

Poverty in India has declined over the last decade

Estimates from a new India consumption survey: The CPHS

May, 2022

Sutirtha Sinha Roy (Poverty and Equity) and Roy Van der Weide (DECRG)



Motivation

- The NSS-2011 is the latest household consumption survey released by India that underlies official estimates of poverty and inequality
 - Estimates of mean consumption per capita derived from leaked 2017 survey cannot be corroborated by estimates from NAS and other survey data
 - Extrapolation methods that combine NSS-2011 with national accounts data are increasingly problematic as the latest NSS becomes increasingly outdated
- In 2014/15, the CPHS came into existence, a household survey collected by the private sector.
- Our objectives are two-fold:
 - Provide a comprehensive assessment of the CPHS
 - Provide estimates of trends in poverty and inequality post-2011 using the CPHS

Background

- Poverty estimates for years with missing surveys ...
 - ... are extrapolated using the distribution of consumption from the nearest survey year ...
 - ... and applying the growth rates from national accounts
- Referred to as “**lining-up**” of estimates for missing survey years
- Caveats:
 - **Assumes distribution-neutral growth** → inequality is considered constant across years
 - **Pass-through rate**: Proportion of the growth in national accounts that are passed through to the growth in consumption observed in surveys is heuristically determined

Measuring poverty and inequality with CMIE's CPHS

- Combine data from CPHS and NSS to estimate poverty rates and consumption inequality in India for 2015 to 2019
- **An overview of CPHS**
 - ❑ New national-level household survey collected by the private sector...
 - Consumption information of 115 items
 - Data on asset ownership, labor market indicators and demographics
 - Sample size roughly of 170,000 households
 - Conducted regularly 3x times a year since 2015
 - ❑ Limitations
 - Under-representation of richest and poorest households in the country
 - ✓ reweighted CPHS to match representativeness observed in nationally representative surveys: NFHS and LFS.
 - Consumption data is not directly comparable to NSS used historically to measure poverty

Discrepancies predating reweighting exercise

Demographics

Households with 2-5 members	Over Represented in CPHS
People below 18 years	Under Represented in CPHS
Hindu households	Over Represented in CPHS

Access to Services and Household Assets

Consumer durables: TV, AC, Refrigerator	Over represented in CPHS
Access: electricity, water, toilet in premises	Over represented in CPHS

Education

Low educated people (15-49)	Severely under represented in CPHS
Literacy among women (15-49)	Severely under represented in CPHS

Labor market indicators

LFP Rates, composition of casual/self-employed/salaried	Lower LFP, more casual workforce in CPHS
Wages (casual and salaried)	Higher in CPHS
Differences across incomes distribution	Are magnified at higher income deciles of CPHS

Consumption

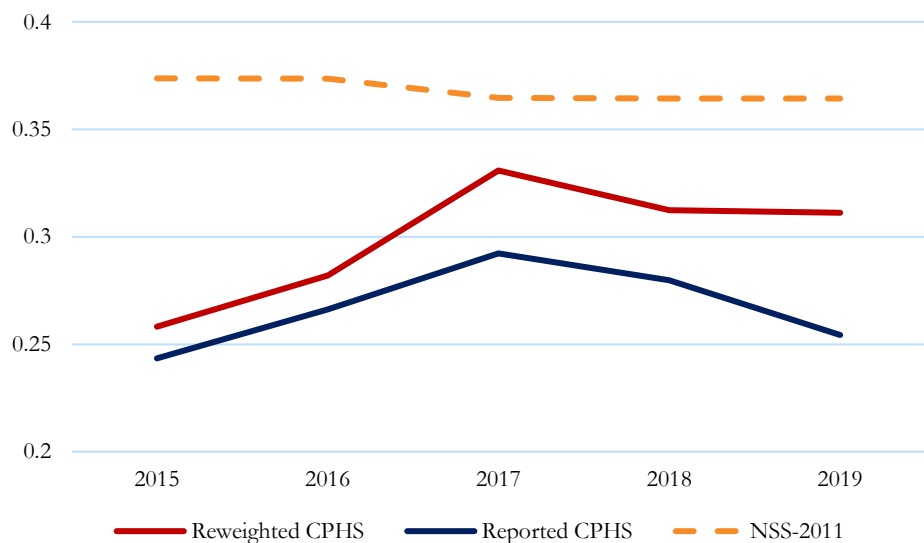
Variance in consumption relative to NSS 2011 CES	Lower in CPHS
Proximity to normality [Excess Kurtosis =0, Skewness = 0]	Higher than in CPHS than in NSS CES 2011

Representation of CPHS and Reweighting

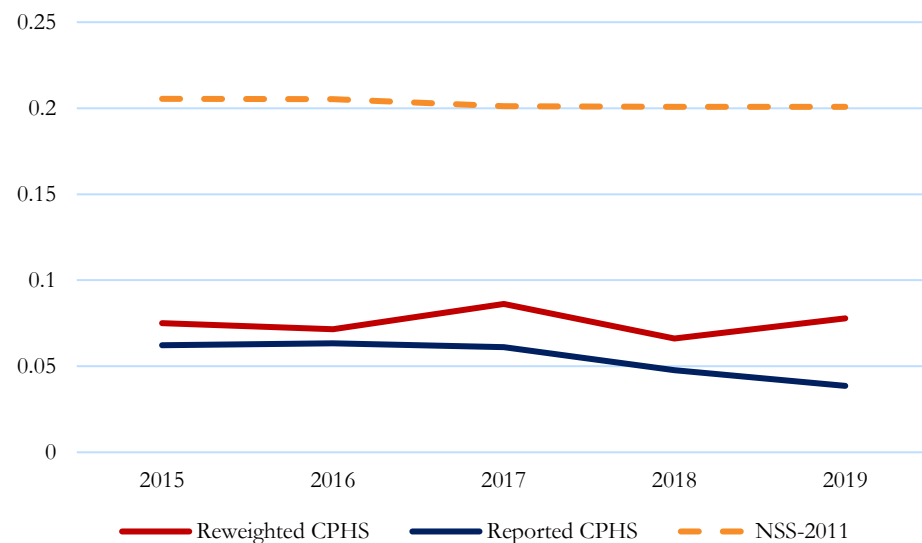
- Maxentropy on benchmark surveys*:
 - NFHS/DHS-4 (2015): assets, education, demographics
 - LFS (2017, 2018, 2019): employment, education, demographics
- Process:
 - Reweight all CPHS rounds based on NFHS-4
 - Reweight 2017-2019 CPHS rounds using contemporaneous LFS surveys
 - At state and rural/urban levels. Winsorize at 0.5th and 99.5th percentile
 - Adjust normalized and adjusted weights at state level to reflect country-level representation
- Test representativeness using :
 - Contemporaneous NSS and LFS surveys
 - NFHS-5 (2019) data released for select states
- https://datacatalog.worldbank.org/search/dataset/0061482/adjusted_sampling_weights_for_cphs_surveys_based_on_sinha_roy_and_van_der_weide_2022_wps_9994

Moments after adjusting CPHS weights

Variance of consumption



Third moment of consumption



Approach 1: Ignoring CPHS consumption data

$$\log(y_i^{NSS}) = c + \beta x_i + u_i,$$

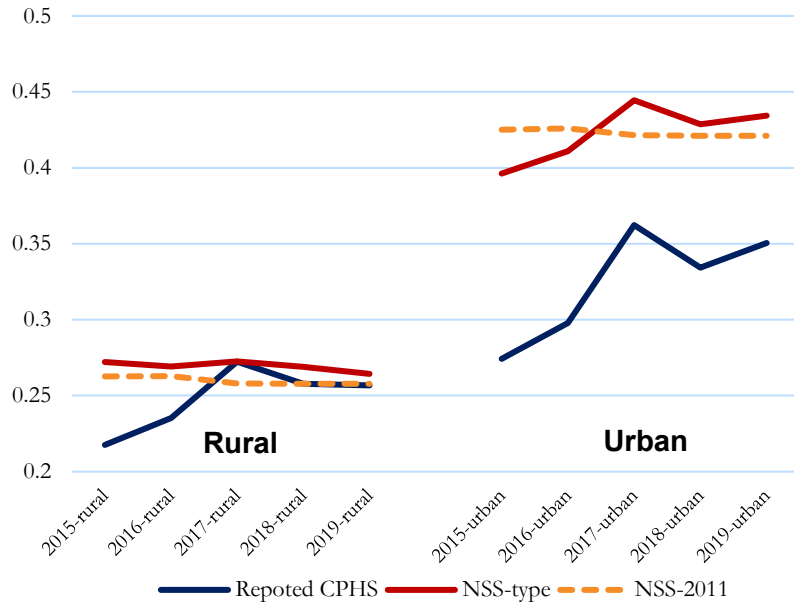
- Covariates x are shared between CPHS and NSS: demographics, education, employment, dwelling characteristics, asset ownership, industry code and dummies capturing consumption of select premium goods
- When imputing NSS-type consumption in the CPHS, the errors are drawn from the empirical distribution (preserving distribution)

Note 1: This ignores the CPHS consumption data

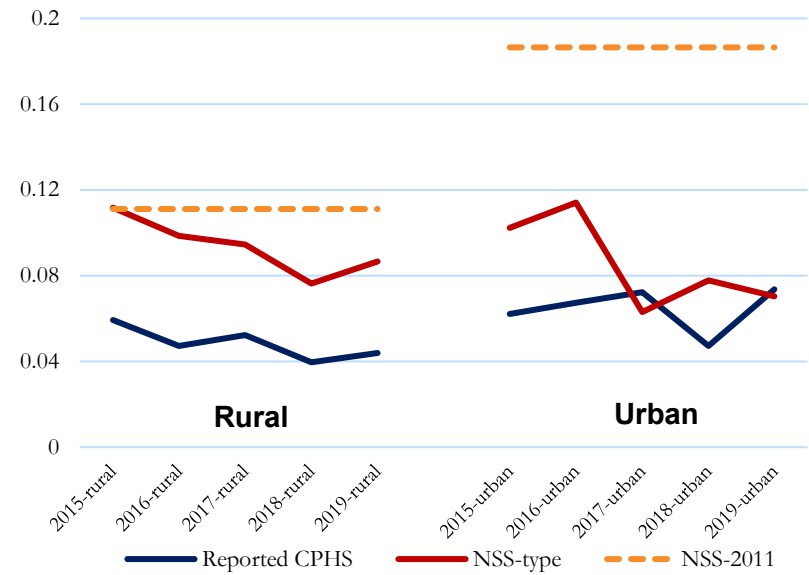
Note 2: Changes in inequality due to time-variation in covariates x

Moments after Approach 1

Variance of consumption



Third Moment of Consumption



Approach 2: Using CPHS consumption data

$$\log(y_i^{CPHS}) = a + b \log(y_i^{NSS}) + \sigma e_i,$$

- e_i is assumed to be normally distributed
- The parameters a , b , and σ are estimated using methods of moments
- Ordering of NSS (right) and CPHS (left) fits stylized facts: 2nd and 3rd moments of CPHS sit in between those of NSS and normal distribution

Note 1: This fully utilizes CPHS consumption data

Note 2: Challenge here is to work out $p(\log(y_i^{NSS}) | \log(y_i^{CPHS}))$

Approach 2 - Continued

- Estimating $p(\log(y_i^{NSS})|\log(y_i^{CPHS}))$
- We fit normal-mixture to unconditional distributions for $\log(y_i^{CPHS})$, from which normal-mixture distribution for $\log(y_i^{NSS})$ can be obtained (given estimate for σ)
- From *Lemma 2* in Elbers and van der Weide (2014) it follows that $p(\log(y_i^{NSS})|\log(y_i^{CPHS}))$ is also a normal mixture (and provides estimators for its parameters)

Approach 2 – Matching Moments

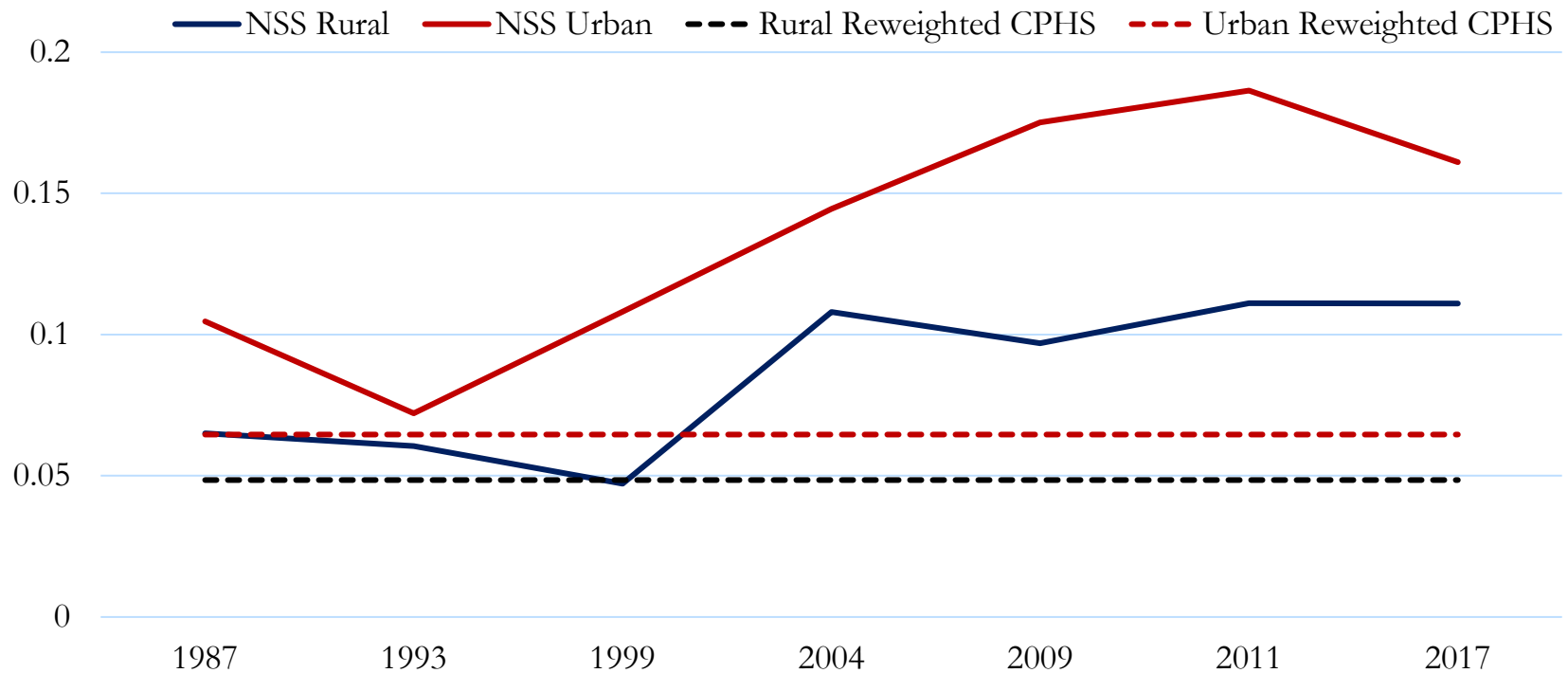
$$\log(y_i^{CPHS}) = a + b \log(y_i^{NSS}) + \sigma e_i$$

- Parameter b : relates to the skewness of the distribution $b = (3m_{cphs}/3m_{nss})^{1/3}$
- Parameter a : relates to the mean of the distribution $a = \mu_{cphs} - b * \mu_{nss}$
- Parameter $s2$: relates to the variance of the distribution $s2 = \sigma^2 = \sigma_{cphs}^2 - b^2 \sigma_{nss}^2$
- Two possible options of fixing a , b and $s2$ as follows:

	a	s2	b
Approach 2 (2011)	Fixed to mean imputed consumption from each year	Fixed to std deviation of NSS-2011	Fixed to skewness of NSS-2011
Approach 2 (2017)	Fixed to mean imputed consumption from each year	Fixed to std deviation of imputed NSS consumption from each year	Fixed to skewness of NSS-2017

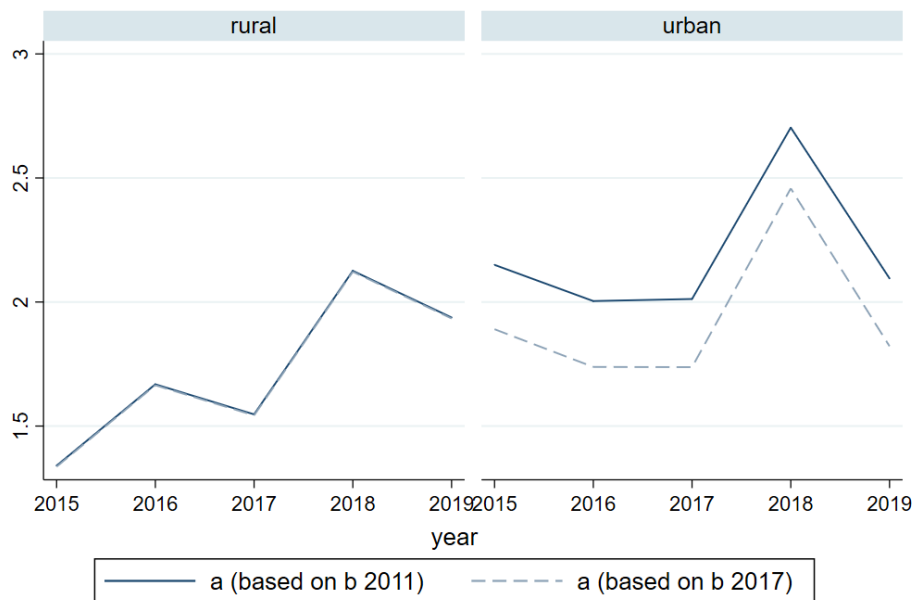
- Fixing b to either NSS-2011 or 2017 is based on the observation that skewness value of NSS surveys have been remarkably stable overtime

Comparing third moment: CPHS and NSS

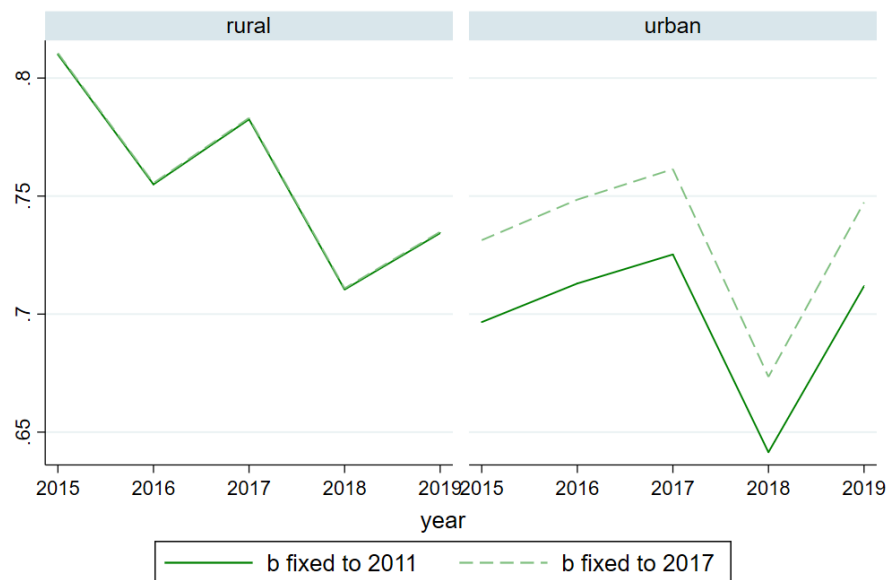


Parameters: a and b

$$a = \mu_{cphs} - b * \mu_{nss}$$



$$b = (skew_{cphs} / skew_{nss})^{1/3}$$

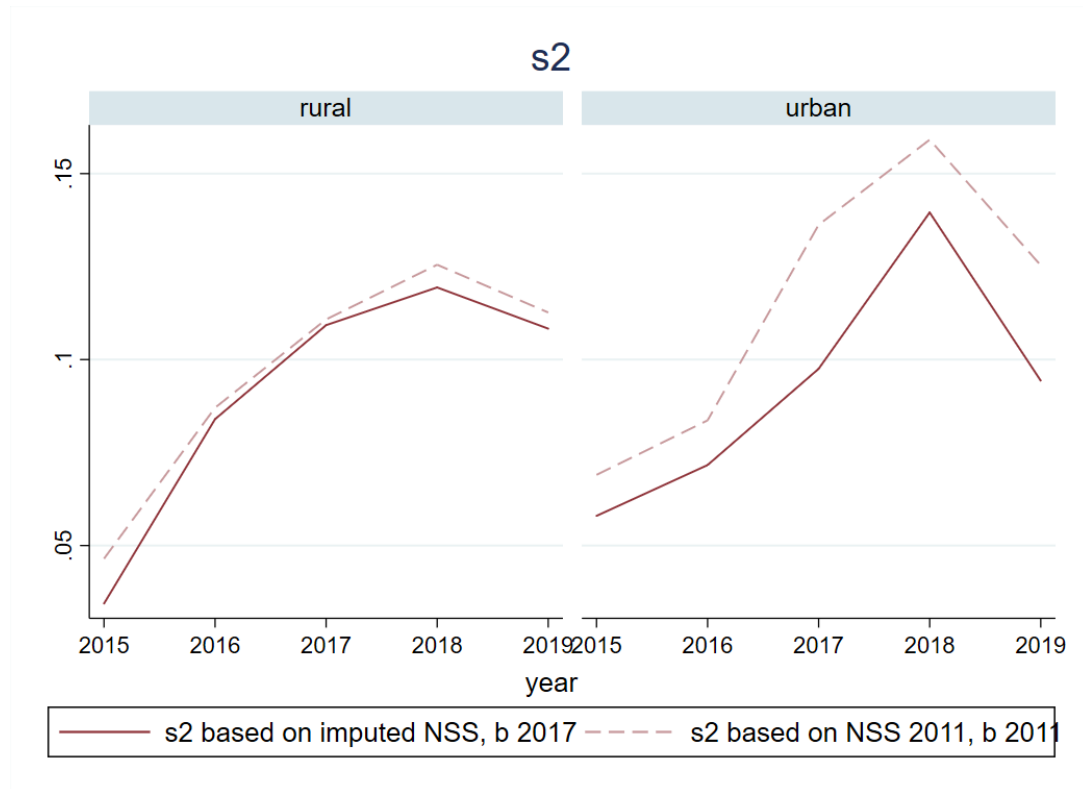


Values of Skew from two NSS rounds

year	Rural	Urban
2011	0.111	0.186
2017	0.111	0.161

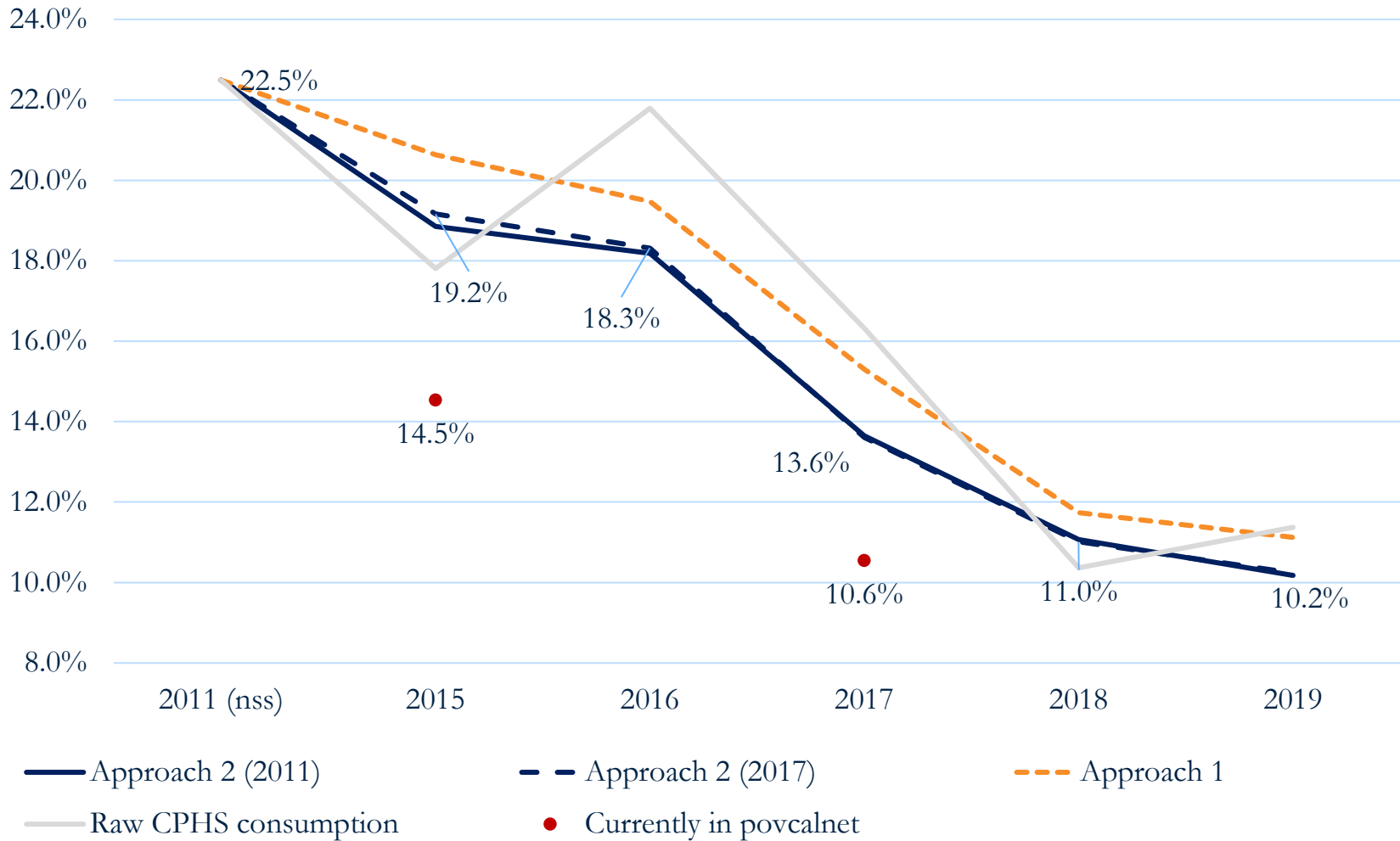
Parameter: s^2

$$s^2 = \sigma^2 = \sigma_{cphs}^2 - b^2 \sigma_{nss}^2$$



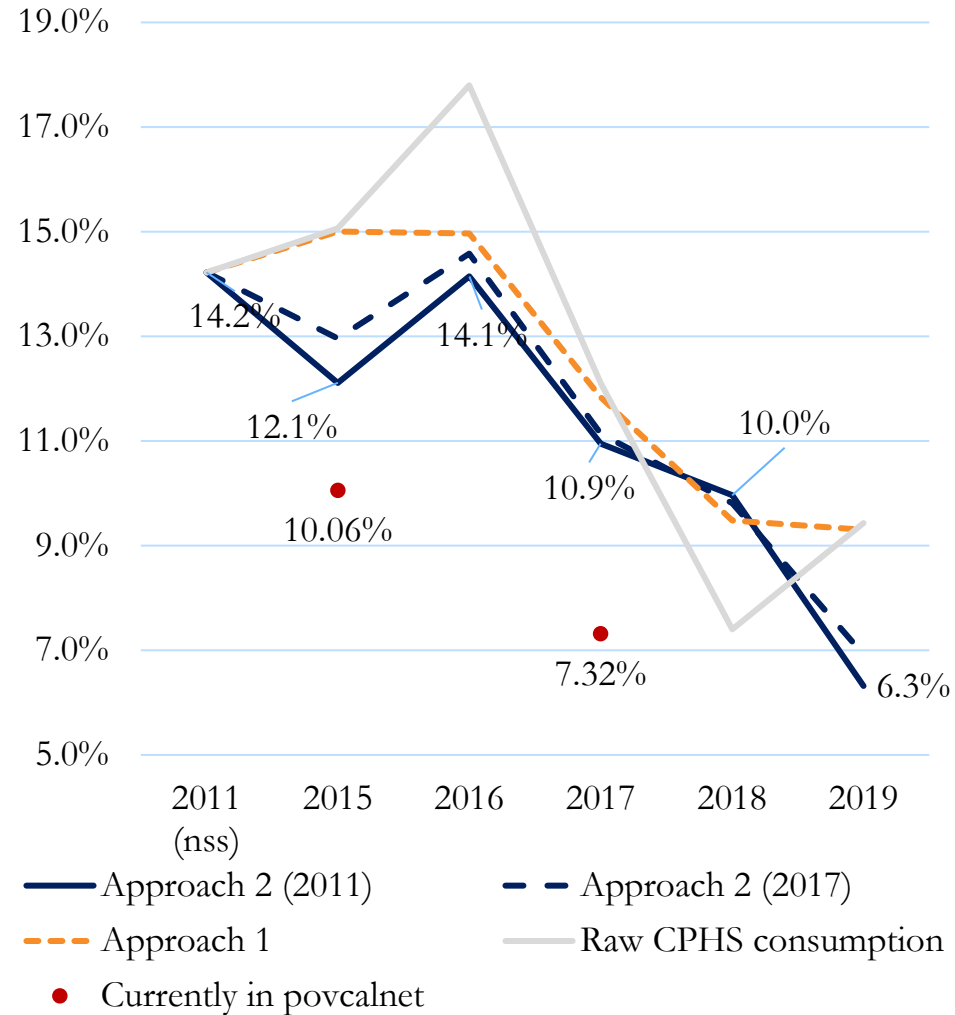
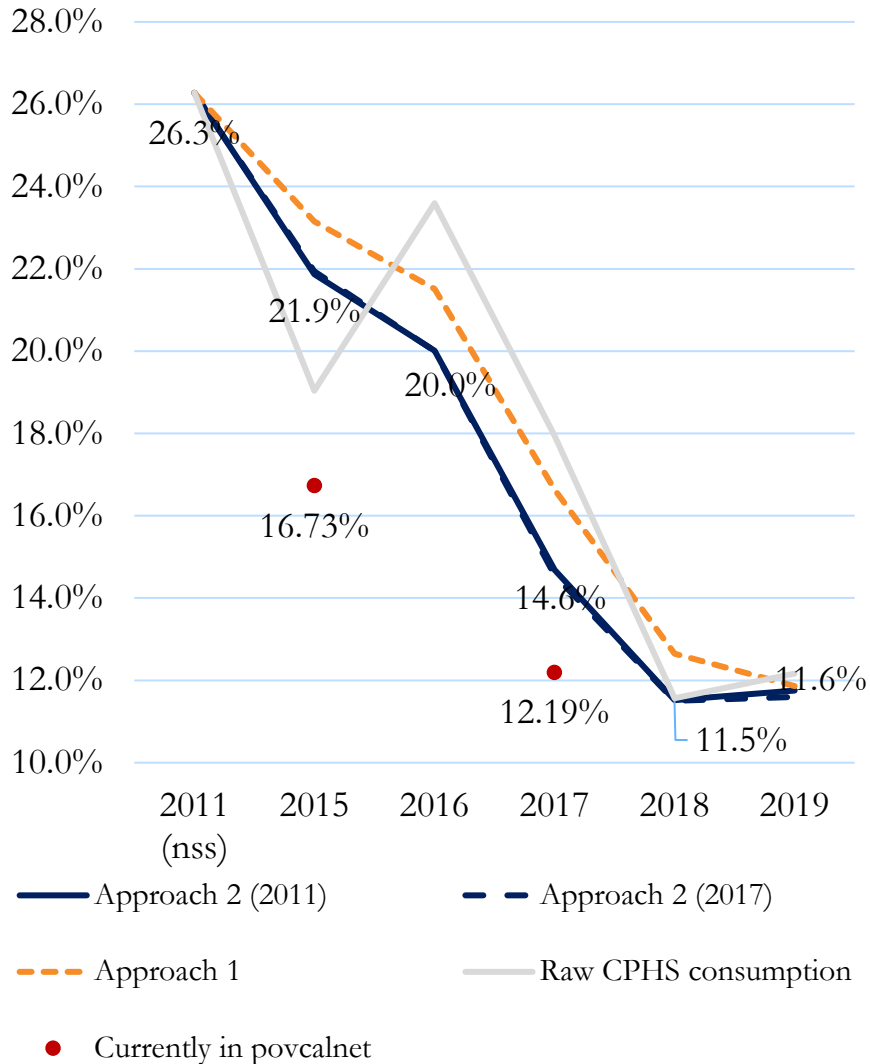
Reduction in Poverty since 2011

Estimates of poverty headcount at the \$1.90 line



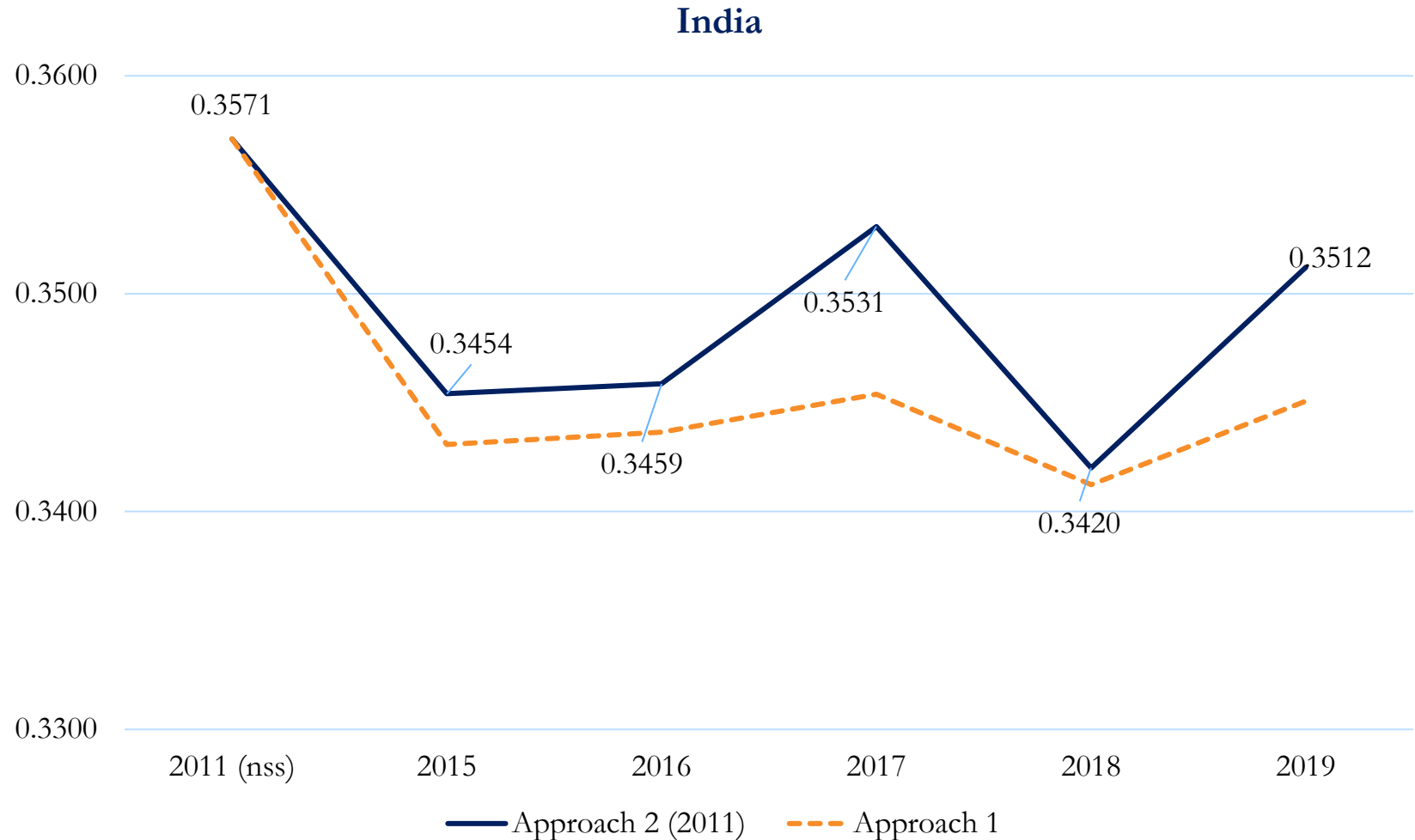
Faster poverty reduction in rural areas

Estimates of poverty headcount at the \$1.90 line



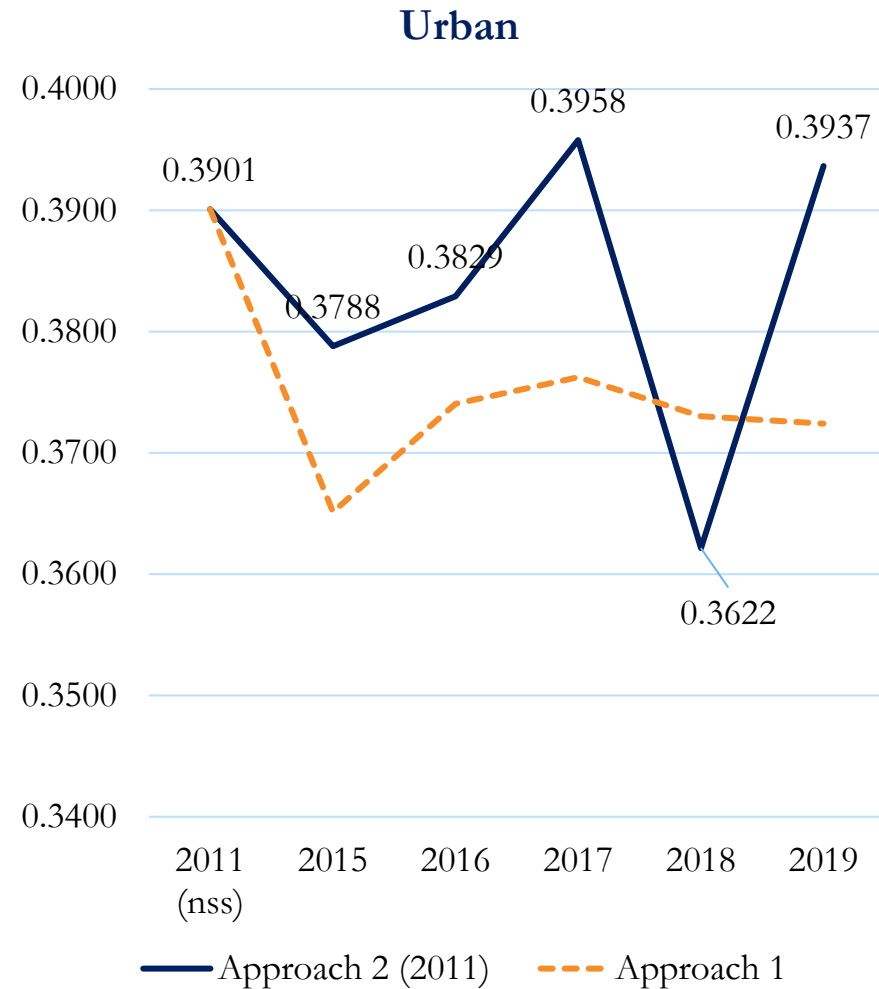
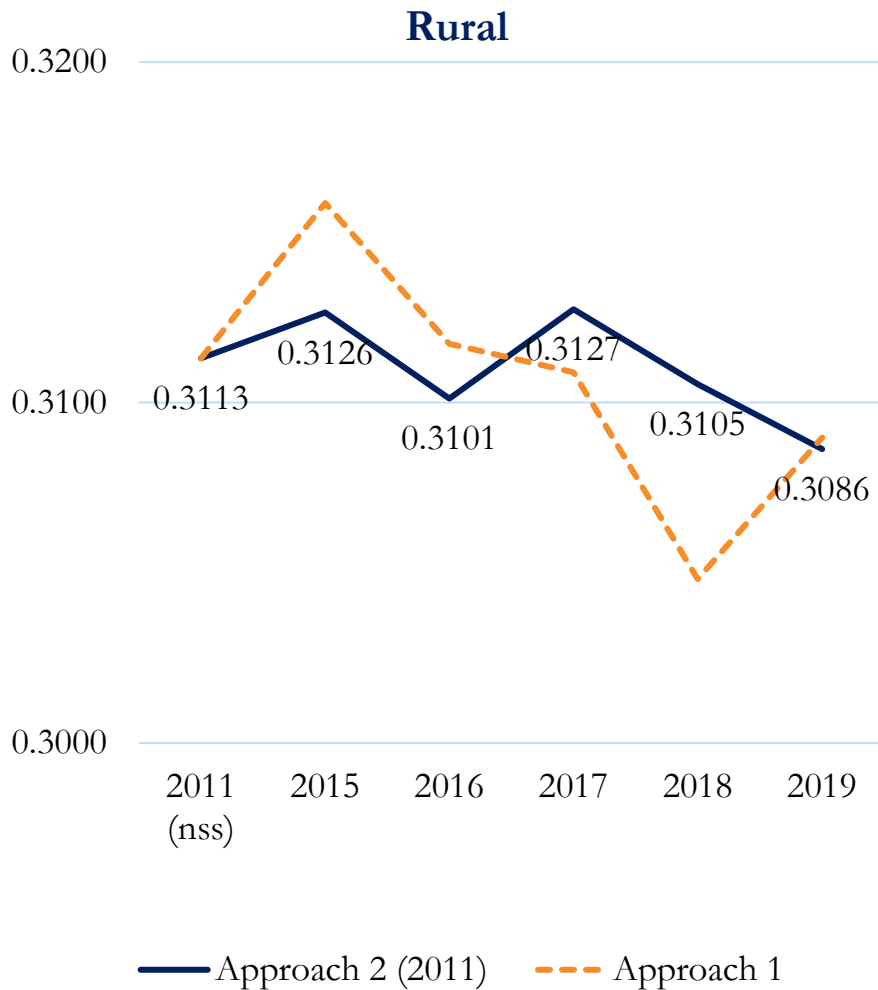
Moderation in Inequality since 2011

Gini measures of inequality



Faster poverty reduction in rural areas

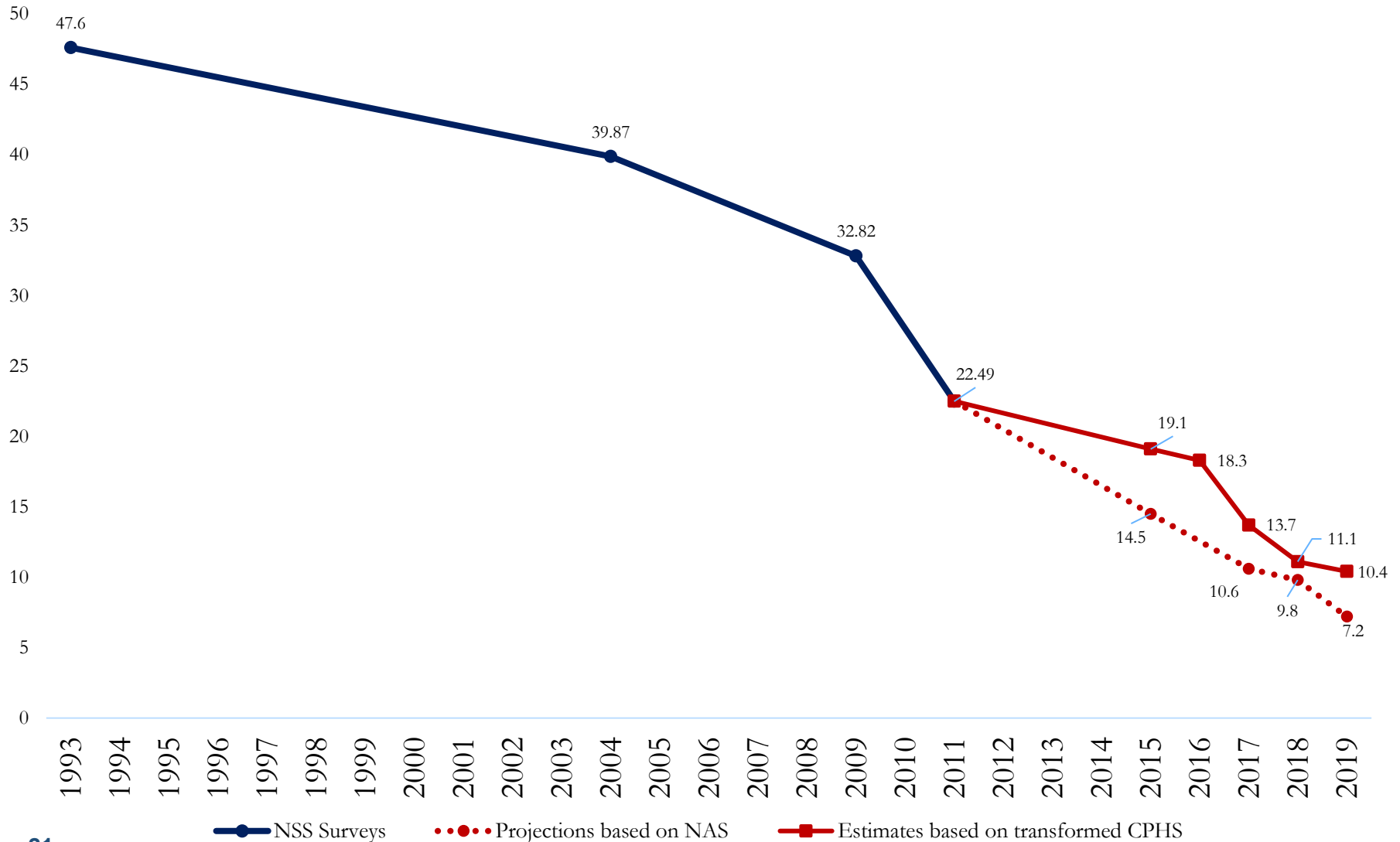
Estimates of poverty headcount at the \$1.90 line



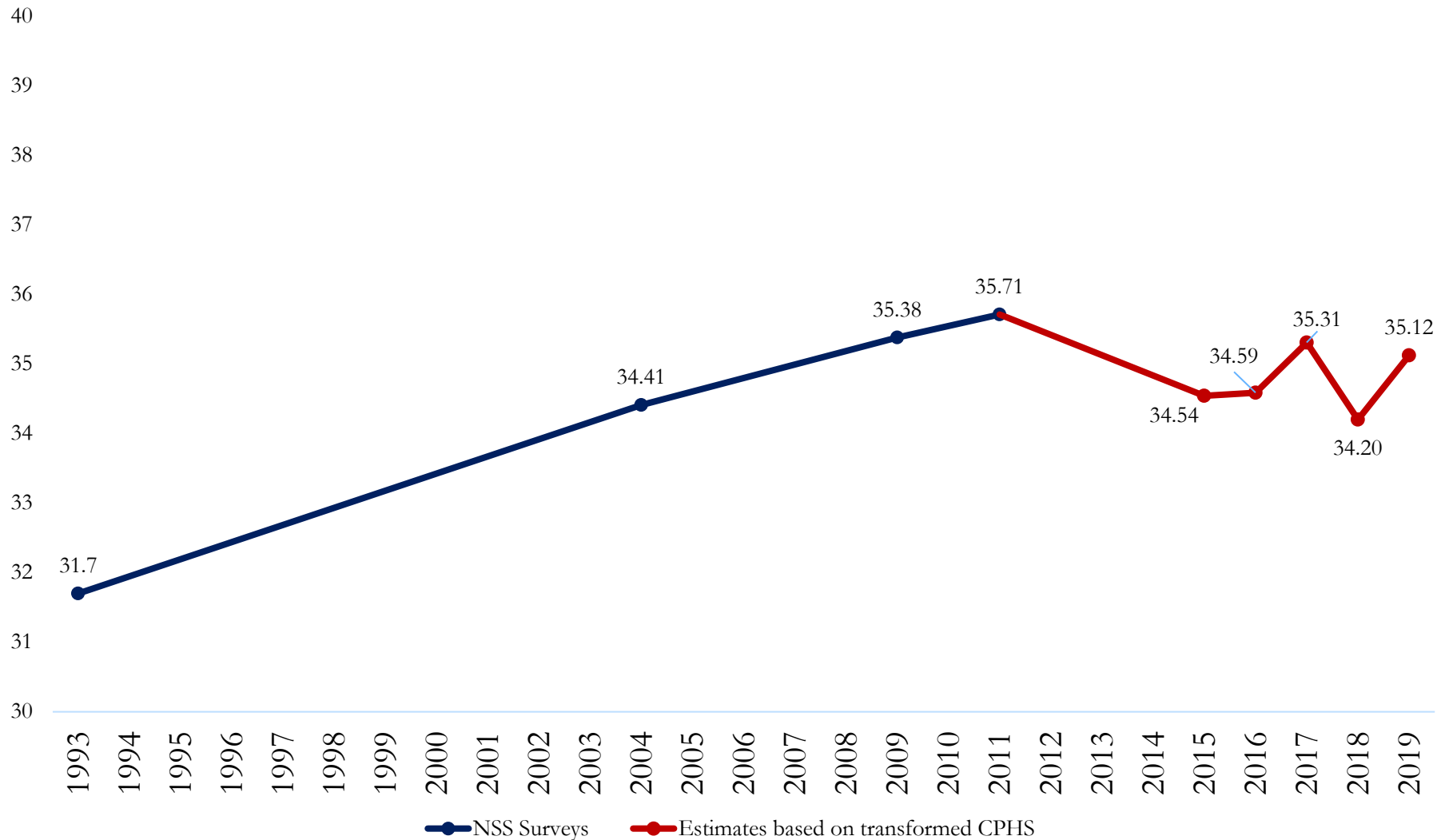
Disclaimer slide regarding inequality

- Household surveys generally undercover households from the top 5 percent of the income distribution
- Consequently, estimates of inequality derived from household survey data (whether NSS or CPHS) capture income disparities between households from bottom-95 percent, say
- When no households from top 5 percent made it into the sample, no amount of reweighting will resolve this issue
 - This motivates the work by Atkinson, Piketty, Alvaredo, Saez, Ravallion and co. who employ income tax records data to obtain estimates of the top tail of the income distribution

The evolution of \$1.90 poverty in India (%)

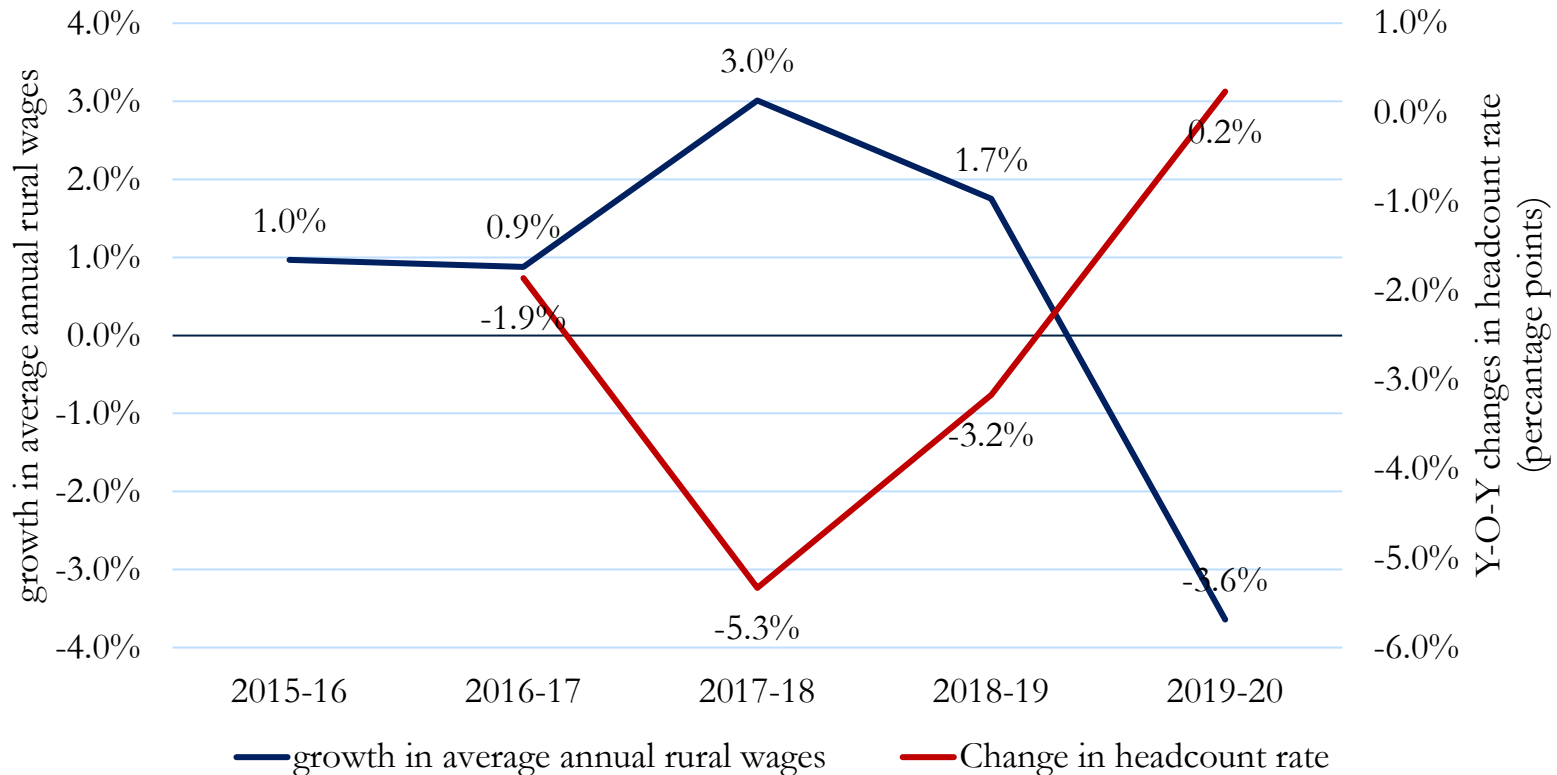


The evolution of Inequality in India



Validation: Should we trust these estimates?

1. In the years following 2015, poverty reduction rates are highest in 2017-2018 and moderated in 2019: **Supported by real wage growth**

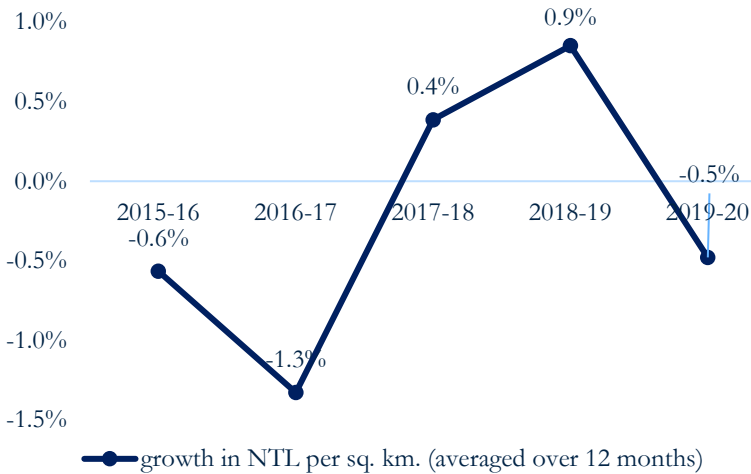


- Faster growth in real rural wages => Faster poverty reduction
- Correlation coefficient = -0.94

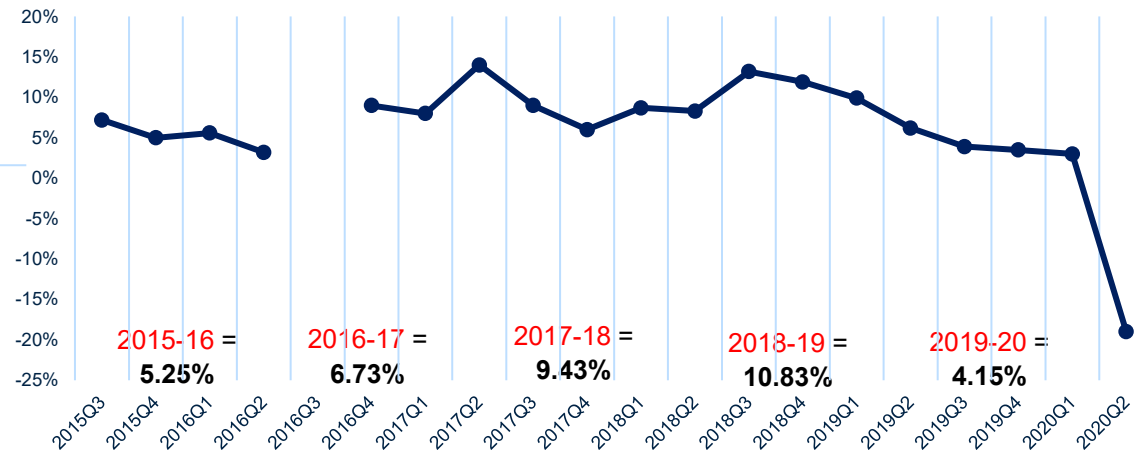
Validation: Should we trust these estimates?

1. In the years following 2015, poverty reduction rates are highest in 2017-2018 and moderated in 2019: **Supported by NTL and Nielsen surveys**

Night Lights



Nielsen retail surveys



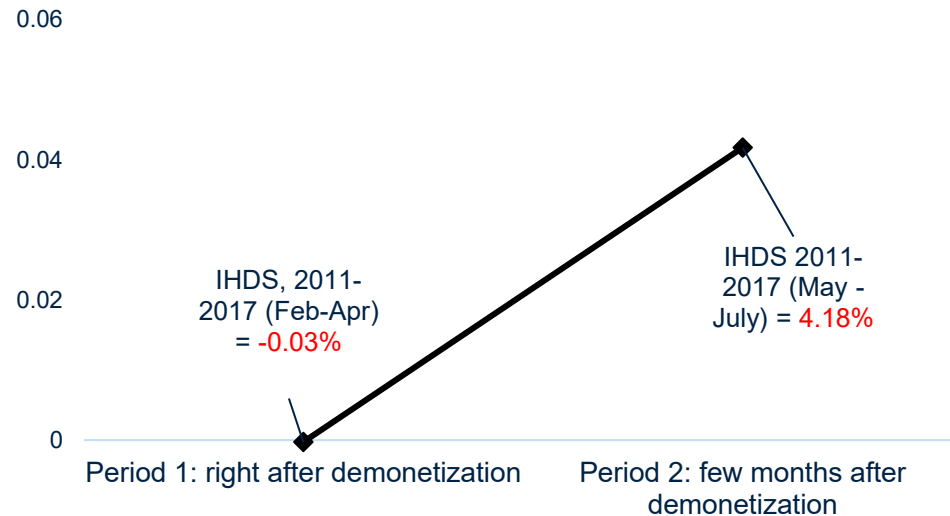
Source: Beyer, Jain and Sinha (2020)

- Nightlights data supports the finding that the fastest reduction in poverty occurred in 2018-19
- Retail store surveys independently conducted by Nielsen support the same finding
- Both survey show a fall in welfare in 2019

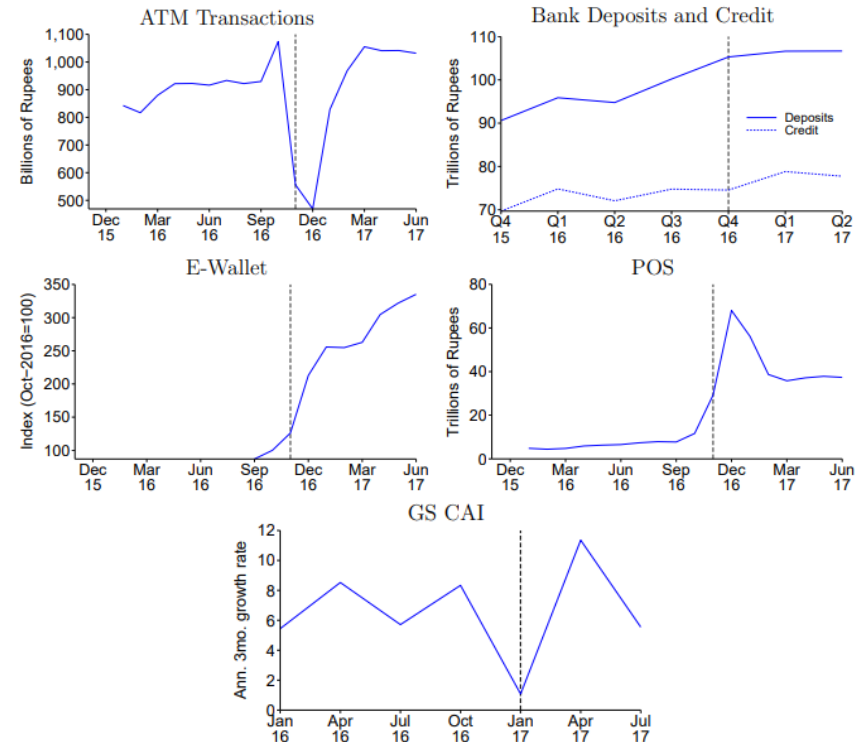
Validation: Should we trust these estimates?

2. Churn around 2016 -- rise in urban poverty in 2016 followed by a rapid rise in consumption in 2017: **Supported by IHDS and banking admin data**

IHDS



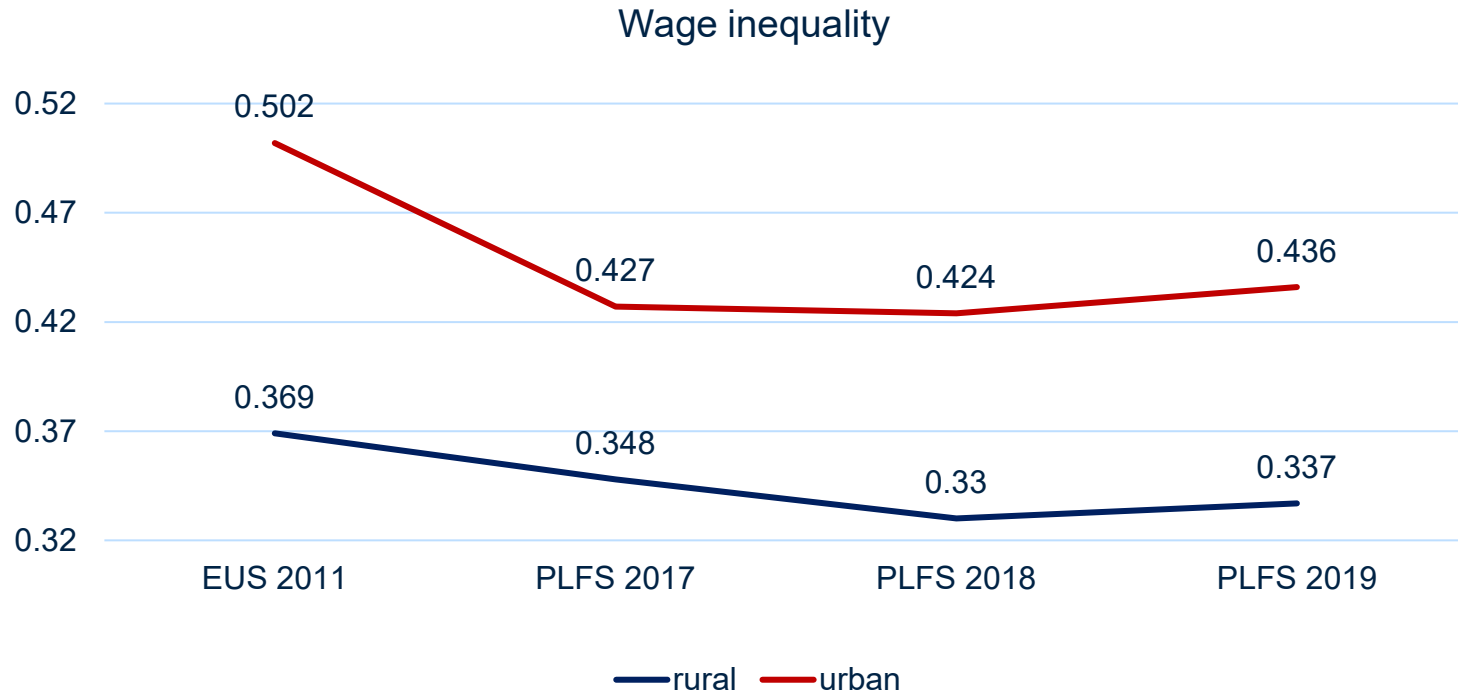
Chodorow-Reich, et. al (2019)



- IHDS-3 result support a drop in poverty overall and churn around demonetization (survey conducted in 3 states)
- Banking data from Chodorow-Reich, et. al (2019) also shows a temporary churn around demonetization followed by quick turnaround (GS CAI = Goldman Sachs Current Activity Indicator)

Validation: Should we trust these estimates?

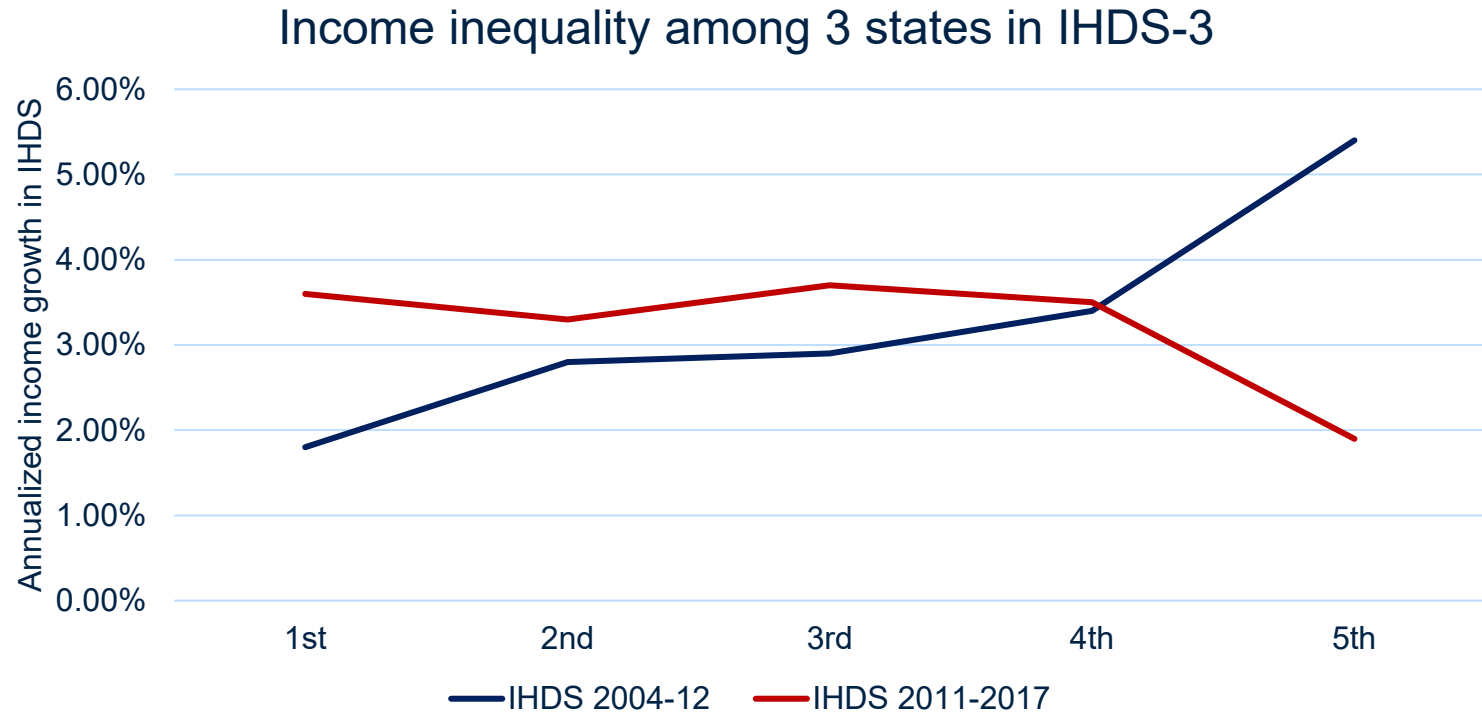
3. No rise in consumption inequality since 2011, but indications of a rise in 2019 :
Supported by fall in wage inequality observed in PLFS



- Wage inequality falls in 2018 and goes up in 2019
- Rise in 2019 wage inequality is higher in urban than rural areas
- Pool salaries and wages of regular wage and casual wage workers. Self employed workers excluded (50% of LF in rural and 40% of LF in urban = self-employed)

Validation: Should we trust these estimates?

3. No rise in consumption inequality since 2011, but indications of a rise in 2019:
Supported by fall in income growth of richest households in IHDS

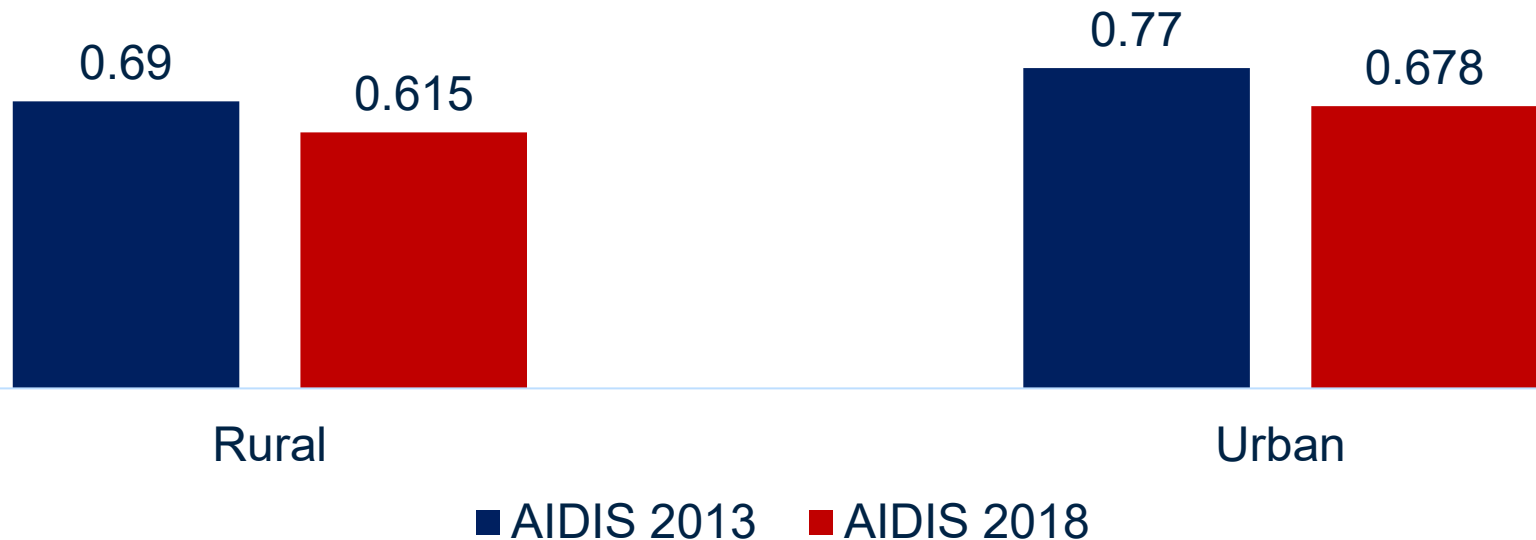


- IHDS 2012 reported a rise in income inequality over 2004. This is due to a larger annualized increase in average incomes of people at the top end of the distribution.
- IHDS 2017, available for 3 states, shows a drop in incomes at the end of the distribution: suggesting a moderation in income inequality.

Validation: Should we trust these estimates?

3. No rise in consumption inequality since 2011, but indications of a rise in 2019:
Supported by fall in wealth inequality Ginis

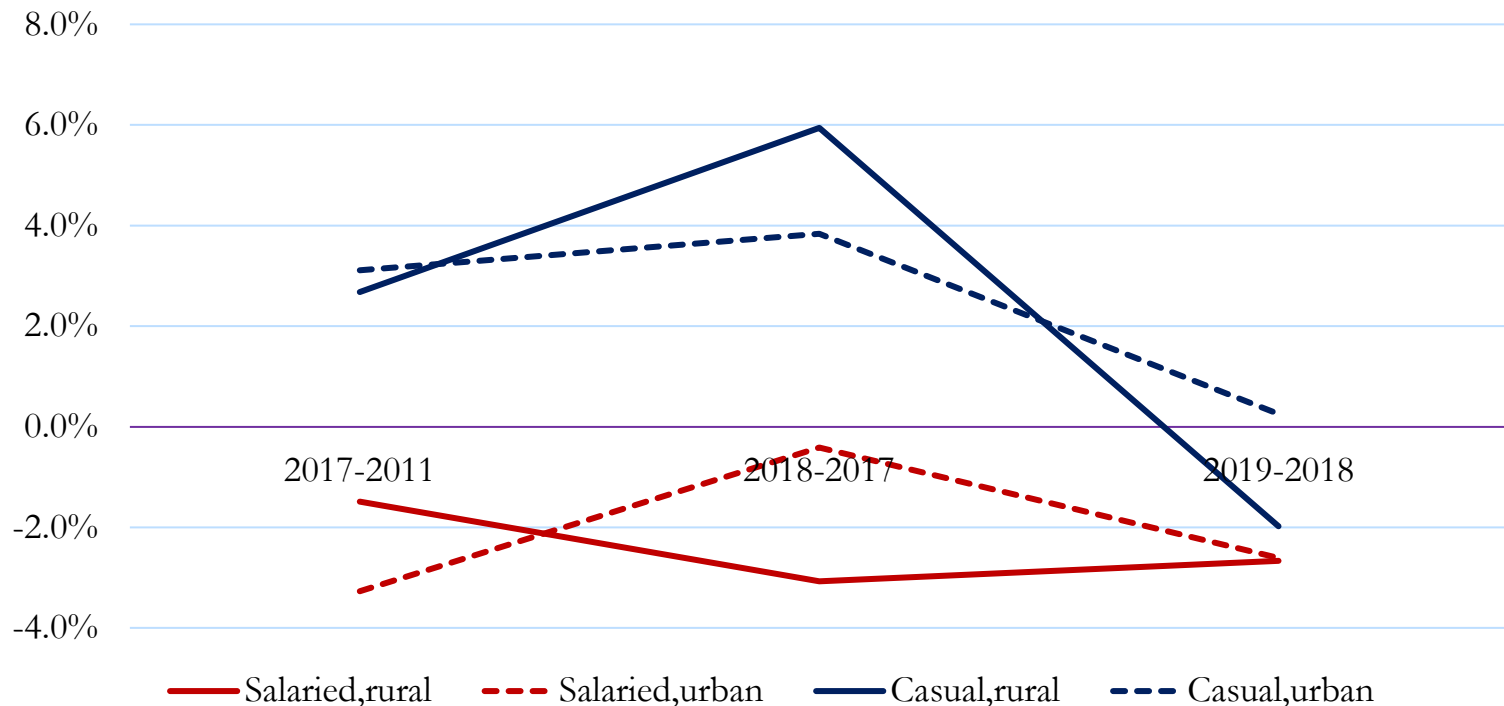
Wealth Inequality



- Data source: All India Asset, Debt and Investment surveys – AIDIS (2013 and 2018)
- Wealth = physical + financial

Validation: Should we trust these estimates?

3. No rise in consumption inequality since 2011, but indications of a rise in 2019 :
Supported by positive casual wage growth but negative salaried growth

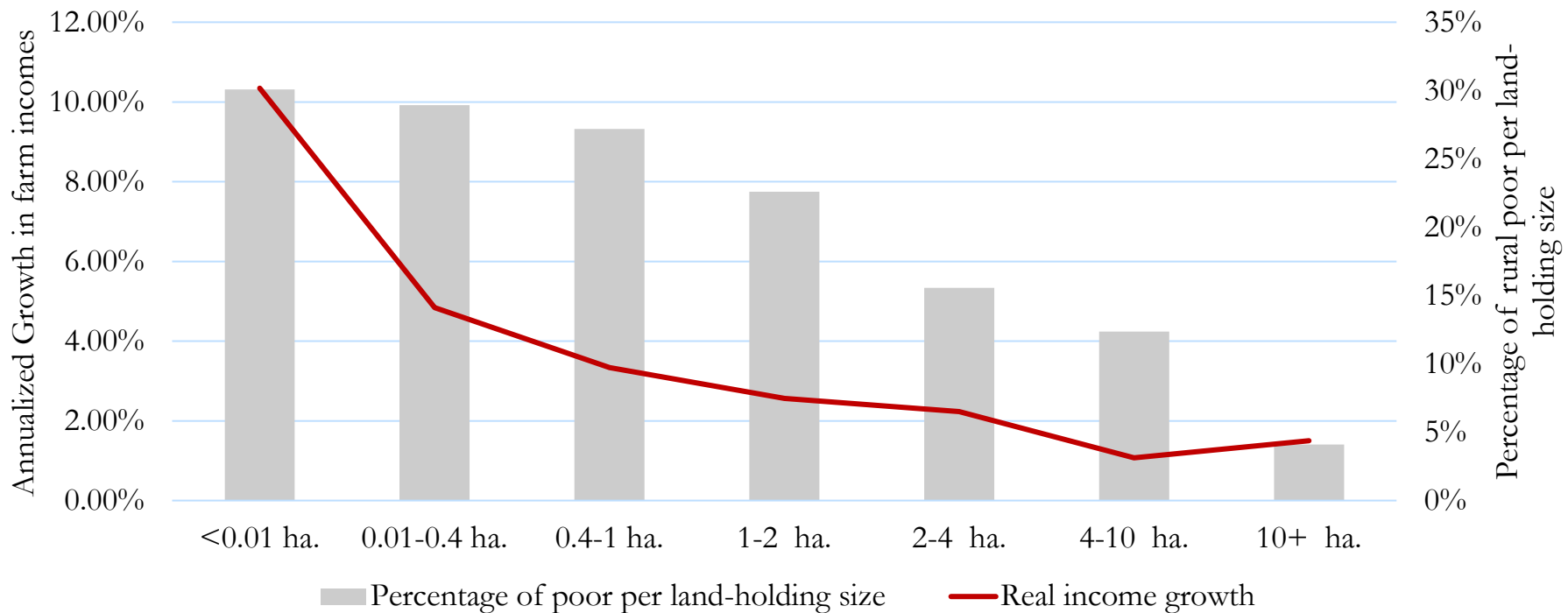


- only 8 % of households from the bottom decile of the consumption distribution in 2011 had a member working in a regular salaried job.
- In contrast, 50% of households from the top decile have at least one salaried member.
- Higher casual wage growth => growth in the bottom part of the distribution and a moderation in inequality

Validation: Should we trust these estimates?

3. No rise in consumption inequality since 2011, but indications of a rise in 2019:

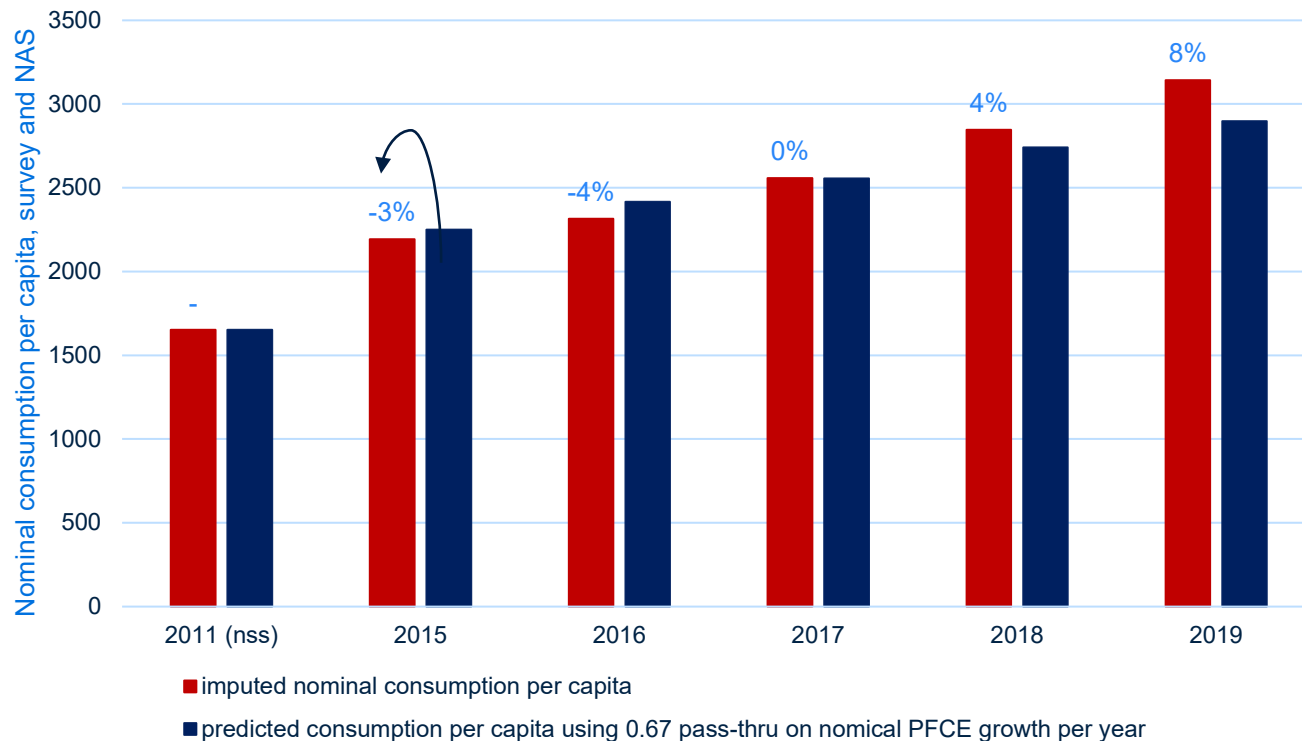
Supported by high income growth for agricultural HHs with smallest land holding size



- Total income = income from wages + net receipt from crop production (out-of-pocket)+ net receipt from farming of animals (out-of-pocket) + net receipt from non-farm business income
- Deflated using CPI AL
- X-axis: size class of land possessed by agricultural households. Ha = hectares.

Validation: Should we trust these estimates?

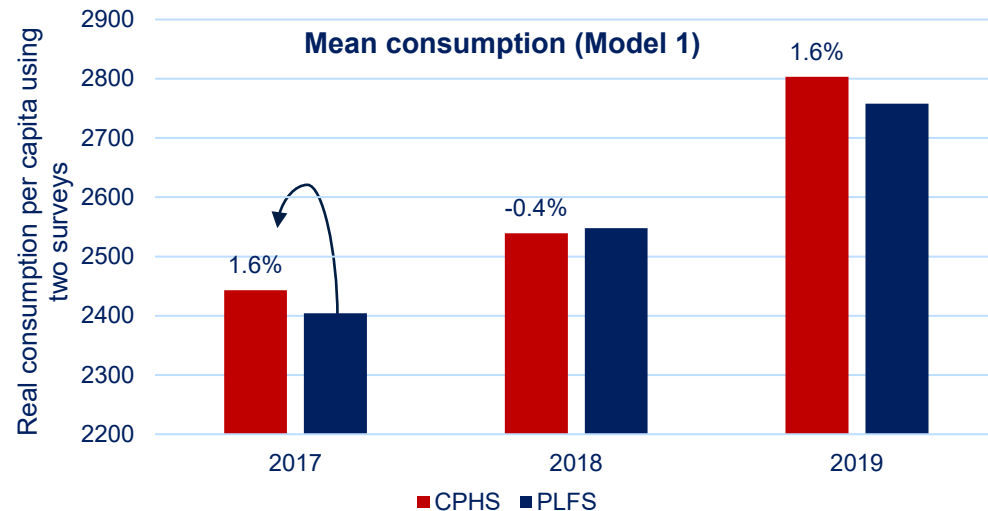
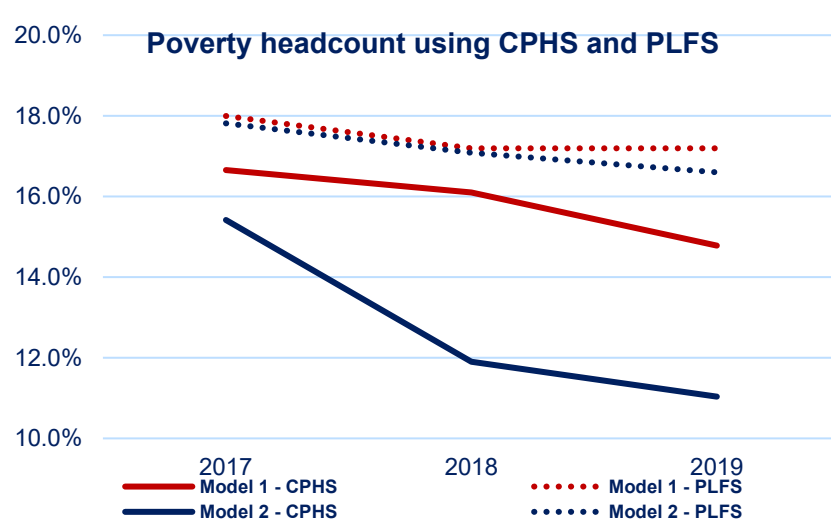
4. Is the mean survey consumption validated?: evident in PFCE growth from NAS



- Yearly consumption predicted using line up method
 - Imputed consumption from CPHS = -3% - 8% of consumption based on line-up method. Relying exclusively on PFCE growth however misses the perturbation in poverty trends around 2016-17
- 31 2019 PFCE includes the covid shock (July'19-June'20); CPHS 2019 misses this period (Apr'19-Mar'20): could explain the 8% gap

Validation: Should we trust these estimates?

4. Is the mean survey consumption validated? similar estimates of average consumption per capita when imputed using PLFS instead of CPHS

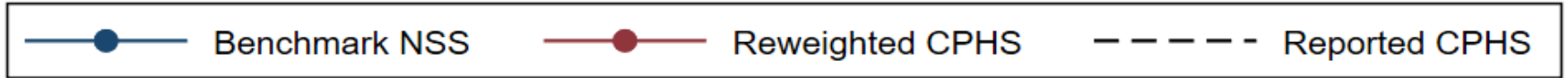
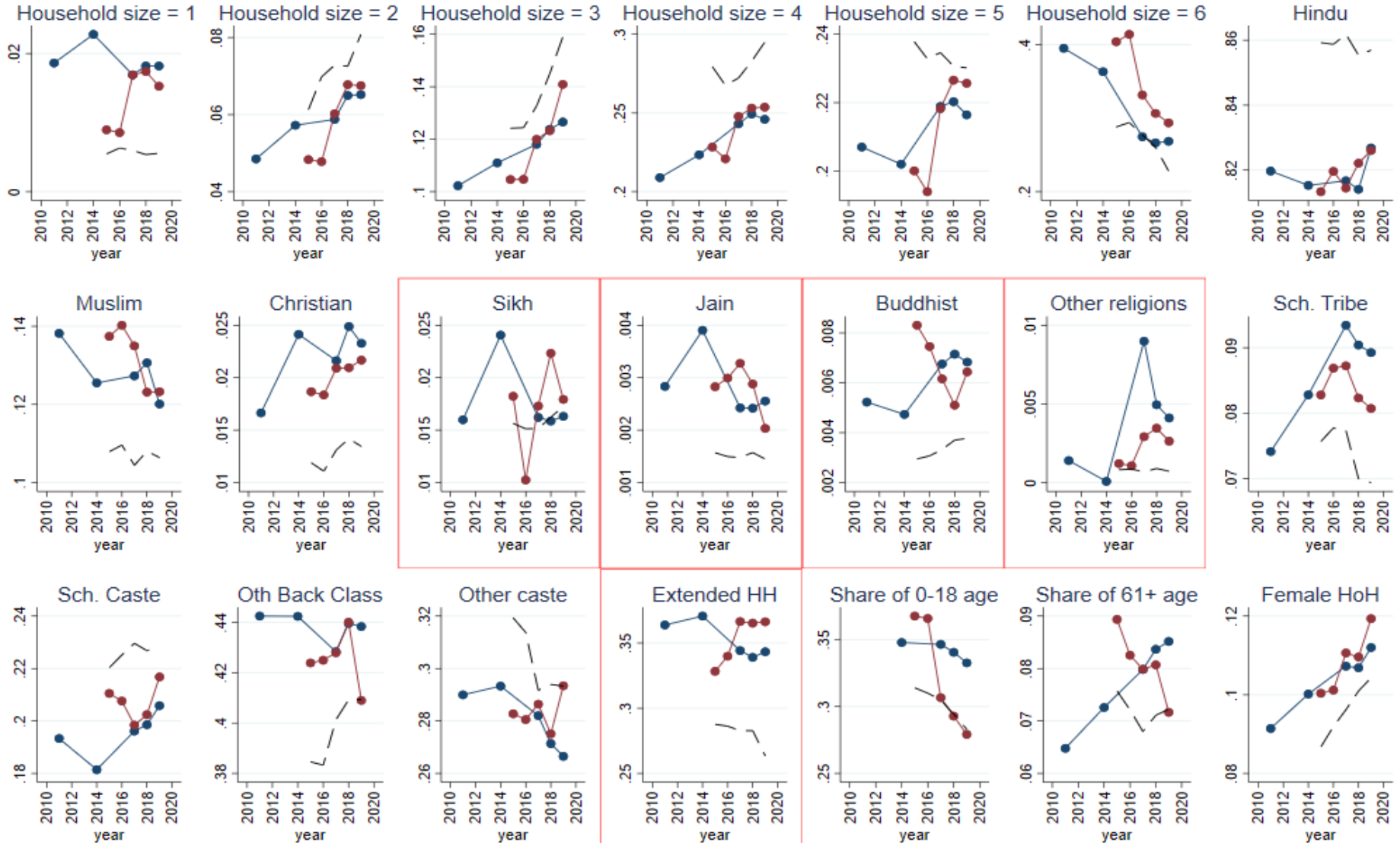


Model 1: Vars common across CPHS, PLFS and NSS-2011, Model 2 – X: Vars common across survey X and NSS-2011 (X:CPHS = demographic + assets vars; X: PLFS = only demographic vars)

- Why not use PLFS instead of CPHS to impute consumption (circumvents some of the challenges of using CPHS)?
- No assets in PLFS; important predictors of household consumption. Estimated consumption significantly lower when vars not included
- Using the same set of imputation variables across PLFS and CPHS we obtain average consumption per capita values that are close to each other

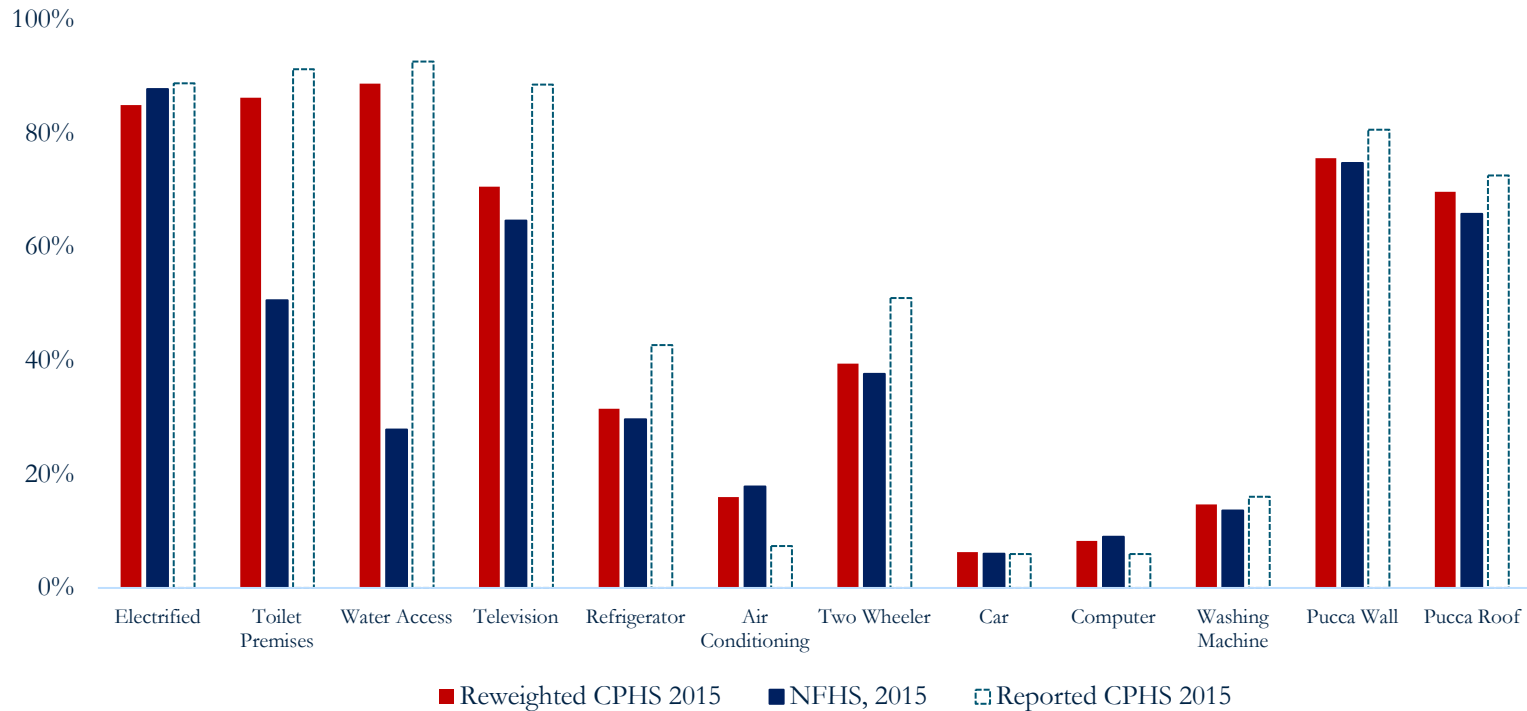
Appendix

Demographics



Access to Services and Household Assets

Panel (a): NFHS and CPHS 2015

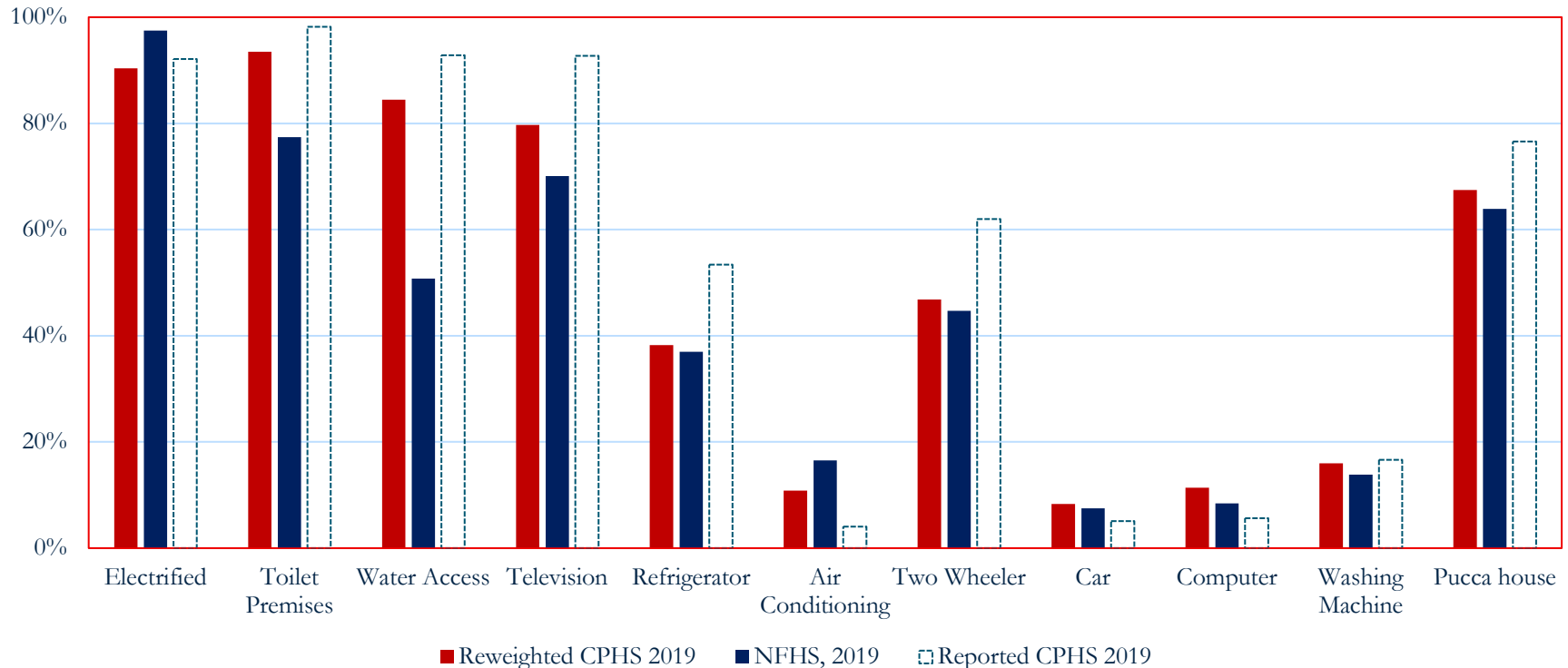


- Definitional differences in electricity, toilet premises, water access potentially driving the differences

Access to Services and Household Assets	Before	After
Consumer durables: TV, AC, Refrigerator	Over represented in CPHS	TV: +7 pp, others balanced rel. to NFHS
Access: electricity, water, toilet in premises	Over represented in CPHS	Over represented in CPHS

Access to Services and Household Assets

Panel (b): NFHS and CPHS 2019



Access to Services and Household Assets	Before	After
Consumer durables: TV, AC, Refrigerator	Over represented in CPHS	TV:+10pp, AC: -6pp, others balanced rel. to NFHS
Access: electricity, water, toilet in premises	Over represented in CPHS	Over represented in CPHS

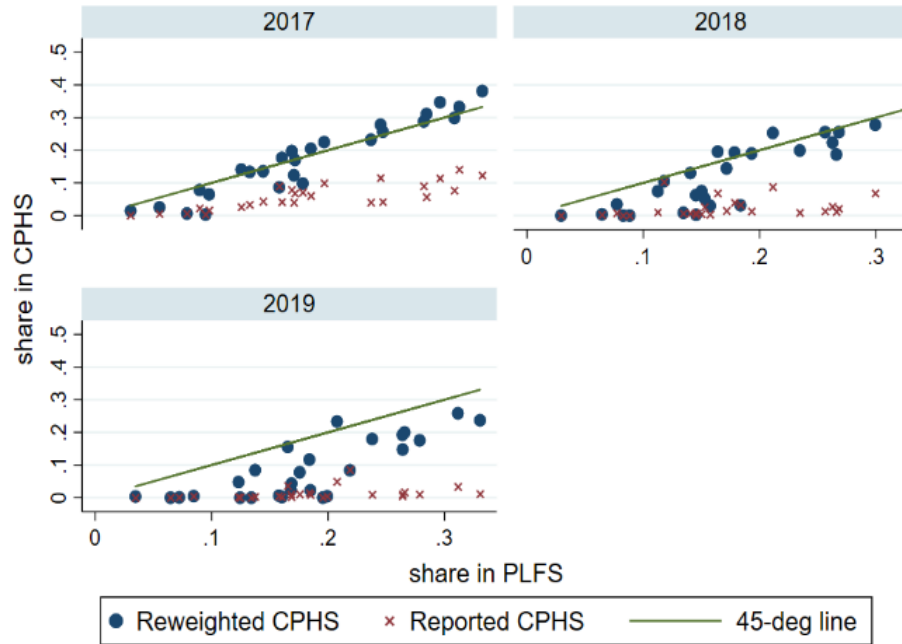
• Sample comprises of 14 states reporting data from NFHS, 2019. Aggregated using households totals

36
from LFS-2019.

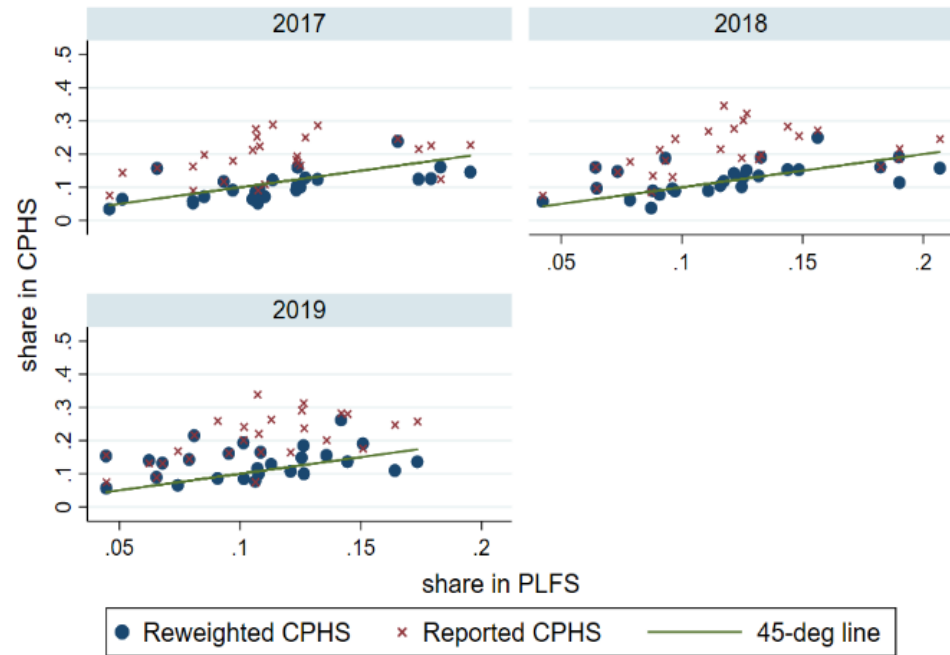
• None of these variables are included in the list of target variables for reweighting in 2019.

Education

Panel (a): Below primary education shares



Panel (b): Primary education shares



- Below primary in CPHS = not literate, PLFS = not literate + below primary
 - Each scatter point represents education levels in a state

Education

Before

After

Low educated people (15-49)

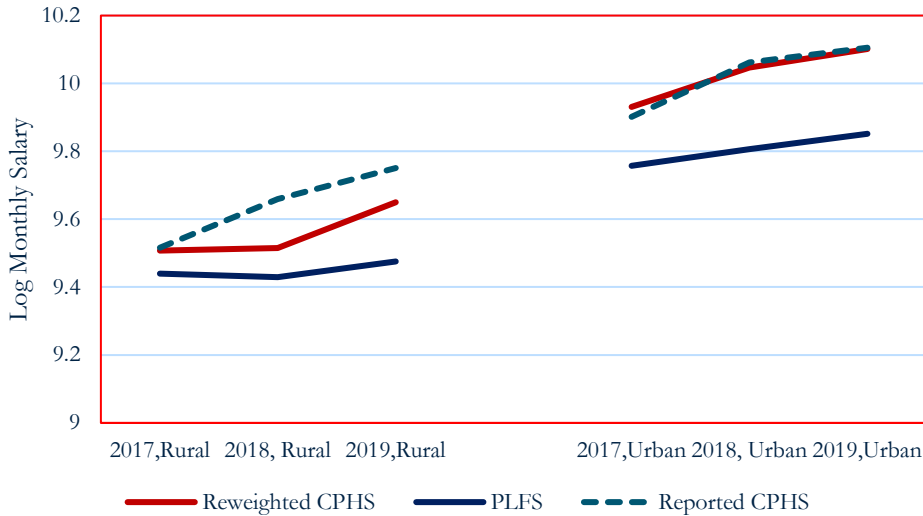
Severely under represented in CPHS

- Balanced in 2017, 2018 balanced rel. to contemporaneous PLFS.
- Lower shares in CPHS 2019 than PLFS 2019 – more likely to be classified as primary educated.

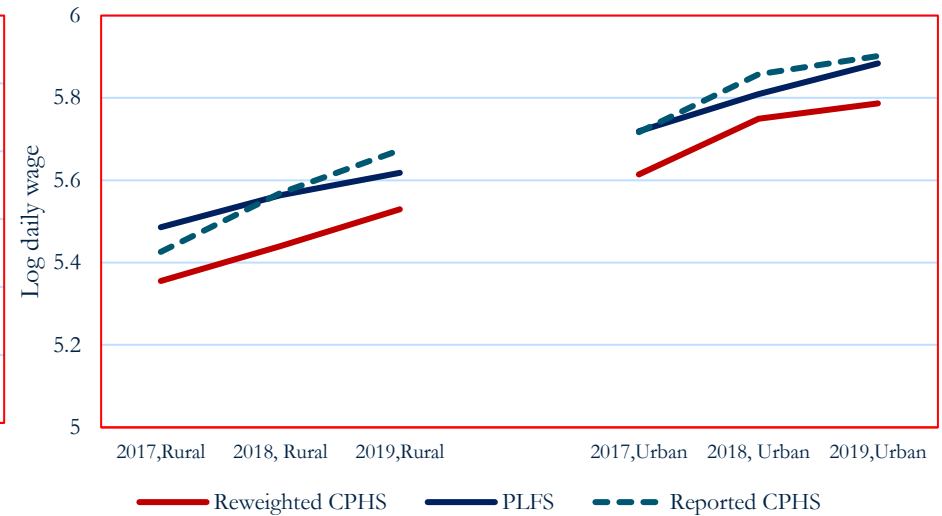
Labor Market Indicators

Average monthly salaries and daily wages

Panel (a): Log Monthly Salaries for Salaried workers



Panel (b): Log Daily Wages for Casual workers



Labor market indicators

Before

After

Labor incomes

Higher in CPHS

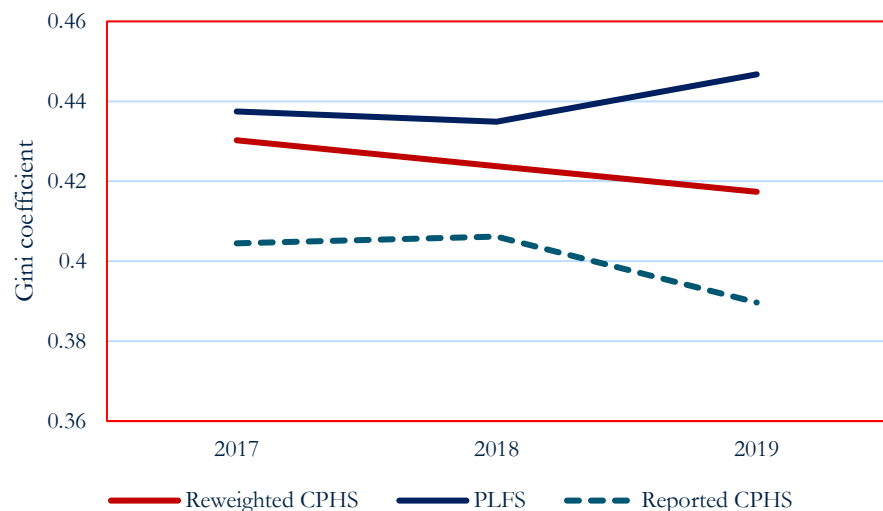
- Gap between PLFS and CPHS salaries and wages closer especially in rural areas.
- Lower casual wages and higher salaries in CPHS than PLFS

- Sample restricted to workers with non-zero wages in both surveys

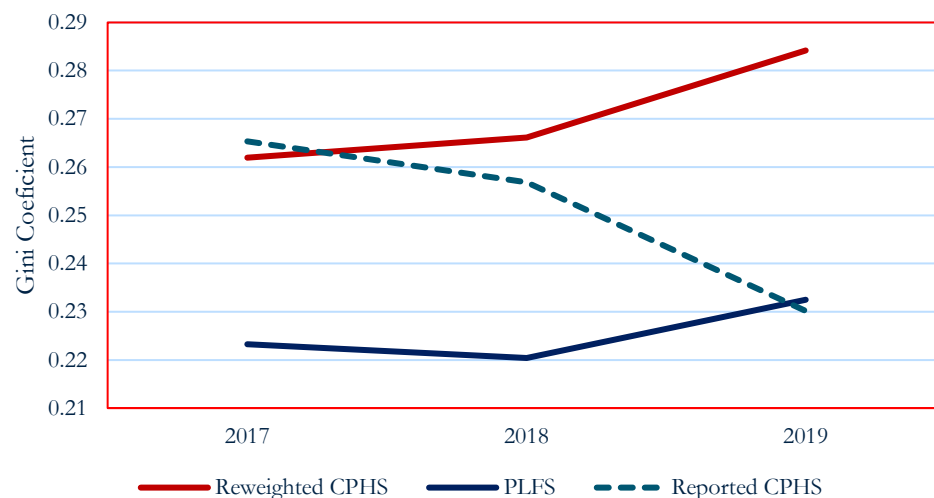
Labor Market Indicators

Inequality in monthly and daily wages

Panel (a): Inequality among salaried earners



Panel (b): Inequality among casual workers



Labor market indicators

Before

After

Labor income inequality

Lower in CPHS

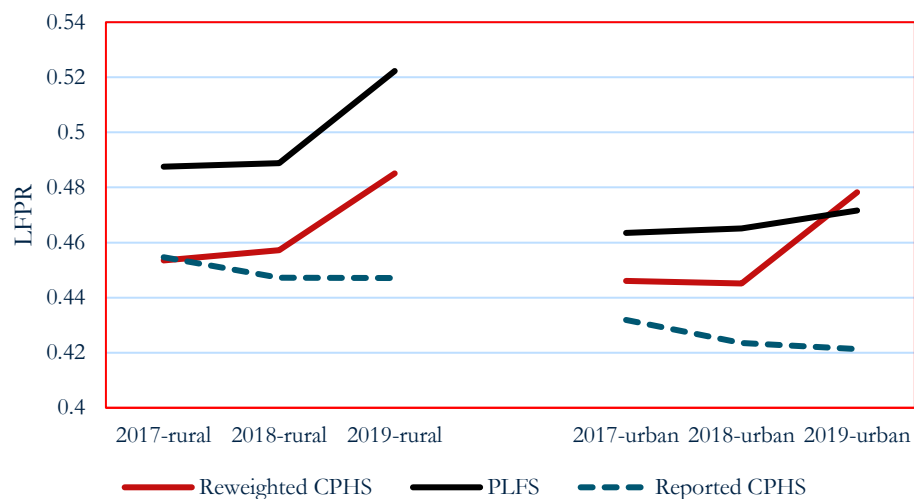
- Gap between PLFS and CPHS earnings inequality narrowed

- Sample restricted to workers with non-zero wages in both surveys

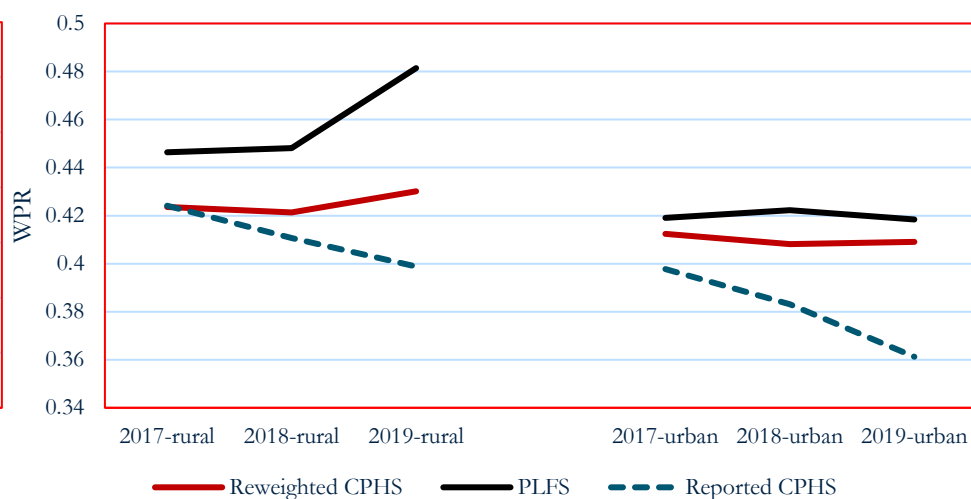
Labor Market Indicators

Key labor market indicators

Panel (a): Labor Force Participation Rates



Panel (b): Worker Population Ratio



Labor market indicators

Before

After

LFP Rates

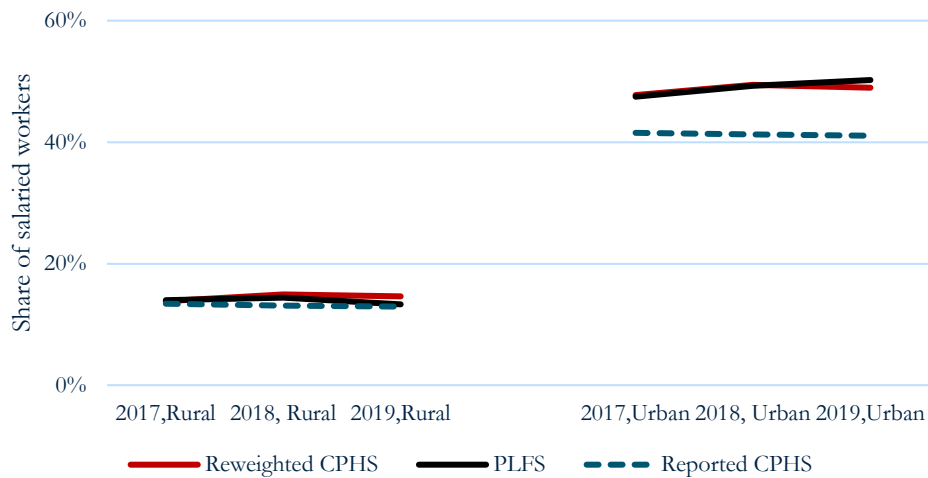
Lower LFP in CPHS

- Comparable LFP across two surveys; matching trends with contemporaneous PLFS.
- Closer female LFPR in two surveys post re-weighting

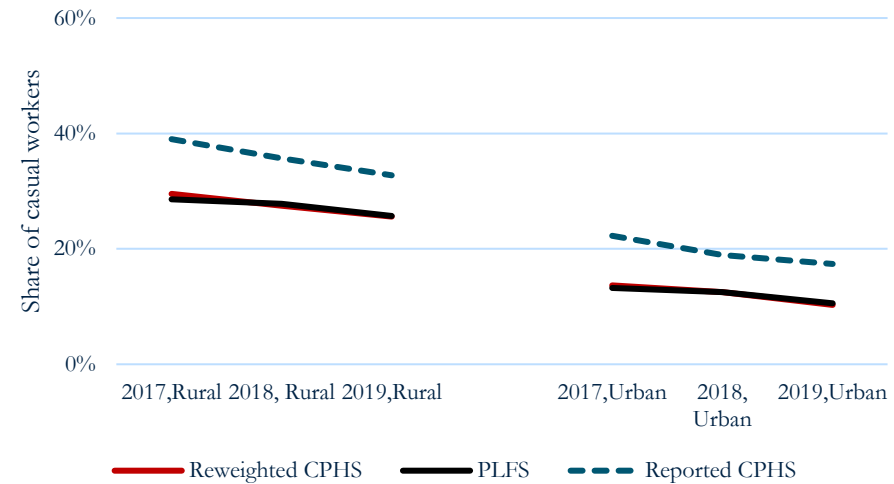
Labor Market Indicators

Composition of work force

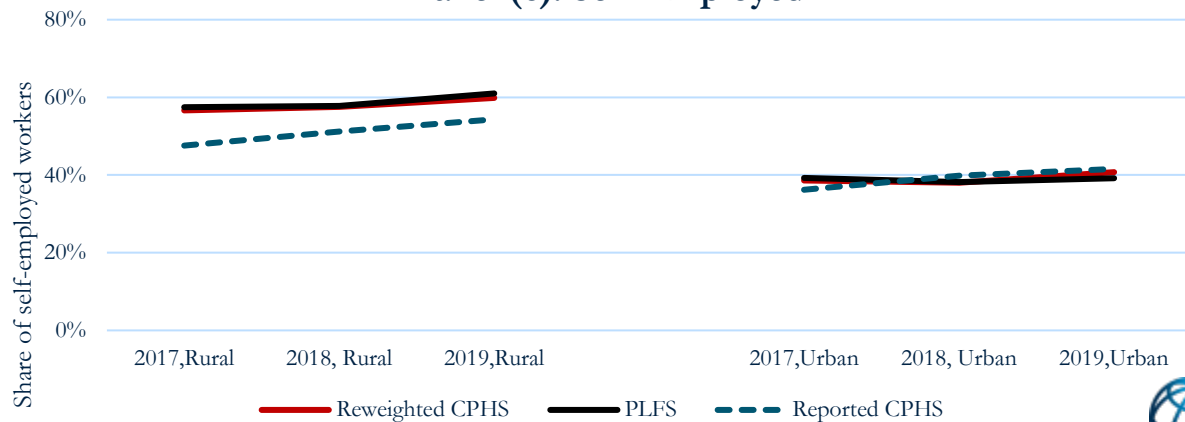
Panel (a): Salaried



Panel (b): Casual wage



Panel (c): Self-Employed



Discussion of Bhalla et al. (2022)

1. Subsidized rice and wheat consumptions are evaluated at market prices of non-subsidized rice and wheat
 - Is that the appropriate price for subsidized food items?
 - Should all consumption of a given item be evaluated at the same price regardless of quality (i.e. should an artisan piece of meat be assigned the same price as a basic cut of meat, even when the market assigns very different prices to these two different cuts of meat)?
2. The consumption basket that defines the poverty line should be evaluated using the same set of prices that are used to evaluate household consumption expenditures
 - One should not use market prices to value household consumption and subsidized prices to evaluate the poverty line

Discussion of Bhalla et al. (2022) – contd.

3. A unit pass-through is estimated using national GDP data which is then applied to state level GDP data to predict mean household consumption

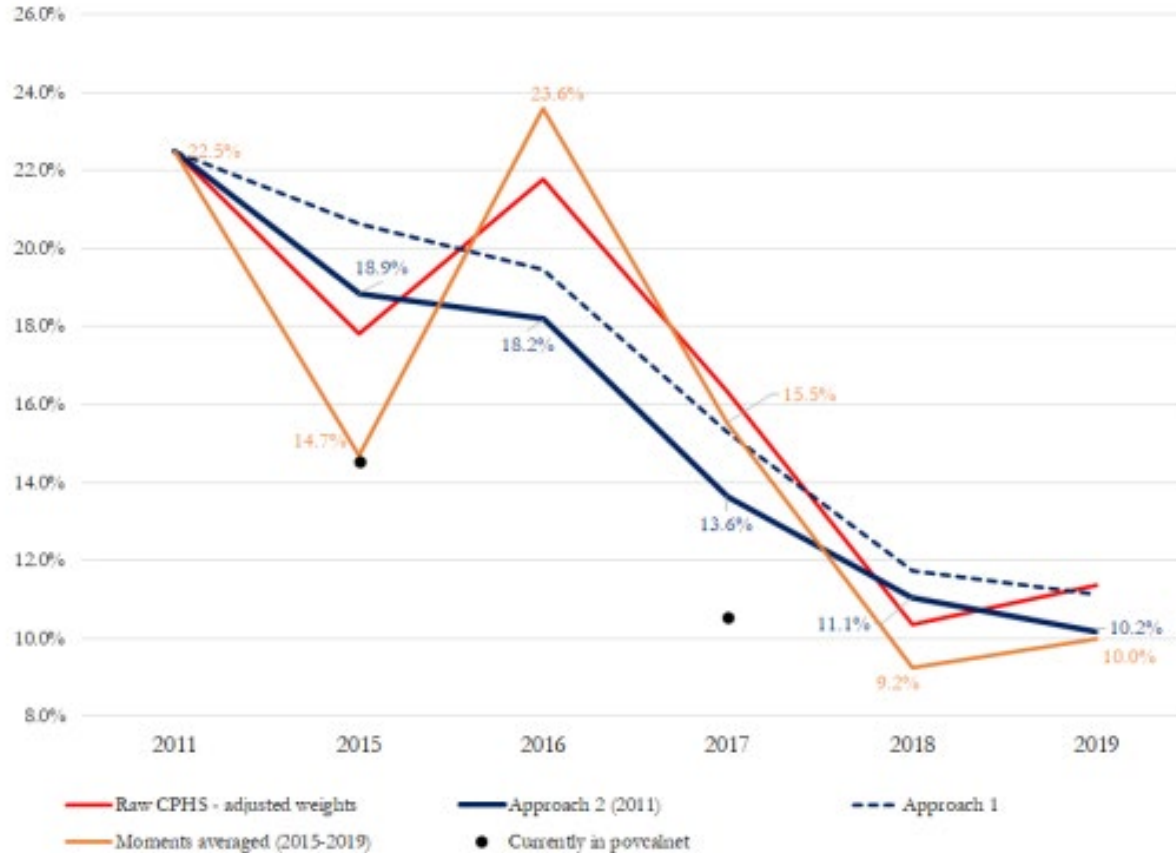
- The state-level GDP and national GDP series grow at different rates
- Regressing household consumption growth on state GDP growth yields a pass-through rate of $0.36 < 1$

4. Urban household consumption and rural household consumption are assumed to grow at the same rate

5. Subsidized food items are assigned to households with exceptionally high levels of efficiency

6. Inequality at state level is assumed to be constant over time

poverty holding, a, b, and s2 constant over time



Real Rural Wages

