Estimating poverty for India after 2011 using private-sector survey data

Luxembourg, December 2024

Sutirtha Sinha Roy and Roy Van der Weide

WORLD BANK GROUP
Poverty & Equity

WORLD BANK GROUP

Background

- World Bank routinely imputes poverty estimates for years with missing surveys …
  - By means of extrapolation 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

- Relies on strong assumptions:
  - 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

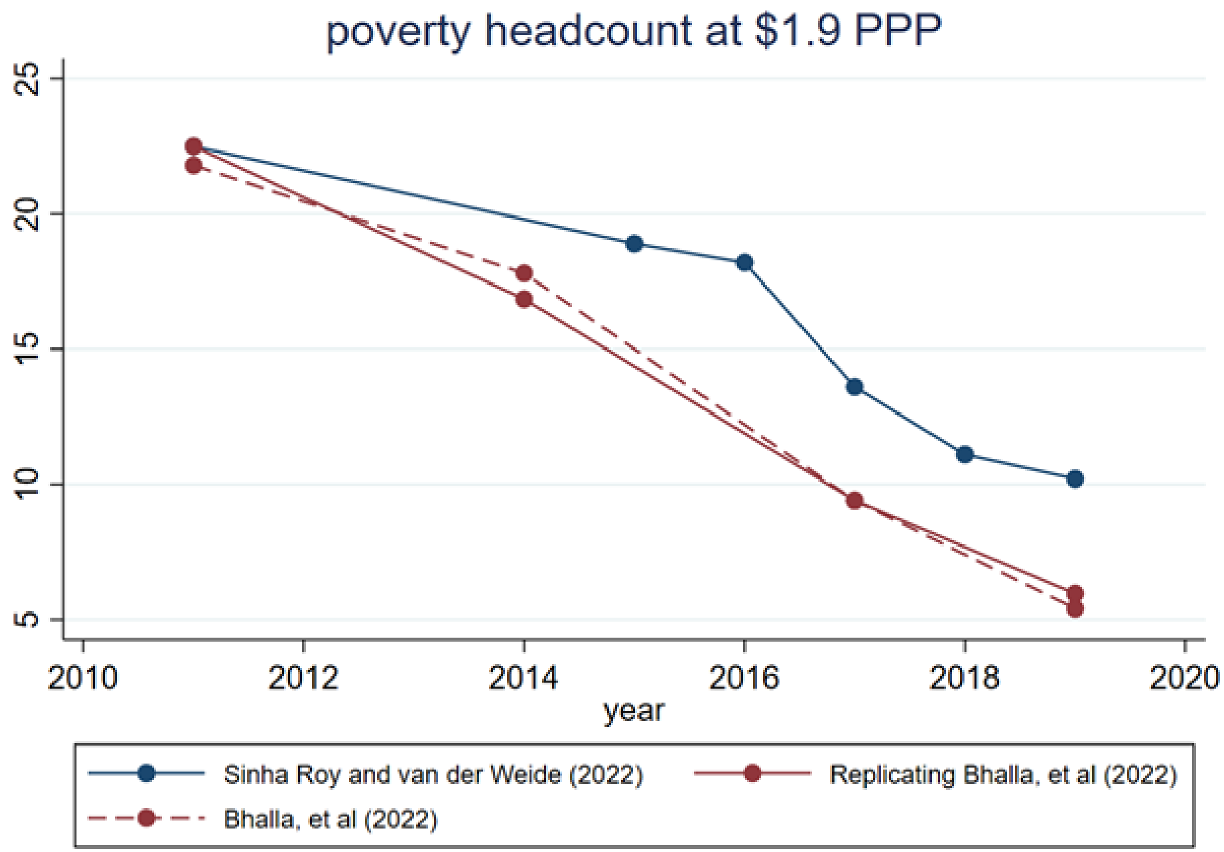
WORLD BANK GROUP

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that combine NSS-2011 with national accounts data are increasingly problematic as the latest NSS becomes increasingly outdated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.

WB and IMF presented new poverty estimates one day apart



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

Approach 1: Ignoring CPHS consumption data

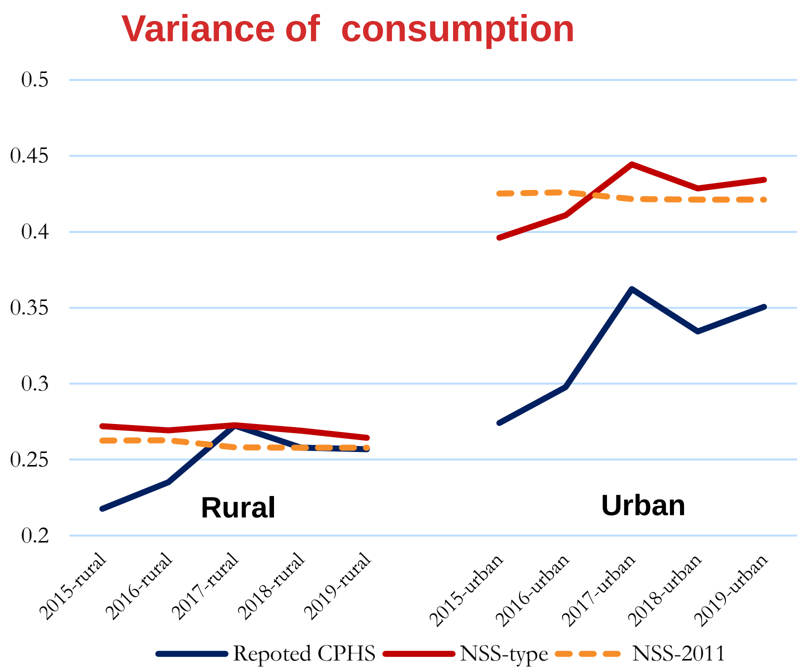
$$log\left(y_i^{NSS}\right) = c + \beta x_i + u_i,$$

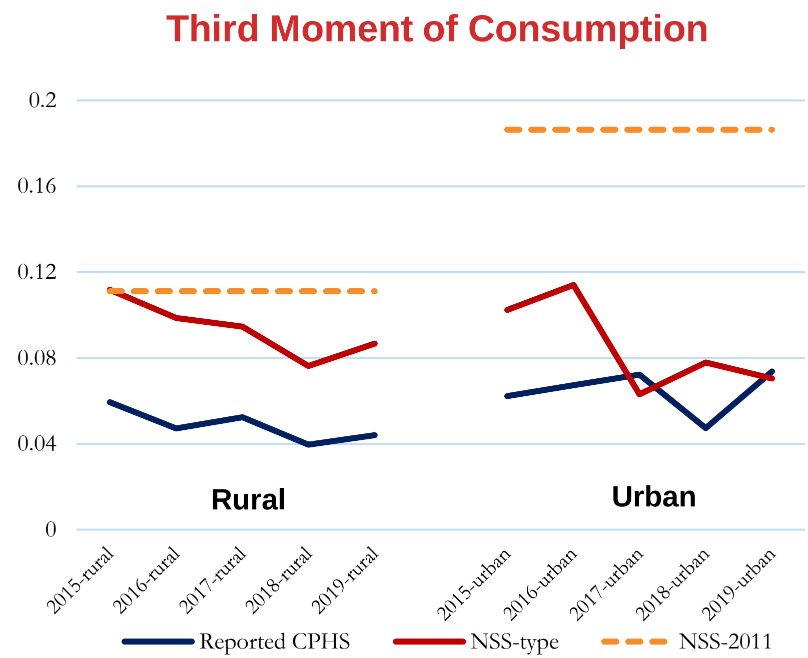
- Covariates x are shared between CPHS and NSS: demographics, education, employment, dwelling characteristics, asset ownership, industry code and dummies capturing consumption of select premium goods
- When imputing NSS-type consumption in the CPHS, the errors are drawn from the empirical distribution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





Approach 2: Using CPHS consumption data

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

- 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 $p\left(\log(y_i^{NSS})|\log(y_i^{CPHS})\right)$

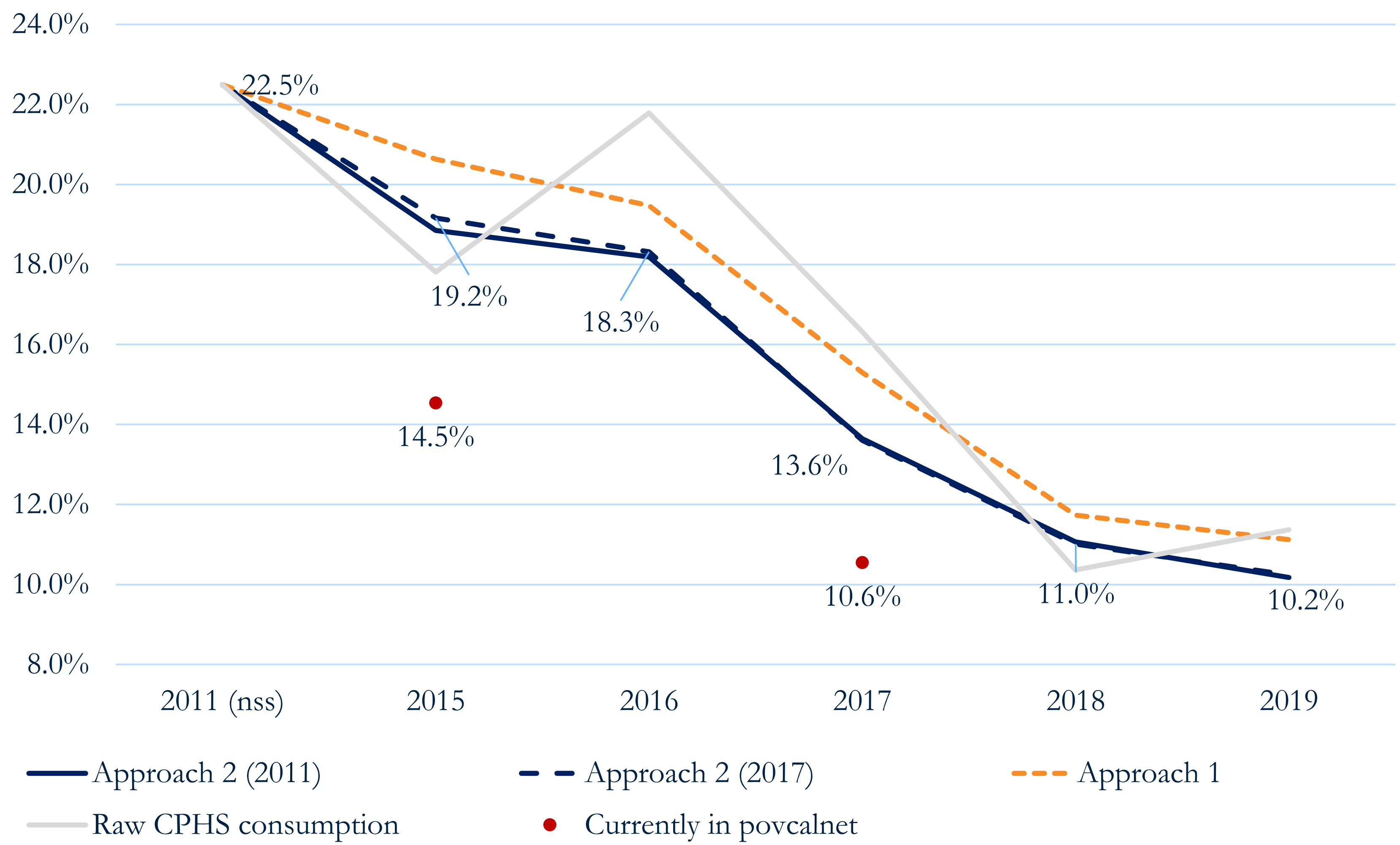
WORLD BANK GROUP

Approach 2 - Continued

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

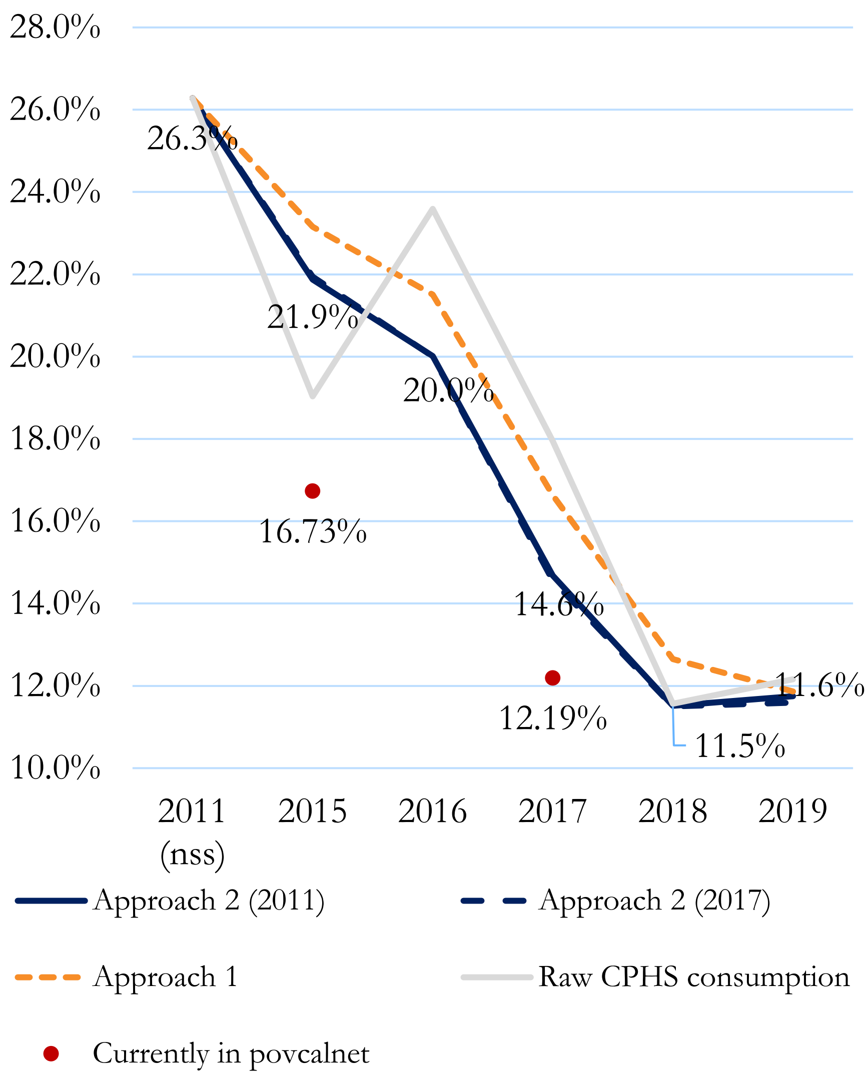
Reduction in Poverty since 2011

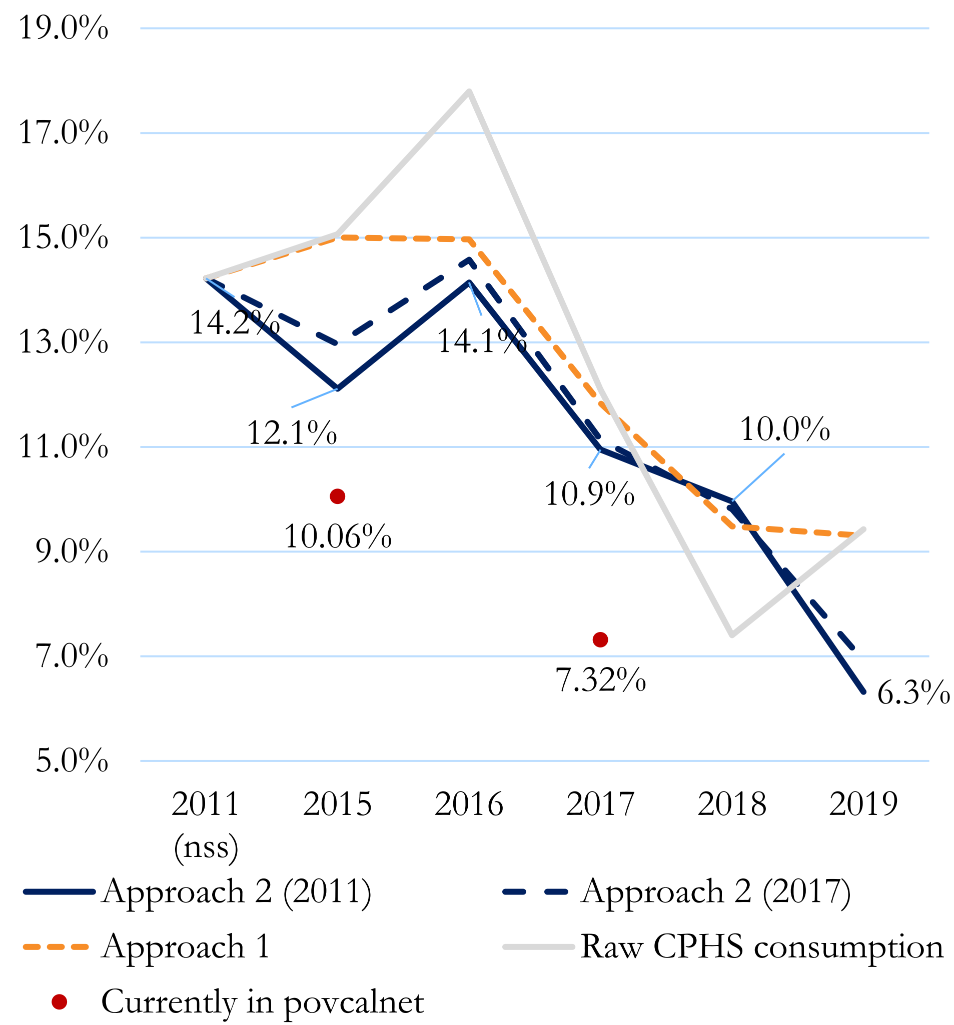
Estimates of poverty headcount at the $1.90 line



Faster poverty reduction in rural areas

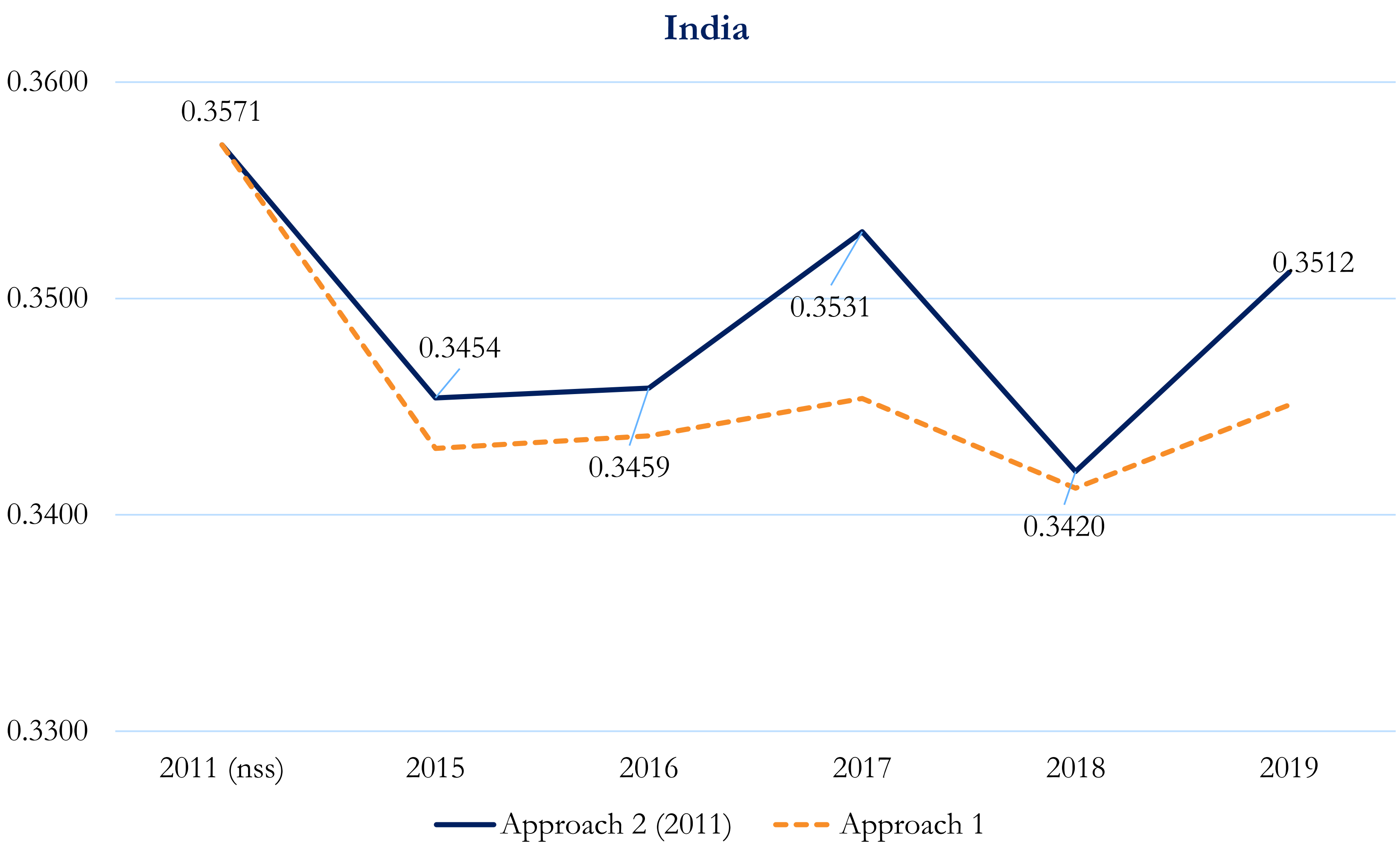
Estimates of poverty headcount at the $1.90 line





Moderation in Inequality since 2011

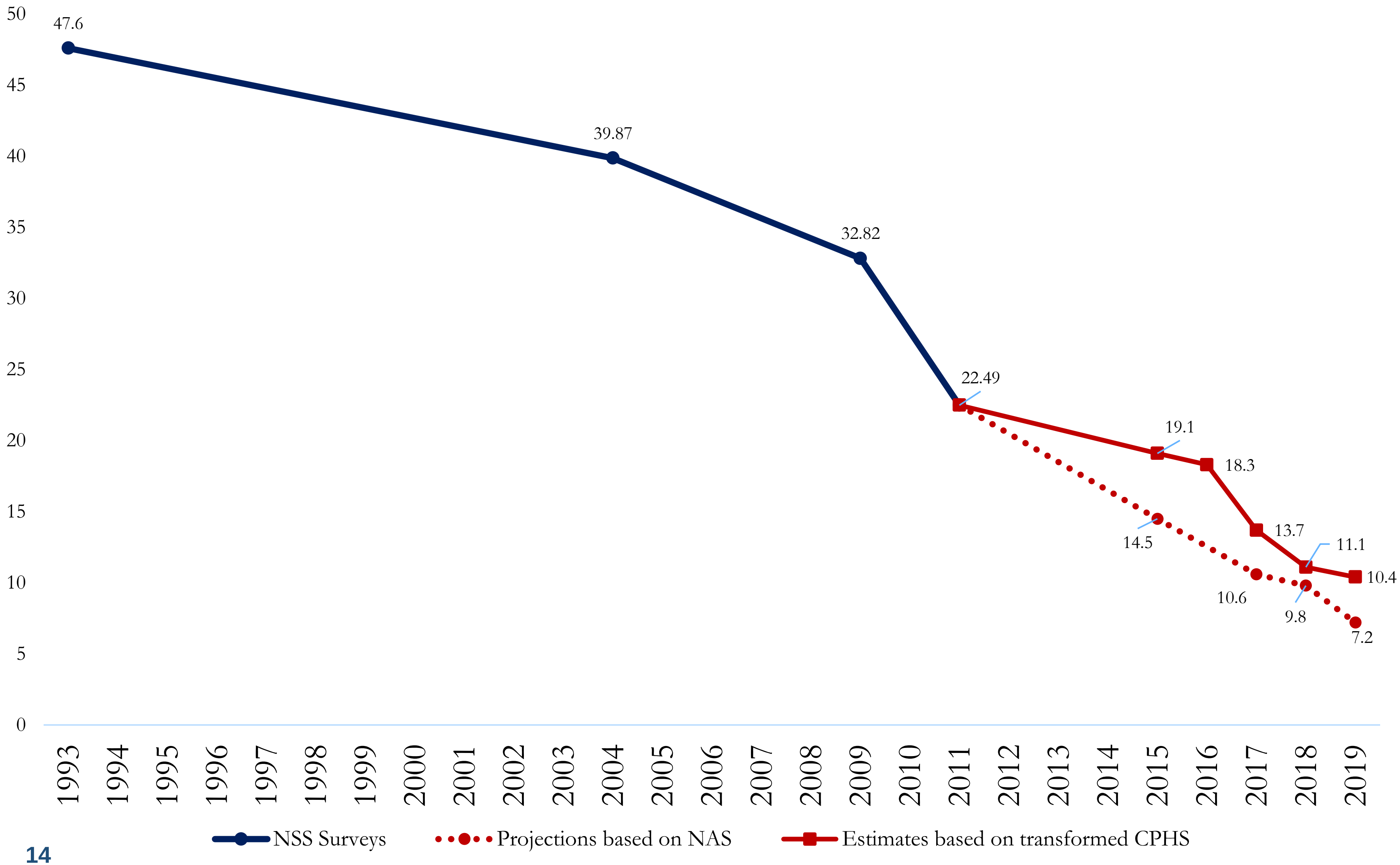
Gini measures of inequality



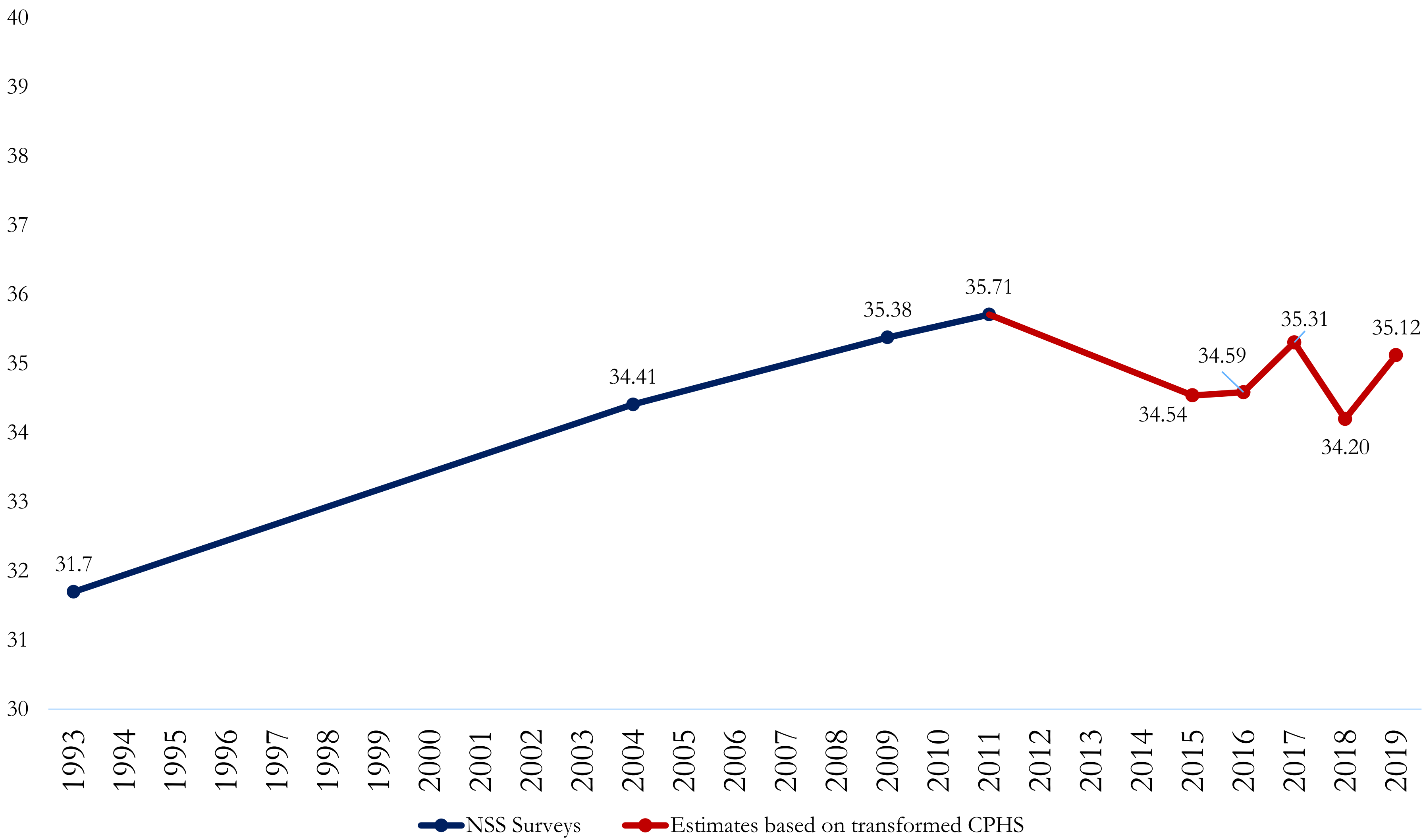
Disclaimer slide regarding inequality

- 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



Under-coverage of poor households

- Dreze and Somanchi (2023) takes a critical view of the reweighting approach as a means of restoring representativeness in the CPHS
- DS evaluate the effectiveness of reweighting by means of a simulation exercise using the government's Periodic Labor Force Surveys (PLFS)
- Their simulation experiment systematically drops poor households from the PLFS to simulate a CPHS-styled survey that underestimates poverty
- It is not clear however through what mechanisms poor households may have been excluded in the CPHS
  - The mechanism matters

A diagnostic tool

Assumption 3 *After reweighting, the truncated sample mean of* $\beta^T x_{ah}$ *matches the population mean:*

$$E_a[\beta^T x_{ah} | x_{ah} \in S] = \mu_a. \quad (11)$$

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

Assumption 4 *The error term* ε_{ah} *has mean zero, is normally distributed, and is orthogonal to* x_{ah}*.*

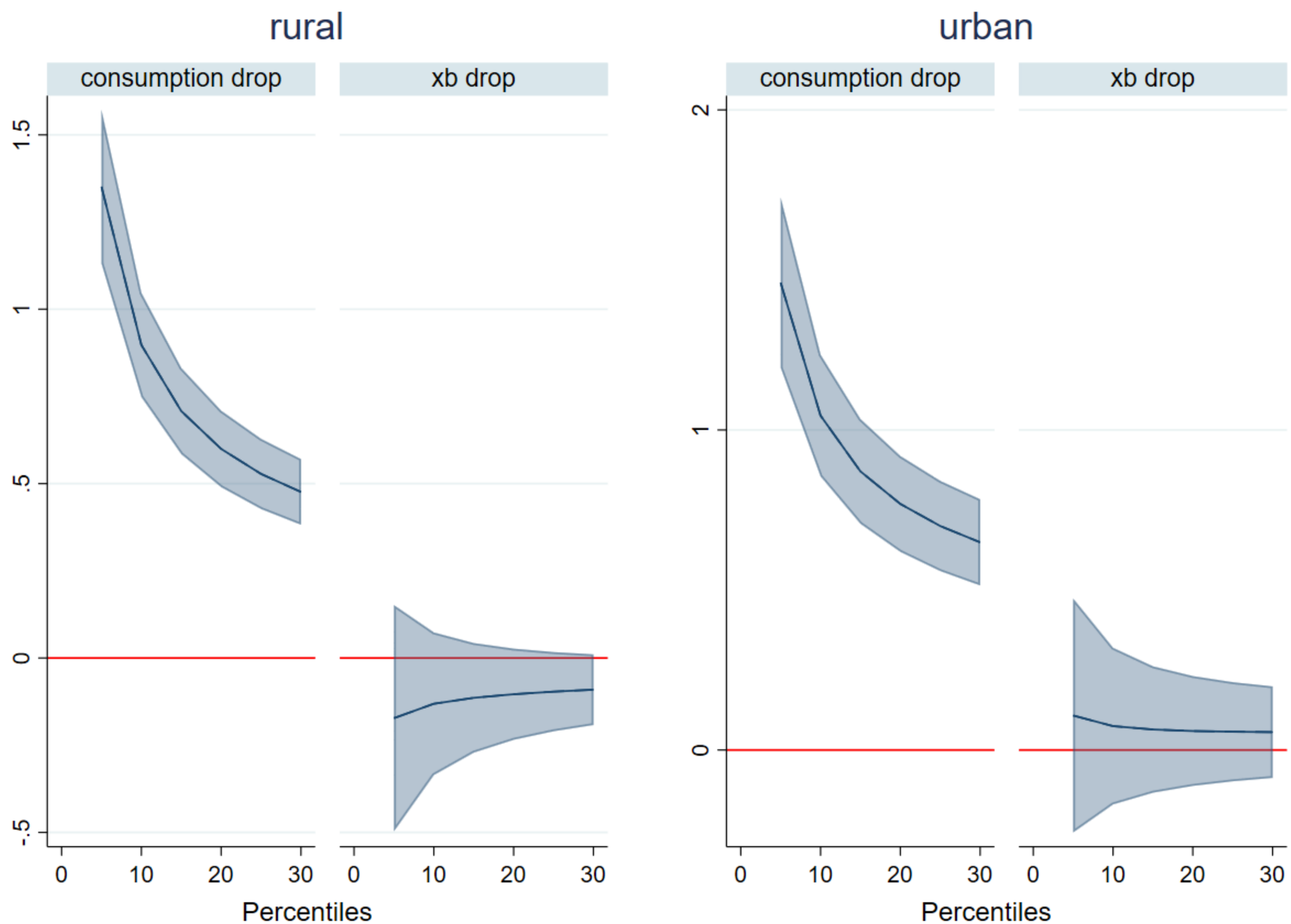
A diagnostic tool

$$
\begin{aligned}
E_{ah}[y_{ah}|y_{ah} > \tau] &= E_a[\mu_{ah}] + \sigma_a E_a\left[\frac{\phi((\tau - \mu_{ah})/\sigma_a)}{1 - \Phi((\tau - \mu_{ah})/\sigma_a)}\right] \\
&= \mu_a + \sigma_a E_a\left[\frac{\phi((\tau - \mu_{ah})/\sigma_a)}{1 - \Phi((\tau - \mu_{ah})/\sigma_a)}\right].
\end{aligned}
$$

Testing the diagnostic through a simulation study

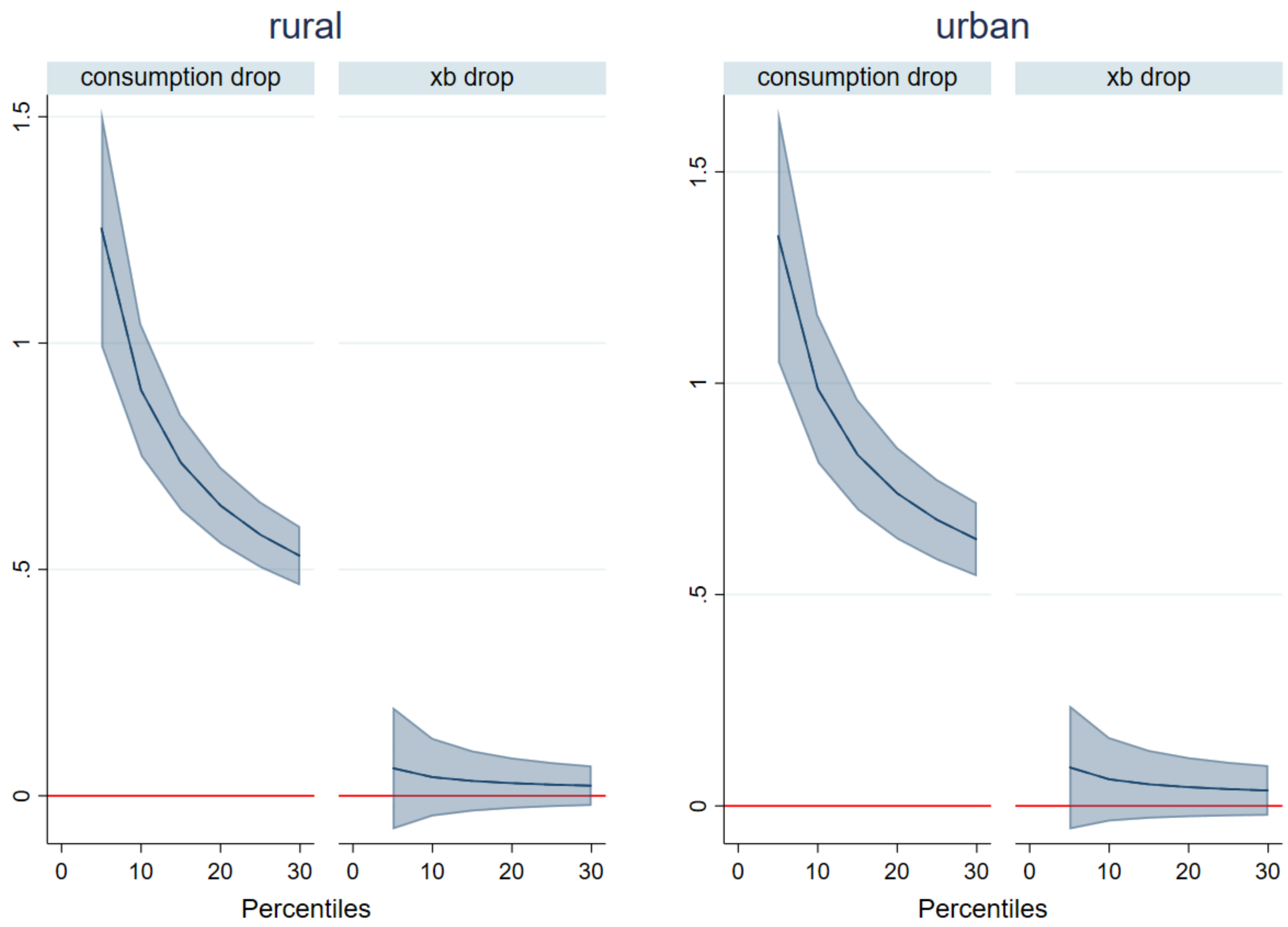
- Using the NSS-2011, we drop poor households either based on their observed consumption levels or correlates of consumption x
- Three scenarios are considered for the simulation of consumption data (which introduce incremental violations of Assumption 4):
  1. Errors are drawn from a normal distribution orthogonally to x
  2. Errors are drawn from empirical distribution orthogonally to x
  3. Observed household consumption data are used (in which case neither normality nor orthogonality can be guaranteed)

Scenario 1



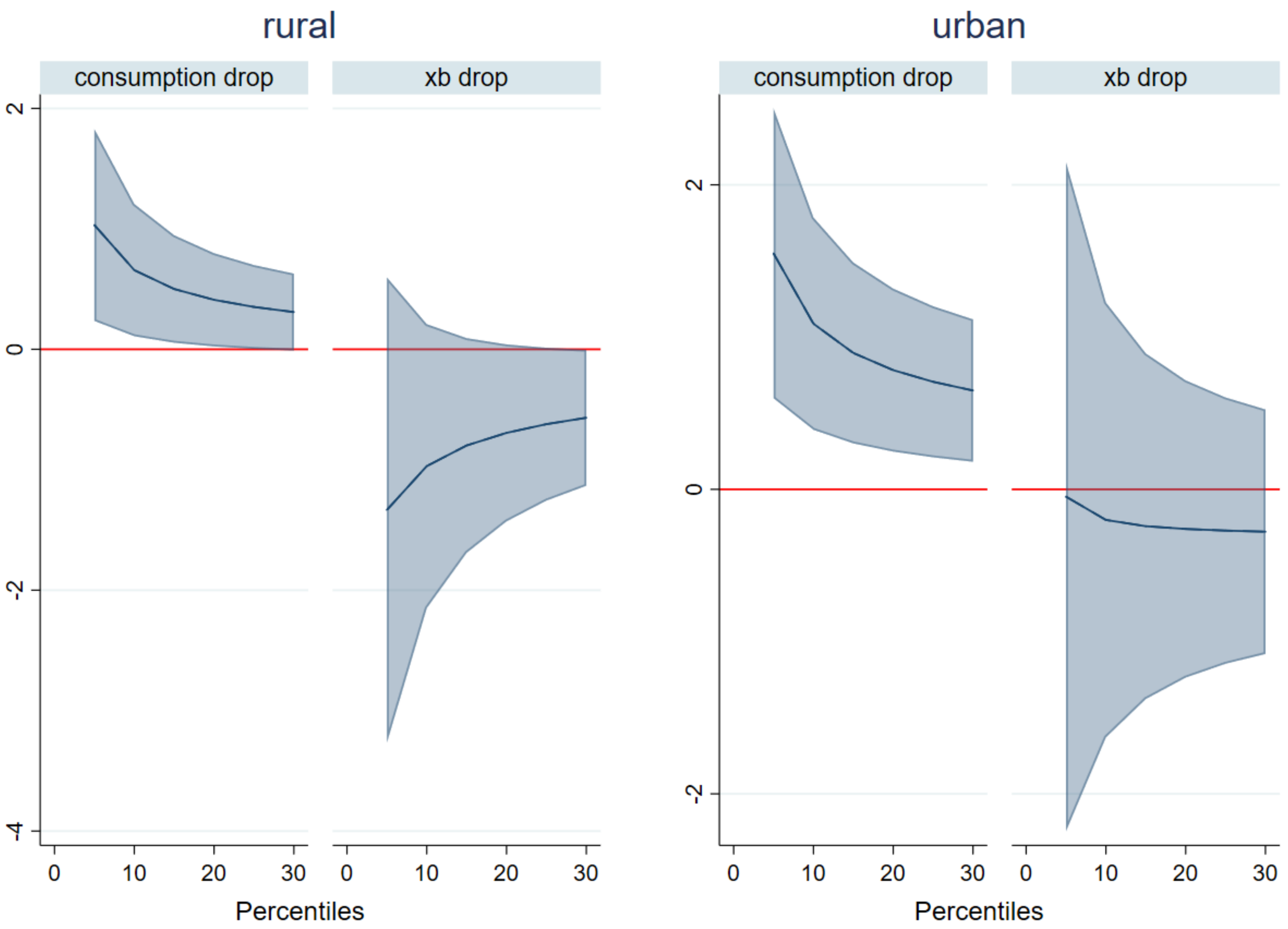
rural
consumption drop
xb drop
urban
consumption drop
xb drop
1.5
1
.5
0
-.5
2
1
0
0
10
20
30
0
10
20
30
0
10
20
30
0
10
20
30
Percentiles
Percentiles

Scenario 2



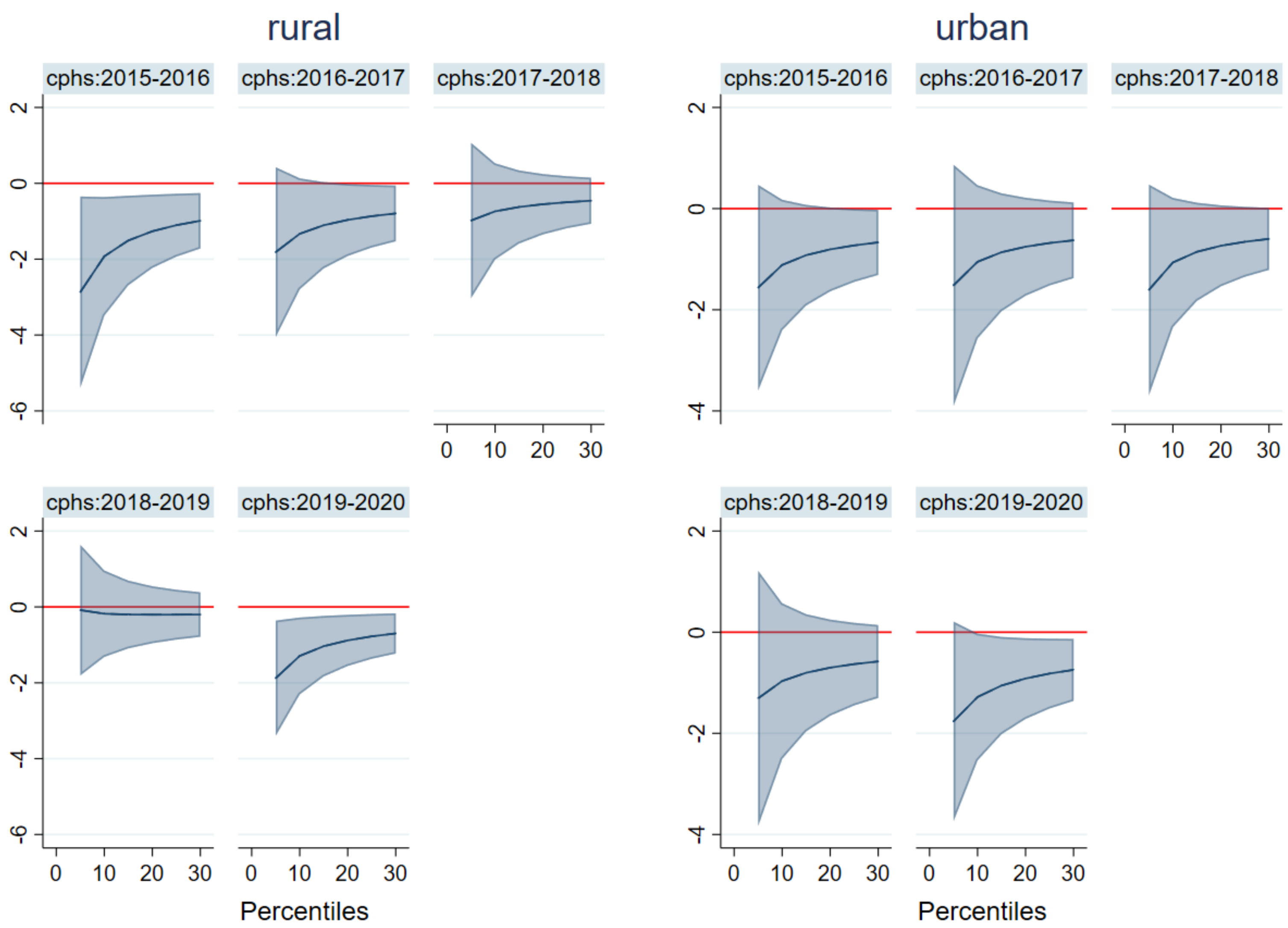
rural
urban
consumption drop
xb drop
consumption drop
xb drop
1.5
1
.5
0
0
10
20
30
Percentiles
Percentiles

Scenario 3



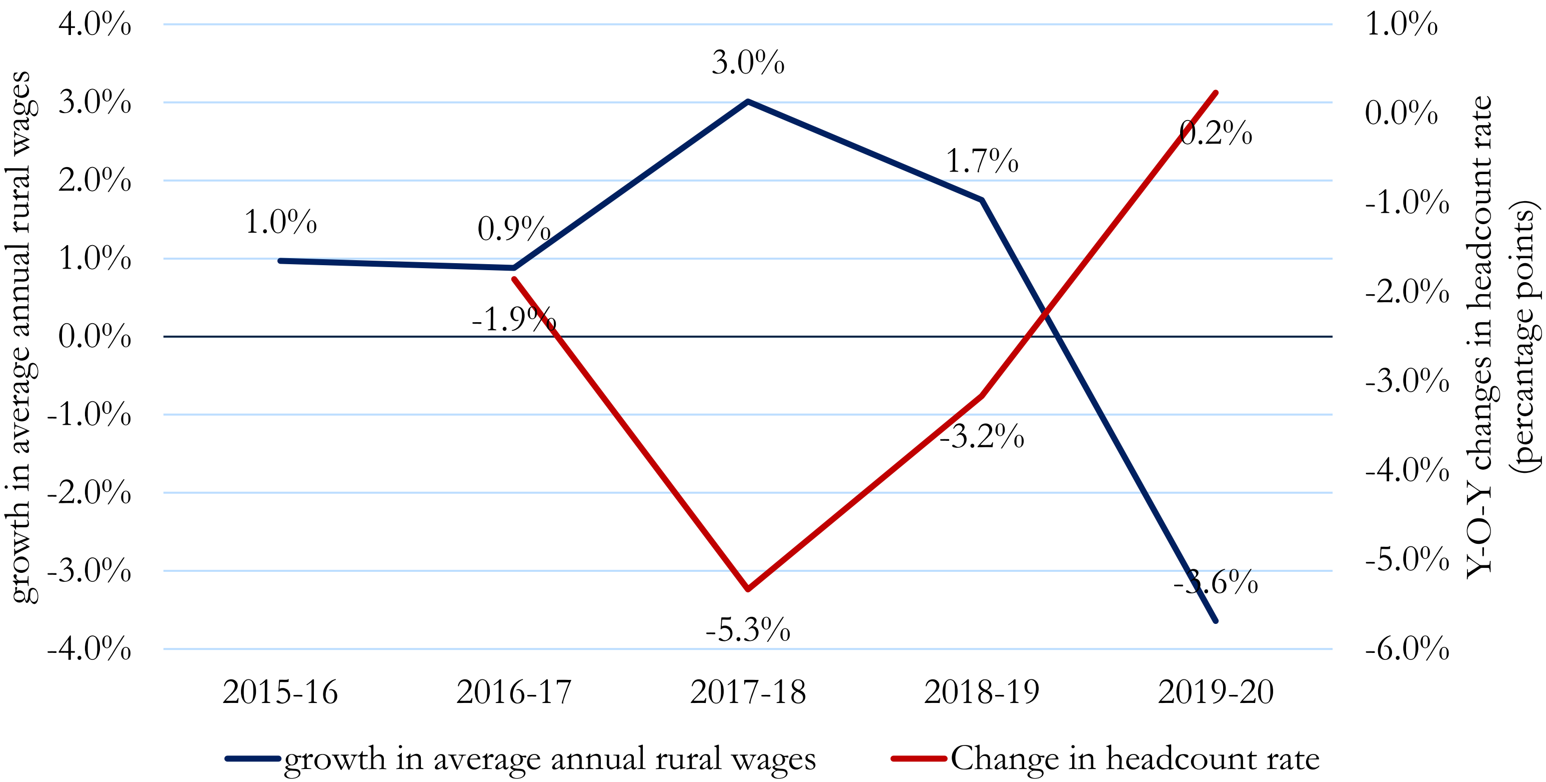
rural
urban
consumption drop
xb drop
consumption drop
xb drop
2
0
-2
-4
2
0
-2
0
10
20
30
Percentiles
Percentiles

Application of diagnostic tool to CPHS



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



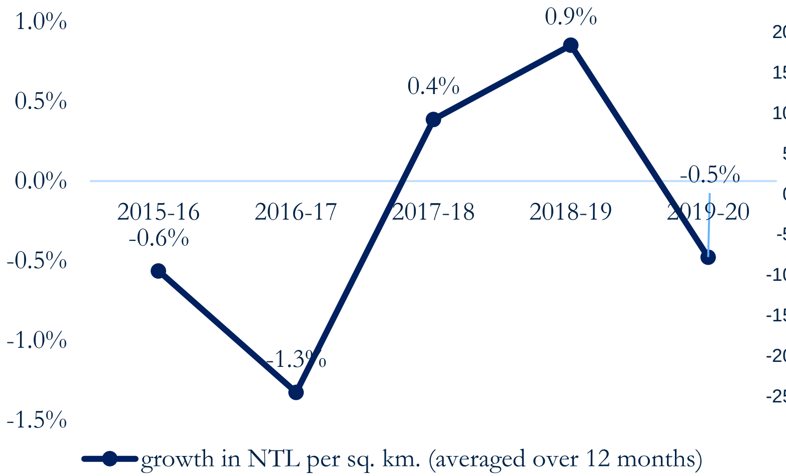
- Faster growth in real rural wages => Faster poverty reduction
- Correlation coefficient = -0.94

WORLD BANK GROUP

Validation: Should we trust these estimates?

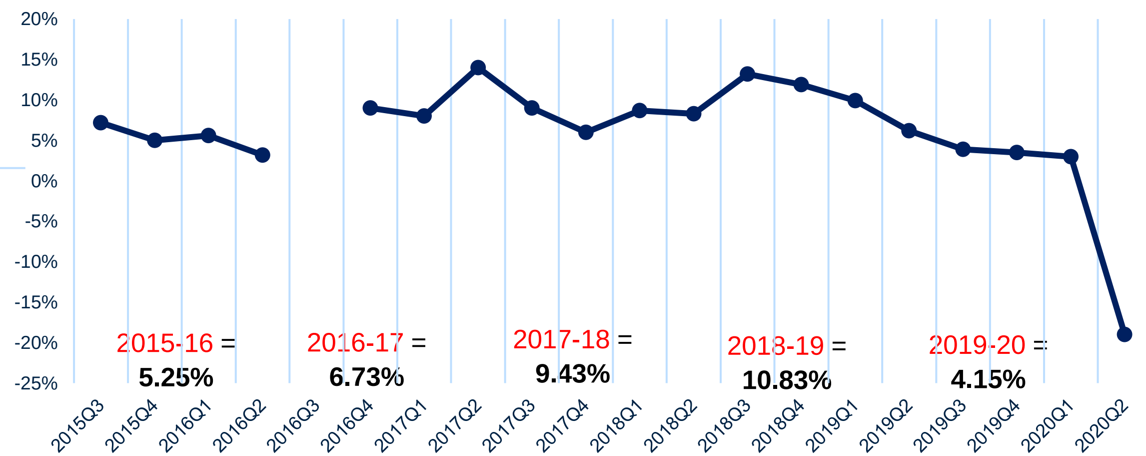
1. In the years following 2015, poverty reduction rates are highest in 2017-2018 and moderated in 2019: Supported by NTL and Nielsen surveys

Night Lights



Source: Beyer, Jain and Sinha (2020)

Nielsen retail surveys



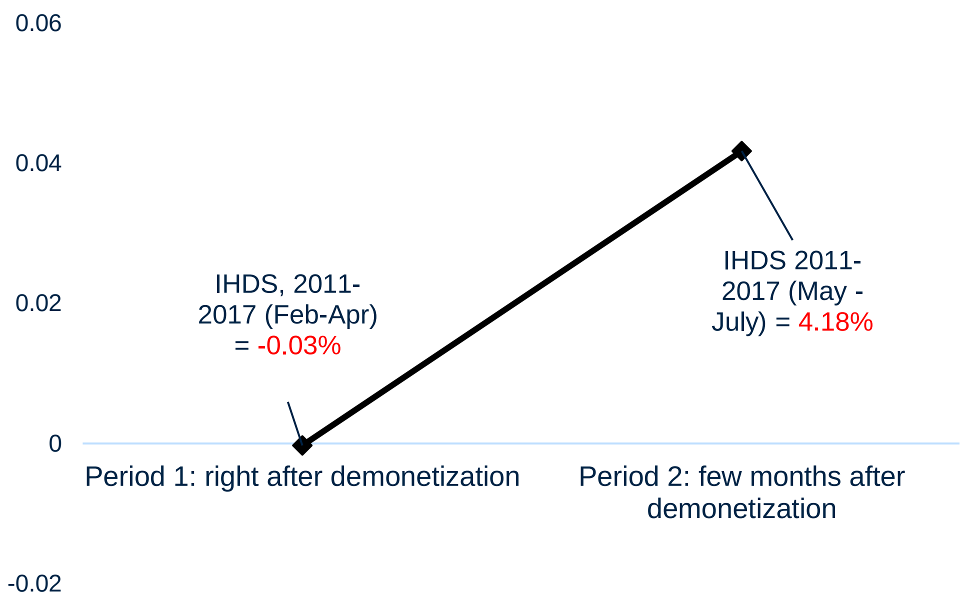
- Nightlights data supports the finding that the fastest reduction in poverty occurred in 2018-19
- Retail store surveys independently conducted by Nielsen support the same finding
- Both survey show a fall in welfare in 2019

WORLD BANK GROUP

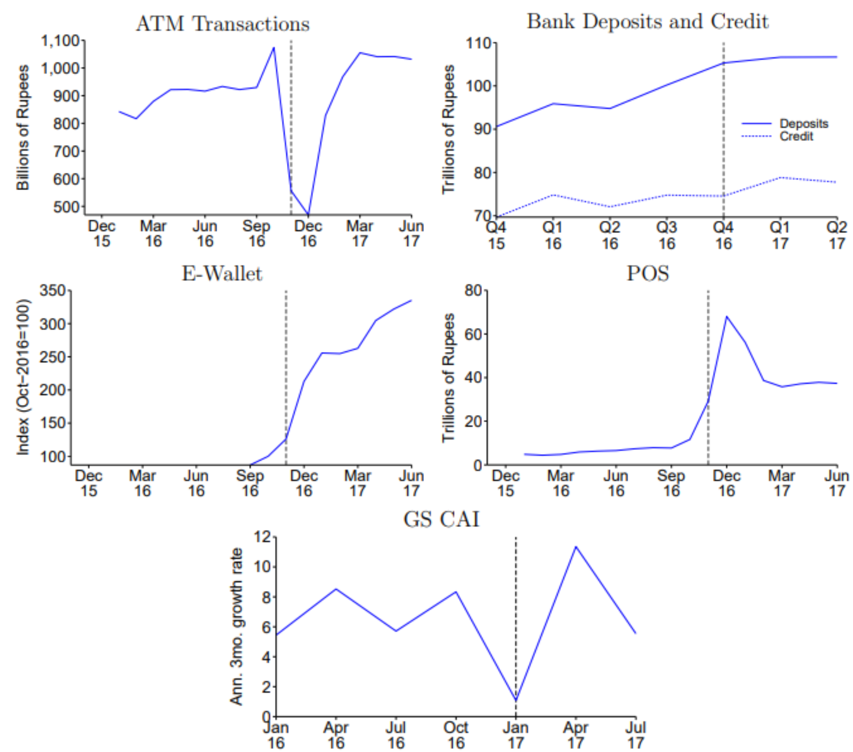
Validation: Should we trust these estimates?

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)

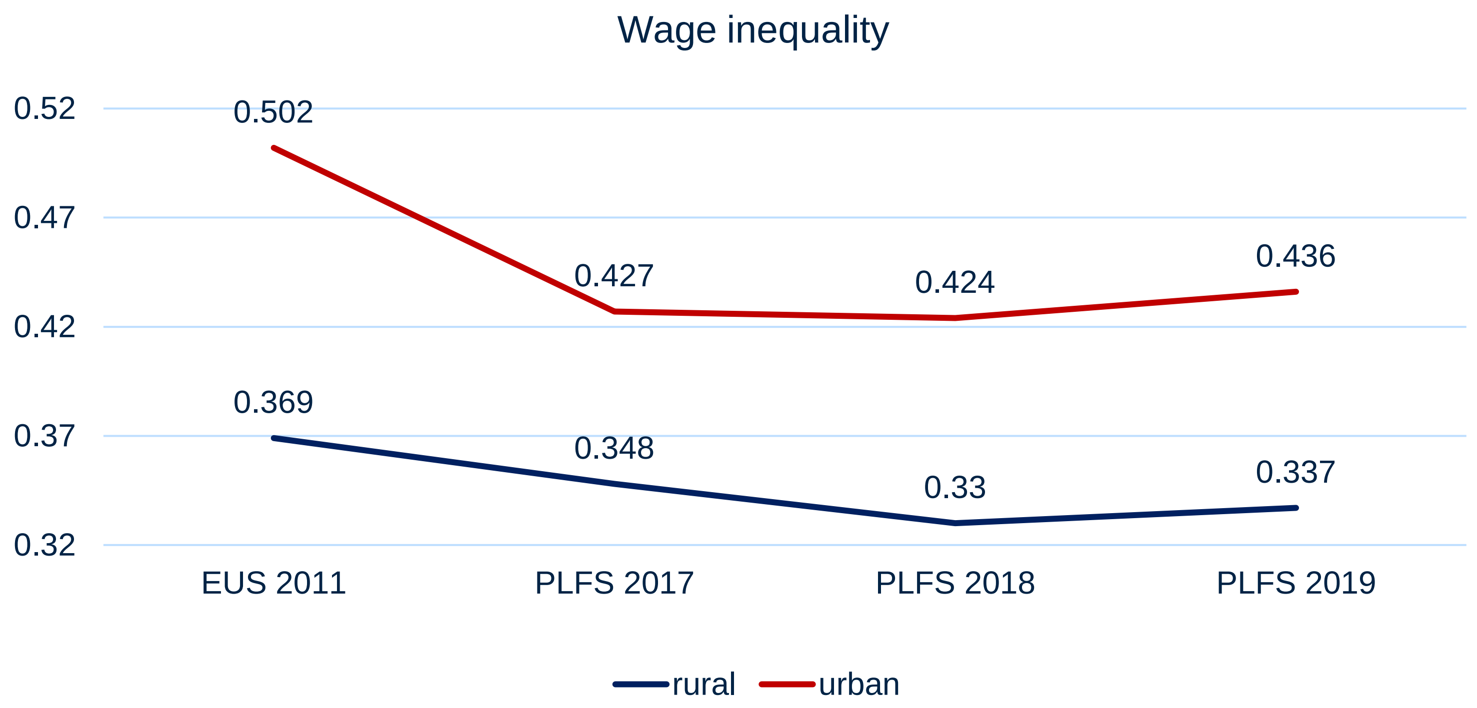


- IHDS-3 result support a drop in poverty overall and churn around demonetization (survey conducted in 3 states)
- Banking data from Chodorow-Reich, et. al (2019) also shows a temporary churn around demonetization followed by quick turnaround (GS CAI = Goldman Sachs Current Activity Indicator)

WORLD BANK GROUP

Validation: Should we trust these estimates?

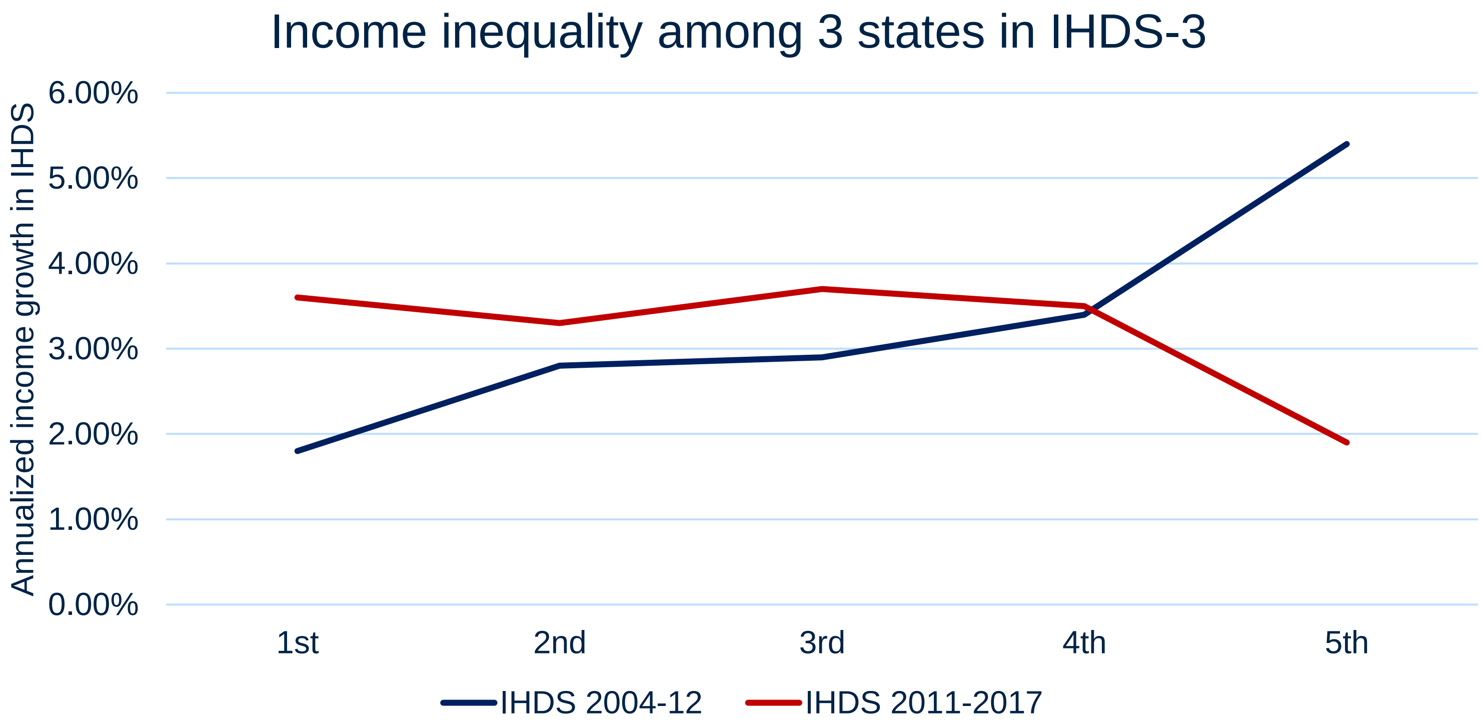
3. No rise in consumption inequality since 2011, but indications of a rise in 2019 :
Supported by fall in wage inequality observed in PLFS



- 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?

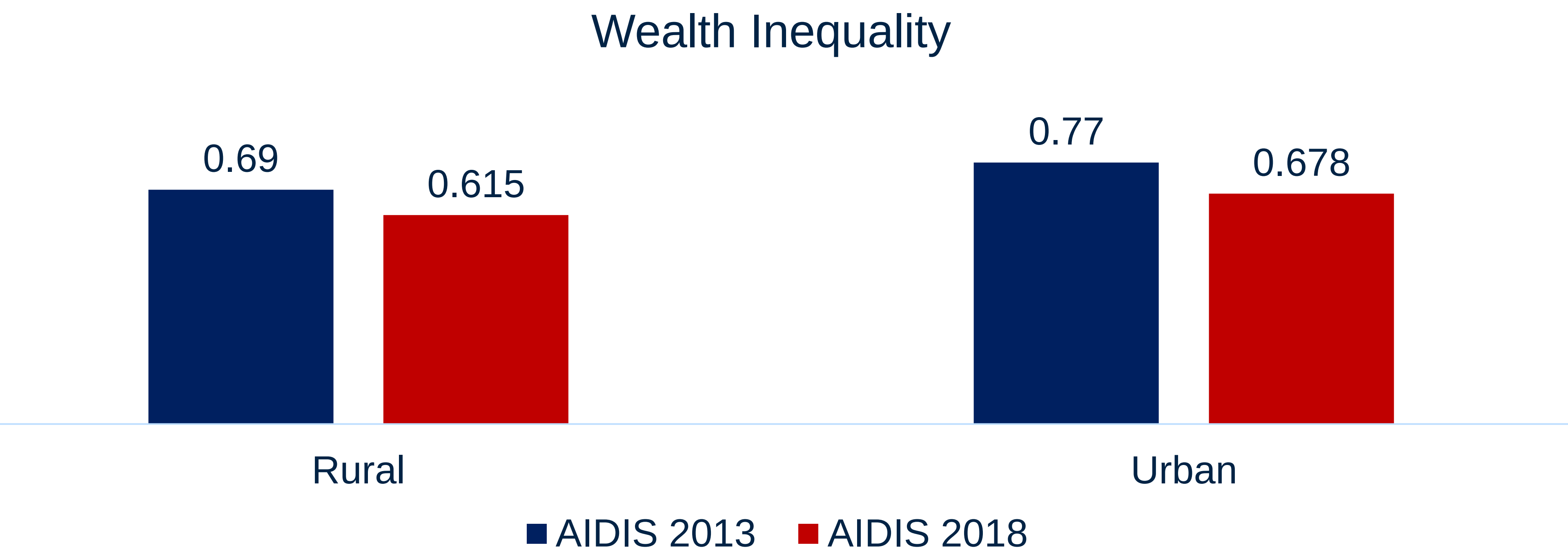
3. No rise in consumption inequality since 2011, but indications of a rise in 2019:
Supported by fall in income growth of richest households in IHDS



- 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?

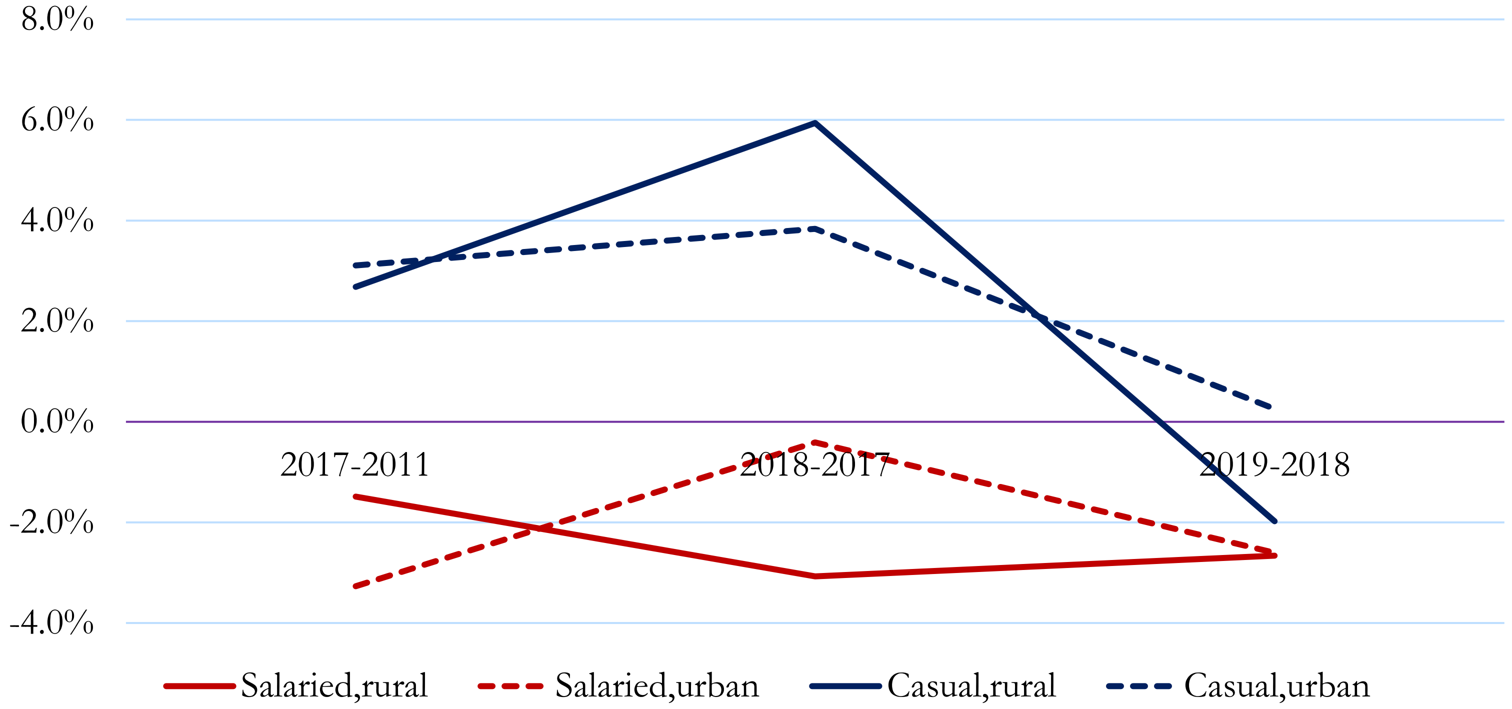
3. No rise in consumption inequality since 2011, but indications of a rise in 2019:
Supported by fall in wealth inequality Ginis



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

Validation: Should we trust these estimates?

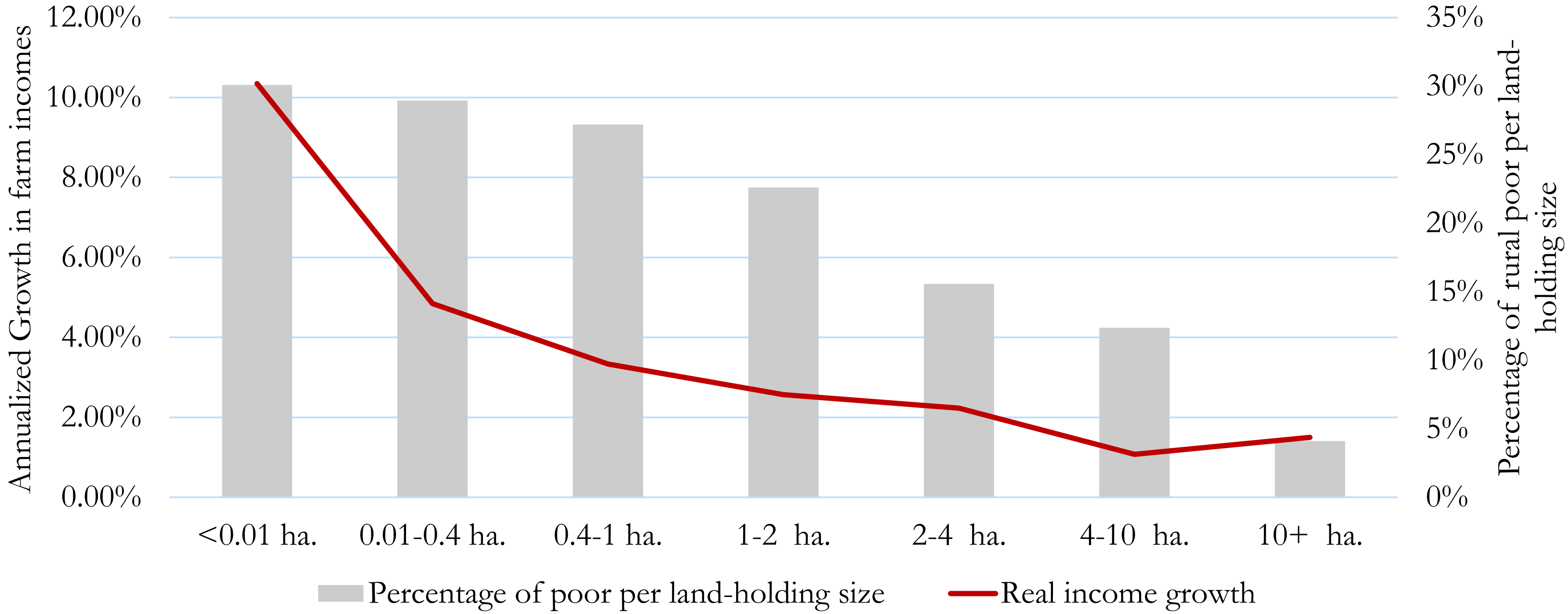
3. No rise in consumption inequality since 2011, but indications of a rise in 2019 :
Supported by positive casual wage growth but negative salaried growth



- 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 i inequality

Validation: Should we trust these estimates?

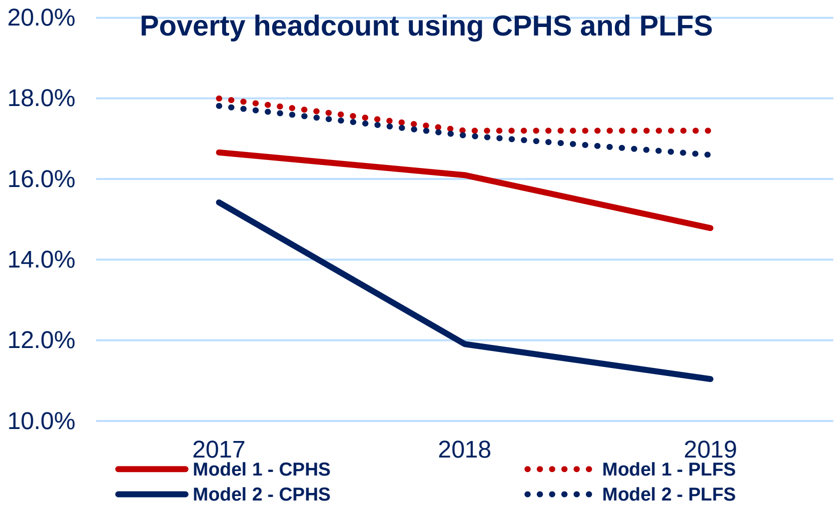
3. No rise in consumption inequality since 2011, but indications of a rise in 2019:
Supported by high income growth for agricultural HHs with smallest land holding size

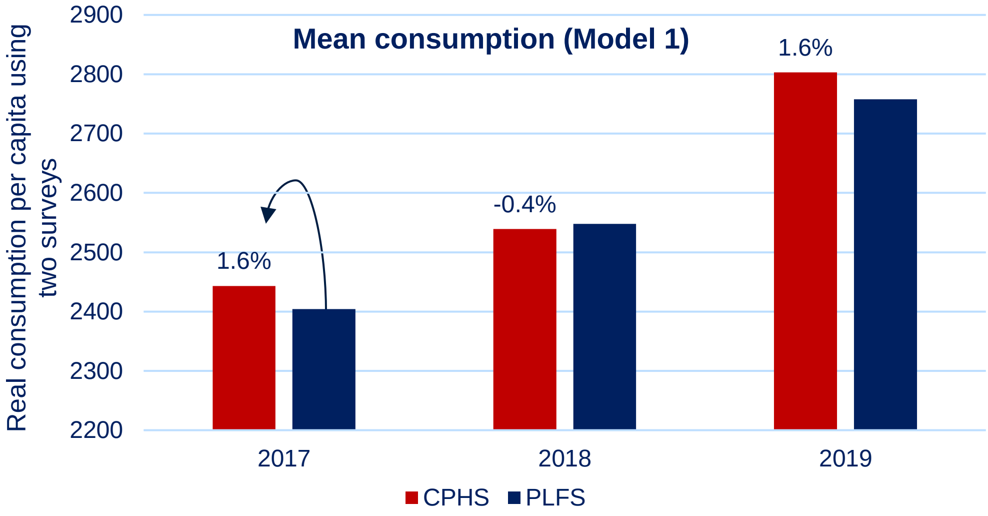


- Total income = income from wages + net receipt from crop production (out-of-pocket)+ net receipt from farming of animals (out-of-pocket) + net receipt from non-farm business income
- Deflated using CPI AL
- X-axis: size class of land possessed by agricultural households. Ha = hectares.

Validation: Should we trust these estimates?

4. Is the mean survey consumption validated? 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

Bhalla et al. (2022) estimates (IMF)

Using nominal GSDP growth rates