What Do We Know about Poverty in India in 2017/18?

Ifeanyi Nzegwu Edochie
Samuel Freije-Rodriguez
Christoph Lakner
Laura Moreno Herrera
David Locke Newhouse
Sutirtha Sinha Roy
Nishant Yonzan

WORLD BANK GROUP
Development Data Group &
Poverty and Equity Global Practice
February 2022

Abstract

This paper nowcasts poverty in India, one of the countries with the largest population below the international poverty line of $1.90 per person per day. Because the latest official household survey dates back to 2011/12, there is considerable uncertainty about recent poverty trends in the country. Applying a pass-through and survey-to-survey methodology, extreme poverty (at the $1.90 poverty line) for India in 2017 is estimated at 10.4 percent with a confidence interval of [8.1, 11.3]. The urban and rural poverty rates are estimated at 7.2 and 12.0 percent, respectively. Across a wide range of publicly available data sources, the paper finds no evidence of an increase in poverty between 2011/12 and 2017/18.

Produced by the Research Support Team

What Do We Know about Poverty in India in 2017/18?

Ifeanyi Nzegwu Edochie, Samuel Freije-Rodriguez, Christoph Lakner,

Laura Moreno Herrera, David Locke Newhouse, Sutirtha Sinha Roy,

Nishant Yonzan [*]

---

1. Introduction

This paper describes several methods to estimate poverty in India in 2017. India is likely the country with the largest number of people living below the international poverty line of $1.90 and its latest publicly available household survey dates to 2011/12, giving rise to considerable uncertainty over the recent trend in global poverty. Because of the decision by the Government of India to withhold the most recent household survey (National Sample Survey 2017/18), we use a range of methods to derive a poverty estimate for India in 2017, which can be incorporated in the global poverty counts.[1] We focus on estimating poverty at the international poverty line of $1.90 (using 2011 purchasing power parities).[2]

We use two main methodologies. The first method uses a survey-to-survey methodology to impute a consumption aggregate into the 2017/2018 Survey on Social Consumption (SCS) on Health. While this survey collects information on covariates that predict consumption, it does not collect a comprehensive consumption aggregate that could be used to measure poverty directly. Our approach is closely related to Newhouse and Vyas (2019) who impute consumption into the 2014/2015 National Sample Survey.[3] This approach builds on the small area estimation methods developed by Elbers et al. (2003), who impute a welfare aggregate into a census. More recently, Douidich et al. (2016) impute a consumption aggregate into a labor force survey to estimate quarterly poverty rates.

The second method assumes that household survey consumption follows the growth in national accounts, adjusted downward by a pass-through factor.[4] The adjustment factor accounts for the fact that survey growth is systematically lower than growth in national accounts, e.g. see Ravallion (2003), Deaton (2005), Pinkovskiy and Sala-i-Martin (2016), Lakner et al. (Forthcoming), Prydz et al. (Forthcoming). The pass-through factor is estimated using a machine-learning algorithm to account for systematic variation in pass-through rates between sub-samples of the data. We report

---

[1] The government decided to indefinitely withhold the survey citing concerns over data quality. See Jha (2019) and Press Information Bureau Government of India, Ministry of Statistics & Programme Implementation issued on November 15, 2019.

[2] We use the revised 2011 PPPs published in May 2020. Following the World Bank's global poverty measures, we use different PPPs for urban and rural areas to account for spatial price differences (Atamanov, et al. 2020). Throughout the paper urban and rural poverty are estimated separately and aggregated to the national estimate using the population weights in the World Development Indicators (WDI).

[3] Using the CES surveys collected in 2004/05, 2009/10 and 2011/12, which collect a consumption aggregate as well as covariates that are also present in the 2014/15 survey, Newhouse and Vyas (2019) estimate several models of household consumption per capita. These models are then used to project household consumption into the 2014/15 CES, which did not collect information on aggregate household consumption, and hence estimate poverty. This poverty estimate underpins the World Bank's global poverty estimate for 2015, see Chen et al. (2018) and World Bank (2018). We use a different set of variables, and different training and target data sets, but a methodology similar to Newhouse and Vyas (2019).

[4] This is similar to the way surveys are brought to a common reference year in the World Bank's global poverty measures, see Chen and Ravallion (2010), Prydz et al. (2019) and World Bank (2015).

a range of poverty estimates that reflect uncertainty in the estimated pass-through rate and the underlying national accounts growth rates, as well as allow for changes in inequality.

Under our preferred specification, using a pass-through rate of 0.67 applied to growth in Household Final Consumption Expenditure in national accounts between 2015 and 2017, we estimate a national extreme poverty rate (those living below the $1.90 poverty line) for 2017 of 10.4 percent.[5] Using a survey-to-survey estimation, the national poverty rate would be slightly smaller (9.9 percent), but its confidence interval, between [8.1, 11.3] percent, includes the estimates from the pass-through method.[6] Our estimates indicate a considerable decline in poverty since 2011/12, when poverty was estimated at 22.5 percent. Important caveats in the methodologies adopted, as well as some robustness checks to control for different assumptions, indicate that poverty rates could be higher than our preferred estimate. But we find no evidence that poverty has actually increased, or the mean declined, between 2011/12 and 2017/18, thus contradicting estimates that have been circulated in the press based on a leaked report on the 2017/18 survey (see Appendix for further details).

The paper discusses three sources of evidence about the evolution of poverty in India. Section 2 uses alternative survey data, from both public and private organizations, to provide descriptive statistics on household mean consumption. Section 3 describes the survey-to-survey imputation method, whiles section 4 describes the results from the pass-through method. Section 5 summarizes and concludes. The Appendix includes additional robustness checks and further details on the methods.

2. Available survey data for India

The Consumption Expenditure Surveys (CES) by the National Statistics Office are the main source of poverty and inequality statistics in India. These surveys have also informed the World Bank's poverty monitoring and are used to track progress towards the Sustainable Development Goal (SDG) number 1, which is focused on ending poverty. The release of the 2017/18 round of the consumption expenditure survey was eagerly anticipated, given that the last available expenditure survey dates to 2011/12. As indicated above, the government decided to withhold these data and hence we explore alternative data sources to provide updated estimates of poverty in India.

Table 1 lists several recent household surveys, all of which are nationally representative and include estimates of household consumption. As indicated above, the CES is the official source for poverty estimation. It includes around 400 questions covering expenditures on a comprehensive

---

[5] This estimate underpins the World Bank's estimate of global poverty in 2017, as reported in World Bank (2020). Also see Castañeda Aguilar et al. (2020).

[6] The interval for the pass-through method [10.0, 10.8], calculated utilizing the confidence interval of the 0.67 pass-through rate, is also within the confidence band of the survey-to-survey method.

array of goods and services.[7] The Survey on Social Consumption (SCS) on Health gathers basic information on health, and the role of public and private health providers. It started on a regular basis since 1995 and the most recent waves correspond to years 2004, 2014 and 2017/18. Similarly, the SCS on Education generates indicators on levels of education, school attendance and incentives received by students. The most recent waves were collected in 2007/08, 2014 (January to June) and 2017/18. In both SCSs, household consumption is captured through a single question on "usual monthly expenditures". Finally, the Periodic Labor Force (PLB) Survey was launched by the NSO in April 2017. This is a continuous survey that collects information about employment and unemployment. Quarterly reports are produced, and only two annual reports have been produced so far: 2017/18 and 2018/19. As in the case of the SCS, it includes a single question on household consumption expenditure.

Two surveys collected by non-government agencies are also available. The India Human Development Survey (IHDS), compiled by several independent research institutions: The National Council of Applied Economic Research (NCAER), the University of Maryland, Indiana University and the University of Michigan. It is a panel survey whose first wave was collected in 2005/06, its second in 2011/12 and the third is scheduled for 2023. In 2017, a subsample round was collected in only three states: Bihar, Rajasthan, Uttarakhand. Finally, the Consumer Pyramids (CP) data set is a continuous survey designed to measure household well-being in India, with a panel survey conducted three times per year since 2014. It is collected by the Center for Monitoring the Indian Economy (CMIE), a private data collection agency.

Using these alternative surveys, the remainder of this section reports summary statistics on recent trends in living standards.

2.1. Official data sources

The SCSs on Education and Health are nationally representative surveys with a sample size of around 65,000 households in the earlier years, and around 100,000 in 2017/18. These surveys include a question on usual household consumption that is not comparable to the more comprehensive consumption aggregates produced for poverty estimation in the CES. The SCSs on Health and Education both show higher average consumption in 2017/18 than in previous waves.[8]

Mean household consumption per capita appears larger in the CES than in SCS, for both urban and rural areas, although it is difficult to draw comparisons since the surveys were fielded in

---

[7] Differences in the recall period of these different items led to different consumption aggregates over time. The 2011/12 survey included three different definitions of the aggregate: the so-called Uniform Reference Period (URP), Mixed Reference Period (MRP) and the Modified Mixed Reference Period (MMRP). The 2017/18 survey only included the MMRP. For details on the consumption aggregates, see http://mospi.nic.in/sites/default/files/publication_reports/KI-68th-HCE.pdf. Also see discussion in the Appendix.

[8] We do not report the evolution of the consumption aggregate in the Periodic Labor Force Survey because there is no comparable survey before 2017. Comparing the PLFS (2017/18) and SCS Education (2014), Himanshu (2019) estimates that real consumption per capita declined by about 4 percent and 0.6 percent in rural and urban India, respectively.

different years and using different questionnaires to collect consumption expenditures. A direct comparison is only possible in 2004/05 (Figure 1 and Figure 2). For that year, the CES reports average consumption expenditures approximately 10 percent higher in rural areas, and 5 percent higher in urban areas, than in SCS. Average consumption does not capture all differences between the two surveys. Comparing the CES and SCS Health in 2004/05, shows that the consumption distributions in the two surveys are very close, with the SCS Health stochastically dominating the CES in the bottom of the distribution but the opposite is true above around $100 per month in 2011 PPP terms (Figure 3, top panels). The comparison of these surveys indicates that the consumption aggregate included in the CES surveys is systematically different than the consumption aggregate captured by the SCS Health and Education surveys. Hence, measures of poverty using the consumption aggregate from the SCS surveys cannot be compared to those using CES surveys.

On the other hand, SCS Health and Education surveys show higher average consumption in year 2017/18 than in previous vintages of the same survey (that is, years 2014 and 2007/08 for the Education survey, and 2014 and 2004/05 for the case of the Health survey). Going beyond the simple averages, a stochastic dominance analysis shows that the distribution of household consumption expenditures of the SCS Health 2017/18 survey is to the right of the distribution of the 2004/05 health survey (Figure 3, middle panels) and the same with respect to the 2014/15 health survey (Figure 3, bottom panels), for any poverty line below $300 per month. This could indicate that household consumption has increased for all those households at the bottom of the distribution and hence poverty is lower in 2017/18 than in previous years.

2.2. Non-official data sources

The subsample of the IHDS survey that was collected in 2017 for three states (Bihar, Rajasthan and Uttarakhand) also shows an increase in mean consumption between 2011/12 and 2017. Real income, consumption and food expenditures grew at an annualized rate of 3.5 percent, 2.7 percent and 1.9 percent, respectively. This is indicative because, historically, growth of consumption expenditure reported in the CES has been faster than in IHDS, although average consumption is higher in IHDS than in CES. For instance, between 2004/05 and 2011/12, the mean real consumption per capita in rural India had average annual growth of 3.3 percent in CES and 2.1 percent in IHDS, as well as 3.8 and 2.9 percent, respectively, in urban areas (see Figure 1 and Figure 2).

The CP survey also shows an upward trend in average real consumption and incomes between 2014 and 2018, although it matters whether the comparison is carried out relative to 2014 or 2015. Comparing with respect to 2014, the growth incidence curves show positive consumption growth throughout the distribution with few exceptions (top panel of Figure 4). In contrast, if comparing with the respect to 2015, households below the 15th percentile experience a decline in real consumption in years 2016 and 2017, which then turns positive for all percentiles in 2018 (middle panel of Figure 4). This is because the bottom of the distribution grows very fast between 2014 and 2015 (bottom panel of Figure 4). The survey data collected by CMIE seems to indicate a

worsening of living conditions for the bottom 15 percent of the population in years 2016 and 2017 with respect to 2015, but improving conditions in year 2018.[9]

The unavailability of CP data from 2011/12 prevents a direct comparisons of consumption growth between CP and CES surveys.[10]

3. Survey-to-survey imputation

As described in the previous section, none of the alternative surveys are fully comparable to the CES of 2011/12. The IHDS uses the same measure of consumption as the official surveys but is not nationally representative in recent years. The SCSs are nationally representative and cover a long period but use a different welfare aggregate. The PLB and CP surveys measure a different welfare aggregate and cover a shorter period, preventing a meaningful assessment of the trend in poverty since 2011/12.

In the absence of a comprehensive welfare aggregate covering the period after 2011/12, we use the survey-to-survey imputation methodology originally proposed by Elbers et al. (2003). We closely follow Newhouse and Vyas (2019), who apply this method to India over an earlier period. This method consists of imputing consumption into a survey without consumption data, based on the relationship between consumption and other household characteristics from a survey with consumption data. With the imputed consumption expenditure in the target survey, it is then possible to estimate poverty. A prerequisite for this method is that the two surveys involved in the exercise have a comparable set of explanatory variables. Here we use the Health SCS 2017/18 that includes a series of demographic, economic and locational characteristics that are also included in the previous rounds of the CES. A comparison of the available CES and SCS Health surveys is included in the Appendix.

3.1. Empirical Methodology

This method predicts the conditional distribution of per capita expenditure, y_{ch}, for household, h, within cluster, c, of the target data set that is missing actual consumption data (in our case the SCS Health 2017/18). The model is estimated in two steps. The first step is to develop an empirical model that predicts the log of per capita household consumption, $\ln(y_{ch})$ from the source (or training) data set, the CES 2011/12 in this case. We adopt a log linear specification relating per capita consumption expenditure to household and district level variables as follows:

---

[9] The underlying causes of this evolution are still subject to study. Regarding changes in inequality, Chodrow-Reich et al. (2020) and Chanda and Cook (2019), find a negative short-term impact of the demonetization introduced in November 2016 among the poorest groups, which then dissipates after several months.

[10] The urban to rural population in CP's sample is distributed by a ratio of 7 to 3; in contrast, India's aggregate urban to rural population is distributed by a ratio of 3 to 7. The estimates of consumption reported in this paper are weighted to correct for the oversampling. Moreover, we exclude expenditures on monthly installments, premiums and pocket monies from CP's consumption aggregate in order to make it as close as possible to CES' basket of items.

$$\ln(y_{ch}) = E[\ln(y_{ch})|x_{ch}^T] + \mu_{ch} = x_{ch}^T\beta + \mu_{ch} \tag{1}$$

where the error term μ follows a normal distribution with mean zero and constant variance, $\mu \sim N(0, \sigma^2)$. This assumption is later relaxed. The set of possible explanatory variables are those common to both training and target data sources, as in Table A.1 and Table A.2 in the Appendix.

We deviate from Newhouse and Vyas (2019) by only including the most recent CES round (2011/12) as training data and excluding the previous rounds in 2004/05 and 2009/10. That paper showed that including a linear time trend substantially improved the accuracy of the prediction when predicting poverty rates in 2004/05 using data from 2009/10 and 2011/12. This suggested that a linear time trend would also give accurate estimates for a projection three years ahead, from 2011/12 to 2014/15. However, validation tests undertaken with the data used in this paper indicated that including a linear time trend, in a model estimated using data from 2009/10 and 2011/12, greatly overpredicted poverty in 2004/05. This is due to a key difference between the data used in this paper and the one used by Newhouse and Vyas (2019), namely the availability of data on some service expenditure items in the latter (see Appendix for further details). Because the real value of these expenditures grew substantially over time, they moderated the estimated impact of the time trend variable and generated a more accurate back-cast of poverty in 2004/05. Because the data considered in this study do not contain data on any expenditure items, relying on a linear time trend to nowcast poverty becomes riskier. This issue is exacerbated by the fact that the prediction from 2011/12 to 2017/18 spans seven years, which is much longer than the three-year gap when projecting from 2011/12 to 2014/15. We therefore assume that the coefficients remain unchanged between 2011/12 and 2017/18. We recognize that this likely understates the extent to which poverty has changed, because it holds the estimated coefficients from 2011 constant, including the intercept.

Similar to Newhouse and Vyas (2019), the x_{ch}^T vector in equation 1 consists of an intercept as well as household and district level demographic variables, labor market indicators as well as district-level rainfall shocks. We include several additional variables, not present in the data used by Newhouse and Vyas (2019) to compensate for the absence of the service expenditure variables in the SCS Health 2017/18. These additional variables include characteristics of the household head such as gender, marital status, and, as explained in the Appendix, the type of cooking fuel.

In order to choose the explanatory variables to be included in the x_{ch}^T vector, we consider two shrinkage or regularization methods: the least absolute shrinkage selection operator (LASSO) regression method and the Stepwise regression algorithm. Both methods reduce the number of predictors to be included in the final specification of the model, with the aim of reducing the variance of the projections at the cost of a negligible increase in the bias of the coefficients. The LASSO algorithm (Tibshirani 1996), solves the residual error minimization problem of the linear model in a manner that only a subset x_{ch}^T of all the x_{ch} potential variables are chosen in the final model used for projections. On the other hand, there are several ways to carry out stepwise regressions. The forward selection starts with no variables and tests each additional variable using

a simple OLS method while the backward elimination starts with all the candidate variables and then deletes each variable that falls below a p-value threshold. We use the backward elimination process while setting the p-value threshold to 0.05. This is chosen over the forward elimination approach because forward elimination depends on the order in which variables are chosen.[11]

Having chosen the set of candidate explanatory variables x_{ch}^T, equation 1 is originally estimated using ordinary least squares. The regressions are weighted using the sampling weights within the surveys. To allow for the possibility of intra-cluster correlations of household expenditures, the random disturbance term is defined as follows:

$$\mu_{ch} = \eta_C + \varepsilon_{ch} \tag{2}$$

where η and ε are assumed independent, uncorrelated with x_{ch}^T and as having different data generating processes. These two components of the error term are assumed to have mean zero and variances σ_η^2 and $\sigma_{\varepsilon,c}^2$, which indicates that the latter is permitted to be heteroskedastic and vary across households in a given cluster, while the former is assumed to be a constant. Clusters are defined as districts, the lowest level of spatial disaggregation that can be matched between CES 2011/12 and SCS Health 2017/18.[12] Our approach allows for the possibility of normal or non-normal heteroskedastic error terms. The variance-covariance matrix of the error term is computed using the methods described in Nguyen et al. (2018).

Given the structure of the errors in equation 2, an OLS estimation of model 1 would underestimate uncertainty. Therefore, in the second step, the model is re-estimated using Generalized Least Squares (GLS) to control for the heterogeneity in the cluster specific errors, so:

$$\ln(y_{ch}) = x_{ch}^T \beta_{GLS} + \mu_{ch} \tag{3}$$

where $\beta \sim N\left(\hat{\beta}_{GLS}, Var(\hat{\beta}_{GLS})\right)$; $\hat{\beta}_{GLS} = \left(x_{ch}^T \hat{\Omega}^{-1} x_{ch}^T\right)^{-1} \left(x_{ch}^T \hat{\Omega}^{-1} \ln(y_{ch})\right)$.

Using a Monte Carlo approach, 100 samples from the training data are drawn to obtain 100 values of the coefficients $\hat{\beta}_{GLS}$ and of the error components $\hat{\eta}_C$ and $\hat{\varepsilon}_{ch}$ (the latter based on assumptions about their distribution and estimates of their variances $\hat{\sigma}_\eta^2$ and $\hat{\sigma}_{\varepsilon,c}^2$ from previous stages).[13] Using these estimates and explanatory variables from the target survey, x_{ch}^R, we obtain 100 imputed values of per capita household consumption for household h in cluster c:

$$\widehat{\ln(y_{ch})} = x_{ch}^R \hat{\beta}_{GLS} + \hat{\eta}_C + \hat{\varepsilon}_{ch} \tag{4}$$

---

[11] For an introduction to variable selection and regularization methods in general, and of the LASSO and Stepwise selection methods in particular, see chapter 6 of James et al. (2013).

[12] Having a smaller number of clusters reduces the likelihood of heteroskedasticity in the cluster component of μ_{ch}.

[13] See Nguyen et al. (2018) and Newhouse and Vyas (2019) for more details on the distributional assumptions.

Poverty rates are calculated for each of the 100 imputations and then averaged across imputations. The standard errors of the poverty estimates are computed following Rubin (2004). All estimates are carried out using version 2 of the Stata SAE package, developed by Nguyen et al. (2018).

In summary, we apply parameters from a model derived using CES 2011/12 to data from the SCS Health survey for 2017/18 to predict Indian poverty rates in 2017/18. We test the robustness of the model specification by varying the variable selection algorithm and the functional form of the rainfall shocks.[14] We test two different functional forms for the rainfall shock which is defined as the quarterly deviation of each district's rainfall from the historical average (between 1981 and 2018). The first functional form of this variable uses the shock and its square. The alternative specification is a simple linear regression on a spline variable created at the 25th, 50th and 75th percentile points of the rainfall shock distribution (i.e. a dummy variable that indicates whether the household lives in a district where the rainfall shock falls in any of the four quartiles of the distribution of rainfall shocks). As previously mentioned, the framework may assume normality or allow for non-normality in the error terms. Our analysis allows for non-normality which is more flexible. All models are estimated for rural and urban areas separately. We run four model specifications, two using LASSO and two using Stepwise selection, where rainfall is specified either as a spline or a quadratic function.

The consumption models explain between 34 and 45 percent of the variance of the dependent variable, which is slightly lower than Newhouse and Vyas (2019). The explanatory variables vary across models because of the use of different variable selection algorithms (i.e. LASSO and stepwise), but sign and significance of the demographic variables do not vary notably across specifications. A full description of the econometric results of these four specifications is shown in the Appendix.

3.2. Poverty Imputation Results

We present the poverty rates that result from the imputation exercise explained in the previous section, and from equation 4 above, in Table 2. The poverty rates from the imputed consumption do not vary significantly across models. In fact, the confidence intervals overlap for all models, in national, urban and rural estimates. The point estimates for the national poverty rate in 2017/18 range from 8.47 percent in model 4 to 8.75 percent in model 2. Point estimates for rural poverty vary from 8.38 percent in model 3 to 9.14 percent in model 4, while urban poverty rates are between 6.85 percent in model 4 and 9.18 percent in model 3. There is no a-priori reason to prefer one model to another, although it seems unlikely that poverty rates in urban and rural India have equaled -as in model 1- or even reversed -as in model 3. Hence models 2 and 4 seem more plausible, which result in poverty being higher in rural than urban areas. In this section, we report

---

[14] Rainfall shocks are the most important predictor of the change in household welfare in Newhouse and Vyas (2019).

further robustness checks to select a preferred model and argue that none of these models is completely satisfactory.

Validation Checks

To validate the results of the survey-to-survey imputation, we use the CES 2011/12 as training data to project poverty rates backward and compare them against the poverty rates observed in Health SCS 2014/15, and CES 2009/10 and CES 2004/05.[15] Table 3 compares the poverty estimates observed in CES 2009/10 and CES 2004/05 to predicted poverty rates based on the consumption model estimated on the CES 2011/12 as training data and CES 2009/10 and CES 2004/05 as target data. In both cases, predicted poverty based on our model is considerably lower than observed poverty. In 2009/10, our estimates hardly vary across models with national poverty rates between 17.23 and 17.65 percent, although the differences are somewhat larger for urban areas (13.04 to 15.65 percent). In all cases, predicted poverty rates are substantially lower than the poverty rates observed in the 2009/2010 survey (31.7 percent nationally) (see Table 3 middle panel). In 2004/05, the predicted national poverty rates range from 28.40 percent to 30.38 percent, which are up to 10 percentage points lower than the observed poverty rate (38.9 percent). The difference is wider in rural areas, whereas some of our estimates for urban areas overlap with the 95 percent confidence interval of the observed poverty rates (see Table 3 top panel).

In 2014/15, we compare the poverty rates that we predict using our four models against the poverty rates predicted by Newhouse and Vyas (2019) (see Table 3 bottom panel). Since the CES 2014/15 does not include actual consumption data, we rely on the estimates by Newhouse and Vyas (2019). In our prediction, we use the Health SCS 2014/15 applied to a consumption model estimated over CES 2011/12. In this comparison, we thus compare predictions across different surveys, while the earlier backcasts compared actual and predicted poverty in the same survey (various rounds of the CES). The 2014/15 CES and SCS, both official nationally representative surveys, show broadly similar socio-economic indicators (Tables 2 and 3). Across all four models, our national poverty estimates (between 16.8 and 20.93 percent) are consistently higher than the estimates from Newhouse and Vyas (2019) (14.6 percent), although the confidence intervals for our estimates in models 2 and 4 would include Newhouse and Vyas' estimates. These differences are mostly driven by rural areas, while for urban areas models 1, 2 and 4 are not very different from Newhouse and Vyas (2019). There is thus an interesting contrast between the three validation tests: While our model underpredicts poverty in 2004/05 and 2009/10, it overpredicts in 2014/15. Again, this is likely because coefficients estimated using 2011 data are applied to earlier data, while in reality the coefficients may vary over time.

---

[15] For an overview of these kinds of validation methods, see James et al. (2013). Similarly, microsimulation exercises are validated by "back-casting" reforms that occurred in the past, for example see Figari et al. (2015).

The inability of any of our models to replicate poverty rates in years 2004/05, 2009/10 and 2014/15 raises concerns whether this method correctly forecasts poverty in 2017/18.[16] This contrast with Newhouse and Vyas (2019) who validate their model by replicating poverty rates in rural areas in 2004/05.[17] Given these limitations, we explore another method to project poverty rates for India in the absence of survey data.

4. Pass-through method

While we should expect that the growth of consumption in national accounts is positively correlated with growth of mean consumption measured from surveys, a substantial literature has found that growth in national accounts does not pass-through one-to-one to household surveys. The difference between these two growth rates is referred to as a pass-through rate. More precisely, and following Ravallion (2003), the pass-through rate is the coefficient estimate β in the regression of the growth in the survey mean, g_{survey}, on the growth in national accounts, g_{NAS}:

$$g_{survey,i} = \beta * g_{NAS,i} + u_i \qquad (5),$$

where i is a growth spell between two survey years, and the residual, u_i, has mean zero. g_{survey} is the growth rate of the survey welfare aggregate, here measured as household income or consumption expenditure per capita. g_{NAS} is the real growth rate in national accounts, for which we consider either Gross Domestic Product (GDP) per capita or Household Final Consumption Expenditure (HFCE) per capita. The pass-through rate, β, captures the rate of growth in consumption that is passed through from national accounts to surveys. If $\beta = 1$, then mean consumption in the survey grows at the same rate as consumption in national accounts. Typically, the literature finds $\beta < 1$ (see more details below), which implies that mean consumption in the survey grows slower than the growth in national accounts.

The literature has discussed several channels for why there are systematic differences in growth of consumption as measured in national accounts and in the survey microdata; for example, see Ravallion (2003), Deaton (2005) and Pinkovskiy and Sala-i-Martin (2016).[18] First, there are

---

[16] The S2S method has been used to generate plausible poverty estimates in other contexts and has often been successfully validated. Any imputation model is forced to make strong assumptions, such as the stability of the consumption model over six years (from 2011/12 to 2017/18) in our case. The availability of additional variables could also improve the predictive performance of our model. In some specifications, we successfully validate the 2014/15 estimates, as well as the 2004/05 estimates in urban areas, but do not accurately backcast the large decline in rural poverty between 2011/12 and 2004/05. This is likely because the estimated parameters, including the intercept term, is fixed at the estimated 2011/12 levels. Including a linear time trend in the model would greatly overestimate the decline in poverty between 2004/05 and 2011/12 in both urban and rural areas. This suggests that incorporating data on expenditures of particular services, which was possible when projecting into 2014/15 (Newhouse and Vyas 2019) but not into 2017/18, helps the linear time trend model perform much better at backcasting.

[17] They fail to replicate poverty rates for urban areas in 2004/05, and make no reference to attempts to replicate 2009/10.

[18] For a general discussion around the differences between household survey data and the data from national accounts, as well as potential adjustments for such differences, among others see Altimir (1987), Bourguignon (2015), and Prydz et al. (Forthcoming).

11

methodological differences between how consumption is measured across the two sources of data. For example, consumption in national accounts, HFCE, is often derived as a residual from GDP. Second, even if they followed the same methodologies, the two series do not have the same scope. In addition to the consumption from households that is captured in the surveys, HFCE includes consumption from non-profit institutions (such as charities, religious organizations, trade unions, and political parties), consumption of financial service intermediaries, and imputed rents for housing (Datt and Ravallion 2002).

For India in particular, researchers have documented a third source of differences. They have noted that the gap between the two sources of data was small during the 1950s and 1960s, but the divergence between the series has grown since. For instance, Kulsehrestha and Kar (2005) note that the gap between the two sources of consumption data was 5 percent (with mean consumption from surveys being lower than that from national accounts) for fiscal year 1957/58, however, this gap had grown to 38 percent by 1993/94. The authors also note that the source of the increase in the discrepancy between consumption in national accounts and consumption in surveys are non-food items. Similarly, Mukherjee and Chatterjhee (1974) find small differences between the two sources of data in the decade leading up to 1963/64, with consumption in surveys on average lower than consumption in national accounts. They also note that the surveys record a lower share of non-food items relative to national accounts. For food items the difference in consumption between the national accounts and surveys has been relatively small (Kulsehrestha and Kar (2005), Sundaram and Tendulkar (2003), Minhas (1988)).

Following the recommendations of the UN System of National Accounts starting in 1993, the CSO in India added a new item to consumption in national accounts, financial intermediation services indirectly measured (FISIM). FISIM is a measure of the value of financial intermediation. It is calculated as the difference between the interest paid by borrowers to banks and the interest received by lenders from banks. Deaton and Kozel (2005) find that the value of FISIM in consumption in national accounts was close to zero percent in 1983/84, but its share had increased to 2.5 percent by 1993/94. They attribute a quarter of the gap in the two series to FISIM. Datt and Ravallion (2002) similarly finds that the discrepancies between national accounts and surveys increase when using the post-1993 definition of consumption in national accounts relative to the pre-1993 definition. They find that consumption in national accounts grew 0.55 percentage points faster annually than the consumption in surveys between 1972 and 1997, while using the newer series, they find a difference of 0.74 percentage points for the same period. While FISIM has added to the discrepancy in the value of consumption between surveys and national accounts, it is less likely to directly affect the living standards of the poor. Hence, when calculating welfare of the poor, it would be ideal to discount the effect of these variables which make consumption larger in national accounts relative to surveys.

4.1. Methodology

The pass-through rate is estimated using equation 5 above. Following Ravallion (2003), the regression is estimated without an intercept. Improving on the earlier literature, we only include growth spells between household survey rounds that are comparable in order to focus on real changes between two survey rounds and ignore any spurious changes. Survey comparability is assessed according to various characteristics, including the method of sampling, the questionnaire design, the methodology used in the construction of welfare aggregates and the price deflation used over time and space.[19] We use the growth in per capita HFCE as opposed to growth in per capita GDP, as HFCE aims to capture household private consumption in national accounts, and so it is, in principle, more aligned with the consumption captured in surveys than GDP (Ravallion (2003), Deaton (2005)). Similarly, to derive reference-year estimates in the absence of a new household survey, PovcalNet uses HFCE over GDP in most countries, including India (Prydz, et al. (2019)).[20]

A crucial question is how to define the relevant sample of survey growth spells. The existing literature has partitioned the sample along various dimensions, for instance by income level, level of inequality or by geographic regions. See, for example, Birdsall et al. (2014), Chen and Ravallion (2010), Chandy et al. (2013) and Corral, et al. (2020). Given that different choices of partitioning variables yield different pass-through rates, a systematic approach of partitioning the data is necessary.

To that end, we follow an approach that is identical to the machine learning algorithm used in Lakner et al. (forthcoming).[21] Here, we use per capita HFCE growth instead of per capita GDP growth as the main independent variable. The partitioning algorithm, referred to as model-based recursive partitioning (MOB), takes equation 1 as the starting point and subsequently adds various input variables interacted with the growth in per capita HFCE, g_{NAS}. Each interaction is added one at a time. For each interaction a Wald test is conducted to determine whether the coefficients on these interactions are statistically significant (at the 5 percent significance level). The input variables are geographical region, a dummy for whether consumption or income is used in the survey, mean consumption, median consumption, the Gini index, population, per capita GDP, and the year of the survey.[22] When a significant interaction is found, the sample is partitioned using that input variable as a splitting variable and the algorithm is applied on each of the sub-samples separately. Splits are only made if at least 10 observations will be in each subsample. For non-

---

[19] The precise assessment of comparability is country-dependent, compiled from the World Bank's economists who are in close dialogue with national data producers and have intimate knowledge of the survey design and methodology. More details on the comparability metadata can be found in Atamanov et al. (2019). The comparability data set can be accessed here: https://datacatalog.worldbank.org/node/506801.

[20] The exception is Sub-Saharan Africa, where GDP per capita is preferred.

[21] The algorithm is a variant of Classification and Regression Tree (CART), pioneered by Breiman, et al. (1984).

[22] We consider two regional definitions. First, the standard World Bank regions, in which all countries are classified according to geography. Second, the regions used by PovcalNet, where most high-income countries form a separate region.

binary interacting variables, all possible splits are tried out and the split with the greatest rejection of equality of the passthrough rates is chosen.

Our source for growth in per capita HFCE and per capita GDP is the World Development Indicators (WDI). We use surveys reported in PovcalNet to calculate growth in survey means. Growth between any two consecutive surveys for a country is referred to as a survey spell. Our main sample consists of a total of 1,671 survey spells if the per capita GDP aggregate is used, and 1,511 spells if per capita HFCE is used. After accounting for survey comparability, 1,429 of 1,671 spells remain when per capita GDP is used and 1,323 of 1,511 spells for per capita HFCE.

4.2. Results

Figure A.1 shows the results of the MOB algorithm. There is significant evidence in favor of the welfare measure in the survey (income or consumption) being relevant for passthrough rates. Using the per capita HFCE growth rates, the MOB algorithm suggests that no other variable besides this yields a significantly different pass-through rate.[23] Table 4 reports the pass-through estimates for each sub-sample in Figure A.1. Observations using income-based surveys have a pass-through rate of 1.00 with a 95 percent confidence interval between 0.89 and 1.12, while observations using consumption-based surveys have a passthrough rate of 0.66 with a 95 percent confidence interval between 0.58 and 0.75.[24] With a p-value of 0.024, we can reject that the coefficients are identical for these two subgroups at a 5 percent significance level.

Given the results from MOB and the fact that the CES are consumption-based surveys, we use the consumption-based partition as the sample to calculate a pass-through rate for India. Our preferred estimate of the pass-through rate thus uses the global sample of comparable consumption survey spells in combination with per capita HFCE growth. There are 471 spells in this sample, and the regression using this sample yields a pass-through rate of 0.67 with a 95 percent confidence interval between [0.59, 0.75].[25] While it is impossible to know the true pass-through rate for India over this period, 95 percent of pass-through rates using global consumption-specific survey spells with the per capita HFCE aggregate will fall within this confidence interval. As determined by the

---

[23] As noted above, Lakner et al. (Forthcoming) use per capita GDP growth as their main dependent variable (as opposed to per capita HFCE growth used here). They find that in addition to partition by data type, further partitions by median income, level of inequality, and geographic regions yield significantly different pass-through rates when using income-based surveys. However, similar to what we find, they find that within the sample of consumption surveys, there are no further significant splits.

[24] Lakner et al. (Forthcoming), using per capita GDP growth rate, calculate a slightly higher pass-through rate of 0.72 for global consumption-specific comparable survey spells.

[25] Table A.5, which is further discussed in the Appendix, compares the estimate from the global consumption-specific comparable sample with estimates using various alternative samples. The difference in the pass-through estimate between the MOB of 0.66, reported in Table 4, and the 0.67 estimate reported here and in Table A.5, is due to the difference in sample size – 457 observations in the former and 471 in the latter. The MOB is more taxing on the data as we use several variables to check for partitioning and there might be missing values in some cases. In this paper, we have used the MOB primarily as a method to confirm the partitions. In what follows, we use 0.67 with a confidence interval of [0.59, 0.75] as the pass-through rate of the preferred sample.

MOB methodology, further partitioning of the global sample into geographic regions is not necessary. When estimating the poverty rate for India, in addition to the poverty derived from the 0.67 pass-through rate, we also present a range of poverty estimates derived using the 95 percent confidence band for the 0.67 pass-through rate.

Our preferred pass-through rate of 0.67 is in line with the broader literature on India. Sen (2000) finds that the ratio of survey mean to consumption in national accounts to be between 0.6 and 0.7 for the period 1972 to 1997. For the same period, Datt and Ravallion (2002) find that the ratio between the survey mean and consumption in national accounts to be between 0.575 and 0.645 – depending upon, as discussed above, which national accounts series (new or old) one uses. Ravallion (2003) estimates this ratio for 1997 in India to be 0.55. The author also finds that, using 22 survey spells of South Asian countries in the 1980s and 1990s, $\beta=0.525$.[26] Deaton and Kozel (2005) estimate the ratio between the two sources to be "currently around" 0.667. The implied pass-through estimate for fiscal year 2014/15 using the poverty rates from the survey-to-survey estimation by Newhouse and Vyas (2018) is around 0.65.[27] The ratio of survey mean to per capita HFCE using all available CES surveys for India is 0.751, and this ratio using comparable spells is 0.765.[28]

4.3. Concerns over National Accounts growth rates

There have been recent debates around the soundness of the official GDP growth rates. A. Subramanian (2019) argues that the official GDP growth for the 2011-2016 period might have been overestimated by as much as 2.5 percentage points due to methodological changes that the Central Statistics Office (CSO) undertook in 2011. According to A. Subramanian (2019), the CSO switched the calculation of national statistics from a volume-based index to a value-based system of accounting. Value-based accounting is sensitive to price fluctuations, and hence a double price-discounting is generally recommended for these indexes. However, the CSO adopted a single deflation in prices of national statistics.[29]

---

[26] Ravallion (2003) finds a pass-through rate of 0.752 with a standard error of 0.563 when employing a regression as in equation (5) but with a non-zero intercept; the author finds a pass-through rate of 0.525 with a standard error of 0.258 on a regression setting the intercept equal to zero. For the global sample consisting of 142 spells, setting the intercept to zero the author finds a pass-through rate of 0.499.

[27] For a detailed discussion on the survey-to-survey methodology see the sub-section above. The national pass-through rate in India for the period 2012-2015 is 0.65. The 0.65 national pass-through rate is calculated as the population-weighted average of the rural pass-through rate (0.699) and urban pass-through rate (0.551). The rural and urban pass-through rates are calibrated using the poverty estimates from Newhouse and Vyas (2018) for 2014/15 (for more details, see Chen et al. (2018)).

[28] The available CES surveys for India are 1977/78, 1983, 1987/88, 1993/94, 2004/05, 2009/10, and 2011/12, of which 1993/94, 2004/05, 2009/10, and 2011/12 surveys are comparable. Using per capita GDP growth instead of per capita HFCE growth yields pass-throughs of 0.639 for the full sample and 0.768 for the comparable spells.

[29] Sengupta (2016) provides similar arguments that highlight the measurement issues created by the choice of accounting methods used by the CSO.

Citing this, Subramanian argues that the official GDP growth rates have been overestimated by 1.1 percentage points when India is compared with other middle-income countries and by 2.5 percentage points when all countries are included in the comparison. This finding has been refuted by Goyal and Kumar (2019) who defend the official statistics citing methodological and data issues in A. Subramanian (2019).

The estimate of poverty we calculate using the pass-through rate is sensitive to the growth in consumption in national accounts. Any downward adjustment of growth, as the one suggested by A. Subramanian (2019), would lead to an increase in poverty. Hence, we also present poverty estimates adjusting growth downward by 1.1 and 2.5 percentage points. It is important to note that A. Subramanian's (2019) findings are for the period 2011-2016 and for the GDP growth rate, while our application of the downward adjustment is for the period 2016-2018 and for the growth in per capita HFCE.[30]

4.4. Poverty rate estimates

Table 5 reports poverty estimates for 2017 using our preferred pass-through rate, as well as a pass-through rate of 1, applied to various national accounts growth rates. These poverty estimates assume that growth is distribution-neutral, i.e. all observations in the survey are scaled up by the same growth rate, similar to the standard extrapolation methods that underpin the World Bank's global poverty numbers (see Prydz et al. (2019)). In all cases, we line up the 2011/12 CES microdata to 2015 using urban/rural-specific growth rates. These growth rates are derived using the implied pass-through rates calibrated from the poverty estimates reported in Newhouse and Vyas (2019).[31] The national poverty estimates are calculated as the weighted sum of the rural and urban poverty rates using the population weights in WDI in the relevant year.[32] The estimates reported in Table 5 use the various pass-through rates after 2015.

In the first row of panel A, we report estimates of poverty derived from the growth in per capita HFCE as reported in WDI using a 0.67 pass-through rate and a range of estimates derived from the 95 percent confidence interval of the 0.67 pass-through rate. In addition, we also report the estimates from applying the raw growth in per capital HFCE (that is, using a pass-through of 1). A higher pass-through implies faster growth in survey consumption and thus a lower poverty rate. The national poverty rate derived from using a 0.67 pass-through is 10.39 percent with a 95 percent confidence interval of [9.97, 10.80], while the national poverty rate derived using a pass-through

---

[30] The average ratio between per capita HFCE and per capita GDP for the years 2011-2018 is 0.560 with a standard deviation of 0.003. For the years 2011-2018, the average annual per capita HFCE growth rate was 6.1 percent, and the average annual per capita GDP growth rate was 5.6 percent for the same period. Source: WDI, World Bank.

[31] A detailed discussion of estimating poverty rates for the years 2012-2015 in India can be found in Chen et al. (2018).

[32] This accounts for changes in rural/urban population shares since the last survey. PovcalNet applies the same methodology to calculate national poverty rates for all countries that have rural/urban surveys, namely China, India and Indonesia. The rural/urban population shares implied by the survey weights may be different from shares in WDI.

of 1 is 8.84 percent.[33] The latter estimate translates to 118 million living under extreme poverty in India in 2017, while the 0.67 pass-through yields 139 million people living under extreme poverty. The confidence band around the 0.67 pass-through suggests that between 134 million and 145 million live in poverty in the nation. Using the 0.67 pass-through, the rural poverty rate is estimated to be 12.02 percent (with a 95 percent confidence interval between [11.52, 12.44]) and the urban poverty rate is estimated to be 7.17 percent (with a 95 percent confidence interval between [6.92, 7.55]).[34]

In rows 2 and 3 of Table 5, we report estimates derived by reducing the official HFCE growth rate by 1.1 and 2.5 percentage points respectively, using the estimates provided by A. Subramanian (2019). The 0.67 pass-through rate yields a national poverty rate between 10.97 percent and 11.75 percent (or between 147 million and 157 million people) depending on the growth adjustment used. If we allow for the uncertainty around the pass-through rate, the national poverty rate could be as high as 12.09 percent, which would imply 162 million living under extreme poverty in 2017. Similarly, the rural poverty rate could be between 13.04 percent and 13.93 percent (which implies between 116 million and 124 million poor) and the urban poverty rate could be between 8.03 percent and 8.47 percent (which implies between 36 million and 37 million poor).

Since the annual per capita GDP growth rates are similar to the annual per capita HFCE growth rates, the poverty rates derived using these two growth rates are fairly similar. The set of estimates employing the growth in per capita GDP are reported in panel B of Table 5.

Figure 5 presents the trends in national poverty for the years 2012 to 2018. These trends are reported for two series presented in Table 5– namely, (a) the trend using official per capita HFCE growth with the 0.67 pass-through rate and the 95 percent confidence interval (see row 1 of Table 2, panel A); and (b) the trend applying a downward adjustment of 2.5 percentage points to the per capita HFCE growth rate with a 0.67 pass-through rate. As highlighted in the figure, the changes in growth rates matter for the poverty rates. A downward adjustment of 2.5 percentage points of growth rates increases the national poverty rate by 1.4 percentage points, which translates to 18 million more people pushed into extreme poverty.

5. Summary and Conclusion

This paper is an attempt to provide up-to-date information on poverty in India in the absence of regular data from the Consumption Expenditure Survey (CES). Most of the alternative data sources

---

[33] Note that the 95% confidence interval is symmetric around the pass-through rate. However, when the pass-through rate and its 95% interval is mapped into the poverty rates, the confidence band around the poverty estimates are not symmetric. This is because the poverty estimates depend on the density around the poverty line.

[34] Using the South Asia consumption-specific pass-through rate of 0.652 Table A.5 with a 95% confidence interval of [0.307, 0.998] yields a national poverty rate of 10.49 percent (with 95% confidence interval between [8.84, 12.43]), a rural poverty rate of 12.13 percent (with 95% confidence interval between [10.21, 14.28]), and an urban poverty rate of 7.23 percent (with 95% confidence interval between [6.13, 8.77]).

indicate an increase in average household consumption per capita, although most of them are either not fully comparable to the official CES, or do not cover a period long enough, or a geographical coverage wide enough, to assess the evolution of household consumption and, more precisely, poverty rates. We then adopt two methods to estimate poverty in 2017/18. First, we use a survey-to-survey methodology to impute consumption into the Social Consumption Survey for Health 2017/18 using a model estimated on the CES 2011/12. Our method builds on Newhouse and Vyas (2019), adding explanatory variables such as household energy consumption and demographic household characteristics, and modifying the functional form of the rainfall shock, but not including a time trend. Second, we project the CES 2011/12 forward using national accounts growth rates combined with a pass-through factor that adjusts for the difference between growth in national accounts and household surveys. Borrowing from Lakner et al. (forthcoming), we use a machine learning algorithm to estimate the pass-through rate and allow for changes in inequality.

Using the survey-to-survey method, we estimate a national poverty rate (at the international poverty line of $1.90 per person per day) of 9.9 percent in 2017, with a 95 percent confidence interval of between 8.1 and 11.3.[35] With the preferred pass-through rate of 0.67, we obtain a national poverty rate of 10.4 percent in 2017, with a confidence interval between 10.0 and 10.8. Despite using very different data sources and methods, the estimated poverty rates are strikingly similar with overlapping confidence intervals. Within urban and rural areas, the differences are somewhat larger, but the confidence intervals again overlap, and there is no evidence that one method is systematically biased in one direction. Using the survey-to-survey method, rural poverty is estimated at 10.5 percent [8.8, 12.0], some 1.5 percentage points lower than the result of the pass-through exercise (12.0 percent, [11.5, 12.4]). In contrast, at 8.5 percent [6.8, 10.1] urban poverty is higher using the survey-to-survey method compared to 7.2 percent [6.9, 7.6] with the pass-through.

Our results are robust to changes in the model used in the survey-to-survey method, plausible alternative pass-through rates and varying the starting year of the pass-through exercise (see Appendix). They are also consistent with the trends using alternative surveys over the same period, which show growth in average household consumption, and in some cases welfare gains across the entire income distribution.

However, neither approach is without limitations. On the one hand, the survey-to-survey method takes advantage of the variation in the survey data to capture changes in the distribution of welfare. But if the imputation is done between periods too far apart, it may fail to capture important changes in the behavior of markets, since the parameters of the consumption model are assumed fixed for a long period of time. Hence, important structural changes in the Indian economy between 2011

---

[35] The survey-to-survey estimates refer to 2017/18, the period of fieldwork for the Health SCS. To compare with the results of the pass-through method, these estimates have been brought back to 2017 using growth in HFCE per capita, following the method described in Chen et al. (2018) for the 2014/15 (Newhouse and Vyas 2019) estimates. The text refers to the results from the preferred model (model 2 in Table A.3).

and 2017 may not be captured by these imputation techniques. In general, this method is more appropriate to estimate poverty for small geographic areas that have no representative samples (e.g. the seminal work by Elbers et al. (2003)), or over short time periods (e.g. Douidich, et al. (2016)).

On the other hand, the pass-through approach assumes that national accounts HFCE growth is accurate and that growth is distribution neutral. Both these assumptions have been the subject of recent debate in India. A. Subramanian (2019) has argued that India's GDP growth from official sources is overstated, although Goyal and Kumar (2019) have disputed his findings. Regarding changes in inequality, Chanda and Cook (2019) and Chodrow-Reich et al. (2020) find a negative short-term impact of the demonetization introduced in November 2016 among the poorest groups, which dissipates after several months. Lahiri (2020), meanwhile, reports a decline in unemployment shortly after demonetization, which may hide an important decline in labor force participation (also see Vyas (2018)).

The poverty rates estimated for 2017 using the pass-through method would be higher if we allow for increasing inequality, for which there is some supportive evidence in the literature cited above as well as the CMIE data between 2016 and 2017. Assuming a 1 percent annual increase in the Gini index between 2015 and 2017 would lead to a poverty rate of 11.3 percent in 2017, a number still within the confidence interval of the survey-to-survey imputation. If the Gini index were to rise 2 percent per year, the poverty rate would climb to 12.4 percent (compared to 10.4 percent with distribution neutrality) in 2017. If the underlying national accounts growth (in terms of either GDP or HFCE) is reduced by 1.1 or 2.5 percentage points for the period 2015-2017, while assuming distribution-neutrality, we estimate a national poverty rate of 11.0 and 11.8 percent, respectively.

All these estimates are subject to strong assumptions; therefore, considerable uncertainty remains about poverty in India in 2017 and the trend in recent years, and this uncertainty can only be resolved if new survey data become available. Using leaked summary statistics of the withheld 2017/18 household survey, S. Subramanian (2019) estimates that poverty increased significantly between 2011/12 and 2017/18. Himanshu (2019) also finds a decline in average consumption using alternative recent survey data. In contrast, Bhalla and Bhasin (2020) claim that poverty declined significantly between 2011/12 and 2017/18. One additional complication is that different welfare aggregates give very different estimates of poverty levels and potentially also the trend. Using the data from the leaked report, similarly to S. Subramanian, we estimate a level of poverty (15.6 percent in 2017) that is still higher than all the estimates using our regular methods.[36] However, leaked data that cannot be verified are not an acceptable source of information for reliable poverty

---

[36] As we explain in more detail in the Appendix (also see Section 2), this is explained by the different consumption aggregates being used. Our main analysis uses the URP aggregate which has been used historically in India and which gives higher levels of poverty than the MMRP aggregate, that is used in the leaked estimates. In other words, projecting a decline in the URP aggregate or an increase in the MMRP aggregate results in levels of poverty that are not too different. This of course does not answer the important question over the direction of poverty in recent years.

estimates. Furthermore, the quality concerns over the 2017/18 survey require further investigation. Across a wide range of publicly available data sources, the paper finds no evidence of an increase in poverty between 2011/12 and 2017/18.

The lack of publicly available data creates doubts among the general public, obstructs scientific debate, and hinders the implementation of sound, empirically based development policies. The imputation methods adopted in this paper are more appropriate to extrapolate poverty across shorter periods when data are not available, or for geographic areas where survey data is not appropriate. They are imperfect substitutes for actual data on standards of living. There is no alternative to timely, quality assured, and transparent data for poverty measurement and for the design and monitoring of anti-poverty policies.

6. Tables

Table 1. Main characteristics of selected household surveys in India

Survey name	Administering agency	Number of questions about consumption	Welfare aggregate	Survey time	Description
Consumption Expenditure survey (CES) 2017/18	NSSO, Government of India	~ 400	MMRP	July 2017 to June 2018	Withheld due to data quality concerns
Consumption Expenditure survey (CES) 2011/12	NSSO, Government of India	~ 400	MMRP, MRP, URP	July 2011 to June 2012	Last available official consumption survey round
Survey on social consumption (Social Consumption 2017/18)	NSSO, Government of India	1	Usual monthly consumption expenditure of the household	July 2017 to June 2018	Education and health specific survey. Sample: 113,823 households
Survey on social consumption (Social Consumption 2014)	NSSO, Government of India	1	Usual monthly consumption expenditure of the household	January to July 2014	Education and health specific survey. Sample: 65,932 households
Periodic Labor Force Survey (PLFS)	NSSO, Government of India	1	Usual monthly consumption expenditure of the household	July 2017 to June 2018	Starts in 2017-18 to replace employment-unemployment surveys. Cross-sectional in rural areas and panel in urban areas. Sample: ~56,000 households
India Human Development Survey (IHDS)	NCAER & University of Maryland, Indiana University and University of Michigan	52	MRP	Two and half rounds: 2004, 2011 and subsample round in 2017	Household panel containing income and expenditure questions. Sample: ~ 41,500 households
Consumer Pyramids (CP)	CMIE, private data collection agency	~ 80	Consumption recall over last three months	Starts in 2014 (and every quarter since)	Starts in 2014. Household level panel with three-monthly period recall. Sample: ~174,000 households

Table 2. Estimated poverty rate in 2017/18 ($1.90 per day poverty line) using survey-to-survey imputation methods, comparing different models

Sector	Model 1	Model 2	Model 3	Model 4
National	8.70	8.75	8.61	8.47
	[7.51, 9.90]	[7.32,10.19]	[7.37, 9.86]	[7.19, 9.76]
Rural	8.92	9.03	8.38	9.14
	[7.36,10.48]	[7.18,10.88]	[6.80, 9.96]	[7.38, 10.9]
Urban	8.18	8.08	9.18	6.85
	[6.34,10.01]	[6.07,10.09]	[7.43,10.94]	[5.50, 8.19]

Table 3. Back-casting comparison of poverty rates ($1.90 per day poverty line) from survey-to-survey methods and actual estimates

	Model 1	Model 2	Model 3	Model 4	Actual
Panel A: 2004/05					
National	29.88	28.64	30.38	28.4	38.90
	[27.16,32.61]	[25.68,31.60]	[27.53,33.24]	[25.62,31.18]	[38.1,39.7]
Rural	32.57	31.52	33.03	32.22	43.40
	[28.97,36.18]	[27.83,35.21]	[29.38,36.67]	[28.69,35.75]	[42.6,44.2]
Urban	21.97	20.16	22.6	17.15	25.40
	[18.88,25.05]	[16.83,23.49]	[19.31,25.90]	[14.75,19.54]	[24.2,26.6]
Panel B: 2009/10					
National	17.49	17.65	17.51	17.23	31.70
	[15.37,19.61]	[15.4,19.89]	[15.58,19.44]	[15.25,19.22]	[30.9,32.5]
Rural	18.38	18.39	18.39	18.79	36.10
	[15.53,21.22]	[15.49,21.29]	[15.83,20.95]	[16.12,21.45]	[35.1,37.1]
Urban	15.09	15.65	15.14	13.04	19.80
	[12.76,17.41]	[12.88,18.41]	[12.87,17.41]	[11.09,14.99]	[18.8,20.8]
Panel C: 2014/15					
National	19.49	17.04	20.93	16.8	14.61
	[16.74,22.23]	[13.84,20.23]	[18.05,23.81]	[13.99,19.61]	[13.04,16.78]
Rural	22.18	19.52	21.95	20.17	16.81
	[18.56,25.79]	[15.38,23.66]	[18.25,25.66]	[16.36,23.97]	[15.24,18.38]
Urban	13.13	11.16	18.52	8.85	10.01
	[9.98,16.27]	[8.10,14.22]	[15.09,21.95]	[6.82,10.88]	[8.44,11.58]

Note: Actual estimates in 2014/15 are estimates from Newhouse and Vyas (2019). Actuals in 2004/05 and 2009/10 are authors' estimates using data from NSS0. 95 percent confidence intervals in square brackets.

Table 4. Estimated pass-through rates for MOB sub-samples

Survey Type	Comparability	Pass-through	95% CI	N
Income	1	1.003	[.890, 1.12]	841
Consumption	1	0.661	[.576, .746]	457

Note: This table reports the pass-through estimates (estimate of coefficient β in equation 1) and their 95 percent confidence intervals for the sub-samples in Figure A.1. The sample size (N) is the number of comparable survey spells that can be included in the MOB estimation.

Table 5. Poverty estimates ($1.90 per day poverty line) for 2017 using pass-through method

	Poverty rate (%)			Number of poor (millions)		
	National	Rural	Urban	National	Rural	Urban
Panel A: HFCE/capita growth rate						
1. Official HFCE/capita growth rate						
passthrough = 1	8.84	10.21	6.13	118.3	90.8	27.6
passthrough = 0.67	10.39	12.02	7.17	139.1	106.8	32.2
95% CI for 0.67 pass-through	[9.97, 10.80]	[11.52, 12.44]	[6.92, 7.55]	[134, 145]	[102, 111]	[31, 34]
2. HFCE/capita growth rate (adjusted downward by 1.1pp)						
passthrough = 1	9.54	11.01	6.63	127.7	97.9	29.8
passthrough = 0.67	10.97	12.65	7.66	146.9	112.4	34.5
95% CI for 0.67 pass-through	[10.61, 11.24]	[12.26, 12.94]	[7.36, 7.89]	[142, 151]	[109, 115]	[33, 35]
3. HFCE/capita growth rate (adjusted downward by 2.5pp)						
passthrough = 1	10.63	12.27	7.40	142.3	109.0	33.3
passthrough = 0.67	11.75	13.50	8.29	157.2	120.0	37.3
95% CI for 0.67 pass-through	[11.36, 12.09]	[13.04, 13.93]	[8.03, 8.47]	[152, 162]	[116, 124]	[36, 37]
Panel B: GDP/capita growth rate						
4. Official GDP/capita growth rate						
passthrough = 1	8.95	10.34	6.19	119.8	91.9	27.8
passthrough = 0.67	10.46	12.09	7.23	140.0	107.5	32.5
95% CI for 0.67 pass-through	[10.08, 10.88]	[11.66, 12.55]	[6.97, 7.59]	[135, 146]	[104, 112]	[31, 34]
5. GDP/capita growth rate (adjusted downward by 1.1pp)						
passthrough = 1	9.67	11.17	6.72	129.5	99.2	30.2
passthrough = 0.67	11.05	12.74	7.70	147.9	113.3	34.6
95% CI for 0.67 pass-through	[10.70, 11.32]	[12.34, 12.99]	[7.45, 8.00]	[143, 151]	[110, 115]	[34, 36]
3. GDP/capita growth rate (adjusted downward by 2.5pp)						
passthrough = 1	10.72	12.36	7.48	143.5	109.9	33.7
passthrough = 0.67	11.84	13.60	8.37	158.5	120.9	37.6
95% CI for 0.67 pass-through	[11.48, 12.17]	[13.15, 14.00]	[8.17, 8.54]	[154, 163]	[117, 124]	[37, 38]

Note: This table reports poverty rates for the $1.90 international poverty line in 2017. The 95 percent confidence interval for the 0.67 pass-through rate is [0.59, 0.75]. Panel A reports results using per capita HFCE growth and panel B reports results using per capita GDP growth.

7. Figures

Figure 1. Mean household per capita consumption expenditure in Rural India, across surveys
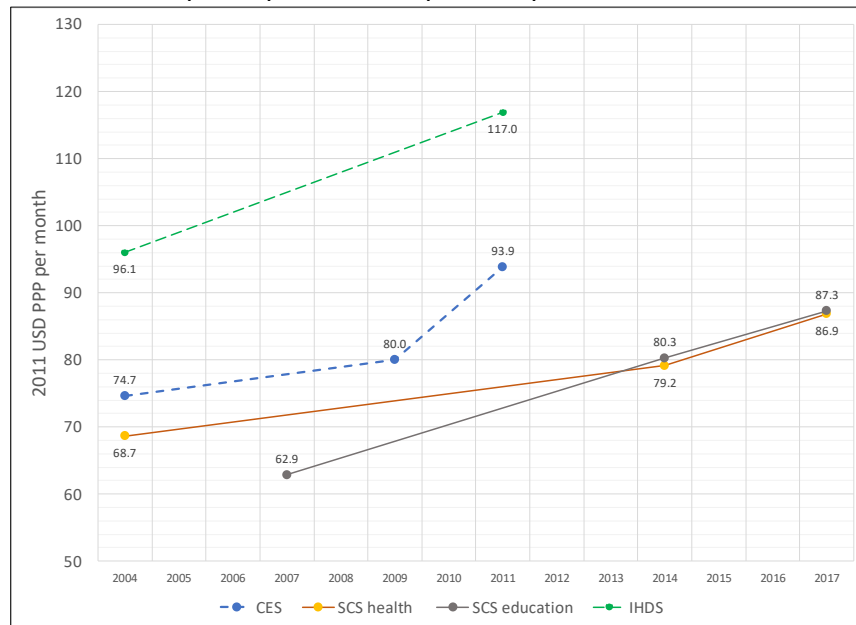
Note: CES refers to Consumption Expenditure Surveys, SCS Health and Education to the Surveys on Social Consumption and IHDS to India Human Development Survey. Consumption expenditures are in 2011 USD PPP using price deflators as in Atamanov et al. (2020).

Figure 2. Mean household per capita consumption expenditure in Urban India, across surveys
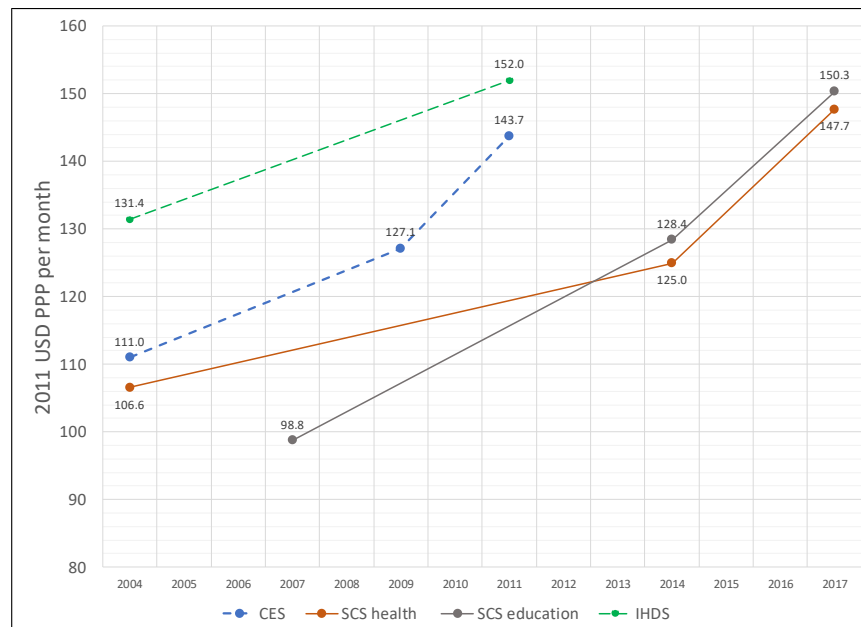
Note: CES refers to Consumption Expenditure Surveys, SCS Health and Education to the Surveys on Social Consumption and IHDS to India Human Development Survey. Consumption expenditures are in 2011 USD PPP using price deflators as in Atamanov et al. (2020).

Figure 3. Stochastic dominance analysis across selected Indian household surveys
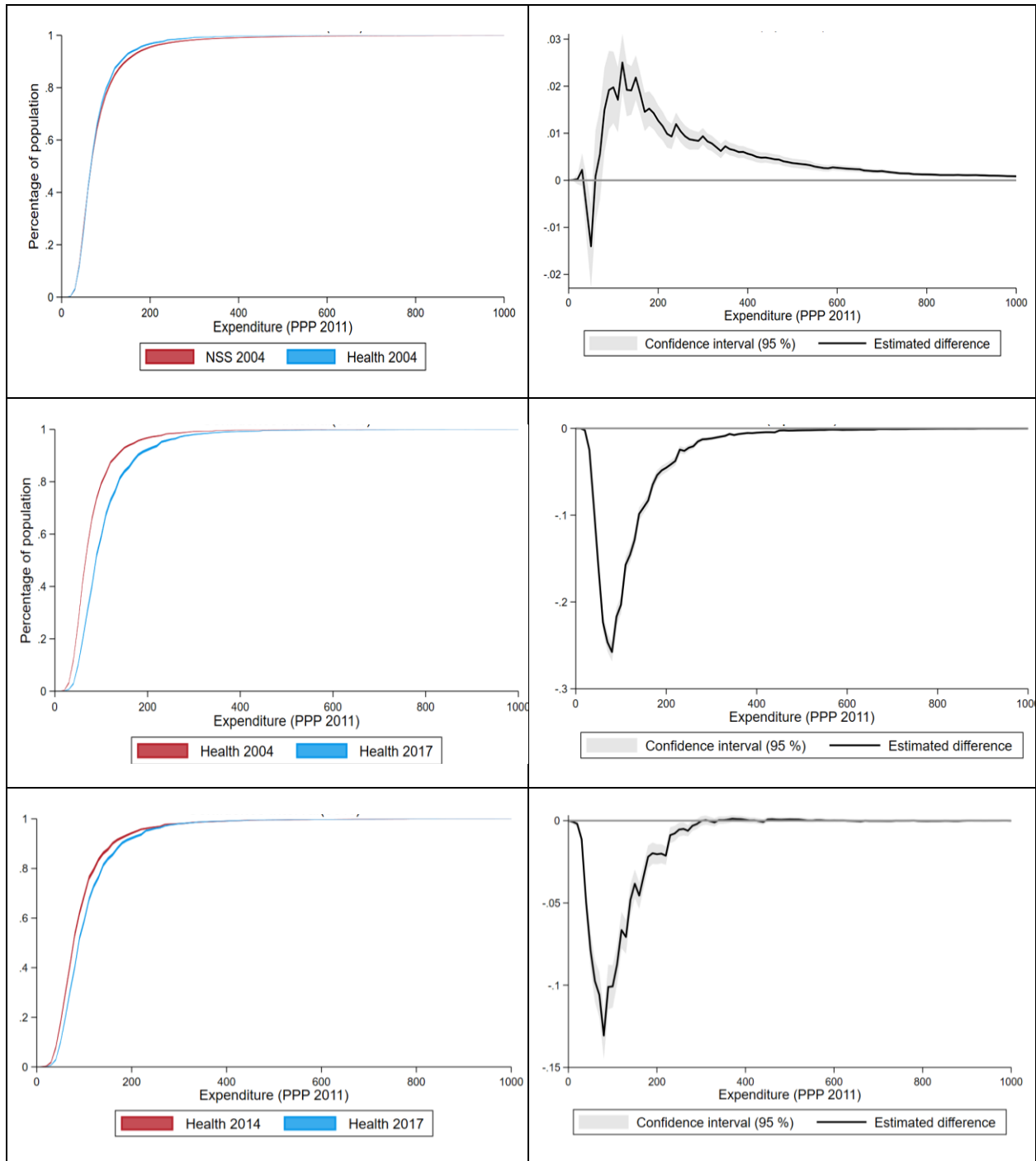
Note: The figures in the left column refer to the cumulative percentage of population (vertical axis) whose per capita household consumption expenditure is below a given expenditure level in 2011 USD PPP per month (horizontal axis), after households have been ranked from lowest to highest expenditure. The curves to the right indicate the difference between each cumulative distribution, and its corresponding confidence interval. These figures are computed using the FGT curves command from the Distributive Analysis Stata Package by Araar and Duclos (2013). The urban and rural 2011 PPP rates are given in Atamanov et al. (2020).

Figure 4. Growth incidence curves from Consumer Pyramids survey data

Note: Vertical axis measures annual percentage change in household consumption expenditures per capita, and horizontal axis refers to the household consumption percentile.

Figure 5. Projections of India's national poverty rate, from year 2015
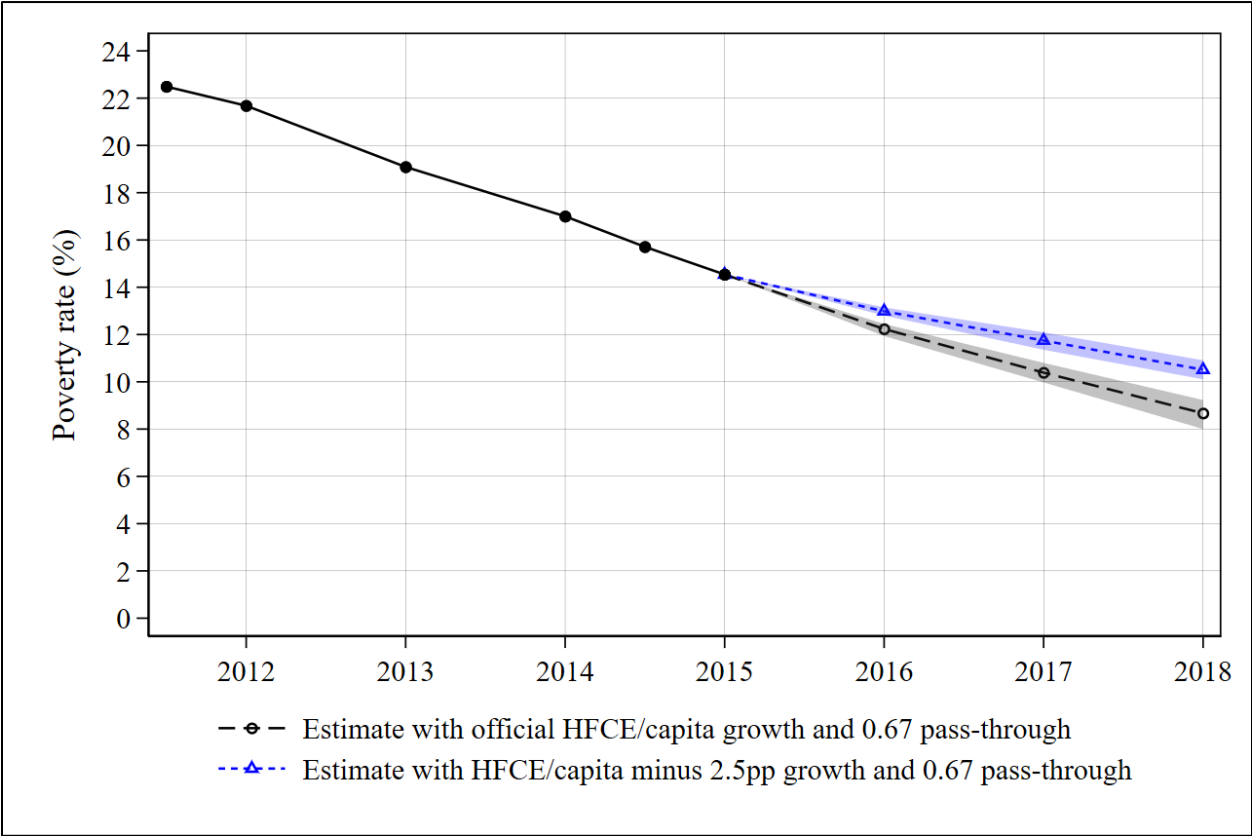
Note: This figure shows the trend in the national poverty rate at the $1.90 per day poverty line for the period 2012-2018. For the 2015-2018 period, we show poverty estimates using the preferred pass-through rate and its 95 percent confidence interval for two growth scenarios: (a) official per capita HFCE growth; and (b) official per capita HFCE growth reduced by 2.5 percentage points. See also Table 5.

Appendix

This Appendix includes a description of the main demographic characteristics of the household surveys used as either training or target data in the survey-to-survey imputation exercises (section A.1) as well as a description of the results of the regression models used for imputation (section A.2). It also includes robustness checks of the pass-through poverty estimates, either by considering different pass-through rates (section A.3), changes in inequality (section A.4) and the use of alternative data sources and welfare aggregates from leaked data (section A.5).

A.1 Comparison across surveys

For the set of variables to be used in the survey-to-survey imputations we compare the means across several surveys (CES in years 2004/05, 2009/10, 2011/12 and 2014/15, and Health SCS in 2014 and 2017/18), for rural and urban samples, respectively. We report the main set of variables used in Newhouse and Vyas (2018) and in Newhouse and Vyas (2019). The CES and Health SCS in 2014/15, which is the year in which both surveys can be compared directly, are very similar across the common set of indicators. This suggests that the differences between the two surveys, for example in terms of questionnaire design, sampling etc., do not have a large impact on these variables. In the remainder of the section, we concentrate on describing the CES surveys before 2011/12 and the SCS data in 2017/2018, which are the surveys used in the survey-to-survey imputation. Due to the absence of a full consumption aggregate, the 2014/15 CES cannot be used to assess the trend in welfare.[37] The other economic and demographic variables are very similar between this survey and the previous one (2011/12 CES) (see Table A.1 and Table A.2).

For the years 2004/05 and 2011/12, the CES shows substantial growth in average per-capita expenditure in rural areas from $75 to $94 per month (in 2011 PPPs), an annual increase of 3.3 percent. This increase in per capita consumption is accompanied by a small decline in household size and some population aging leading to a decline in the dependency ratio (the ratio of the population aged less than 15 and above 64 to the overall population). The changes in household head characteristics by gender, social group or religion are all within 1 percentage point. The share of household heads working in agriculture falls from 63 to 56 percent, while the proportion of self-employed increases from 51 to 53 percent. The trends in urban areas are similar. Average per-capita consumption grows at 3.4 percent per year, from $111 to $144 (in 2011 PPPs). The dependency ratio also falls here, and changes in the gender, caste and religion of the household head are small (around 2 percentage points).

In the Health SCS, household size continues to decline with the proportion of households with more than 5 people falling from 45 percent to 43 percent in rural areas and from 35 percent to 32 percent in the urban sector between 2014 and 2017/18. The population also appears to be aging

---

[37] This survey was not meant to be a full Consumption Expenditure Survey, rather it was supposed to be a survey of durable goods and services consumption. Formally, it is the 72nd round survey, but for simplicity we refer to it as 2014/15 CES.

slightly with the share under age 15 falling to 30 percent in rural areas and to 24 percent in urban areas. This has led to a decline in the dependency ratio in both areas.

The distribution of household heads by gender, religion and caste in the Health SCS 2017/2018 does not differ substantially from the distribution of these characteristics in CES 2011/12. The share of household heads self-reporting as Hindu remains at around 77 percent in urban areas and 83 percent in rural areas. The distribution of heads by caste does not change between these two surveys, either in rural or urban areas. In contrast, the proportion of female heads is somewhat lower in the SCS than the CES (8 and 10 percent in rural and urban samples, compared with 12 percent in CES). The demographic characteristics in the SCS show a continuation of the trends seen in the CES data for earlier years. However, the distribution of household heads by employment characteristics shows greater divergence between CES 2011/12 and Health SCS 2017/2018, particularly in urban areas. The proportion of household heads working in agriculture is higher for both rural and urban areas in the SCS (61 and 6 percent, respectively) than in CES 2011/12 (56 and 4 percent). The proportion of self-employed has also increased between the two surveys: from 50 to 56 percent in rural areas, and from 34 to 40 percent in urban areas.

At the bottom of Table A.1 and Table A.2 we present the means of the population-weighted district level rainfall shocks that are included in the model. Following Newhouse and Vyas (2019), we compute these values by taking the quarterly deviation of each district's rainfall from the historical mean (between 1981 and 2018).[38] In 2017/18, rainfall appears to be higher than the historical average with the exception of the first quarter of 2018, in both rural and urban areas. This is consistent with the 2017 annual report of the Monsoon Department, which indicates that for all of India, rainfall in Q1 and Q2 was comparable to its long-term average. But there were highly localized and extreme rainfall events which may drive district level averages up or down.[39]

Newhouse and Vyas (2019) used household expenditures on recreational services and transport as explanatory variables that could be found in both CES 2014/15 and earlier CES rounds. These consumption variables are not captured in the Health SCS, so we use a different consumption variable that is common across the CES and SCS. Table A.1 and Table A.2 also report the primary source of cooking fuel, which shows a rapid transition from the use of firewood and chips to liquefied petroleum gas (LPG). The proportion of rural households using firewood and chips fell from 67 percent in CES 2011/12 to about 51 percent in the Health SCS 2017/2018, while LPG use

---

[38] With the exception of the period between April to June, our rainfall shocks variable in SCS 2014/15 approximates the results used for CES 2014/15 by Newhouse and Vyas (2019). The slight differences occur as a result of different districts being included in the CES and SCS surveys.
[39] Government of India, Ministry of Earth Sciences, India Metereological Department (2017).

rose from 15 percent to 41 percent. A similar trend can be observed in urban areas with the use of firewood and chips falling by 7 percentage points while LPG use rose 21 percentage points.[40]

A.2 Econometric results for survey-to-survey imputation

As explained in the main text, we run four models using CES 2011/12 as training data, and then used the coefficients to predict poverty in 2017/18, using data from SCS Health 2017/18 and parameters from the consumption model. The models have a different set of explanatory variables because we run four model specifications, two using LASSO and two using Stepwise selection, where rainfall is specified either as a spline or a quadratic function, but sign and significance of the demographic variables do not vary notably across specifications. Full econometric results, for urban and rural areas, are included in Table A.3.

The percentage of children and young adults is associated with lower household consumption per capita, while the percentage of prime-age adults (e.g. between 25 and 49) has a positive impact. Consequently, the dependency ratio has a negative and usually significant impact on household consumption. Household size is also associated with lower household consumption per capita across all models. The gender and marital status of the household head are not statistically significant across all models. But a low-caste household head is associated with lower household consumption in all models.

The estimated effect of employment characteristics is also similar across all models. Households whose head is working in high-skilled occupations have the highest consumption per capita, followed by middle-skilled occupations and low-skilled occupations (omitted category). A household head working as a regular wage worker or self-employed is associated with higher consumption than casual laborer (omitted category). Finally, in rural areas, households whose head works in agriculture or industry have lower consumption than families whose head works in services. All these results are not too surprising given common assumptions about returns to skills, job function and industry of employment.[41]

The inclusion of variables associated with rainfall varies widely across models. However, there is some regularity. Rainfall in the first quarter (that is from January to March) has always negative coefficients, sometimes significant; while rainfall in the fourth quarter (October to December) has always a positive coefficient, often significant. Rainfall in the third quarter (July to September) is never statistically significant different from zero, while rainfall in the second quarter (April to

---

[40] The move towards LPG may have been facilitated by a national LPG distribution program called Ujjwala, which was established in 2016. The program's objective was to distribute 50 million LPG connections to women of BPL families. The evidence from the SCS is consistent with administrative data which suggest that the program seems to have been successful in expanding access to LPG.

[41] In urban areas, however, all models show average consumption being higher for households whose heads work in agriculture, followed by those in services and those in manufacturing having the lowest (the exception being model 3, where all sectors have lower returns than services). This may be explained by a low prevalence of agricultural workers in urban areas (around 6 percent), which may thus be a highly selected sample (e.g. landowners living in urban areas).

June) is always positive, sometimes significant. The squares and splines of these variables show no apparent pattern. The pattern of the rainfall variable alone seems consistent with positive rainfall shocks in the fourth quarter being associated with higher household consumption, whereas the opposite association emerges if it occurs in the first quarter. The monsoon months in India are from June to September. Overall, we should expect higher farm output when rainfall is adequate during the monsoon period. However, this period can vary across locations, and it is difficult to isolate the sign for the quarters when rainfall is more than the long-term period average. This is because these deviations are highly localized and their effects depend on local cropping patterns, timing of crops, etc.

Finally, the primary cooking fuel also shows some regularity across models. Cooking with electricity or LPG is positively and significantly associated with household consumption in rural households, having usually the largest or second-largest effect. Models differ regarding the sign and significance for all the other fuels.

We use a version of the Kitagawa-Oaxaca-Blinder decomposition (Kitagawa (1955), Oaxaca (1973) and Blinder (1973)) to understand what variables account for most of the change in estimated poverty between 2011/12 and 2017/18. Typically, this decomposition breaks down the mean difference across two groups (here households in the 2011/12 CES and 2017/18 SCS, respectively) into components that are explained by differences in endowments and differences in returns to these endowments. Within the survey-to-survey exercise, the coefficients which predict household consumption in 2017/18 are those estimated from the model in 2011/12, so there is no change in the returns and all observed changes are driven by changes in endowments (i.e. the vector of explanatory variables x_{ch}).

Table A.4 shows the contribution of a variable to the difference in log (i.e. the growth rate) in per capita consumption between 2011/12 and 2017/18 (that is, the sum of the impact of each component adds up to the total change at the end of the table and the share of the contribution would be the ratio between the impact of a given variable and the total change). The imputed consumption in 2017/18 implies a growth of 32.4 percent in rural areas and 21.3 percent in urban areas over the period. The rainfall shock in the fourth quarter accounts for 34.5 percent of the change in log welfare in rural areas and 30 percent in urban areas. This is explained by the positive coefficient of this variable in the models of Table A.1 and Table A.2 and the above average rainfall recorded in this quarter.

The use of LPG is the second most important variable in rural areas (accounting for 24 percent of the change) but has no impact in urban areas. This is explained by the large increase in the share of rural households using LPG as the main cooking fuel (from 15 percent to 41 percent). Beyond these two variables, demographic variables are also important. The reduction in the share of children (age 0 to 14) explains about 11 percent of the consumption increase in urban areas, and 6 percent in rural areas. Approximately two thirds of the projected growth in per capita household

consumption are explained by these three variables: rainfall in the fourth quarter, use of LPG as cooking fuel and proportion of children in households.

A.3 Robustness checks for pass-through rate

Table A.5 reports several alternative pass-through rates estimated using different samples of growth spells from household surveys. Panel A reports the preferred estimate calculated using the global consumption-specific comparable surveys. Panels B, C, and D aggregate the surveys according to various geographic regions, similar to what has sometimes been done in the literature. The table reports estimates of the pass-through rate (column 5) and a 95 percent confidence interval around these estimates (column 6) calculated using regression (5) in the main text with various samples of the data. In particular, the universe of surveys for all countries and years is divided along four dimensions: (a) by geographic region (global, South Asia or India), (b) by national accounts measure (growth in GDP or HFCE per capita), (c) by welfare measure used in the survey (consumption only, or both consumption and income surveys) (the Indian surveys use consumption), and (d) whether only comparable surveys are used (comparability equals 1).

The pass-through rate using the various samples ranges from 0.65 to 0.93. The estimates from various samples of India-specific surveys ranges between 0.69 to 0.93. However, it is important to note that these India-specific estimates are based on very few survey spells (6 in total, of which only 3 are comparable). This is one reason that there is large statistical uncertainty around the India-specific pass-through estimate, with the 95% confidence interval ranging from -1.05 to 2.55.

The estimates including all South Asian countries range between 0.65 and 0.74. We prefer the global estimates over the South Asian estimates, because the MOB algorithm reported in the main text showed no evidence that the sample should be split by geographic region. Furthermore, the comparable South Asia sample largely consists of surveys from Bangladesh and Pakistan, and it is not clear that this would be any more informative for India than a global estimate. More importantly, the relevant South Asia estimate (consumption-specific comparable surveys using per capita HFCE growth) is 0.65, which is extremely close to our preferred estimate of 0.67. More generally, all the South Asia-specific estimates are within the 95 percent CI of our preferred estimate.

A second robustness check refers to using a different start period. All previous estimates in Table 5 refer to poverty estimates for 2017, starting from year 2015. That is, starting from a poverty estimate that is based on previous work by Newhouse and Vyas (2018) as implemented in Chen et al. (2018). Alternatively, the pass-through coefficient and methodology described above could be applied from the last period for which micro-data from an official survey is publicly available, that is 2011/12.

Figure A.2 shows poverty estimates for years 2012-2018 using the official per capital HFCE growth rate with 0.67 pass-through applied to the 2011/12 survey and poverty rates using the 95 percent confidence interval around the 0.67 pass-through rate. It also reports estimates for 2014/15

from Newhouse and Vyas (2019) and the preferred estimate with 95 percent confidence interval for 2017. The 2017 preferred estimate is calculated using official per capital HFCE growth rate with 0.67 pass-through applied for years 2016-2018.

Interestingly, the poverty estimates from Newhouse and Vyas (2019) include within their confidence interval the poverty estimates that would be derived from our new pass-through exercise starting from 2011/12. Moreover, the baseline poverty estimates for 2017 that use a pass-through rate of 0.67 and official HFCE growth (10.4 percent national, 12.0 percent rural and 7.2 percent urban, see Table 5) fall within the confidence interval of the exercise shown in Figure A.2. Comparing the point estimates between Table 5 and Figure A.2, the national poverty rate is only 0.1 percentage point higher, while rural poverty is 0.4 percentage point lower, and urban poverty is 0.5 percentage point higher. This evidence shows that poverty estimates under the pass-through methodology would not be significantly affected by a change in the start period.

A.4 Distribution-sensitive poverty projection

The poverty nowcasts presented in sections above, assumed distribution-neutral growth. Changes to inequality would result in different poverty estimates, e.g. an increase in inequality would lead to greater poverty. Since we have no empirical evidence how inequality has changed over this period, we consider four different scenarios for changes in the Gini index: annual increases of 2 percent and 1 percent and annual decreases of 1 percent and 2 percent. Following Lakner et al. (2014) and Lakner et al. (forthcoming), we use the povsim simulation tool to nowcast poverty.[42] The rural Gini index in India in 2015 was 0.311, the urban Gini index was 0.390, and the implied national Gini index was 0.337.[43] For example, the four inequality scenarios discussed above would yield a rural Gini index in 2017 of 0.324, 0.317, 0.305, 0.299, respectively.

Table A.6 presents the poverty rates and the number of poor in 2017 derived by changing the Gini index. As our baseline, we have used the preferred scenario with official per capita HFCE growth rates and a 0.67 pass-through rate (reported in column 1). Whereas a distribution-neutral scenario would lead to a national poverty rate of 10.4 percent, an increase in inequality could increase poverty to between 11.3 percent (increasing the Gini index by 1 percent annually) and 12.4 percent (increasing the Gini index by 2 percent annually), or an equivalent decrease in inequality could decrease poverty to between 9.4 percent (decreasing the Gini index by 1 percent annually) and 8.6 percent (decreasing the Gini index by 2 percent annually). In terms of the number of poor, the distribution-neutral scenario predicts 139 million living in extreme poverty in 2017, the changes in the Gini index predicts that this number could range between 115 million (decreasing the Gini

---

[42] The povsim Stata package is available from the authors. Given a change in the Gini index, a growth rate in the mean and a functional form of the growth incidence curve (GIC), it simulates distributional changes in a welfare distribution. In this application, we use a linear GIC which is a relatively conservative specification.

[43] Note that the Gini index at the national level in 2015 is 0.337, while it is 0.357 in 2011/12. This change in the Gini index in 2015 relative to the 2011/12 micro data is due to: (a) the difference in urban/rural growth rates between 2011/12 and 2015 (as explained above, these are calibrated on the urban/rural poverty rates estimated by Newhouse and Vyas (2019)), and (b) the change in the shares of urban/rural population between 2011/12 and 2015.

index by 2 percent annually) and 166 million (increasing the Gini index by 2 percent annually). Figure A.3 presents the national, rural, and urban poverty rates respectively in the top panel, and similarly, the number of poor in the bottom panel for the years 2011-2018.

A.5 Pass-through estimates using CES 2017/18 leaked data

In a series of articles, the Business Standard daily released consumption growth rates that are purported to be based on a leaked report using the 2017/18 CES survey, which was not released by the authorities citing concerns over data quality.[44] The newspaper articles suggest that household consumption expenditure in rural areas decreased by as much as 8.8 percent, while in urban areas it increased by 2 percent over the 2011/12 to 2017/18 period. If substantiated, the consumption growth rates reported in Business Standard signal an increase in poverty for the first time in India in four decades. This contrasts with all other sources of data and methodologies reported in this paper that suggest a declining trend in poverty over the 2011/12-2017/18 period.

Panel C of Table A.7 reports poverty estimates for 2017 calculated by applying the consumption growth rates reported in the Business Standard daily to the 2011/12 micro data. These growth rates are reported separately for rural/urban for the 2011/12-2017/18 period. For 17 of 28 states rural/urban growth rates are reported in addition to the all-India growth rates. To model the distributional changes as flexibly as possible, we use the rural/urban state-level growth rates wherever possible. The states whose growth rates are not reported are treated as a residual category for which the growth rate (separately by rural/urban) is given by the following expression:

$$g_m^j = \frac{g^j - s_{n,t-1}^j g_n^j}{s_{m,t-1}^j} \qquad (2),$$

where j indicates rural or urban, and m and n represent missing and non-missing groups, respectively. $s_{m,t-1}^j$ is the consumption share for the group of states with missing consumption growth rates, estimated from the 2011/12 survey using the appropriate sampling weights. g_n^j is the consumption growth rate for the states with growth rates reported by the Business Standard, aggregated using the sampling weights of the 2011/12 survey. Expression (2) assumes that population growth in the missing and the non-missing groups is the same; this is a required assumption since data on rural/urban population at the state-level are not available for 2017/18.

Panel A of Table A.7 reports the poverty estimates from 2011/12, and Panel B reports the estimates calculated using the official per capita HFCE growth rates with a 0.67 pass-through reported in Table 2 in the main text. It is important to note that the 2011/12 CES data report several welfare aggregates which result in substantial differences in poverty rates. Panel A and B use the Uniform Reference Period (URP) aggregate, which is currently used by PovcalNet, since it is the aggregate used in the historic data, thus allowing for a long time series. The 2017/18 CES round discontinued the URP aggregate and only used the Mixed Modified Reference Period (MMRP) aggregate,

---

[44] These articles were published in the Business Standard daily. See for example Jha (2019).

which is also collected in the 2011/12 data.[45] We assume that the growth rates reported in the Business Standard daily are using the MMRP aggregate (since it is the only aggregate available in both survey rounds) and thus apply them to the MMRP aggregate in the 2011/12 survey. We estimate the national poverty rate for 2017 based on the leaked growth rates to be 15.59 percent, which translates to 210 million people living in extreme poverty in India (Panel C).[46] These estimates are somewhat larger than our upper bound for national poverty reported using all the methods reported in this paper. However, these estimates are subject to substantial caveats, because leaked data that we are unable to verify is not an acceptable source of information for reliable poverty estimates.

---

[45] The difference between MMRP and URP is the design of the survey questionnaire, see World Bank (2018). The national poverty rate for 2011/12 using the URP aggregate is 22.49 percent, while using the MMRP aggregate it is 13.25 percent.

[46] The leaked growth rates bring the 2011/12 distribution forward to 2017/18. The estimate for 2017 is computed by bringing the 2017/18 numbers backwards by half a year using HFCE per capita growth. This is identical to how the 2017 results from the survey-to-survey method are derived (for details see Section 3 in the main text). Using the all-India rural/urban growth rates reported in the Business Standard, rather that the state-level rural/urban growth rates reported in Table A.2, would yield a rural poverty rate of 19.88 percent, an urban poverty rate of 8.84 percent, and a national poverty rate of 16.17 percent in 2017.

Appendix Tables

Table A.1. Descriptive statistics (rural samples)

Indicator	CES 2004/05	CES 2009/10	CES 2011/12	CES 2014/15	SCS Health 2014	SCS Health 2017/18
Mean HH per capita expenditure (2011 USD PPP per month)	74.7	80.0	93.9		79.2	86.9
Household size						
1 or 2	0.05	0.06	0.06	0.06	0.05	0.06
3	0.08	0.09	0.09	0.09	0.09	0.09
4	0.16	0.19	0.19	0.20	0.19	0.21
5	0.19	0.20	0.21	0.21	0.21	0.22
6 and more	0.51	0.46	0.45	0.44	0.45	0.43
(District average)						
1 or 2	0.05	0.06	0.06	0.06	0.06	0.06
3	0.08	0.09	0.1	0.1	0.09	0.09
4	0.17	0.19	0.2	0.21	0.2	0.21
5	0.20	0.2	0.21	0.21	0.21	0.21
6 and more	0.50	0.46	0.44	0.43	0.44	0.42
Household age structure						
0-14	0.38	0.35	0.34	0.33	0.31	0.3
15-24	0.16	0.16	0.16	0.16	0.19	0.17
25-34	0.15	0.15	0.15	0.15	0.15	0.16
35-49	0.17	0.19	0.19	0.20	0.19	0.2
50-64	0.10	0.11	0.11	0.11	0.12	0.13
65 and over	0.04	0.04	0.05	0.05	0.05	0.04
(District average)						
0-14	0.33	0.30	0.30	0.28	0.30	0.27
15-24	0.16	0.17	0.17	0.17	0.19	0.19
25-34	0.15	0.15	0.15	0.16	0.16	0.17
35-49	0.18	0.19	0.19	0.20	0.19	0.2
50-64	0.10	0.11	0.11	0.11	0.12	0.13
65 and over	0.08	0.08	0.08	0.08	0.05	0.04
Household head: religion						
Hindu	0.84	0.84	0.83	0.83	0.83	0.83
Other	0.16	0.16	0.17	0.17	0.17	0.17
(District average)						
Hindu	0.82	0.82	0.82	0.81	0.82	0.82
Other	0.18	0.18	0.18	0.19	0.18	0.18
Household head: social group						
Scheduled caste	0.74	0.76	0.77	0.79	0.77	0.79
Others	0.26	0.24	0.23	0.21	0.23	0.21
(District average)						
Scheduled caste	0.71	0.73	0.74	0.76	0.74	0.76
Others	0.29	0.27	0.26	0.24	0.26	0.24
Household head: employment type						
Regular wage			0.09	0.08	0.10	0.09
Self-employed	0.51	0.47	0.50	0.53	0.57	0.56

Casual labor			0.33	0.31	0.30	0.32
Others			0.08	0.08	0.04	0.03
(District average)						
Regular wage			0.06	0.06	0.14	0.14
Self-employed	0.54	0.51	0.53	0.55	0.55	0.54
Casual labor			0.26	0.24	0.27	0.28
Others			0.15	0.15	0.05	0.04
Household head: principal industry						
Agriculture	0.63	0.60	0.56	0.60	0.60	0.61
Industry	0.15	0.18	0.20	0.18	0.22	0.21
Others	0.22	0.23	0.24	0.22	0.18	0.18
(District average)						
Agriculture	0.55	0.51	0.48	0.51	0.48	0.48
Industry	0.17	0.20	0.22	0.21	0.23	0.22
Others	0.28	0.29	0.30	0.28	0.29	0.3
Household head: gender						
Female	0.11	0.11	0.12	0.12	0.09	0.08
Male	0.89	0.89	0.88	0.88	0.91	0.92
(District average)						
Female	0.11	0.11	0.12	0.12	0.09	0.09
Male	0.89	0.89	0.88	0.88	0.91	0.91
Household head: marital status						
Married	0.86	0.86	0.85		0.89	0.89
Not married	0.14	0.14	0.15		0.11	0.11
(District average)						
Married	0.86	0.86	0.85		0.89	0.89
Not married	0.14	0.14	0.15		0.11	0.11
Dependency ratio (mean)						
Sample average	0.4	0.37	0.36	0.35	0.38	0.34
District average	0.39	0.36	0.35	0.34	0.37	0.34
Primary energy source for cooking						
Liquefied petroleum gas	0.09	0.12	0.15		0.17	0.41
Firewood & chips	0.75	0.76	0.67		0.70	0.51
No cooking	0.01	0.02	0.01		0.00	0.00
Others	0.03	0.02	0.05		0.02	0.01
Dung cake	0.09	0.06	0.10		0.10	0.06
Coke, coal	0.01	0.01	0.01		0.01	0.01
Electricity	0.00	0.00	0.00		0.00	0.00
Kerosene	0.01	0.01	0.01		0.00	0.00
Charcoal	0.00	0.00	0.00		0.00	0.00
Gobar gas	0.00	0.00	0.00		0.00	0.00
(District average)						
Liquefied petroleum gas	0.09	0.12	0.15		0.27	0.50
Firewood & chips	0.75	0.76	0.67		0.6	0.43
No cooking	0.01	0.02	0.01		0.00	0.00
Others	0.03	0.02	0.05		0.01	0.01
Dung cake	0.09	0.06	0.10		0.08	0.05
Coke, coal	0.01	0.01	0.01		0.02	0.01
Electricity	0.00	0.00	0.00		0.00	0.00
Kerosene	0.01	0.01	0.01		0.01	0.00
Charcoal	0.00	0.00	0.00		0.00	0.00
Gobar gas	0.00	0.00	0.00		0.00	0.00

District rainfall shock (deviation from historical average)

	CES 2004/05	CES 2009/10	CES 2011/12	CES 2014/15	SCS Health 2014	SCS Health 2017/18
July - September	-0.22	-0.1	0.42	0.02	0.01	0.28
July - September (squared)	0.27	0.21	0.38	0.22	0.22	0.32
October - December	-0.30	0.26	-0.62	-0.07	-0.08	0.13
October - December (squared)	0.23	0.43	0.51	0.26	0.25	0.63
January - March	0.30	-0.3	-0.22	0.63	0.66	-0.13
January - March (squared)	0.43	0.22	0.29	1.15	1.17	0.45
April - June	-0.21	-0.14	-0.15	0.34	0.31	0.35
April - June (squared)	0.35	0.43	0.17	0.34	0.32	0.37

Note: For categorical variables, table denotes the share in each category.

Table A.2. Descriptive statistics (urban samples)

Indicator	CES 2004/05	CES 2009/10	CES 2011/12	CES 2014/15	SCS Health 2014	SCS Health 2017/18
Mean HH per capita expenditure (2011 USD PPP per month)	111.0	127.2	143.7		124.8	147.7
Household size						
1 or 2	0.07	0.08	0.08	0.09	0.08	0.10
3	0.10	0.11	0.12	0.13	0.12	0.13
4	0.22	0.24	0.24	0.25	0.25	0.26
5	0.21	0.20	0.20	0.20	0.20	0.19
6 and more	0.41	0.37	0.36	0.33	0.35	0.32
(District average)						
1 or 2	0.06	0.07	0.08	0.08	0.07	0.09
3	0.09	0.10	0.11	0.12	0.11	0.12
4	0.20	0.23	0.23	0.24	0.23	0.25
5	0.20	0.21	0.21	0.20	0.21	0.20
6 and more	0.44	0.39	0.38	0.36	0.37	0.35
Household age structure						
0-14	0.31	0.29	0.27	0.26	0.25	0.24
15-24	0.19	0.18	0.18	0.18	0.18	0.17
25-34	0.17	0.17	0.17	0.17	0.18	0.18
35-49	0.20	0.20	0.21	0.22	0.21	0.21
50-64	0.10	0.11	0.11	0.12	0.12	0.14
65 and over	0.03	0.05	0.06	0.05	0.05	0.05
(District average)						
0-14	0.33	0.30	0.29	0.28	0.27	0.24
15-24	0.17	0.17	0.17	0.17	0.18	0.19
25-34	0.16	0.17	0.17	0.17	0.17	0.18
35-49	0.19	0.20	0.20	0.21	0.20	0.21
50-64	0.10	0.11	0.11	0.12	0.12	0.14
65 and over	0.05	0.05	0.06	0.05	0.05	0.05
Household head: religion						
Hindu	0.78	0.78	0.77	0.76	0.77	0.77
Other	0.22	0.22	0.23	0.24	0.23	0.23
(District average)						
Hindu	0.81	0.81	0.81	0.79	0.80	0.80
Other	0.19	0.19	0.19	0.21	0.20	0.20
Household head: social group						
Scheduled caste	0.54	0.57	0.60	0.61	0.61	0.61

Others	0.46	0.43	0.40	0.39	0.39	0.39
(District average)						
Scheduled caste	0.63	0.65	0.67	0.68	0.67	0.68
Others	0.37	0.35	0.33	0.32	0.33	0.32
Household head: employment type						
Regular wage	0.39	0.37	0.40	0.41	0.38	0.38
Self-employed	0.38	0.36	0.34	0.31	0.41	0.40
Casual labor	0.12	0.14	0.13	0.15	0.15	0.15
Others	0.11	0.13	0.13	0.12	0.06	0.07
(District average)						
Regular wage	0.20	0.19	0.21	0.23	0.27	0.27
Self-employed	0.48	0.45	0.46	0.44	0.46	0.45
Casual labor	0.04	0.06	0.05	0.06	0.22	0.23
Others	0.29	0.30	0.28	0.26	0.05	0.05
Household head: principal industry						
Agriculture	0.06	0.05	0.04	0.05	0.07	0.06
Industry	0.31	0.30	0.3	0.29	0.35	0.32
Others	0.64	0.65	0.66	0.66	0.58	0.62
(District average)						
Agriculture	0.37	0.34	0.31	0.31	0.30	0.30
Industry	0.24	0.25	0.26	0.26	0.28	0.26
Others	0.40	0.41	0.43	0.43	0.42	0.44
Household head: gender						
Female	0.10	0.11	0.12	0.14	0.11	0.11
Male	0.90	0.89	0.88	0.86	0.89	0.89
(District average)						
Female	0.10	0.11	0.12	0.14	0.10	0.10
Male	0.90	0.89	0.88	0.86	0.90	0.90
Household head: marital status						
Married	0.82	0.81	0.80		0.87	0.86
Not married	0.18	0.19	0.20		0.13	0.14
District average)						
Married	0.82	0.81	0.80		0.88	0.87
Not married	0.18	0.19	0.20		0.12	0.13
Dependency ratio (mean)						
Sample average	0.36	0.33	0.32	0.31	0.32	0.30
District average	0.36	0.33	0.32	0.31	0.34	0.31
Primary energy source for cooking						
Liquefied petroleum gas	0.57	0.65	0.68		0.74	0.89
Firewood & chips	0.22	0.18	0.14		0.18	0.07
No cooking	0.05	0.07	0.07		0.00	0.01
Others	0.01	0.01	0.01		0.00	0.01
Dung cake	0.02	0.01	0.01		0.02	0.01
Coke, coal	0.03	0.02	0.02		0.02	0.01
Electricity	0.00	0.00	0.00		0.00	0.00
Kerosene	0.10	0.06	0.06		0.03	0.01
Charcoal	0.00	0.00	0.00		0.00	0.00
Gobar Gas	0.00	0.00	0.00		0.00	0.00
(District average)						
Liquefied petroleum gas	0.57	0.65	0.68		0.50	0.69
Firewood & chips	0.22	0.18	0.14		0.39	0.25
No cooking	0.05	0.07	0.07		0.00	0.01
Others	0.01	0.01	0.01		0.00	0.01

Dung cake	0.02	0.01	0.01		0.05	0.03
Coke, coal	0.03	0.02	0.02		0.02	0.01
Electricity	0.00	0.00	0.00		0.00	0.00
Kerosene	0.10	0.06	0.06		0.02	0.00
Charcoal	0.00	0.00	0.00		0.00	0.00
Gobar gas	0.00	0.00	0.00		0.00	0.00
District rainfall shock (deviation from historical average)						
July - September	-0.2	-0.04	0.5	0.02	0.01	0.28
July - September (squared)	0.34	0.17	0.44	0.28	0.22	0.32
October - December	-0.26	0.32	-0.51	-0.07	-0.08	0.13
October - December (squared)	0.21	0.55	0.42	0.17	0.25	0.63
January - March	0.13	-0.25	-0.33	0.62	0.66	-0.13
January - March (squared)	0.29	0.18	0.37	1.14	1.17	0.45
April - June	-0.04	-0.07	-0.18	0.42	0.31	0.35
April - June (squared)	0.28	0.35	0.19	0.40	0.32	0.37

Note: For categorical variables, table denotes the share in each category.

Table A.3. Regression model (models 1 - 4)

Variables	Model 1 Rural Coeff.		S.E.	Model 1 Urban Coeff.		S.E.	Model 2 Rural Coeff.		S.E.	Model 2 Urban Coeff.		S.E.	Model 3 Rural Coeff.		S.E.	Model 3 Urban Coeff.		S.E.	Model 4 Rural Coeff.		S.E.	Model 4 Urban Coeff.		S.E.
Household age structure																								
Age 0-14	-0.33	***	0.03	-0.51	***	0.04	-0.32	***	0.03	-0.49	***	0.05	-0.27	***	0.02	-0.24	***	0.02	-0.28	***	0.02	-0.37	***	0.02
Age 15-24	-0.15	***	0.03	-0.2	***	0.04	-0.19	***	0.06	-0.19	***	0.04												
Age 25-34	0.07	***	0.03	0		0.04	0.03		0.05	0.03		0.04	0.2	***	0.02	0.15	***	0.02	0.15	***	0.02			
Age 35-49	0.14	***	0.02	0.1	***	0.04	0.08		0.06	0.12	***	0.04	0.26	***	0.02	0.18	***	0.02	0.18	***	0.02			
Age 50-64				0.03		0.03	-0.06		0.05	0.03		0.04	0.12	***	0.02	0.06	***	0.02						
(District averages):																								
Age 0-14	-0.83	*	0.47	-0.92	**	0.42	-0.68	*	0.37	-0.85	**	0.41	-0.71	**	0.35	-0.48		0.35	-0.63	***	0.24	-1.17	***	0.41
Age 15-24	-0.6		0.5				-0.39		0.41	0.44		0.44							-0.5		0.33			
Age 25-34	-0.89	*	0.54	0.16		0.41	-0.68		0.42	0.59	*	0.32	-0.48		0.33	-0.52		0.33	-0.74	**	0.37			
Age 35-49	-0.94	*	0.55	-0.56		0.45	-0.68	*	0.39				-0.48		0.32	-0.52		0.32	-0.73	**	0.33	-0.77	***	0.3
Age 50-64	-0.28		0.65	0.39		0.5				0.98	**	0.43												
Household size	-0.03	***	0	-0.05	***	0	-0.03	***	0	-0.05	***	0	-0.03	***	0	-0.05	***	0	-0.03	***	0	-0.06	***	0
Dependency ratio																								
Household ratio	-0.15	***	0.02				-0.21	***	0.05				-0.06	***	0.02	-0.1	***	0.02	-0.1	***	0.02			
District average				1.03	**	0.43	0.12		0.49	1.4	***	0.45	0.34		0.39	0.16		0.39				1.16	***	0.44
Characteristics of the household head																								
Low caste	-0.11	***	0.01	-0.16	***	0.01	-0.11	***	0.01	-0.15	***	0.01	-0.12	***	0.01	-0.12	***	0.01	-0.12	***	0.01	-0.16	***	0.01
Male	-0.01		0.01	0.01		0.02	-0.01		0.01	0		0.02	0		0.01	0.02	**	0.01	0		0.01	0.04	***	0.01
Married	0.01		0.01	0.03	*	0.02	0.01		0.01	0.03		0.02												
(District averages):																								
Low caste	-0.06		0.06	-0.04		0.06	-0.08		0.05				-0.07		0.05	-0.11	**	0.06	-0.08		0.05	-0.07		0.05
Male	-0.01		0.28	-0.04		0.29	0.02		0.28													0.08		0.23
Married	0.12		0.3	0.25		0.34	0.11		0.3															
Household head skill level																								
High skill	0.12	***	0.01	0.3	***	0.01	0.12	***	0.01	0.29	***	0.01	0.12	***	0.01	0.13	***	0.01	0.12	***	0.01	0.3	***	0.01
Middle skill	0.04	***	0.01	0.06	***	0.01	0.04	***	0.01	0.06	***	0.01	0.04	***	0.01	0.05	***	0.01	0.04	***	0.01	0.07	***	0.01
(District averages):																								
High skill	-0.52	***	0.15				-0.48	***	0.15	-0.12		0.15	-0.5	***	0.15	-0.36	**	0.16	-0.48	***	0.15			

	Coef		SE	Coef		SE	Coef		SE	Coef		SE	Coef		SE	Coef		SE	Coef		SE	Coef		SE
Middle skill	-0.51	***	0.2	-0.08		0.19	-0.49	***	0.2				-0.51	***	0.19	-0.44	**	0.2	-0.5	***	0.19			
Household head principal industry																								
Agriculture	-0.07	***	0.01	0.1	***	0.02	-0.07	***	0.01	0.1	***	0.02	-0.07	***	0.01	-0.07	***	0.01	-0.07	***	0.01	0.1	***	0.02
Industry	-0.07	***	0.01	-0.03	***	0.01	-0.07	***	0.01	-0.03	***	0.01	-0.07	***	0.01	-0.06	***	0.01	-0.07	***	0.01	-0.04	***	0.01
(District averages):																								
Agriculture	-0.35	***	0.09	-0.3	***	0.08	-0.31	***	0.08	-0.32	***	0.08	-0.33	***	0.08	-0.27	***	0.08	-0.31	***	0.08	-0.19	***	0.08
Industry																								
Household head employment																								
Regular wage				0.11	***	0.01				0.11	***	0.01										0.1	***	0.01
Self-employed	0.11	***	0.01	0		0.01	0.12	***	0.01	0		0.01	0.12	***	0.01	0.13	***	0.01	0.12	***	0.01	-0.01		0.01
(District averages):																								
Regular wage				0.48	***	0.16				0.45	***	0.17										0.44	***	0.14
Self-employed	0.21	***	0.08				0.18	**	0.08				0.19	***	0.08	0.13	*	0.08	0.18	**	0.08			
Household primary cooking energy																								
Charcoal				-0.21		0.14				-0.21	**	0.11										-0.32	***	0.13
Coke, coal				-0.35	***	0.03				-0.34	***	0.02	0.13	***	0.03	0.15	***	0.03	0.12	***	0.03	-0.46	***	0.03
Dung cake	0.04	**	0.02	-0.23	***	0.04	0.02	*	0.01	-0.23	***	0.04	0.07	***	0.02	0.08	***	0.02	0.07	***	0.02	-0.33	***	0.04
Electricity	0.51	***	0.09	0.02		0.06	0.45	***	0.13	0.04		0.08	0.53	***	0.09	0.83	***	0.08	0.51	***	0.13			
Firewood	0.02		0.02	-0.32	***	0.01				-0.32	***	0.01	0.05	***	0.02	0.06	***	0.02	0.05	***	0.02	-0.42	***	0.03
Gobar gas																								
Kerosene	0.06		0.04	-0.23	***	0.02				-0.22	***	0.02	0.07	*	0.04	0.13	***	0.04	0.07	*	0.04	-0.34	***	0.03
LPG	0.32	***	0.02				0.3	***	0.01				0.35	***	0.02	0.37	***	0.02	0.35	***	0.02	-0.08	***	0.02
No cooking	0.61	***	0.05	0.27	***	0.03	0.53	***	0.07	0.26	***	0.03	0.58	***	0.08	0.51	***	0.06	0.54	***	0.08			
(District averages):																								
Charcoal	-5.05	*	2.89	-2.31		4.37	-5.53	**	2.83	-4.48		4.19	-5.59	**	2.74	-6.22	**	2.97	-5.72	**	2.77	-4.66		4.3
Coke, coal	-0.45	**	0.23				-0.45	**	0.22				-0.56	***	0.22	-0.49	**	0.23	-0.57	***	0.22			
Dung cake	-0.23		0.17	0.36	***	0.12	-0.2		0.17	0.29	**	0.15	-0.3	*	0.17	-0.17		0.17	-0.27	*	0.17	0.16	**	0.08
Electricity	-0.92		1.86	-0.29		1.68	0.02		1.89	-0.28		1.68												
Firewood	-0.31	**	0.15	0.16	*	0.09	-0.28		0.15	0.13		0.12	-0.37	***	0.15	-0.27	*	0.15	-0.37	***	0.15			
Gobar gas																								
Kerosene	-0.87	*	0.47	0.14		0.37	-0.91	**	0.46	0.05		0.37	-0.93	**	0.46	-0.99	**	0.47	-1	**	0.46	-0.19		0.36
LPG	0.12		0.17				0.17		0.17	0.01		0.16	0.05		0.17	0.21		0.17	0.08		0.16			
No cooking	-0.71		0.58	0.81		0.9	-0.66		0.78	1.02		1.22	-0.65		0.55	-0.93	***	0.32	-0.8	***	0.31			

Rainfall shocks

	(1) Coef	Sig	SE	(2) Coef	Sig	SE	(3) Coef	Sig	SE	(4) Coef	Sig	SE	(5) Coef	Sig	SE	(6) Coef	Sig	SE	(7) Coef	Sig	SE	(8) Coef	Sig	SE
Rainfall Q1	-0.1		0.11	-0.16		0.12	-0.13	***	0.02	-0.07	***	0.03							-0.13	***	0.02			
Rainfall Q1 squared							0.07	**	0.03	0.13	***	0.04							0.07	**	0.03	0.16	***	0.04
Rainfall Q2	0.16	**	0.08	0.08		0.06	0.04	*	0.02				0.13		0.1	0.17	*	0.1	0.04	*	0.02			
Rainfall Q2 squared							-0.1	***	0.04	-0.03		0.05							-0.11	***	0.04			
Rainfall Q3	0.01		0.08				0.02		0.04	-0.03		0.04	-0.01		0.07	-0.02		0.07	0.02		0.04	-0.06		0.04
Rainfall Q3 squared							0.08	***	0.03	0.11	***	0.04							0.08	***	0.03	0.14	***	0.04
Rainfall Q4	0.2		0.17				0.18	***	0.03	0.09	**	0.05	0.18	***	0.03	0.17	***	0.03	0.18	***	0.03	0.16	***	0.03
Rainfall Q4 squared										-0.02		0.05												
Rainfall Q1 knot 1	-0.22		0.2	-0.08		0.15							-0.39	***	0.08	-0.33	***	0.08						
Rainfall Q1 knot 2	0.16		0.24										0.31	***	0.11	0.21	**	0.11						
Rainfall Q1 knot 3	0.1		0.18	0.31	***	0.09																		
Rainfall Q1 knot 4																								
Rainfall Q2 knot 1													0.1		0.2	0.08		0.2						
Rainfall Q2 knot 2	-0.18		0.41	-0.1		0.09							-0.28	**	0.15	-0.34	**	0.15						
Rainfall Q2 knot 3	-0.02		0.44																					
Rainfall Q2 knot 4																								
Rainfall Q3 knot 1	0.02		0.15	0.17	**	0.08							0.11		0.1	0.13		0.11						
Rainfall Q3 knot 2	0.16		0.2	-0.45	**	0.19																		
Rainfall Q3 knot 3	-0.01		0.17	0.65	***	0.19							0.11		0.09	0.1		0.09						
Rainfall Q3 knot 4																								
Rainfall Q4 knot 1	-0.03		0.18	0.12	***	0.04																		
Rainfall Q4 knot 2																								
Rainfall Q4 knot 3																								
Rainfall Q4 knot 4																								
Constant	5.93	***	0.48	4.98	***	0.4	5.57	***	0.39	4.7	***	0.25	5.53	***	0.24	5.56	***	0.24	5.68	***	0.24	5.43	***	0.24
N	59501			41706			59501			41706			59502			59502			59502			41707		
Adj R-squared	0.34			0.45			0.34			0.44			0.34			0.37			0.34			0.44		
R-squared	0.34			0.45			0.34			0.44			0.34			0.37			0.34			0.44		
RMSE	0.41			0.48			0.41			0.48			0.41			0.43			0.41			0.49		
F-stat	553.77			672.34			631.9			722.39			661.24			745.18			702.64			1046.11		

Table A.4. Drivers of the differences between predicted average log consumption in 2011/12 and 2017/18

Variable	Urban	Rural
Age 0-14	0.023	0.019
Age 15-24	-0.001	0.004
Age 25-34	-0.001	0.001
Age 35-49	0.001	0.001
Age 50-64	0	-
(District average) Age 0-14	0.053	0.053
(District average) Age 15-24	0.009	0.019
4,(District average) Age 25-34	0.008	0.014
(District average) Age 35-49	-	0.006
(District average) Age 50-64	0.018	0.011
Household size	0.005	0.002
Dependency ratio	-	0.003
(District average) Dependency ratio	-0.02	0.002
Household head: age	-0.001	0
Household head: male	0	0
Household head: married	0	0
Household head: low caste	-0.003	0.002
(District average) Male	-	0
(District average) Married	-	0.001
(District average) Low Caste	-	0.002
Household head: high skill	0.016	0.001
Household head: middle skill	0	0
(District average) High Skill	-0.005	0.007
(District average) Middle Skill	-	0.001
Household head: regular wage	0.001	-
Household head: self-employed	-0.001	0.003
(District average) Regular wage	0.038	-
(District average) Self-employed	-	0.001
Household head: agriculture	0	0.002
Household head: industry	-0.001	0.001
(District average) Agriculture	0.003	0
Cooking energy: charcoal	0	-
Cooking energy: coke, coal	0.006	-
Cooking energy: dung cake	0.002	0.002
Cooking energy: electricity	0	0
Cooking energy: firewood	0.034	-
Cooking energy: kerosene	0.01	-
Cooking energy: liquefied petroleum gas	-	0.078
Cooking energy: no cooking	-0.003	0.001
(District average) Charcoal	-0.001	0.001
(District average) Dung cake	-0.009	0.009

(District average) Coke, coal	-	0.003
(District average) Electricity	0	0
(District average) Firewood	-0.021	0.057
(District average) Kerosene	0.008	0.013
(District average) LPG	-0.009	0.046
(District average) No cooking	-0.002	0
Rainfall Q1	-0.008	0.009
Rainfall Q1 squared	0.008	0.006
Rainfall Q2	-	0.023
Rainfall Q2 squared	-0.014	0.03
Rainfall Q3	0.008	0.002
Rainfall Q3 squared	-0.012	0.008
Rainfall Q4	0.064	0.112
Rainfall Q4 squared	-0.004	-
Log Difference	0.213	0.324
Log Welfare 2011/12	4.738	4.399
Log Welfare 2017/18	4.951	4.723

Note: Estimates based on model 2 in Table A.3.

Table A.5. Pass-through rates for various samples

	(1) Aggregate	(2) Welfare Measure	(3) Comparability	(4) Pass-through	(5) 95% CI	(6) N
Panel A: Prefered global sample						
	HFCE/capita	C	1	0.670	[.588, .752]	471
Panel B: India-specific sample						
	HFCE/capita	C	0	0.890	[.314, 1.47]	6
	HFCE/capita	C	1	0.929	[-.562, 2.42]	3
	GDP/capita	C	0	0.689	[.048, 1.33]	6
	GDP/capita	C	1	0.748	[-1.05, 2.55]	3
Panel C: South Asia-specific sample						
	HFCE/capita	C	0	0.664	[.406, .923]	36
	HFCE/capita	C	1	0.652	[.307, .998]	25
	GDP/capita	C	0	0.744	[.448, 1.04]	39
	GDP/capita	C	1	0.720	[.398, 1.04]	26
Panel D: Global sample						
	HFCE/capita	C	0	0.707	[.611, .802]	575
	GDP/capita	C	0	0.751	[.635, .868]	688
	GDP/capita	C	1	0.746	[.646, .846]	540
	HFCE/capita	C/I	0	0.820	[.750, .890]	1511
	HFCE/capita	C/I	1	0.804	[.737, .871]	1323
	GDP/capita	C/I	0	0.856	[.776, .936]	1671
	GDP/capita	C/I	1	0.866	[.791, .941]	1429

Note: This table reports the pass-through rates for various sub-samples partitioned by geographic region, national accounts aggregate, welfare measure, and comparability of the household survey. It also reports the 95 percent confidence interval and sample size for each pass-through rate. The estimates are for coefficient β in the following regression: $g_{survey,i} = \beta * g_{NAS,i} + u_i$, where i is a growth spell between two survey years, g_{survey} is the growth in the survey mean and g_{NAS} is the growth in national accounts (either per capita HFCE or GDP). Survey type *C* refers to consumption-based and type *I* refers to income-based surveys.

Table A.6. Poverty in 2017 with changes to the Gini index

	(1) No Change	(2) Gini -2%	(3) Gini -1%	(4) Gini +1%	(5) Gini +2%
Panel A: Poverty rate (%)					
National	10.39	8.59	9.39	11.25	12.42
Rural	12.02	10.20	10.90	12.90	14.00
Urban	7.17	5.40	6.40	8.00	9.30
Panel B: Number of poor (millions)					
National	139.1	115.0	125.7	150.6	166.3
Rural	106.8	90.7	96.9	114.7	124.4
Urban	32.2	24.3	28.8	36.0	41.8

Note: This table reports poverty rates (percent) and number of poor (millions) using four scenarios – decreasing Gini index by 2 percent (column 2) and 1 percent (column 3) and increasing Gini index by 1 percent (column 4) and 2 percent (column 5). Column 1 reports statistics using the distribution-neutral growth scenario. Official per capita HFCE growth with a 0.67 pass-through is used throughout. See also Table 5 in the main text.

Table A.7. Poverty estimates using 2017/18 CES consumption growth rates

	Poverty rate (%)			Number of poor (millions)		
	National	Rural	Urban	National	Rural	Urban
Panel A: 2011/12 headcount						
Uniform Reference Period	22.49	26.28	14.22	282.9	226.6	56.3
Panel B: 2017 estimate using offical HFCE/capita growth rate with 0.67 pass-through						
Uniform Reference Period	10.39	12.02	7.17	139.1	106.8	32.2
Panel C: 2017 estimates using leaked growth rates						
Mixed Modified Reference Period	15.59	19.11	8.63	209.8	170.2	39.3

Note: The estimate reported in panel A is calculated using the 2011/12 CES survey. The estimate in panel B is the preferred poverty estimate calculated using the official HFCE per capita growth rate with 0.67 pass-through rate (also see Table 2 in the main text). The estimates from the leaked growth rates, reported in the Business Standard daily, are presented in Panel C. Note that the leaked growth rate uses the Mixed Modified Reference Period (MMRP) welfare aggregate, whereas the estimates in panels A and B are calculated using the Uniform Reference Period (URP).

Appendix Figures

Figure A.1. Decision tree of pass-through rates
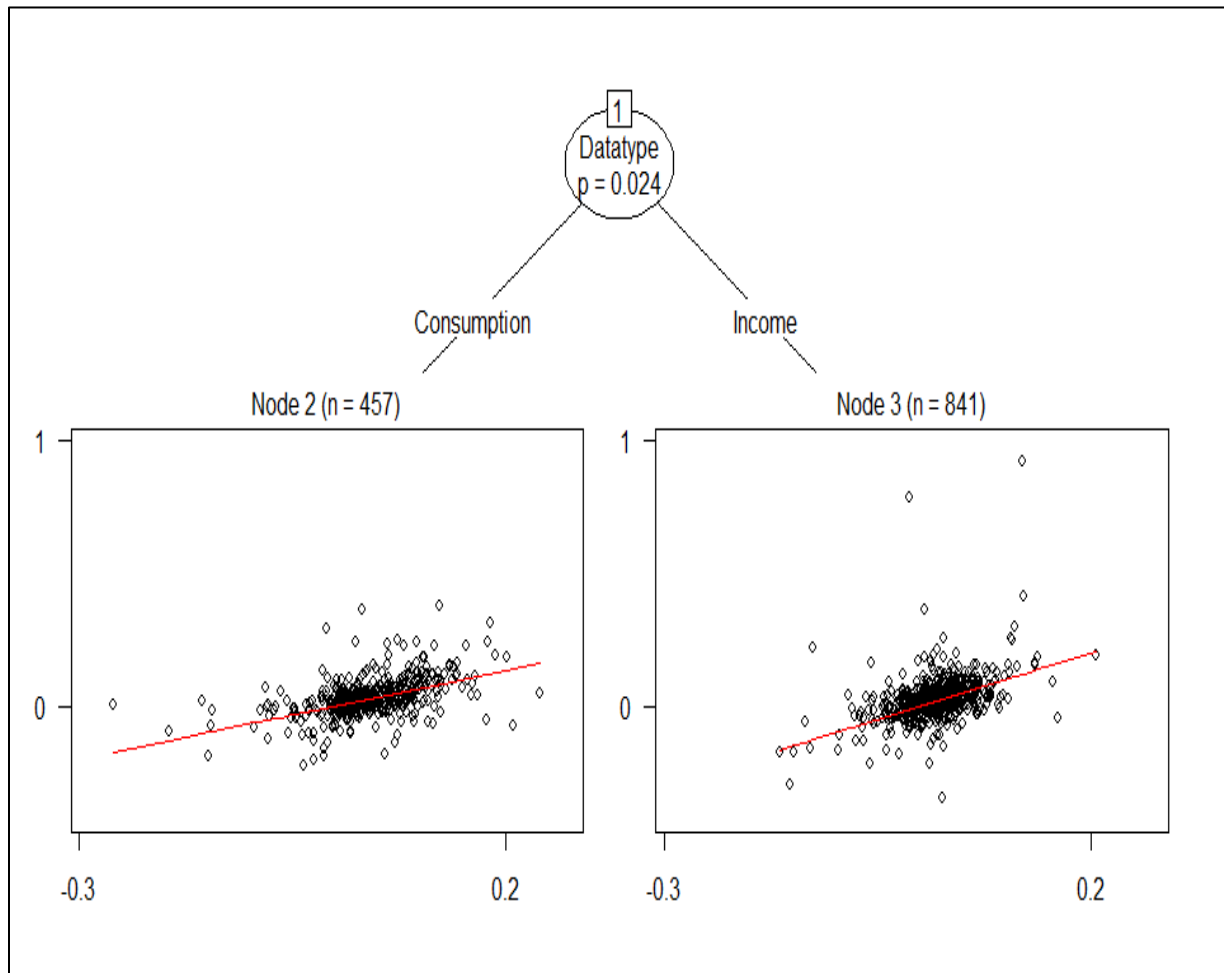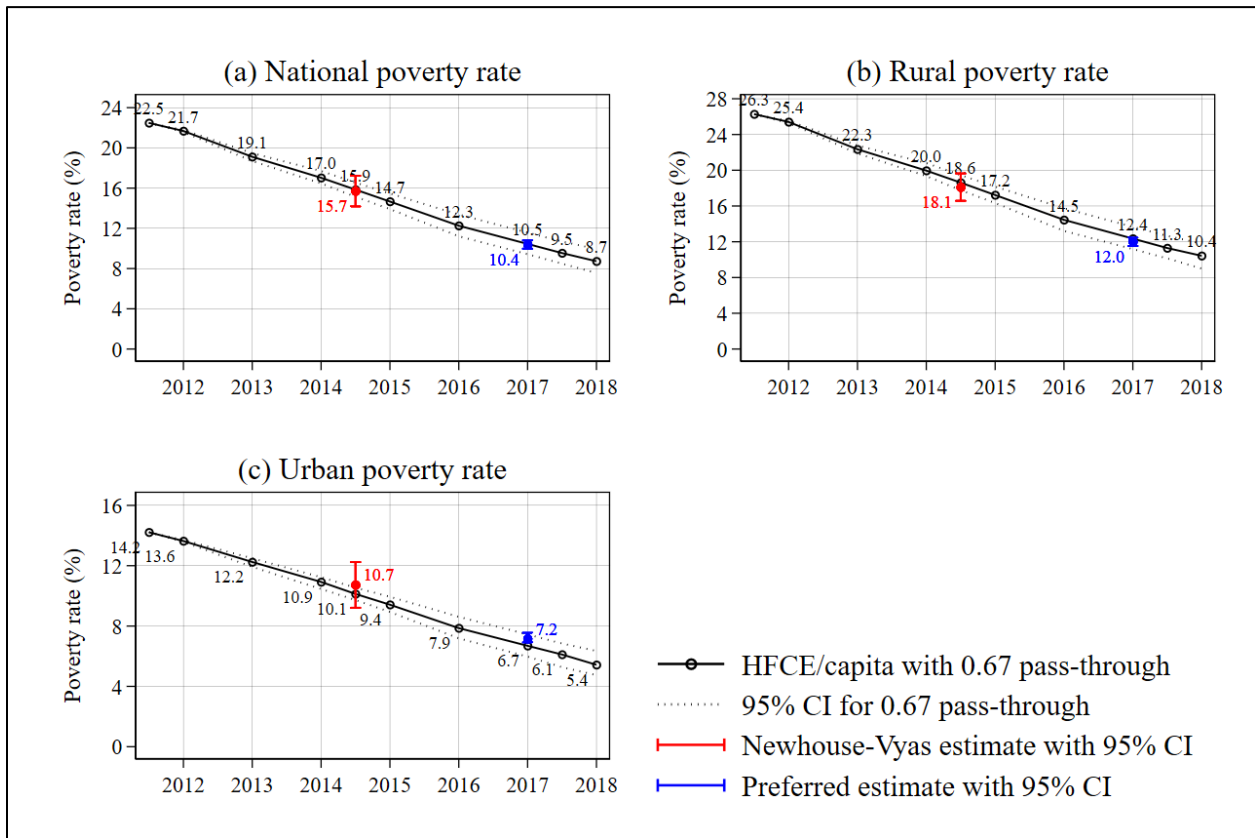
Figure A.2. Poverty nowcasts using 0.67 pass-through on official per capita HFCE growth rate for the entire 2012-2018 period

Note: This figure reports poverty estimates for years 2012-2018 using the official per capita HFCE growth rate with 0.67 pass-through applied to the 2011/12 survey and poverty rates using the 95 percent confidence interval around the 0.67 pass-through rate. It also reports estimates for 2014/15 from Newhouse-Vyas (2018) and the preferred estimate with 95 percent confidence interval for 2017. The 2017 preferred estimate is calculated using official per capita HFCE growth rate with 0.67 pass-through applied after for years 2016-2018. Also see Table 2 in the main text.

Figure A.3. Poverty trends with changes to the Gini index
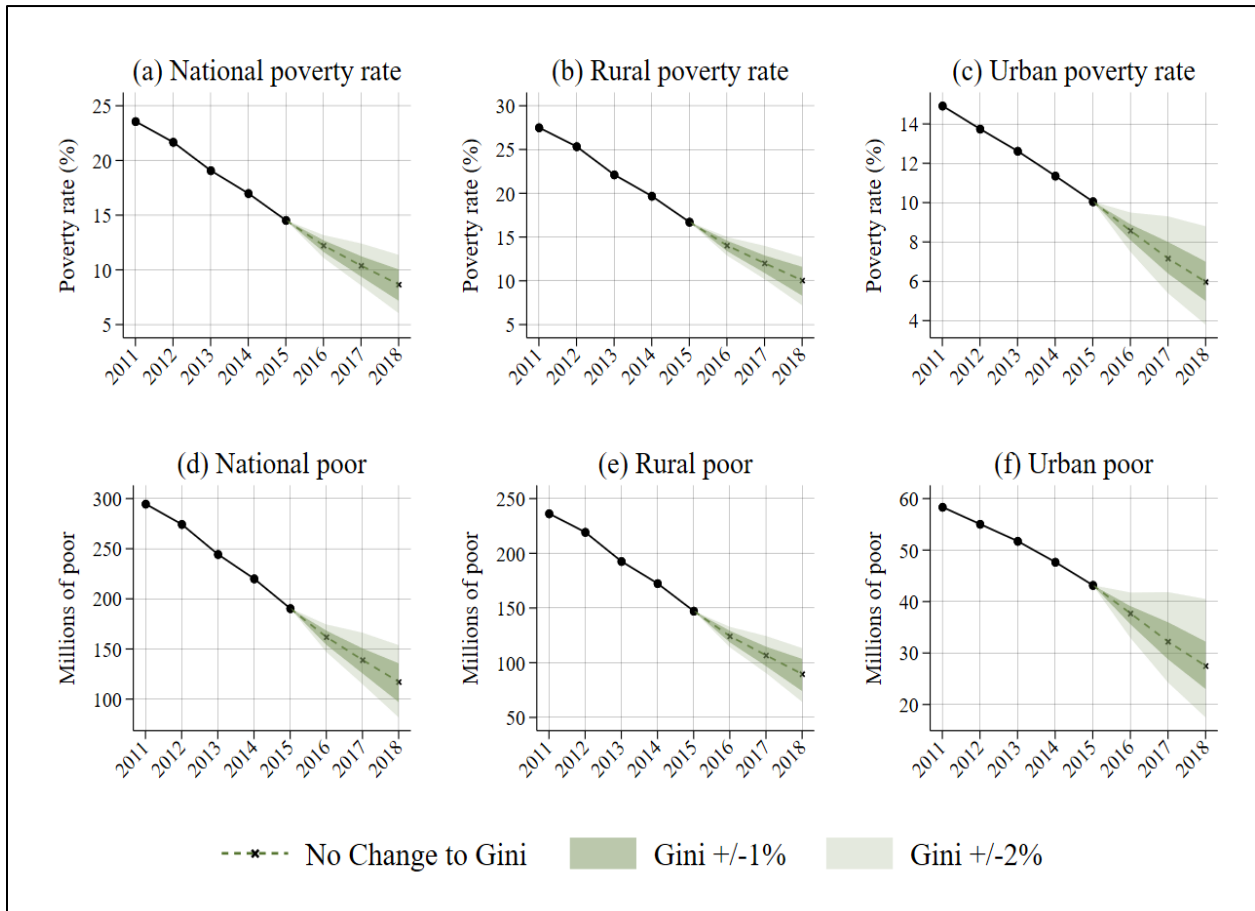
Note: This figure presents the poverty rates and number of poor at the national, rural, and urban levels for years 2011-2018. The source for poverty rates for the lineup years 2012-2015 is PovcalNet. For years 2016-2018, the figure shows poverty rates and number of poor by changing the Gini index by +/- 1 percent and +/- 2 percent annually, while using a 0.67 pass-through on official per capita HFCE growth.

Bibliography

Altimir, O. 1987. "Income distribution statistics in Latin American and their reliability." *Review of Income and Wealth* 33 (2).

Araar, A., and J. Y. Duclos. 2013. *DASP: Distributive Analysis Data Package.* User Manual. DASP Version 2.3, University Laval.

Atamanov, A., C. Lakner, D.G. Mahler, S. Tetteh Baah, and J. Yang. 2020. "The effect of new PPP estimates on global poverty: a first look." *Global Poverty Monitoring Technical Notes* (The World Bank) 12.

Atamanov, A., R. A. C. Aguilar, C. Diaz-Bonilla, D. Jolliffe, C. Lakner, D. G. Mahler, J. Montes, et al. 2019. "September 2019 PovcalNet Update: What's New." *Global Poverty Monitoring Technical Note* 10.

Bhalla, S.S., and K. Bhasin. 2020. "Separating fact from economic fiction: growth slowed beyond expectations starting late 2018." *Indian Express*, January 25. https://indianexpress.com/article/opinion/columns/economy-slowdown-gdp-growth-5-trillion-nirmala-sitharman-6234056/.

Birdsall, N., N. Lustig, and C. J. Meyer. 2014. "The strugglers: The new poor in Latin America." *World Development* 60: 132-146.

Blinder, A.S. 1973. "Wage discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 8 (4): 438-455.

Bourguignon, F. 2015. "Appraising income inequality databases in Latin America." *Journal of Economic Inequality* 13 (4): 557-578.

Breiman, L., J. Friedman, C. Stone, and R. Olshen. 1984. *Classification and Regression Trees.* Taylor and Francis.

Castañeda A., R.A., T. Fujs, D. Jolliffe, C. Lakner, D. Gerszon Mahler, M.C. Nguyen, M. Schoch, et al. 2020. *September 2020 PovcalNet Update: What's New.* Washington DC: World Bank.

Chanda, A., and C.J. Cook. 2019. "Who gained from India's demonetization? Insights from satellites and surveys." *Department of Economics Working Paper* (Louisiana Stata University) 16.

Chandy, L., N. Ledlie, and V. Penciakova. 2013. "The final countdown: Prospects for ending extreme poverty by 2030." *Global Views Policy Paper* (The Brookings Institution).

Chen, S., and M. Ravallion. 2010. "The Develping World is Poorer than We Thought, but No Less Successful in the Fight Against Poverty." *The Quarterly Journal of Economics* 125 (4): 1577-1625.

Chen, S., D.M. Jolliffe, C. Lakner, K. Lee, D.G. Mahler, R. Mungai, M.C. Nguyen, et al. 2018. "PovcalNet update: What's new." *Global Poverty Monitoring Technical Note* (The Wrold Bank) 2.

Chodrow-Reich, G, G. Gopinath, P. Mishra, and A. Narayanan. 2020. "Cash and the Economy: Evidence from India's Demonetization." *Quarterly Journal of Economics* 135 (1): 57-103.

Corral, P., A. Irwin, N. Krishnan, D.G. Mahler, and T. Vishwanath. 2020. *On the frontlines of the fight against poverty.* Washington, DC: World Bank.

Datt, G., and M. Ravallion. 2002. "Is India's Economic growth eaving the poor behind?" *Journal of Economic Perspectives* 16 (3): 89-108.

Deaton, A., and V. Kozel. 2005. "Data and Dogma: the great Indian poverty debate." *The World Bank Research Observer* 20 (2): 177-199.

Deaton, A. 2005. "Measuring Poverty in a growing world (or measuring growth in a poor world)." *Review of Economics and Statistics* 87 (1): 1-19.

Douidich, M., A. Ezzrari, R. Van der Weide, and P. Verme. 2016. "Estimating quarterly poverty rates using labor force surveys: a primer." *The World Bank Economic Review* 30 (3): 475-500.

Elbers, C., J.O. Lanjouw, and P. Lanjouw. 2003. "Micro-level Estimation of Poverty and Inequality." *Econometrica* 71: 355-364.

Figari, F., A. Paulus, and H. Sutherland. 2015. *Microsimulation and Policy Analysis.* Vol. 2B, in *Handbook of Income Distribution*, 2141-2221. Elsevier B.V.

Government of India, Ministry of Earth Sciences, India Metereological Department. 2017. *Annual Report.* New Delhi.

Goyal, A., and A. Kumar. 2019. "Indian Growth is not Overestimated: Mr. Subramanian you Got it Wrong." *Indira Gandhi Institute of Development Research* Working paper (19).

Himanshu. 2019. "Opinion: What happened to poverty during the first term of Modi?" *Live Mint*, August 15. https://www.livemint.com/opinion/columns/opinion-what-happened-to-poverty-during-the-first-term-of-modi-1565886742501.html.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. *An Introduction to Statistical Learning.* New York: Springer.

Jha, S. 2019. "Consumer spend sees first fall in 4 decades on weak rural demand: NSO data." *Business Standard*, November 15. https://www.business-standard.com/article/economy-policy/consumer-spend-sees-first-fall-in-4-decades-on-weak-rural-demand-nso-data-119111401975_1.html.

—. 2019. "Govt scraps NSO's consumer expenditure survey over "Data Quality." *Business Standard*, November 16. https://www.business-standard.com/article/economy-

policy/govt-scraps-nso-s-consumer-expenditure-survey-over-data-quality-119111501838_1.html.

Kitagawa, E. 1955. "Components of a DIfference Between Two Rates." *Journal of the American Statistical Association* 50 (272): 1168-1194.

Kulsehrestha, A.C., and A. Kar. 2005. "Consumer Expenditure from the National Accounts and National Sample Survey." In *The Great Indian Poverty Debate*, by A. Deaton and V. Kozel. New Delhi: Macmillan.

Lahiri, A. 2020. "The great Indian demonetization." *Journal of Economic Perspectives* 34 (1): 55-74.

Lakner, C., D. G. Mahler, M. Negre, and E. B. Prydz. [Forthcoming]. "How Much Does Reducing Inequality Matter for Global Poverty?" *Journal of Economic Inequality*.

Lakner, C., M. Negre, and E. B. Prydz. 2014. "Twinning the goals: how can promoting shared prosperity help to reduce the global poor?" *World Bank Policy Research Working Paper* 7106.

Minhas, B.S. 1988. "Validation of large-scale sample survey data: Case of NSS Household Consumption Expenditure." *Sankhya Series B* 50 ((supp.)): 1-63.

Mukherjee, M., and G.S. Chatterjhee. 1974. "On the validity of NSS estimate of consumtpion expenditure." In *Poverty and Income Distribution in India*, by P. Bardhan and T.N. Srinivasan. Calcutta: Statistical Publishing Society.

Newhouse, D., and P. Vyas. 2019. "Estimating Poverty in India without Expenditure Data. A Survey-to-survey imputation approach." *World Bank Policy Research Working Paper* 8878.

Newhouse, D., and P. Vyas. 2018. "Nowcasting Poverty in India for 2014-15: A survey-to-survey imputaiton approach." *Global Poverty Monitoring Technical Note* (World Bank) 6.

Nguyen, M.C., P. Corral, J.P. Azevedo, and Q. Zhao. 2018. "Sae: A STATA Package for Unit Level Small area Estimation." *World Bank Policy Research Working Paper* 8630.

Oaxaca, R. 1973. "Male-Female Wage differentials in Urban Labor Markets." *International Economic Review* 693-709.

Pinkovskiy, M., and X. Sala-i-Martin. 2016. "Lights, Camera ... Income! Illuminating the national accounts-household surveys debate." *The Quarterly Journal of Economics* 131 (2): 579-631.

Prydz, E. B., D. M. Jolliffe, and U. Serajuddin. [Forthcoming]. "Mind the gap: disparities in assessments of living standards using national accounts and surveys." *Review of Income and Wealth*.

Prydz, E. B., D. M. Jolliffe, C. Lakner, D. G. Mahler, and P. Sangraula. 2019. "National Accounts Data used in Global Poverty Measurement." *Global Poverty Monitoring Technical Note* (World Bank) (No. 8).

Ravallion, M. 2003. "Measuring Aggregate Welfare in Developing Countries: How well do national accounts and surveys agree?" *The Review of Economics and Statistics* 85 (3): 645-652.

Rubin, D.B. 2004. *Multiple Imputation for Nonresponse in Surveys.* John Wiley and Sons.

Sen, A. 2000. "Estimates of Consumer Expenditure and its Distribution: Statistical Priorities after the NSS 55th Round." *Economic and Political Weekly*, 4499-4501.

Sengupta, R. 2016. "IGIDR Conference on GDP measurement issues. A summary of discussions." Mumbai: Indira Gandhi INstitute of Development and Research. http://www.igidr.ac.in/pdf/conference/GDPConferenceSummary_IGIDRAug5.pdf.

Subramanian, A. 2019. "India's GDP Mis-estimation: Likelihood, Magnitudes, Mechanisms and Implications." *CID Faculty Working Paper* 354.

Subramanian, S. 2019. "What is happening to Rural Welfare, Poverty and Inequality in India?" *India Forum*, December 12. https://www.theindiaforum.in/article/what-happened-rural-welfare-poverty-and-inequality-india-between-2011-12-and-2017-18.

Sundaram, K., and S. Tendulkar. 2003. "Poverty has declined in the 1990s: A resolution of the comparability problems in NSS consumer expenditure data." *Economic and Political Weekly* 327-337.

Tibshirani, R. 1996. ""Regression Shrinkage and Selection Via the Lasso"." *Journal of the Royal Statistical Society: Series B (Methodological)* 58: 267-288.

Vyas, M. 2018. "Using fast frequency household survey data to estimate the impact of demonetization on employment." *Review of Market Integration* 10 (3): 159-183.

World Bank. 2015. *A Measured Approach to Ending Poverty and Boosting Shared Prosperity.* Washington, DC: World Bank.

World Bank. 2018. *Poverty and Shared Prosperity 2018: Piecing together the poverty puzzle.* Washington, DC: World Bank.

World Bank. 2020. *Poverty and Shared Prosperity 2020: Reversals of Fortune.* Washington, DC: World Bank.