}{1 - \Phi((\tau - \beta_a^T x_{ah})/\sigma_a)} \right]. \tag{11}$$

The left-hand-side of eq. (11) can be readily estimated by replacing expected values with sample means (using adjusted survey weights). Let this variable be denoted by $\hat{b}_a$.

Suppose that the value of τ is known. Then, given estimates for the parameters β_a and σ_a (denoted by $\hat{\beta}_a$ and $\hat{\sigma}_a$), the right-hand-side of eq. (11) can be obtained by evaluating the sample mean of $\hat{\sigma}_a \phi((\tau - \hat{\beta}_a^T x_{ah})/\hat{\sigma}_a)/(1 - \Phi((\tau - \hat{\beta}_a^T x_{ah})/\hat{\sigma}_a))$ for each area a. Let this variable be denoted by $\hat{q}_a(\tau)$.

This suggests that regressing $\hat{b}_a$ on $\hat{q}_a(\tau)$ could be adopted as a diagnostic tool. In the event households have been excluded based on their consumption expenditure levels, then the resulting regression coefficient is predicted to equal 1. Alternatively, if households have been excluded based on non-consumption attributes, then the regression coefficient is predicted to equal 0. In practice, however, one does not know the value of τ. We will address this by evaluating $\hat{q}_a(\tau)$ for different choices of τ and run the above-mentioned regression for each. The uncertainty in τ means that the regression coefficient need not necessarily equal 1 in the case where households are dropped based on their consumption levels. We will interpret regression coefficients that are positive and significant as evidence favoring the scenario where households are excluded based on their consumption levels and regression coefficients that are not positive and significant as evidence favoring the scenario where households are dropped based on non-consumption attributes.

3.2 A simulation experiment

We execute a DS-2023 styled simulation experiment to distinguish between two possible sources of bias in CPHS: (1) exclusions based on precisely observed value of household consumption; or (2) households excluded based on variables that are correlated with main street access and eventually consumption. To evaluate the effectiveness of the diagnostic tool described in Section 3.1, we use the NSS-2011 to simulate the different ways in which poor households can be excluded from the sample. Households are dropped either based on their consumption level or based on non-consumption attributes that are correlated with consumption and assess whether the regression described in Section 3.1 is able to distinguish between these two scenarios. For ease of exposition, the two alternative scenarios are implemented as follow. Household are excluded from the sample when: (1) their consumption expenditure is below the 25th percentile of the consumption distribution (for urban and rural India separately), or (2) their value of $\beta^T x_{ah}$ is below the 25th percentile of the $\beta^T x_{ah}$ distribution (for urban and rural India separately).

To assess the significance of Assumption 2, we consider three different choices for simulating household (log) consumption expenditures:

1. $y_{ah} = \beta_a^T x_{ah} + \sigma_a \varepsilon_{ah}$, where ε_{ah} is drawn from the standard normal distribution independent from x_{ah}.

2. $y_{ah} = \beta_a^T x_{ah} + \sigma_a \varepsilon_{ah}$, where ε_{ah} is drawn without replacement from the empirical errors observed in the NSS-2011 (imposing independence from x_{ah} and zero mean across areas a).

3. Using the observed household (log) consumption data for y_{ah}.

The first option satisfies all aspects of Assumption 2. The second option satisfies the zero mean and orthogonality to x_{ah} requirements but does not guarantee that the error term is normally distributed. The third option does not guarantee any of the requirements from Assumption 2.

After selected households are dropped, we use the resulting sample to construct $\hat{\beta}$ for urban and rural using OLS regression. $\hat{\sigma}_a$ is obtained by evaluating the standard deviation of the residuals for each state-urban/rural separately. Next we construct the variable $z_{ah} = \hat{\sigma}_a \phi((\tau - \hat{\beta}_a^T x_{ah})/\hat{\sigma}_a)/(1 - \Phi((\tau - \hat{\beta}_a^T x_{ah})/\hat{\sigma}_a))$ for different choices of τ (selected percentiles of either y_{ah} or $\hat{\beta}_a x_{ah}$ evaluated for urban and rural separately), and evaluate the following mean values (that will feature as dependent and independent variables in the diagnostic regression):

$$\hat{b}_a = \bar{y}_a - \hat{\beta}_a \bar{x}_a \tag{12}$$

$$\hat{q}_a = \bar{z}_a, \tag{13}$$

where $\bar{y}_a$, $\bar{x}_a$ and $\bar{z}_a$ denote sample mean values of y_{ah}, x_{ah} and z_{ah}, respectively, evaluated at the state-urban/rural level. Finally, we regress $\hat{b}_a$ on $\hat{q}_a$.

Figures 1 through 3 plot the regression coefficients from the regressions of $\hat{b}_a$ on $\hat{q}_a$ against the choice of percentile of either y_{ah} or $\hat{\beta}_a x_{ah}$ (corresponding to different values of τ) for the three alternative assumptions concerning the simulated household (log) consumption data. Assumption 2 is satisfied in Figure 1, but violated in Figures 2 and 3 (with Figure 3 corresponding to the larger violation of Assumption 2). The incremental violations of Assumption 2 notwithstanding, which are seen to incrementally weaken the results, the diagnostic is successfully able to distinguish between households being excluded based on their consumption levels versus their non-consumption attributes.

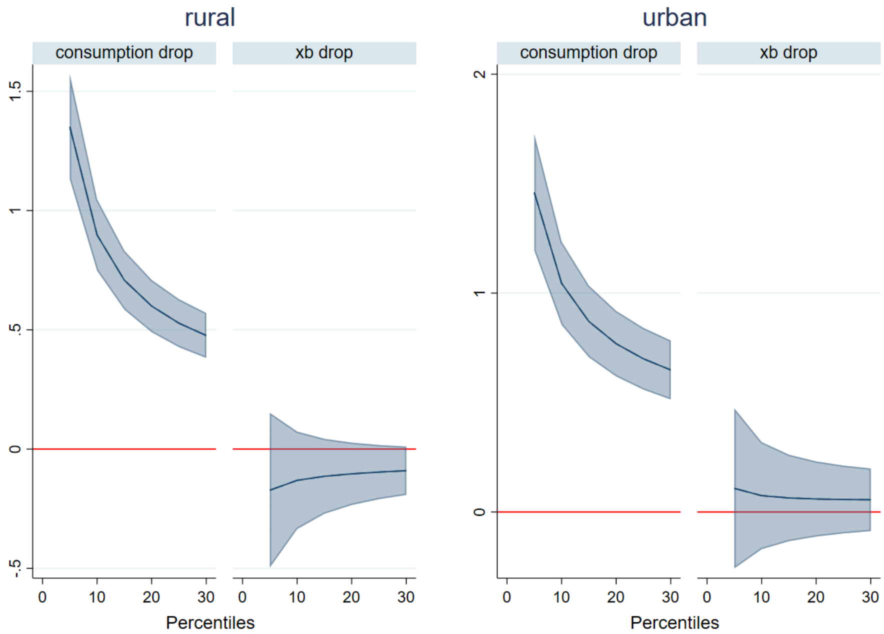


Figure 1: Regression coefficient from regression of $\hat{b}_a$ on $\hat{q}_a(\tau)$ for different choices of τ (selected percentiles evaluated for urban and rural separately). Household consumption per capita is simulated using actual data for $\beta^T x_{ah}$ from the NSS-2011 and normally distributed errors. The variance of the errors vary across state-urban/rural (and are obtained from the NSS-2011).

3.3 Diagnosing the CPHS

This section is organized into two parts. Both parts aim to assess which of the two scenarios by which households may have been excluded from the CPHS sample is more plausible (i.e., the plausibility that households are excluded from the CPHS sample based on their consumption expenditure levels versus their non-consumption attributes such as remoteness or main street access). The exercise presented in the first subsection applies the diagnostic regression from Sections 3.1 and 3.2 to the CPHS data. In the second subsection, we impute main street access into the CPHS and compare it to the levels observed in another NSS survey that provides representative estimates of household level access to the main street in 2018.

3.3.1 Diagnostic regression applied to CPHS

Suppose that poor household are excluded from the CPHS sample either based on their consumption levels (scenario 1) or based on non-consumption attributes (scenario 2). The diagnostic regression analysis described in Section 3.1 and tested by means of a sim-

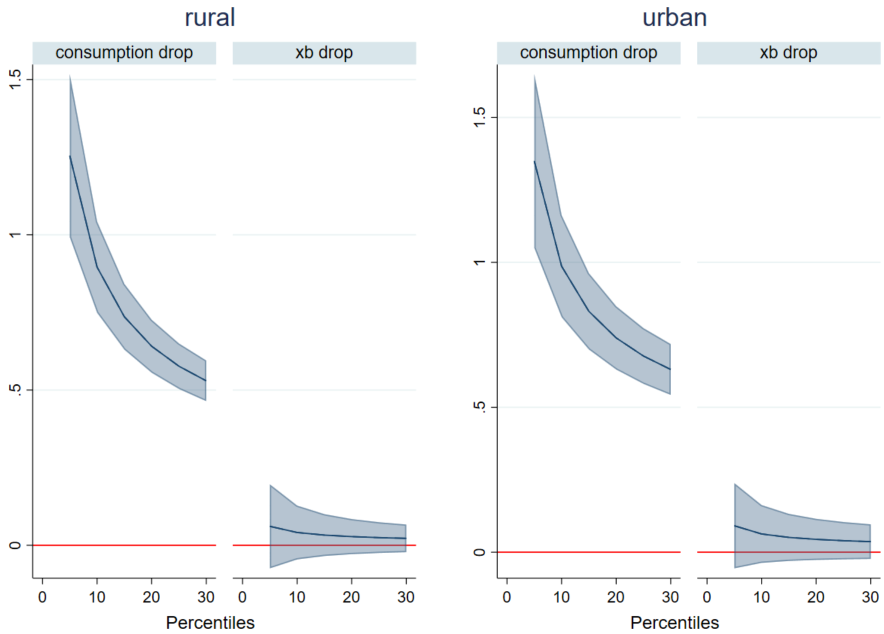


Figure 2: Regression coefficient from regression of $\hat{b}_a$ on $\hat{q}_a(\tau)$ for different choices of τ (selected percentiles evaluated for urban and rural separately). Household consumption per capita is simulated using actual data for $\beta^T x_{ah}$ and empirical errors from the NSS-2011. The errors are drawn randomly without replacement at the state-urban/rural level.

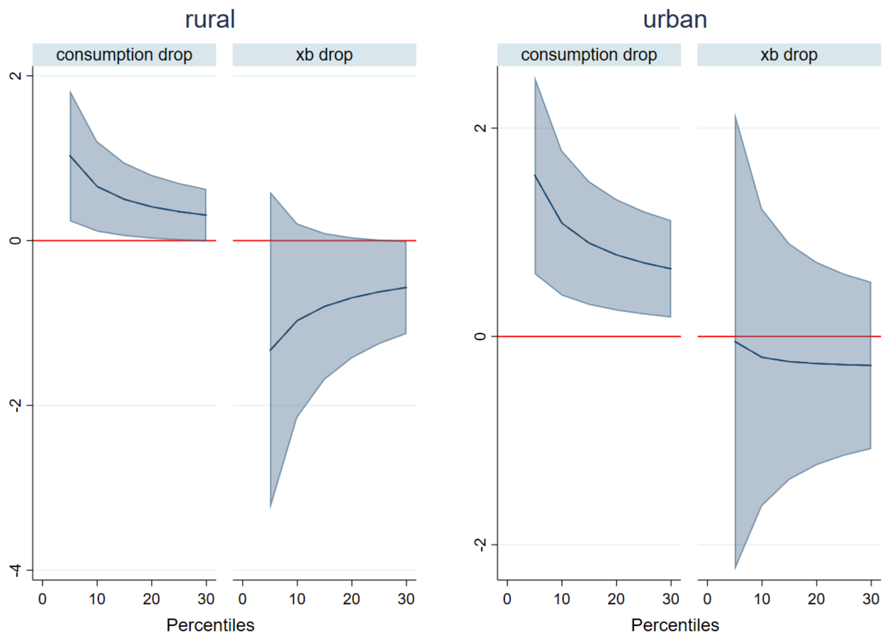


Figure 3: Regression coefficient from regression of $\hat{b}_a$ on $\hat{q}_a(\tau)$ for different choices of τ (selected percentiles evaluated for urban and rural separately). Actual household consumption per capita from the NSS-2011 is used.

ulation study in Section 3.2 is employed to infer whether one of these two scenarios is more plausible than the other. Where the simulation study from Section 3.2 considers different ways of excluding households from the NSS sample and different ways of simulating household consumption data, this empirical application works with the sample of households from the CPHS. It is presumed that selected poor households have already been excluded from this sample.

The diagnostic regression results are presented in Figure 4. We fail to obtain a significantly positive regression coefficient in any of the specifications and years. In other words, we fail to obtain evidence in favor of households having been excluded on the basis of their consumption levels. While a null result does not provide conclusive evidence, these results are consistent with a scenario where households may have been under-sampled on this basis of non-consumption attributes (such as remoteness).

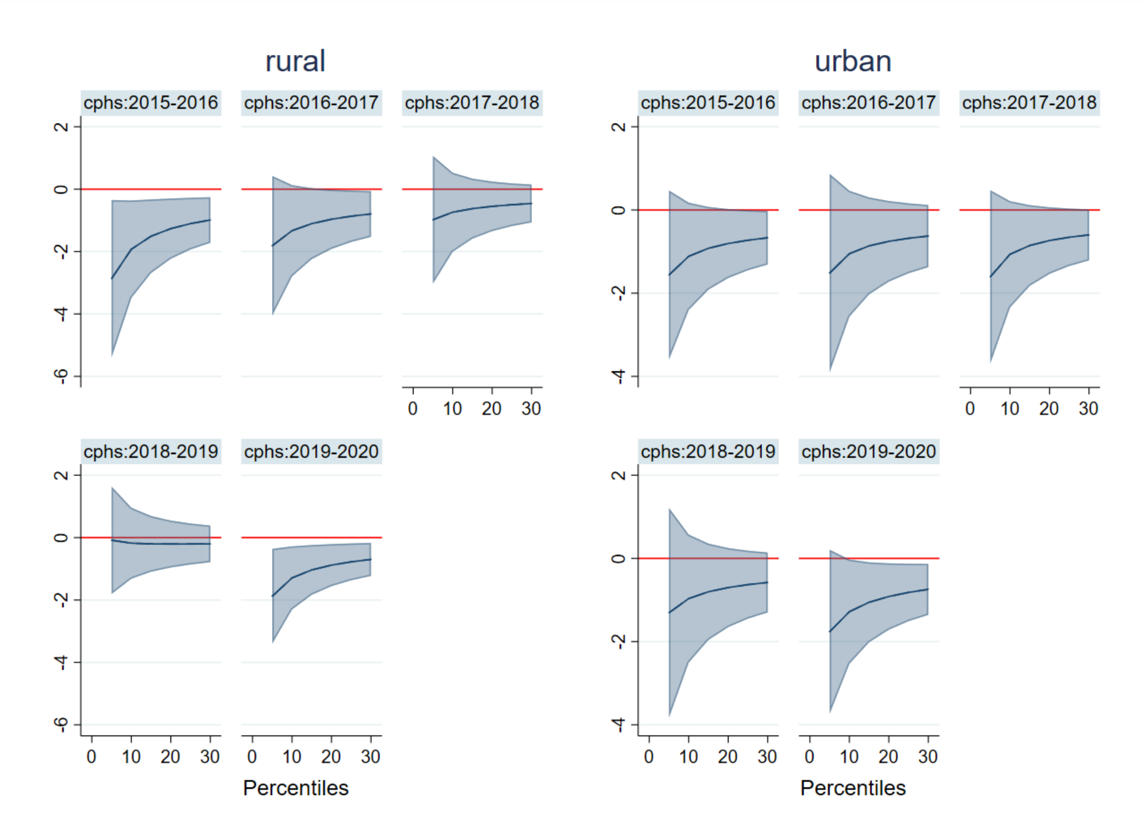


Figure 4: Regression coefficient from regression of $\hat{b}_a$ on $\hat{q}_a(\tau)$ for different choices of τ (selected percentiles evaluated for urban and rural separately). Actual household consumption per capita from the CPHS is used.

3.3.2 Does the CPHS indeed have a "main street" bias?

Based on the 76$^{\text{th}}$ round of NSS, the share of India's population in 2018 with direct access to motorable roads with and without street lights is estimated at 31.8% and 23.4%. Direct access to *non*-motorable roads with and without street lights is estimated at 7.8% and 26.1%, respectively. An additional 10.6% of the population did not have access to roads, lanes, or constructed paths in 2018.

Are households with direct access to motorable roads overrepresented in the CPHS as conjectured in DS-2021? Since the CPHS does not collect information on direct access to the main street, we impute this information into the CPHS. The imputation model uses predictors of main street access that are shared between NSS' 76$^{\text{th}}$ round survey and the CPHS. The model is trained on data from the NSS' 76$^{\text{th}}$ round and applied to obtain multiple imputed values of main street access in the CPHS using Stata's MI command.[6] The model, presented in Table 1, shows that the probability of direct access to a motorable road with street lights is higher among smaller-sized families and increases with the share of older members above 61 years of age and more educated members. The probability of main street access is also higher among Jain and Buddhist households and lower for households belonging to Schedule Tribes.

To verify the validity of the model, we also use it to impute main street access in NSS' 2011 consumption expenditure survey and the three rounds of the PLFS between 2017 and 2019. Since the PLFS' 2018 and NSS' 76$^{\text{th}}$ round surveys were conducted almost contemporaneously, and both surveys are nationally representative, comparing the imputed values in the PLFS' 2018 to the observed road access shares from the NSS' 76$^{\text{th}}$ round provides an opportunity to assess the out-of-sample performance of the model. Furthermore, we expect household access to motorable roads with street lights to improve over time. In other words, main street access imputed into NSS' 2011 survey is expected to be lower than the values observed in the more recent NSS' 76$^{\text{th}}$ round, the CPHS, and PLFS surveys.

Figure 5 summarizes the results from this exercise. First, consistent with expectations, the model is seen to reproduce an increasing trend in access to motorable roads and street lights over time. Second, road access shares observed in the 76$^{\text{th}}$ round are close to the

[6]We estimate five imputation models, one for each of the main street accessibility categories in the 76$^{\text{th}}$ round of NSS. For each road category, 50 imputed values are produced based on 50 draws of the model parameters from their asymptotic distribution and 50 draws of the model errors (assuming normally-distributed error terms). The point estimates reported in Figure 5 are obtained by averaging across these 50 repetitions.

estimates based on imputed values in the PLFS-2018. Both of these observations confirm the validity of the imputation model. Third, estimates of direct main street access based on imputed values in the CPHS are found to be significantly higher than the values observed in the 76$^{\text{th}}$ round and the contemporaneous rounds of the PLFS. Based on the CPHS-2018, direct access to motorable roads with light is estimated at 36.2% of the population. According to the NSS 76$^{\text{th}}$ round and the PLFS, these population shares are 4.4 and 3.6 percentage points lower. This is not a trivial difference. Despite an annualized per capita growth of 5.7% between 2011 and 2018, the share of the population with street lights is estimated to have risen only by 3.1%. Fourth, our estimates suggest that the under-sampling of households in the CPHS may have more to do with the presence of street lights than with motorable features of the road. Access shares in the CPHS for roads without street lights are lower for all three road types when compared to the values in the PLFS surveys conducted in the same year. This finding may be relevant for the implementation of future rounds of the CPHS. Fifth, and finally, the results suggest that the use of adjusted CPHS weights as advocated in SR goes a long way in closing the gap in main street access. The gap in the share of motorable roads with street lights between the CPHS and PLFS is reduced by about 90%, 80% and 60% for the years between 2017 and 2019.

4 Reweighting the survey: A simulation study

SR adopt the max-entropy approach advocated by Jaynes (1957) to systematically reweigh the CPHS and restore representativeness into the survey. This reweighting procedure consists of two steps. First, information on households assets, education and demographic characteristics obtained from the National Family and Health Survey (NFHS) of 2015 is used to reweigh all rounds of CPHS conducted between 2015 to 2019[7]. In the second step, we use demographic, education and labor market information from PLFS rounds of 2017, 2018 and 2019 to adjust the sampling weights in contemporaneous rounds of CPHS[8].

[7]The following variables are used to reweigh in the first step: dummy variables for ownership of air conditioners, cars, computers, refrigerators, television sets, two-wheelers, washing machine; dummies for household sizes 1 and 2, sizes 3 and 5; dummy variables for hindu, muslim, scheduled caste, schedule tribe, other backward classes households; total number of members less than 10 years old, over 60 years old; and, total members with below primary level of education, primary level and secondary level of education.

[8]The following variables are used in the second step: dummy variables for female-headed household; scheduled caste, scheduled tribe and other backward classes households; dummy variables for household sizes 1 to 5; total members working in casual, salaried and self-employed jobs; total number of members

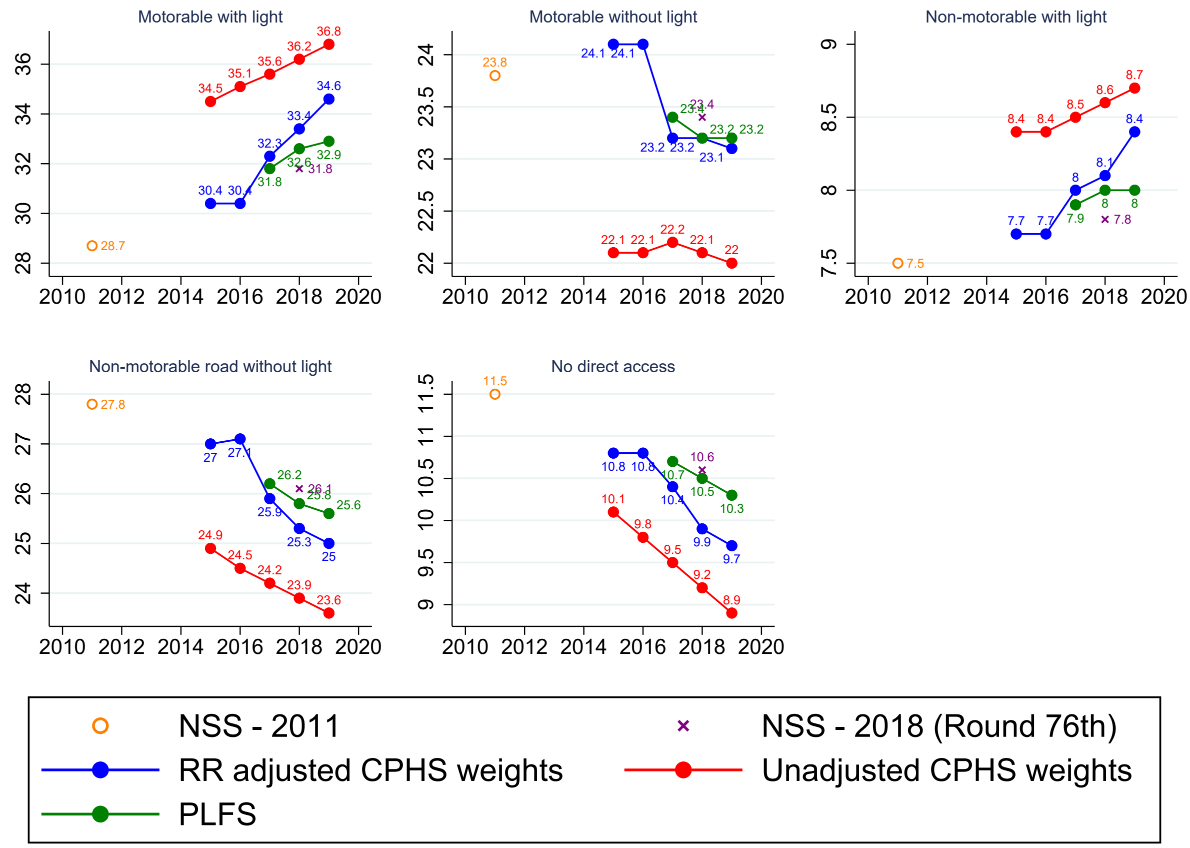


Figure 5: Observed and Predicted values of main street access from NSS' 76$^{\text{th}}$ round, NSS' 2011 consumption expenditure survey, PLFS and the CPHS. Notes: The NSS' 76$^{\text{th}}$ round estimate is based on actual survey data while the remaining estimates are derived using Multiple Imputations on a model trained on NSS' 76$^{\text{th}}$ survey data. RR adjusted CPHS weights reflect the sampling weights used in SR. Unadjusted CPHS weights are the raw household-level weights provided by CMIE divided by household size.

The second step of the reweighting procedure accounts for changes in socio-economic indicators over time.

In selecting the target variables on which to reweigh, we prioritize indicators where the gap between CPHS and nationally representative surveys are most egregious. An example of such a target variable is the share of undereducated adults (comprising of illiterate and below primary levels of education). We reserve a set of indicators that are common to CPHS and NFHS or PLFS to conduct extensive validation of the reweighted results.

The maxentropy reweighting procedure minimizes the distances between the weighted

less than 10 years old, over 60 years old; and, total members with below primary level of education, primary level of education and secondary level of education.

means of target variables obtained from PLFS or NFHS and the CPHS. As earlier noted, we execute the reweighting procedure separately for rural and urban samples of each state. Following existing practices (e.g. Chen et al., 2018; Haziza and Beaumont, 2017; Kolenikov, 2014), the adjusted individual level weights obtained from maxentropy are winsorized at the 0.25^{th} and 99.75^{th} percentile level. To achieve national level representation, we multiply the resulting normalized weights obtained from the maxentropy procedure by the rural and urban population of each state. The population estimates are obtained from the NFHS-2015 for 2015 and 2016 rounds; and from the PLFS 2017 to 2019 for the remaining year. Finally, household-level weights are constructed by dividing the adjusted individual-level weights by the household size.

4.1 When households are excluded based on consumption

4.1.1 Design of simulation experiment

DS-2023's simulation experiment uses the PLFS as a point of departure. It then drops households from this survey to simulate a sample that is not representative of the full population by design. Specifically, the experiment considers scenarios that create biased samples by excluding households based on their household consumption per capita levels (using a variety of criteria). The exclusions make the simulated survey biased against the poor in an effort to mimic the biases that are assumed to underlie the CPHS. Finally, the authors adopt the re-weighting approach put forward in SR to evaluate whether the adjusted sampling weights can reproduce the estimates of poverty and mean consumption observed before the exclusion of poor households.

DS-2023 consider four scenarios as part of their study (not to be confused with the two scenarios described in the preceding sections): Scenario 1 randomly drops half of all households whose consumption is below the 40th percentile of the distribution. Scenario 2 varies the consumption cut-off and the number of observations excluded (70% of households from the lowest 10% of the consumption distribution are dropped, 50% from the bottom 20%, and 30% of households from the bottom 30% of the distribution). In Scenario 3, DS-2023 exclude all households with less than 10% of the consumption distribution. Scenario 4 considers the case where households are excluded based on criteria other than their consumption expenditure levels. Specifically, after excluding half of all households in the bottom 40% of the consumption distribution, an additional 20% of Muslim, Scheduled Caste (SC), and Scheduled Tribe (ST) are excluded from the sample as well as 20% of households with casual labor as the primary source of income and 30% of households below the lowest quartile of education.[9] As less accessible households, casual workers, and minority households tend to be poorer, excluding these households from the CPHS could similarly bias the survey against the poor.

Reweighting is found to perform poorly in these cases. The gaps between true and reweighted mean consumption and poverty headcount levels are reported to vary between 5.3% and 7.7% and between 13% and 30% respectively. DS-2023 also consider a variation on Scenario 4 where the first step, excluding households based on their consumption per capita level, is skipped, in which case households are excluded based only on religion, caste, and other non-consumption characteristics. Reweighting is found to resolve a large share of the resulting bias in this case. This suggests that when households are excluded from the CPHS based on strong correlates of consumption, rather than values of consumption observed prior to survey implementation, the reweighting approach is better equipped to restore representativeness.

4.1.2 Results

Figures 6 and 7 present our results. We are able to reproduce the findings from DS-2023 using the 76$^{\text{th}}$ round of the NSS instead of the PLFS. Using the full nationally representative survey we obtain a poverty headcount rate of 20.5% and mean per capita expenditure (mpce) level of Rs. 2500.2. When 50% of households from the bottom 40% of the distribution are excluded randomly (Scenario 1 from DS-2023), the biased sample returns a headcount poverty rate (mpce value) of 13% (Rs. 2752.7) with raw uncorrected weights. Similar gaps are observed for the other three scenarios. Re-weighting produces marginally better results – more for average consumption than the headcount poverty estimate – which mirrors DS-2023's findings. The smallest gain in terms of bias reduction is observed in the case of Scenario 2 (also consistent with DS-2023's findings).

4.2 When households are excluded based on remoteness

4.2.1 Design of simulation experiment

We consider a variation on DS-2023's experiment that aims to simulate the "main street" bias described in DS-2021. We use the 76$^{\text{th}}$ round of NSS surveys, namely Schedule 1.2

[9] We do not simulate this step to make the analysis simple and tractable.

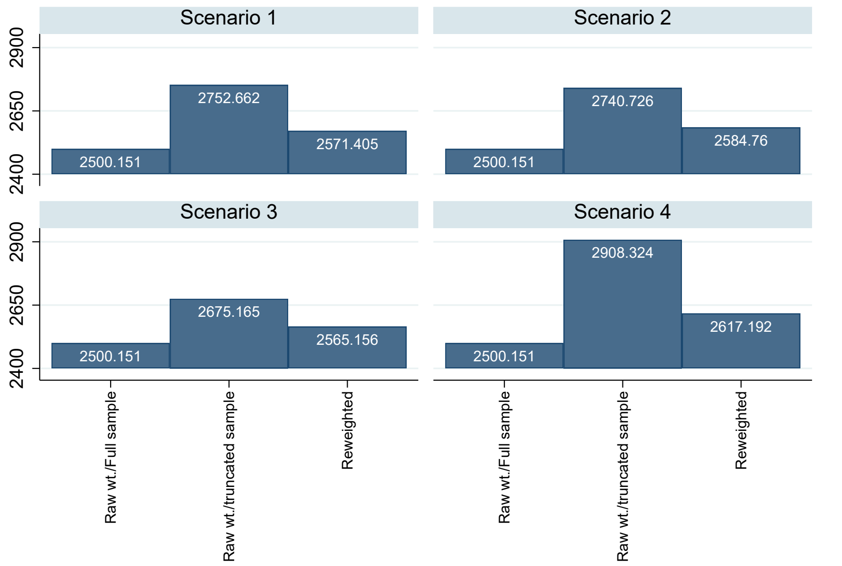


Figure 6: Mean Consumption Per Capita; Consumption based simulations
Notes: "Raw wt./Full sample" reflect unbiased estimates from NSS' 76th round; "Raw wt./Truncated sample" denote estimates based on simulated sample using unadjusted weight; and, "Reweighted" show estimates based on max-entropy adjusted sampling weight.

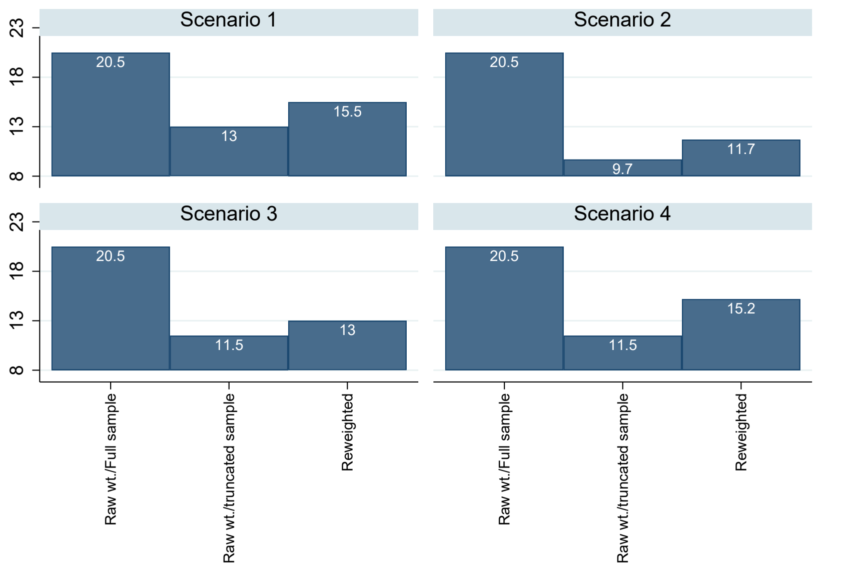


Figure 7: Poverty Headcount Estimates; Consumption based simulations.
Notes: "Raw wt./Full sample" reflect unbiased estimates from NSS' 76th round; "Raw wt./Truncated sample" denote estimates based on simulated sample using unadjusted weight; and, "Reweighted" show estimates based on max-entropy adjusted sampling weight.

on "Drinking Water, Sanitation, Hygiene, and Housing Conditions". The survey was conducted between July 2018 and December 2018 across India and contains a sample size of 106,838 households. We restrict the sample to the 27 states that are covered by the CPHS. The household consumption expenditure variable included in the survey is similar to the consumption variable included in the PLFS that is used in DS's analysis. Important for our analysis, the survey also includes a question on direct household access to a road, lane, or constructed path with five categorical responses.

We use these data to set up a simulation similar to DS-2023. Specifically, we modify DS-2023's four scenarios as follows:

- Scenario 1: Randomly drop 70% of households that have either direct access to a *non-motorable road without streelight* OR have *no direct opening to road,lane or constructed path*
- Scenario 2: Randomly drop 90% of households with *no direct opening to road, lane or constructed path*, 60% with access to *non-motorable road without streetlight* and 50% that are accessible by a *non-motorable road with streetlight*
- Scenario 3: Drop all households that have either accessible only by a *non-motorable road without streelight* OR have *no direct opening to road, lane or constructed path*
- Scenario 4: First, randomly drop 90% of households that accessible only by a *non-motorable road without streelight* OR have *no direct opening to road,lane or constructed path.* Of the remaining households, randomly drop 50% of Muslims, Scheduled Caste (SC), and Scheduled Tribe (ST) households.

Following DS-2023, we draw 100 random samples for Scenarios 1, 2, and 4, resulting in 301 simulations in total. Next, we examine whether re-weighting can successfully restore representativeness in samples that are subjected to main street bias. The sampling weights are adjusted in two steps. First, we restore the population distribution at the state-sector level by multiplying the sampling weights by a scalar such that the sum of the re-scaled individual-level weights at the state-sector level matches that of the nationally representative survey (before selected households are dropped).[10] Following SR, the second step of readjustment applies the maxentropy reweighting approach that

[10]Without this step, certain state-sectors may be overrepresented in the survey, which is not the case in the CPHS.

is critically evaluated in DS-2023. We use the following target variables for re-weighting: (log) age of the household head, the share of members with middle to high school level of education, and dummy variables indicating whether the household has any member with diploma to post-graduate level of education, is female-headed, has an extended family and lives in a rented unit [11].

4.2.2 Results

Survey samples with "main street" bias (obtained by randomly excluding households with main street access in NSS) exhibit higher average household consumption per capita and lower headcount poverty rates relative to the nationally representative sample (see Figures 8-9). This stems from the fact that access to main street is strongly correlated with household income and consumption levels. The bias in mean consumption is most pronounced in the case of Scenarios 3 and 4. Re-weighting in this case however, is seen to come a long way in resolving the biases. The gaps are reduced to 0.6%, 0.9%, 1%, and 1.7% in mpce across the four scenarios (see Figure 8).

Re-weighting is found to be equally effective in resolving the biases observed in the poverty headcount estimates (see Figure 9). The remaining gaps in headcount are less than 1 percentage point (on a base of 20.5%). In case of Scenario 4, where the bias in the headcount rate equals 6 percentage points, the gap is reduced to 1.5 percentage points.

5 Concluding remarks

Dreze and Somanchi, among others, have expressed reservations about the representativeness of CPHS samples. The CPHS is conjectured to over-sample households in more accessible locations and thereby undercover the poor. We conduct two simulation experiments to evaluate whether reweighting can restore representativeness in surveys that are biased against the poor (i.e., under-sampled poor households). In the first experiment, we randomly drop households from a representative survey based on their consumption expenditure levels, where households with lower consumption levels are dropped with higher probabilities. In the second experiment, we randomly drop households based on how accessible they are, where households without access to the main road and/or a road

[11]We implement this reweighting strategy at the national level to fully replicate DS-2023's analysis. In SR, our reweighting algorithm was undertaken at the state-sector level which is likely to close existing gaps further.

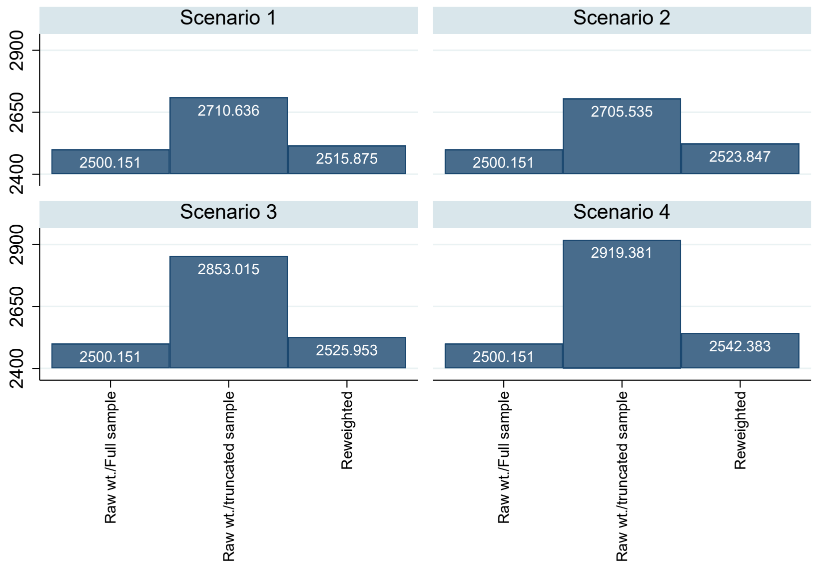


Figure 8: Mean Consumption Per Capita; Simulations based on Main-Street. Notes: "Raw wt./Full sample" reflects unbiased estimate from NSS' 76$^{\text{th}}$ round; "Raw wt./Truncated sample" denotes estimate based on simulated sample using unadjusted weight; and, "Reweighted" shows estimate based on max-entropy adjusted sampling weight.

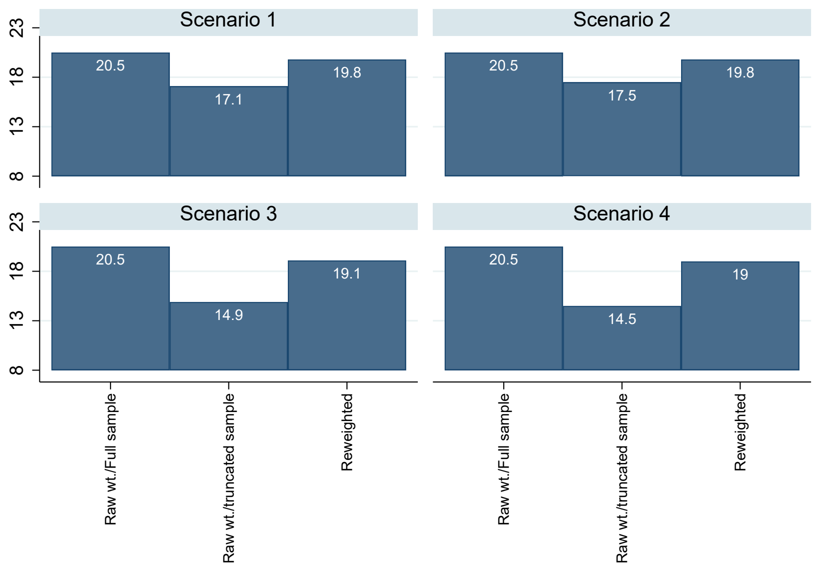


Figure 9: Poverty Headcount Estimate; Simulations based on Main-Street. Notes: "Raw wt./Full sample" reflect unbiased estimates from NSS' 76$^{\text{th}}$ round; "Raw wt./Truncated sample" denote estimates based on simulated sample using unadjusted weight; and, "Reweighted" show estimates based on max-entropy adjusted sampling weight.

with street lights are dropped with higher probabilities. As access to the main road is strongly correlated with household income and consumption levels, both experiments end up under-sampling poor households (biasing the survey sample against the poor).

Our findings indicate that reweighting procedures are ill-equipped to resolve biases in consumption expenditure and poverty statistics in cases where households have been excluded from the survey based on their household consumption values (consistent with Dreze and Somanchi, 2023). By contrast, reweighting is found to be highly effective in closing the gaps in poverty and mean consumption expenditure statistics when surveys exclude households based on characteristics such as the household's access to the main road. The observed biases are reduced by about 80%, i.e., a bias of 5 percentage points in the poverty headcount rate is reduced to a bias of less than 1 percentage point.

We argue that "main street" bias, the exclusion of households based on access to the main street (along with other characteristics that are strongly correlated with household consumption such as caste, occupation, etc.), is the more plausible reason for the observed biases in the CPHS, as originally conjectured in Dreze and Somanch (2021). Survey firms such as the CMIE do not observe household consumption before implementing the surveys. This conjecture finds support in the data. After imputing road access variables into the CPHS, we estimate that access to the main road is significantly higher in the CPHS than it is in nationally representative benchmark surveys such as the NSS' 76$^{\text{th}}$ round and the PLFS. Furthermore, a diagnostic regression analysis developed to detect whether households have been dropped based on their consumption expenditure levels, fails to detect evidence in favor of exclusions based on consumption levels when applied to the CPHS data.

	(1) Motor+light	(2) Motor—light	(3) Non-motor+light	(4) Non-motor—light	(5) No road
Household size=1 or 2	0.140***	-0.0525***	0.0133*	-0.0712***	-0.0297***
	(16.01)	(-6.51)	(2.49)	(-8.88)	(-5.37)
Household size=3 to 5	0.0776***	-0.0315***	0.0167***	-0.0421***	-0.0207***
	(15.08)	(-6.12)	(5.33)	(-8.12)	(-5.76)
Multigeneration family	0.0405***	0.00234	0.00585	-0.0272***	-0.0215***
	(8.10)	(0.48)	(1.89)	(-5.53)	(-6.25)
Extended family	0.0423***	-0.0265***	0.0145*	-0.0259***	-0.00429
	(4.84)	(-3.48)	(2.40)	(-3.41)	(-0.79)
Sh of mem: 0 to 18 year	-0.138***	0.0503***	-0.0340***	0.0786***	0.0435***
	(-15.44)	(5.58)	(-6.38)	(8.62)	(6.87)
Sh of mem: 61+ year	0.0853***	-0.0242*	-0.00608	-0.0557***	0.000718
	(7.36)	(-2.30)	(-0.87)	(-5.27)	(0.10)
If any member over middle school	0.0439***	-0.0155*	-0.00407	-0.0308***	0.00648

	(1) Motor+light	(2) Motor—light	(3) Non-motor+light	(4) Non-motor—light	(5) No road
	(5.87)	(-1.96)	(-0.85)	(-3.73)	(1.04)
Sh of members over middle school	0.200***	-0.00465	0.00966	-0.125***	-0.0803***
	(22.00)	(-0.52)	(1.65)	(-13.96)	(-12.96)
If any member over diploma	-0.134***	0.0401***	-0.00300	0.0761***	0.0211***
	(-23.65)	(7.83)	(-0.86)	(16.18)	(7.02)
Muslim	-0.00258	-0.00617	-0.00160	0.0172**	-0.00689
	(-0.45)	(-1.13)	(-0.46)	(3.04)	(-1.79)
Christian	0.129***	-0.0814***	-0.00968	-0.0259*	-0.0125
	(10.17)	(-9.20)	(-1.39)	(-2.52)	(-1.43)
Sikh	-0.0845***	0.216***	-0.0461***	-0.0150	-0.0707***
	(-5.75)	(12.05)	(-7.49)	(-1.00)	(-13.45)
Jain	0.242***	-0.110***	-0.0147	-0.0602*	-0.0574***
	(7.62)	(-5.02)	(-0.70)	(-2.54)	(-6.54)
Buddhist	0.153***	-0.0145	-0.0198	-0.110***	-0.00916
	(5.82)	(-0.62)	(-1.41)	(-5.76)	(-0.54)
Other religions	-0.0451**	0.00804	-0.00626	0.0177	0.0256
	(-2.73)	(0.48)	(-0.61)	(0.98)	(1.87)
Scheduled Castes	0.0841***	-0.0348***	0.0331***	0.00239	-0.0848***
	(13.24)	(-4.68)	(8.72)	(0.32)	(-13.39)
Other Backward Classes	0.126***	-0.0257***	0.0316***	-0.0287***	-0.103***
	(21.78)	(-3.74)	(9.20)	(-4.14)	(-17.43)
Other Castes	0.150***	-0.0648***	0.0367***	-0.0229**	-0.0991***
	(22.49)	(-8.94)	(9.01)	(-3.13)	(-16.36)
Constant	0.156***	0.250***	0.0471***	0.314***	0.233***
	(13.72)	(21.82)	(6.56)	(27.01)	(26.44)
Observations	97115	97115	97115	97115	97115
Adjusted R^2	0.096	0.014	0.005	0.031	0.031

Table 1: Estimating household probability of road access
Notes: t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Source: NSS' 76$^{\text{th}}$ round survey

References

Afridi, Farzana, Arora, Abhishek, Dhar, Diva and Mahajan, Kanika (2023). Women's work, social norms and the marriage market. Technical Report 15948. IZA Discussion Working Paper Series.

Agarwal, Bina (2021). Livelihoods in covid times: Gendered perils and new pathways in india. *World Development*, **139,** number 105312.

Beyer, Robert CM, Franco-Bedoya, Sebastian and Galdo, Virgilio (2021). Examining the economic impact of covid-19 in india through daily electricity consumption and nighttime light intensity. *World Development*, **140,** number 105287.

Bhattacharya, Shrayana and Roy, Sutirtha Sinha (2021). Intent to implementation: Tracking india's social protection response to covid-19. *World Bank*.

Chanda, Areendam and Cook, C. Justin (2020). Was india's demonetization redistributive? insights from satellites and surveys. *Journal of Macroeconomics*, **73,** 103438.

Chen, Shaohua, Jolliffe, Dean Mitchell, Lakner, Christoph, Lee, Kihoon, Mahler, Daniel Gerszon, Mungai, Rose, Nguyen, Minh Cong, Prydz, Espen Beer, Sangraula, Prem, Sharma, Dhiraj, Yang, Judy and Zhao, Qinghua (2018). Povcalnet update: What's new. Global Poverty Monitoring Technical Note 2. The World Bank.

Chodorow-Reich, Gabriel, Gopinath, Gita, Mishra, Prachi and Narayanan, Abhinav (2020). Cash and the economy: Evidence from india's demonetization. *Quarterly Journal of Economics*, **135,** number 1, 57–103.

Deshpande, Ashwini (2020). The covid-19 pandemic and gendered division of paid and unpaid work: Evidence from india. Discussion Paper No. 13815. IZA.

Dreze, Jean and Somanchi, Anmol (2021). View: New barometer of india's economy fails to reflect deprivations of poor households. *The Economic Times*.

Dreze, Jean and Somanchi, Anmol (2023). Weighty evidence? poverty estimation with missing data. *Ideas for India*.

Ghatak, Maitreesh, Kotwal, Ashok and Ramaswami, Bharat (2020). What would make india's growth sustainable. Blog. The India Forum.

Gupta, Arpit, Malani, Anup and Woda, Bartek (2021a). Explaining the income and consumption effects of covid in india. NBER Working Papers w28935. National Bureau of Economic Research.

Gupta, Arpit, Malani, Anup and Woda, Bartek (2021b). Inequality in india declined during covid. NBER Working Papers w29597. National Bureau of Economic Research.

Haziza, David and Beaumont, Jean-François (2017). Construction of weights in surveys: A review. *Statistical Science*, **32.2,** 206–226.

Jaynes, E.T. (1957). Information theory and statistical mechanics. *Physical review*, **106.4,** number 620.

Kolenikov, Stanislav (2014). Calibrating survey data using iterative proportional fitting (raking). *The Stata Journal*, **14,** 22–59.

Pais, Jesim and Rawal, Vikas (2021). Cmie's consumer pyramids household surveys: An assessment. Blog September 3. The India Forum.

Somanchi, Anmol (2021). Missing the poor, big time: A critical assessment of the consumer pyramids household survey. Web 11 Aug. SocArXiv.

Srija, A. and Singh, Jitendra (2021). View: It's time for cmie to rethink how they determine labour market data. *The Economic Times*.