I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17136

Imputing Poverty Indicators without Consumption Data: An Exploratory Analysis

Hai-Anh H. Dang
Talip Kilic
Kseniya Abanokova
Calogero Carletto

JULY 2024

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17136

Imputing Poverty Indicators without Consumption Data: An Exploratory Analysis

Hai-Anh H. Dang
World Bank, GLO, IZA, Indiana University, and London School of Economics and Political Science

Talip Kilic
World Bank

Kseniya Abanokova
World Bank and National Research University

Calogero Carletto
World Bank

JULY 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Imputing Poverty Indicators without Consumption Data: An Exploratory Analysis*

Accurate poverty measurement relies on household consumption data, but such data are often inadequate, outdated or display inconsistencies over time in poorer countries. To address these data challenges, we employ survey-to-survey imputation to produce estimates for several poverty indicators including headcount poverty, extreme poverty, poverty gap, near-poverty rates, as well as mean consumption levels and the entire consumption distribution. Analyzing 22 multi-topic household surveys conducted over the past decade in Bangladesh, Ethiopia, Malawi, Nigeria, Tanzania, and Vietnam, we find encouraging results. Adding either household utility expenditures or food expenditures to basic imputation models with household-level demographic, employment, and asset variables could improve the probability of imputation accuracy between 0.1 and 0.4. Adding predictors from geospatial data could further increase imputation accuracy. The analysis also shows that a larger time interval between surveys is associated with a lower probability of predicting some poverty indicators, and that a better imputation model goodness-of-fit (R2) does not necessarily help. The results offer cost-saving inputs into future survey design.

JEL Classification: C15, I32, O15

Keywords: consumption, poverty, survey-to-survey imputation, household surveys, Vietnam, Ethiopia, Malawi, Nigeria, Tanzania, Sub-Saharan Africa

Corresponding author:
Hai-Anh H. Dang
World Bank Development Data Group
1818 H Street
Washington, DC
USA
E-mail: hdang@worldbank.org

* We would like to thank Madeleine Gauthier, Nathan Ives, Hoa Nguyen, Lars Osberg, Anne Swindale, Petra Todd, and seminar participants at Australian National University, the Center for the Study of African Economies conference (Oxford), IARIW-TNBS conference (Arusha), Econometric Society's Summer Meeting (Vanderbilt), and Statistics Canada's International Methodology Symposium for their helpful discussion and feedback on the earlier drafts. We are grateful for the funding from the United States Agency for International Development (USAID).

1. Introduction

Accurate poverty measurement is the prerequisite for policies aiming at reducing poverty. Yet, development practitioners face the typical challenges that the available household survey data that underlying poverty estimates are either inadequate (e.g., do not offer nationally representative estimates) or outdated (e.g., do not offer timely estimates of poverty trends). Worse still, in the few countries where survey capacity is well established, data were known to turn out to exhibit varying degrees of incompatibilities over time due to changes with survey design (Deaton and Kozel, 2005). These data challenges could likely hinder effective policy implementation, especially for poorer countries with low statistical capacity (Devarajan, 2013; Jerven, 2019).[1]

To address these challenges, alternative methods to obtain poverty estimates that rely on data imputation (instead of direct data collection through surveys) have become increasingly common (World Bank, 2021; Dang and Lanjouw, 2023).[2] Building on the seminal technique that imputes from a household consumption survey into a census to generate poverty maps (Elbers *et al.,* 2003), recent studies have imputed from a household consumption survey into another survey to provide poverty estimates.[3] The central idea is to build an imputation model using appropriate predictor variables from an existing older consumption survey, which can be subsequently applied to the same variables in a more recent survey (that does not collect consumption data) to provide poverty estimates for the latter survey.

[1] Serajuddin *et al.* (2015) show that over the period 2002- 2011, of the 155 countries for which the World Bank monitors poverty data using the World Development Indicators (WDI) database, almost one-fifth (i.e., 28) have only one poverty data point and as many as 29 countries do not have any poverty data point in the same period. Furthermore, poorer countries have fewer surveys: a 10-percent increase in a country's household consumption level is associated with almost one-third (i.e., 0.3) more surveys (Dang, Jolliffe, and Carletto, 2019). The ongoing Covid-19 pandemic could increase poverty and further exacerbate these data deprivations and digital divides for poor countries (Naude and Vinuesa, 2020).

[2] Imputation techniques are regularly used by international organizations and national statistical agencies to fill in missing data gaps such as education statistics (UOE, 2020) and income data (US Census Bureau, 2017).

[3] The poverty-mapping technique combines a household consumption survey and a non-consumption census, which allows us to provide poverty estimates at a more disaggregated level than available in the household survey.

Building on Elbers *et al.*'s (2003) method, recent studies have innovated in various aspects. These include combining data between a household consumption survey and a different survey (Stifel and Christiaensen, 2007; Douidich *et al.*, 2016), modelling techniques for the error terms or standard errors (Tarozzi, 2007; Mathiassen, 2009; Dang, Lanjouw, and Serajuddin, 2017), and experimenting with survey design and selecting suitable variables (Kilic and Sohnesen, 2019; Christiaensen, Ligon, and Sohnesen, 2022; Dang *et al.*, forthcoming). Most recently, poverty imputation has been employed to provide estimates for hard-to-find refugee population groups that are not typically captured in the standard household survey (Altındağ et al., 2021; Beltramo *et al.*, 2024; Dang and Verme, 2023).

Reviewing some key studies in the past 20 years covering poor and middle-income countries ranging from India, Jordan, and Sub-Saharan African countries to Vietnam, Dang *et al.* (forthcoming) observe that imputation-based poverty estimates can perform reasonably well against the survey-based poverty estimates using actual consumption data. Further analyzing data from 14 rounds of multi-topic household surveys conducted over the past decade in Ethiopia, Malawi, Nigeria, Tanzania, and Vietnam, the authors find that rather parsimonious imputation models consisting of household-level demographic and employment variables and household utility expenditures could provide accurate estimates, which even fall within the more rigorous precision criteria of being within one standard error of the true poverty rates in many cases.

This paper makes several new contributions to the literature on survey-to-survey imputation of poverty estimates, both conceptually and empirically. On the conceptual front, we significantly expand this literature to various common poverty indicators such as i) near-poverty (vulnerability) status, ii) extreme poverty, iii) poverty gap, and iv) other Foster, Greer and Thorbecke (FGT) poverty indices. Furthermore, we also examine the performance of the imputed consumption distribution against the distribution of the actual household consumption data, which underlie these poverty indicators. These extensions set our paper apart from the existing literature, which almost exclusively focuses on the headcount poverty rate. Indeed, to our knowledge, this is the first study that attempts to provide a comprehensive and systematic examination of these various poverty indicators as well as the entire consumption distribution.

Empirically, for illustrations we harmonize and rigorously analyze data from 22 recent rounds of multi-topic household surveys conducted over the past decade in Bangladesh, Ethiopia, Malawi, Nigeria, Tanzania, and Vietnam. These six countries span three regions (i.e., Sub-Saharan Africa, South Asia, and Southeast Asia) and different income levels (i.e., low-income to lower-middle-income), thus exhibit more heterogeneity regarding income levels, geographical variations, and population sizes than previous studies. To our knowledge, our study offers an application of survey-to-survey imputation to the most comprehensive dataset that has been analyzed to date. Consequently, our findings would make a useful contribution to future survey-to-survey imputation efforts.[4]

We find that (imputation) model heterogeneity exists, with certain models performing better for some poverty indicators and the consumption distribution only. In particular, two models perform better than the others. One model consists of adding food expenditures to household demographic, employment characteristics, and house assets (Model 3), and the other model consists of adding household utility consumption expenditures (including electricity, water, and garbage) to household demographic and employment characteristics (Model 9). Model 3 works reasonably well for headcount poverty, extreme poverty, poverty gap, and consumption mean,

[4] The existing study with the most comprehensive dataset is Dang *et al.* (forthcoming), which analyzes data from 14 survey rounds in Ethiopia, Malawi, Nigeria, Tanzania, and Vietnam. This study focuses on headcount poverty alone.

raising the probability of accurate imputation for these indicators by around 0.3 (compared to a reference model with just household demographic and employment characteristics). Compared to Model 3, Model 9 performs slightly better for headcount poverty, raising the probability of imputation accuracy by 0.4. It also raises the probability of imputation accuracy for near-poverty, extreme poverty, poverty gap, and mean consumption by around 0.1-0.2.

Further adding agricultural soil quality information to Model 9 results in higher imputation accuracy (and stronger statistical significance) for headcount poverty, increasing the probability of imputation accuracy by 0.5. Models 3 and 9 also perform better than the other for imputing the consumption distribution. Finally, a larger time interval between the base survey and the target survey is associated with lower imputation accuracy, but a better model goodness-of-fit (R^2) does not appear to help. Further robustness analysis shows that the proposed method works better than some common machine learning techniques.

This paper consists of six sections. We discuss the analytical framework in the next section before describing the data in Section 3. We subsequently present in Section 4 the main estimation results using the latest survey rounds for each country before summarizing the results using all the available survey rounds for all the countries (Section 4.1). We further extend the analysis to more general setting (Section 4.2), such as such as using other FGT indexes that are more sensitive to the poor and estimates for the entire consumption distribution before discussing a more specific application, within-year imputation. We offer meta-analysis results on model selection in Section 5 and finally conclude in Section 6.

2. Analytical Framework

2.1. Imputation Model

A household maximizes utility subject to an income budget constraint that includes choice variables such as quantities of goods, durables, and leisure (or labor supply) (Deaton and Muellbauer, 1980). This results in the common practice that total household consumption is constructed as an aggregate of consumption of different items such as food, non-food (including clothing, education, and/or health expenses), durable goods, and housing (Deaton and Zaidi, 2002). It follows that a model of (log) household consumption per capita (y_j) is typically estimated using the following reduced-form linear model for survey *j*, for *j*= 1, 2,

$$y_j = \beta_j{}' x_j + \mu_j \tag{1}$$

where x_j can include household variables such as the household head's age, sex, education, occupation, ethnicity, religion, and language—which can represent household tastes.[5] x_j can also include household assets or incomes, and μ_{ij} is the error term (see, e.g., Elbers *et al.*, 2003; Ravallion, 2016).

We employ the survey-to-census imputation framework that was first introduced by Elbers *et al.* (2003), which was subsequently refined by Dang *et al.*'s (2017) for survey-to-survey imputation. This method has been validated and applied to data from poor and middle-income countries in different regions ranging from India, Jordan, Tunisia, and Sub-Saharan African countries to Vietnam (Beegle *et al.*, 2016; Cuesta and Ibarra, 2017; Dang and Lanjouw, 2023). Recent applications of this method include providing poverty estimates for refugees (Dang and Verme, 2023; Beltramo *et al.*, 2024).

---

[5] More generally, *j* can be larger than 2 and can indicate any type of relevant surveys that collect household data sufficiently relevant for imputation purposes such as labor force surveys or demographic and health surveys. To make the notation less cluttered, we do not show the subscript for households in the equations. It is also standard practice with household survey analysis to transform the consumption variable to logarithmic scale to help improve the model fit.

We briefly describe the method next. For better accuracy, the error term μ_j is further broken down into two components, a cluster random effects (υ_{cj}) and an idiosyncratic error term (ε_j). Conditional on the x_j characteristics, the cluster random effects and the error term are assumed uncorrelated with each other and to follow a normal distribution such that $\upsilon_{cj}|x_j \sim N(0, \sigma^2_{\upsilon_j})$ and $\varepsilon_j|x_j \sim N(0, \sigma^2_{\varepsilon_j})$. We relax this assumption later and employ an alternative approach where we use the empirical distribution of the error terms instead.

Household consumption (or income) data exist in one survey but are missing in the other survey, thus without loss of generality, let survey 1 and survey 2 respectively represent the survey with and without household consumption data, and y_1 represent household consumption in survey 1. More generally, these two surveys can be either in the same period or in different periods. Our objective is thus to impute the missing consumption data in survey 2, given that consumption data is available in survey 1 *only*, and the survey characteristics x_j are available in *both* surveys. Note that while we do have consumption data for survey 2, for validation purposes, we assume that household consumption data in this survey round were unavailable.

Writing out Equation (1) we have

$$y_1 = \beta_1{}' x_1 + \upsilon_{c1} + \varepsilon_1 \qquad (2)$$

Equation (2) provides a standard linear random effects model that can be estimated using most available statistical packages. Applying the parameters obtained from Equation (2) to the variables in survey 2, the imputed household consumption in this survey round is given by[6]

[6] This assumes that the returns to the characteristics x_j are captured by equations (1) and (2) and precludes the (perhaps exceptionally) rare situations where there could be no correlation between these characteristics and household consumption due to unexpected upheavals in the economy or calamitous disasters. Contexts where there are sudden changes to the economic structures (e.g., overnight regime change) may also introduce noise into the comparability of the estimated parameters, but (variants of) this imputation approach has been found to be rather robust to such changes; see our discussion later.

$$y_2^1 = \beta_1{}'x_2 + \upsilon_{c1} + \varepsilon_1 \quad (3)$$

While equations (1) and (2) can also be specified as a simple OLS model (i.e., with the random effects υ_{cj} being subsumed into the error terms), modelling the random effects explicitly would help improve the precision of estimation results. Indeed, the advantage of the random effects model over the OLS model is that the former can better capture the between-cluster variations thanks to the additional information offered by the random effects. This role of υ_{cj} is especially important under our estimation framework since the random effects are instrumental not only in estimating β_j but also our estimates of poverty in survey 2 as a component of the predicted household consumption. Put differently, υ_{cj} is utilized for both for the point estimate of poverty in survey 2 and its standard errors.

We are most interested in the poverty estimates for survey 2, where the consumption data are missing. Let z_2 be the poverty line in period 2; if y_2 existed the poverty rate P_2 in this period could be estimated with the following quantity

$$P(y_2 \leq z_2) \quad (4)$$

where *P(.)* is the probability (or poverty) function that gives the percentage of the population that are under the poverty line z_2 in survey 2. Since poverty has an inverse relationship with household consumption (i.e., richer households are less likely to be poor), this function is generally non-increasing in household consumption.[7]

We further make the following assumptions that underlie the theoretical framework, which we will relax and offer validation tests for in subsequent sections.

---

[7] If we impose a more restrictive assumption that follows the standard normal distribution and combine the estimation of equations (2) and (3) in the same step, then the probit model that directly estimates poverty results (i.e., the estimating equation is, with j= 1, 2, where is the cumulative normal distribution). Note that we also assume homoscedasticity of the error terms for simplicity.

Assumption 1: Let x_j denote the values of the variables observed in survey j, for j= 1, 2, and let X_j denote the corresponding measurements in the population. Then x_j are consistent measures of X_j for all j (i.e., x_j=X_j for all j).

Assumption 1 is crucial for imputation and ensures that the sampled data in survey 1 and survey 2 are each representative of the target population. Put differently, this assumption implies that, for two contemporaneous (i.e., implemented in the same time period) surveys, measurements of the same characteristics *x* are identical (except for potential sampling errors) since they are consistent measures of the population values; for two non-contemporaneous surveys, these estimates from the two surveys are consistent and comparable over time. While surveys of the same design (and sample frame) are more likely to be comparable and can thus satisfy Assumption 1, there is no *a priori* guarantee that these surveys can provide comparable estimate across two different time periods, or even the same estimates in the same time periods. Examples where Assumption 1 may be violated include the cases where national statistical agencies change the questionnaire for the same survey over time, or where one considers different surveys that focus on different population groups (e.g., the average household size may differ between a household survey and a labor force survey depending on the specific definition that is used). Violation of Assumption 1 rules out the straightforward application of survey-to-survey imputation technique and would require that additional assumptions be made on the relevance of the estimated parameters from one survey to the other.

Assumption 2: Let ΔP and Δx respectively represent the changes in poverty rates and the explanatory variables x over time, and Θ_j the set of parameters ($\beta_j, \sigma^2_{\upsilon_{cj}}, \sigma^2_{\varepsilon_j}$) that map the variables x into the household consumption space in period j where the consumption data are available. Then $\Delta P = P(\Delta x|\Theta_j)$, where P(.) is the given poverty function.

Assumption 2 implies that, given Θ_j or the estimated consumption parameters from survey 1, the changes in the explanatory variables *x* between the two periods can capture the change in poverty rate in the next period. More intuitively, given the commonly observed variables in the two surveys and their linkage to household consumption, this assumption allows the imputation of the missing household consumption for survey 2. In practical terms it implies that the change in poverty rates over time is attributable to changes in the explanatory variables x rather than the returns to characteristics (or economic structure) and the unexplained characteristics (or random shocks)—which are respectively represented by β_1 and $(\upsilon_{c1}, \varepsilon_1)$. Clearly, this is a testable assumption if household consumption is available for both of the periods under consideration.

As discussed earlier, previous studies commonly assume that the distributions of the household consumption parameters β_1, υ_{c1} and ε_1 in equations (2) and (3) based on the data in survey (or period) 1 remain the same for the data in survey (period) 2. Assumption 2 is less restrictive since it allows the distributions of these estimated parameters to change over time, as long as the changes in the variables x *alone* can correctly capture the change in poverty rate. Assumption 2 only requires that overall, the parts of the consumption distributions below the poverty line for both periods (that can be explained by the changes in x in our model) be equal and not all the percentiles along the consumption distributions be equal as implied by the assumption made in existing studies; this result is formally stated in Corollary 1.2 below.

Given Assumptions 1 and 2, Dang *et al.* (2017) provide the following proposition that lays out the estimation framework.

Proposition 1: Imputation framework

Given Assumptions 1 and 2, the poverty rate for period 2 can be predicted using the estimated consumption parameters based on survey 1 and the data in survey 2. In particular, let P(.) be the poverty function and y_2^1 be defined as $\beta_1' x_2 + \upsilon_{c1} + \varepsilon_1$, we have

$$P(y_2) = P(y_2^1) \tag{5}$$

Corollary 1.1

Let $\hat{\beta}_1, \hat{\sigma}^2_{\upsilon_{c1}}$ and $\hat{\sigma}^2_{\varepsilon_1}$ represent the estimated parameters obtained from equation (2) and let $\hat{y}^1_{2,s} = \hat{\beta}_1{}'x_2 + \tilde{\hat{\upsilon}}_{c1,s} + \tilde{\hat{\varepsilon}}_{1,s}$, where $\tilde{\hat{\upsilon}}_{c1,s}$ and $\tilde{\hat{\varepsilon}}_{1,s}$ represent the s^{th} random draw from their estimated distributions, for s= 1,…, S. The poverty rate P_2 in period 2 can be estimated as

$$\hat{P}_2 = \frac{1}{S}\sum_{s=1}^{S} P(\hat{y}^1_{2,s} \leq z_1) \tag{6}$$

Corollary 1.2

Instead of Assumption 2, assume the traditional but more restrictive assumption that the consumption model parameters and the distributions of the error terms in equation 1 remain the same in period 2 (that is $\beta_1 \equiv \beta_2$, and υ_{c1} and ε_1 have the same distributions as υ_{c2} and ε_2 respectively). Given Assumption 1 and this stricter assumption, we have

$$W(y_2) = W(y^1_2) \tag{7}$$

where W(.) is a general one-to-one mapping welfare function, which includes the poverty function P(.) as a special case.

Proof.
See Dang *et al.* (2017).

2.2. Welfare Indicators

The poverty indicators that we estimate generally belong to the Foster, Greer and Thorbecke (FGT) (1984) class. Consider N - a population of income-receiving units (persons or households), $i = 1, \dots N$, with income y_i and weight w_i. Let $N = \sum_{i=1}^{n} w_i$, when the data are unweighted $w_i = 1$ and $N = n$. The poverty line is z and the income gap up to the poverty line for person i is $max(0, z - y_i)$. The FGT class of poverty indices is given by

$$FGT(y; \alpha) = \sum_{i=1}^{N} \frac{w_i}{N} \left[\frac{(z-y_i)}{z}\right]^{\alpha} I_i \tag{8}$$

where $I_i = 1$ if $y_i \leq z$ and $I_i = 0$ otherwise. α is a given parameter, whose first three non-negative integer values are most commonly used. In particular, $FGT(y; 0)$ is the headcount poverty ratio, $FGT(y; 1)$ is the (average normalized) poverty gap, and $FGT(y; 2)$ is the (average normalized) poverty gap squared. The larger α is, the greater the degree of poverty aversion is (i.e., more weights are placed on poorer individuals).

The poverty gap measurement as defined by USAID is a modified version of $FGT(y;1)$, which only applies to the poor population

$$FGT(y;1)^{USAID} = \sum_{i=1}^{N_p} \frac{w_i}{N_p} \left[\frac{(z-y_i)}{z}\right] I_i \tag{9}$$

where N_p is the number of poor people (i.e., those with income below poverty line z). Hereafter we refer to this indicator as the USAID poverty gap.

The near-poverty (vulnerability) rate represents the proportion of the population with an income above the poverty line but below the vulnerable line V

$$P_o = \frac{w_i}{N} \sum_{i=1}^{N} I_{i,v}(z < y_i \leq V) \tag{10}$$

where $I_{i,v} = 1$ if $z < y_i \leq V$ is true and $I_i = 0$ otherwise. V is defined as 1.25 times of the poverty line for our analysis.

Finally, in addition to the FGT indexes, we also provide imputed estimates of the general distribution of household consumption y_i, which underlies estimation of all the poverty indicators discussed above. This outcome extends our focus on the poorer part of the consumption distribution to the whole distribution.

Some additional remarks are useful. First, since the FGT class of poverty indices is monotonic (Foster *et al.*, 2010), it satisfies the one-to-one mapping condition for the welfare function *W(.)* in Corollary (1.2). Consequently, the imputed poverty estimates are asymptotically equivalent to the true poverty indicators in Equations (8) to (10). Second, since it is not straightforward to obtain analytical formulae for the standard errors for estimates of the poverty indicators in Equations (8) to (10), we provide the bootstrap standard errors for these estimates.[8] Third, the poverty line and

[8] Dang *et al.* (2017) offer an analytical formula for the standard error of the estimated headcount poverty rate, which provide similar estimates to those based on the bootstrap standard errors that we obtain.

the extreme poverty line vary in different countries, so we employ those that are commonly used for each country. We come back to more discussion on the (extreme) poverty lines in the next section.

3. Data

We analyze multi-topic household survey data from a total of 22 survey rounds from five different countries: Bangladesh (3), Ethiopia (1), Malawi (5), Nigeria (3), Tanzania (6), and Vietnam (4), with the number of survey rounds for each country being noted in parenthesis. In the four Sub-Saharan African countries (Ethiopia, Malawi, Nigeria, and Tanzania), the data originate from the nationally-representative, multi-topic household surveys that have been implemented by the respective national statistical office with support from the World Bank Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) initiative. Being similar to the LSMS-type surveys supported by the World Bank, the surveys from Vietnam are implemented biennially by the country's General Statistical Office (GSO) with technical support from the World Bank. These surveys are generally regarded as being of high quality and are regularly employed by the national governments, international organizations, and academic researchers to provide estimates on household welfare.[9]

The data sets include

i. the Bangladesh Integrated Household Survey (BIHS) 2011/12, 2015, and 2018/19
ii. the Ethiopia Socioeconomic Survey (ESS), 2018/19 round
iii. the Malawi Integrated Household Survey (IHS), 2010/11, 2016/17, 2019/20 rounds

[9] For example, Baulch (2011) considers the VHLSSs as having high quality data and heavily use these surveys for poverty analysis. Other researchers analyze the LSMS-ISA surveys for various topics such as agricultural input uses (Sheahan and Barrett, 2017) or temperature shocks and household consumption (Letta, Montalbano, and Tol, 2018).

iv. the Malawi Integrated Household Panel Survey (IHPS), 2010 and 2013 rounds

v. the Nigeria General Household Survey (GHS)–Panel, 2010/11, 2012/13 and 2018/19 rounds[10]

vi. the Tanzania National Panel Survey (TZNPS) 2008/09, 2010/11, 2012/13, 2014/15, 2019/20, and 2020/21 rounds.

vii. the Vietnam Household Living Standards Survey (VHLSS) 2010, 2012, 2014, and 2016 rounds.

The sample sizes hover around 3,000 to 5,000 households in each survey round for the LSMS-ISA surveys (including Nigeria and Tanzania), 5,500-7,000 households for the BIHSs and the ESS, 9,300 households for the VHLSSs, and over 12,000 households for the Malawi IHS. The consumption data are deflated in the same survey year's prices and are comparable across survey rounds for each country.[11] The objective is to produce the imputation-based welfare estimates of interest as if we did not have consumption data and then evaluate these imputation-based estimates against those based on the actual survey data (i.e., the "true" welfare rates).

For the poverty line, we use the national poverty lines for Ethiopia, Malawi, Tanzania, and Vietnam and the international poverty lines of $1.90 (in 2011 Purchasing Power Parity (PPP) prices) for Bangladesh and Nigeria.[12] The extreme poverty line is defined as US$1.25 (2011 PPP)

[10] We did not include the 2015/16 round for Nigeria for comparability issues (e.g., the total consumption aggregate for this round does not include healthcare expenditures and it is not adjusted using temporal and spatial price deflators).

[11] In particular, for Bangladesh, Tanzania, and Vietnam, consumption data are deflated to 2018/19 prices, 2020/21 prices, 2010 prices respectively. For the Malawi IHPSs and IHSs, consumption data are deflated to 2013 prices and 2010/11 prices respectively.

[12] For Bangladesh and Nigeria, we employ the international poverty line for analysis since official poverty data for these countries are based on different data sources that are not available to us, such as the Bangladesh Household Income Expenditure Survey (HIES) and the Nigeria Living Standards Surveys (NLSSs). However, note that Nigeria's national poverty line calculated using NLSS 2018/19 is close to the international poverty line of $1.90 per person per day in 2011 PPP (Lain and Vishwanath, 2022). Furthermore, data comparability issues also exist with various rounds of the Bangladesh HIESs (Fernandez *et al.*, 2024).

per day per capita for Bangladesh and Nigeria and half of the national poverty line for Vietnam and Ethiopia. For Malawi and Tanzania, we use the national food poverty lines as the extreme poverty lines.

We prepare and add several geospatial variables for Malawi, Nigeria, Tanzania, and Vietnam, including the distances from the commune center to various important locations (e.g., the nearest major road and the nearest international land border crossing) and agricultural soil quality. These data are obtained from various sources including the Food and Agriculture Organization (FAO) and are provided together with the LSMS-ISA public use data sets. The exception is Vietnam where we process these data separately and we could add nightlight intensity data for this country.[13]

There are two data limitations with the BIHSs. The BIHS questionnaires provide inconsistent variables for utilities expenditures (Appendix A, Table A.11), and there are no geo-spatial data for this country. Consequently, we do not estimate certain models (Models 8 and 9) for Bangladesh and exclude this country from the discussion on overall imputation accuracy (Section 4.1) and subsequent meta-analysis (Section 5).

The survey rounds listed above share the same sampling frame for each country and are generally regarded as comparable over time by most data users. This satisfies Assumption 1 that the sampled data in round 1 and round 2 are representative of the same population in each period. As LSMS-type surveys, these surveys are also comparable across countries. We provide both across-year and within-year imputation results for all the countries, except for Ethiopia, where we can only analyze one survey round and test within-year imputation.

[13] See Tanzania's National Bureau of Statistics (2011) for more discussion on the geospatial variables in the context of this country. For Vietnam, we collect and process data from various public data sources including Harmonized World Soil Database, Open Street Map, and NOAA Climate Data.

4. Estimation Results

4.1. Main Results

Main results

To examine the sensitivity of imputation accuracy to various predictor variables, we build the estimation models on a cumulative basis, with the later models sequentially adding more variables to the basic models (Model 1 or Model 2). On the whole, we employ nine core imputation models across five countries.[14] Model 1 is the most parsimonious (or basic) model and consists of household size, household heads' age and gender, household heads' highest completed levels of schooling, a dummy variable indicating whether the head belongs to the ethnic majority group, the shares of household members in the age ranges 0-14, 15-24, 25-59 and 60 and older, a dummy variable indicating whether the head worked in the past 12 months, and a dummy variable indicating urban residence. Model 2 adds household asset variables and house (dwelling) characteristics to Model 1. Household assets include variables indicating whether the household has a car, motorbike, bicycle, desk phone, mobile phone, DVD player, television set, computer, refrigerator, air conditioner, washing machine, or electric fan. House characteristics include the construction materials for the house's roof and wall and the type of water and toilet the household has access.[15] Models 1 and 2 include standard variables available in most LSMS-type surveys and other types of micro surveys.

[14] For misspecified regressions, adding more variables may result in larger inconsistency (Snijders and Bosker, 1994; De Luca, Magnus, and Peracchi, 2018). As such, it is useful to examine imputation accuracy for different models.

[15] For Vietnam, house wall material is assigned numerical values using the following categories: 6 "cement", 5 "brick", 4 "iron/wood", 3 "earth/straw", 2 "bamboo/board", and 1 "others". Toilet type is assigned numerical values using the following categories: 6 "septic", 5 "suilabh", 4 "double septic", 3 "fish bridge", 2 "others", and 1 "none".

Model 3 adds total food expenditures to Model 2, and Model 4 adds total non-food expenditures to Model 2. Models 5 to 8 add to Model 2, respectively, durables expenditures, health expenditures, education expenditures, and utilities expenditures (such as on electricity, water, and garbage). All these expenditures are on a per capita (or per adult equivalent) basis and are converted to logarithmic form. Finally, Model 9 adds utilities expenditures to Model 1.

The specific predictors used in the imputation models for Equation (2) for each country are provided in Appendix A, Tables A.1 to A.6. For comparison purposes and robustness checks, we use two estimation methods with different assumptions about the error terms. Method 1 uses the normal linear regression model (assuming that the distribution of the error terms follows a normal distribution), and Method 2 uses the empirical distribution of the error terms. Both methods include the random effects at the primary sampling unit for each country.

Table 1 (Panel A) provides the imputed poverty rates for 2018/19 for Bangladesh using the 2015 round as the base survey. The estimation results show that all the imputation models, except for Models 1 and 3, provide headcount poverty estimates that are statistically not significantly different (or fall inside the 95 percent confidence interval (CI)) from the "true poverty rate" of 7.3 percent for 2018/19 (i.e., the poverty rate that is estimated using the actual consumption data for 2018/19). Regarding the near-poverty rate, the most basic model (Model 1), as well as Model 3 offer estimates that lie within the 95 percent CI of the true near-poverty rate of 12.2 percent. In fact, the estimates from Model 1 fall inside one standard error of the true near-poverty rate. The estimated extreme poverty rate and poverty gap for Models 3, 4 and 6 fall inside the 95 percent CI from the true rates of, respectively, 0.6 and 1.1 percent for 2018/19. For extreme poverty, Model 4, which controls for non-food expenditure, produces estimates that fall inside one standard error of the true rates. For the USAID poverty gap, Models 3 and 4, which respectively include food or non-food expenditures, work well with estimates that even fall inside one standard error of the true rate.

We turn next to the results for other countries, shown respectively in Table 1 (Panels B to E) for Malawi, Nigeria, Tanzania, and Vietnam[16]. Table 1 (Panel B) provides the predicted poverty rates for 2019/20 for Malawi using the 2016/17 round as the base survey. The estimation results show that six out of nine imputation models, including Model 9, provide headcount poverty estimates that are statistically not significantly different from the true rate of 51.1 percent for 2019/20. These results are consistent with those found by Dang *et al.* (forthcoming). Regarding the near-poverty rate, except for Model 3, all the other models offer estimates that lie within the 95 percent CI of the true near-poverty rate of 14.1 percent. In fact, almost all these estimates, except for Model 8, even fall inside one standard error of the true near-poverty rate. Again, the predicted extreme poverty rate and poverty gap mostly mirror the estimates for the headcount poverty rate, adding Model 5 to the set of the models with estimates that fall inside the 95 percent CI of the true rate of, respectively, 20.6 and 17.1 percent for 2019/20. Except for Model 3, eight out of nine models yield good estimates for the USAID poverty gap, falling inside the 95 percent CI of the true rate. In fact, almost all of these models (i.e., seven of eight models) produce estimates that fall inside one standard error of the true rate of poverty gap.

Notably, fewer models work for Nigeria, which may be due to a longer time gap between the base and target surveys in Nigeria (Table 1, Panel C). While no model works for headcount poverty and extreme poverty, eight out of nine models for the near-poverty rate, including Model 9, produce estimates that fall inside the 95 percent CI of the true rate of 13.7 percent in 2018/19.

[16] Since we only have data for one survey round for Ethiopia, we are unable to provide similar estimates for this country.

Models 1, 2, 7, 8, and 9 provide good estimates for the USAID poverty gap, with estimates even falling inside one standard error of the true rate.

On the other hand, almost all models work for headcount poverty, near- and extreme poverty rates and the USAID poverty gap in Tanzania, except for Model 3 with poverty gap (Table 1, Panel D). For Vietnam, Models 4, 5, and 9 each work for two to three out of four indicators only (Table 1, Panel E). More models work for the USAID poverty gap than for the other four indicators in Vietnam with the estimates from Models 1 to 5 and 7 falling inside the 95 percent CI from the true rate.

The results with imputing for mean consumption show a mixed pattern, with certain models performing better for some countries only (Table 2). Model 3 performs well for Bangladesh and Malawi and Model 9 performs well for Vietnam, while Model 5 performs well for Tanzania and Vietnam. In Tanzania, five out of nine models produce estimates of mean consumption per capita that fall inside the 95 percent CI from the true mean and Model 6 even falls inside one standard error of the true rate.

As an alternative to the normal linear regression model, we employ the empirical distributions of the error terms. The results shown in Appendix A, Tables A.7 and A.8 are qualitatively similar.

Overall imputation accuracy

The results discussed in Tables 1 and 2 use the most recent pair of survey rounds for each country. But we implemented imputation for the other older surveys for all the countries and years available and we also added geospatial variables where data are available. Given the various across-year imputation model variants that we tested for different countries and years, it is useful to summarize the results graphically. We plot in Figure 1 the *imputation accuracy* for 15 different models (of which the last two models with nightlight data are for Vietnam alone), which is defined as the share of the estimates that are not statistically significantly different from the true poverty rate for a model. The measure is computed across all instances of a given model's estimation with a unique pair of a base survey and a target survey in a given country. These models include the core Models 1 to 9 (shown in Tables 1 to 6) and four additional models where we further add geospatial variables to Models 2 and 9.

Regarding headcount poverty, Figure 1 suggests that for the first nine models, Models 3 and 9 perform better than average with an imputation accuracy of, respectively, 65 and 69 percent, followed by Model 8 (50 percent). Adding agricultural soil quality and geospatial characteristics, such as soil index and distance to facilities, significantly improved the prediction of Model 9 up to 70 and 75 percent, respectively, but it does not help to improve Model 3. On the other hand, adding geospatial nightlight information to Model 2 increases the accuracy of the prediction up to 67 percent for Vietnam. Moreover, adding nightlight information to Model 9 increases accuracy up to 83 percent for Vietnam.

Regarding near-poverty, both Model 3 and Model 9 perform better than average with an imputation accuracy of, respectively, 77 and 69 percent, followed by Models 8 and 5 (both up to 65 percent). However, adding geospatial characteristics, such as soil quality and distance to facilities to Model 9, marginally improves imputation accuracy up to 70 percent. Adding nightlight to Model 9 for Vietnam does not help to improve Model 9.

Regarding extreme poverty, Model 3 has the highest imputation accuracy for across all the different models tested – about 69 percent, followed by Model 9 (54 percent) and Model 4 (46 percent). Model 3 also has the highest imputation accuracy for the poverty gap, as it raises the imputation accuracy above the average model performance to 65 percent.

Unlike the indicators discussed above, multiple models perform better than average for the USAD poverty gap. Models 8 and 3 have the highest imputation accuracy of 65 percent, followed by Models 1, 2, 7 and 9 (62 percent) and Models 5 (58 percent).

We further plot in Figure 2 the imputation accuracy for mean consumption. Model 3, again, is the best performer that achieves an imputation accuracy rate of 65 percent, followed by Model 9 with an imputation accuracy rate of 46 percent. Adding soil quality and distance to facilities to Model 9 increase imputation accuracy for this model to 50 percent and adding nightlight to Model 9 increases imputation accuracy to 100 percent for Vietnam.

Machine learning as alternative

We consider machine learning (ML) as an alternative imputation method.[17] The standard ML procedures split a data sample into a training sample (to estimate the imputation model) and an estimation sample (to obtain out-of-sample predictions). In our context, the base survey and the target survey respectively correspond to the training sample and the estimation sample. Employing three common ML techniques, LASSO, Elastic Net, and Random Forest, we show the estimation results in Appendix B, Tables B.1 and B.2 for Tanzania and Vietnam, respectively. The ML poverty estimates do not work for both countries, except for the estimates of consumption mean that are within one standard error of the true mean for LASSO and Elastic Net in Malawi.[18] These

[17] See Mullainathan and Spiess (2017) and Athey and Imbens (2019) for recent reviews of ML in economics.

[18] Lasso linear model and Elastic net linear model are trained in the first round and tested against the second round. Lambda in LASSO is selected by 10-fold cross-validation for out of sample prediction. Alpha and lambda in Elastic Net are selected by 10-fold cross-validation for out of sample prediction. The final selected variables and prediction models with statistics for Lasso and Elastic Net using postselection coefficient estimates are shown in Table B.3 for Tanzania and Table B.4 for Vietnam (Appendix B). Random forest model is trained in the first round and tested against the second round. The number of sub-trees is set at 1000. Both out-of-bag error and validation error are used to determine the best possible model. Importance matrix of the variables is shown in Table B.5 for Tanzania and Table B.6 for Vietnam.

inconsistent results are similar to those obtained earlier for poverty imputation in Dang *et al.* (forthcoming).

4.2. Further Extensions

The results shown in the preceding section focus on FGT indexes with $\alpha \leq 2$ (i.e., headcount poverty, poverty gap, and poverty gap squared) and mean consumption. We further investigate whether these results still hold in more general settings. In particular, we generally consider the entire consumption distribution instead of just the mean consumption. We further provide estimation results for α going up to higher values (i.e., up to 5). We also consider within-year imputation results.

Entire consumption distribution

We plot in Figure 3 the imputation accuracy for different percentiles of the consumption distribution (including the 5^{th}, 10^{th}, 25^{th}, 50^{th} (or median), 75^{th}, 90^{th}, and 95^{th} percentiles), using the latest survey round for each country. While Model 1 works for the upper part of the (consumption) distributions for Bangladesh and Malawi, producing the estimates for the 75^{th} and higher percentiles that are within the 95 percent CI of the true figures (i.e., the gray bandwidths), it mostly works for the lower part of the distribution in Nigeria and Tanzania and does not work in Vietnam. Model 2 works for the lowest 5^{th} percentile in Bangladesh and Nigeria, and for the lower parts of the distribution in Tanzania, and for the highest 95^{th} percentile in Tanzania and Malawi, but it does not work in Vietnam. Model 3 works for the distribution from the 10^{th} to 50^{th} percentiles for Bangladesh, from the 10^{th} to 95^{th} percentiles for Malawi and Tanzania, and for the 75^{th} and 95^{th} percentiles for Nigeria. Model 3 does not work in Vietnam. Models 4 and 6 mostly work in

Tanzania, but also in the lowest 5th and 10th percentiles for Bangladesh and both – the highest 5th and 10th percentiles for Vietnam and the highest 95th percentile in Malawi. Model 5 works for the full distribution in Vietnam, but mostly works in the lower parts of the distributions for Tanzania and Bangladesh. Models 7, 8 and 9 work well from the 5th to 50th percentiles of the distributions for Tanzania and for the 5th percentile in Nigeria, and Model 9 works from the 5th to 50th percentile for Tanzania and from 5th to 25th percentile for Vietnam.

In summary, Figure 3 suggests that Models 3 and Model 9 seem to work better than the other models. Figure A.1 in Appendix A provides a summary of the number of models that offer estimates that are not statistically different from the true estimates.

For a more in-depth look into these models' performance, we plot in Figure 4 the results for Model 3, which covers all countries. Except for Vietnam, Model 3 works reasonably well for predicting consumption values for the entire distributions for all countries with the estimates overlapping with the true rates and their 95 percent CI (i.e., the dotted red line and gray bandwidth). We also plot the results for Model 9 in Appendix A, Figure A.2, which excludes Bangladesh. This figure suggests that Model 9 works well for predicting consumption values for the entire distribution for Tanzania and Vietnam, while it works starting from the 75th percentile and higher for Malawi and for the lowest part of the distribution in Nigeria. We return to more discussion on the meta-analysis of model performance in Section 5.

Other FGT indexes

As discussed earlier, a larger α in the FGT poverty index suggests more poverty aversion. We consider Tanzania as an example where we let α go up to 5 and plot the results in Figure 5. This figure shows that all the imputation models, except for Model 3, work for the FGT indexes with these different values of α. When considering the USAID poverty gap, all the models work for all values of α, except that Model 3 does not work for α falling between 3 and 5 (Appendix A, Figure A.3).

Within-year imputation

For the within-year imputation, we divide the estimation sample into two random halves for each country.[19] We subsequently use one random half as the base survey and impute from this base survey into the other random half, which serves as the target survey. The estimation results suggest that the within-year imputation works well for most models for every country. Summarizing the results for Bangladesh, Ethiopia, Vietnam, Malawi, Nigeria, and Tanzania, Figure 6 indicates that the estimates fall within the 95 percent CI of the true poverty rates for the majority of the models. Specifically, out of the nine models considered, at least seven models work for all the countries for all the outcomes, except for Malawi and Nigeria (regarding USAID poverty gap) and Vietnam (regarding headcount poverty). In Bangladesh, at least four models (out of seven) work for all the outcomes. Yet, for these exceptions, at least three models still work. We offer more detailed results for each country in Figures A.4-A.9 in Appendix A.

These results have several practical implications for survey implementation for poverty imputation. First, in contexts where there is only a single base survey at hand, it could be tempting to carry out a similar within-survey imputation exercise and decide on the best performing model to be used for across-year imputation. But we would strongly advise against this approach. The reason is that while all the tested models appear to be achieving comparable within-year imputation

[19] We pretend that each household survey offers the universe of households for each country and implement the random sampling method on the sampled households to obtain the random halves. The poverty rates using the actual consumption data for these random halves are thus not identical, but are very close, to those using all the sampled households.

performance, only a subset of the models can fulfill across-year imputation needs and provide poverty estimates that are not statistically significantly different from the true poverty rates.

Second, on the other hand, these results provide further supportive evidence for those in earlier studies (see, e.g., Dang and Verme (2023) in the context for refugees) that within-year imputation may potentially offer a promising direction to obtain poverty estimates at lower costs for various situations. For example, data may not be collected for a location due to reasons beyond one's control such as inaccessible roads or unexpected natural calamities (i.e., flood, storms, or landslides), or conflict and violence. Or it can simply be that prohibitively expensive survey costs can prevent data collection at a specific location. In these cases, if the welfare variable exists for another geographical location that is comparable to the location without these data, we can employ our proposed technique to provide imputation-based poverty estimates for the latter location.[20]

5. Meta-analysis

The analysis shown in Figures 1 and 2 is obtained by simply averaging across the imputation models the results across the countries, the years, as well as other variables (e.g., region or estimation methods). To further take into account the potential contributions from these model characteristics, we estimate the following logit regression with country fixed effects

$$P_{kn} = F(\sum_{k=1}^{K} \gamma' m_k + \tau_n + \omega_{kn}) \qquad (11)$$

where P_{kn} is a binary variable that equals 1 if the poverty estimate is not statistically significantly different from the true poverty rate and 0 otherwise, for *k= 1,.., K* models and *n= 1,.., N* countries.

[20] To ensure that geographical locations are comparable, we may need to bring in additional information on other aspects (e.g., some qualitative information about income levels or poverty rates for these regions).

F(.) is the logit function (i.e., $F(a) = \frac{1}{1+e^{-a}}$). m_k are the dummy variables indicating the imputation models, τ_n are the country dummy variables, and ω_{kn} is the error term.

The dynamics between a country dummy variable and the performance of the imputation models can be captured to varying extents by the characteristics of the imputation models. Consequently, to shed more light on these differences, we can replace the country dummy variables with the model characteristics, to estimate the following alternative equation:

$$P_{kn} = F(\sum_{k=1}^{K} \delta' m_k + \theta' Z + \varphi_{kn}) \qquad (12)$$

where Z are the model characteristics such as the true poverty rate in the target survey, the (logarithm of) sample size of the base survey, the time difference between the base survey and the target survey, the number of pairs of survey rounds available for analysis, the model goodness-of-fit (as measured by R^2), and the estimation method (normal linear regression model or the empirical distribution of the error terms). But the model characteristics can only offer a guide to model selection, since these model characteristics likely represent a correlational—rather than causal—and *ex post* relationship with the imputation outcomes. While the estimation results would be strongest if they agree under both equations, our preferred equation for interpretation is Equation (11) that clearly lays out the models *a priori*, particularly where the estimates are different.[21]

For easier interpretation, Table 3 shows the marginal effects from the logit regressions for Equations (11) and (12), using Model 1 as the reference model. The associated regression results

[21] This concern is particularly relevant to the estimated model parameters (versus the exogenous model parameters given by the data). As an example, the correlation between the model goodness-of-fit statistics R^2 (or the correlation between the predicted consumption and the actual consumption for the target survey ρ(y,y)) with the model numbers is around -0.34 and strongly statistically significant for the whole country sample. As such, we do not include them in the regressions for Equations (5) and (6).

are presented in Appendix A, Table A.9.[22] We estimate robust standard errors clustered at the country level for both equations.

Several interesting findings stand out from Table 3. First, regarding the specific imputation models to use, differences exist by the type of indicator. Out of the main nine models, Model 9 performs the best for headcount poverty; it raises the probability of accurate imputation by 0.4 for both Specification 1 (the preferred Equation (11)) and Specification 2 (Equation (12)), with the results for both specifications being statistically significant at the five percent level. Except for the USAID poverty gap, Model 9 also works for most of the remaining indicators to varying degrees. In particular, this model raises the probability of imputation accuracy for near-poverty, extreme poverty, and poverty gap by around 0.1-0.2 (with Specification 1), but the differences are marginally statistically different at the 10 percent level, except for extreme poverty where the difference is strongly statistically different at the five percent level. Notably, Model 9 with Specification 2 works well for consumption mean, where it increases the probability of imputation accuracy by 0.2 and this result is strongly statistically different at the one percent level. Further adding agricultural soil quality information to Model 9 (i.e., creating Model 13) results in much higher imputation accuracy (and stronger statistical significance) for headcount poverty, increasing the probability of imputation accuracy by 0.5 for both Specifications.

Model 3 works for headcount poverty, extreme poverty, poverty gap, and consumption mean, raising the probability of accurate imputation for these indicators by around 0.3 (under Specification 1). The results are strongly statistically different at the five percent level or less. For

[22] Alternatively, we can employ an ordered logit regression instead where the outcome variable is defined as taking the values of 1 or 2 if the poverty estimate falls within the 95 percent CIs or one standard error around the true poverty rate, and 0 otherwise. The results, shown in Appendix A, Table A.10, are qualitatively similar but have less statistical significance. For example, the pseudo-R^2 for headcount poverty and near-poverty in this table are about half (or less) of those for the logit regressions shown in Appendix A, Table 1.9 for Specifications 1 and 2.

the USAID poverty gap, while Model 3 also offers strong statistical significance under Specification 2, it does not work under Specification 1.

Several models work for USAID poverty gap in Specification 2 but do not work in Specification 1. These include Models 4 to 8, but the results are statistically significant at the five percent level for Models 4, 5, and 8 and are marginally statistically significant at the 10 percent level for Models 6 and 7.

Finally, the estimation results using the estimated model parameters (Specification 2) indicate that a larger time interval between the base survey and the target survey is generally associated with lower imputation accuracy for headcount poverty, extreme poverty, and poverty gap. More (higher) extreme poverty, poverty gap and consumption mean are positively associated with imputation accuracy. While more survey rounds help increase imputation accuracy for the poverty rate and poverty gap (possibly through higher data quality and/ or local survey staff capacity due to more surveys being implemented), a higher model goodness-of-fit (R^2) does not help. However, as discussed earlier, the relationship between the estimated model parameters and the imputation accuracy is, at best correlational, so these results should be regarded as indicative and should be further investigated.

6. Conclusion

We make several new conceptual and empirical contributions to the literature on survey-to-survey imputation of poverty estimates. Conceptually, we significantly expand this literature to various poverty indicators including near-poverty (vulnerability) status, extreme poverty, poverty gap, and other FGT poverty indexes. These extensions extend the existing literature, which almost exclusively focuses on the headcount poverty rate. Furthermore, we also examine the performance of imputed household consumption, which underlies these poverty indicators.

Empirically, we harmonize and rigorously analyze data from 22 recent rounds of multi-topic household surveys conducted over the past decade in Bangladesh, Ethiopia, Malawi, Nigeria, Tanzania, and Vietnam. These six countries span three regions (i.e., Sub-Saharan Africa South Asia, and Southeast Asia) and different income levels (i.e., low-income to lower-middle-income) and offer the most comprehensive dataset that has been analyzed to date.

We find that survey-to-survey imputation provides encouraging results. However, imputation model heterogeneity exists, with certain models performing better for some poverty indicators only. In particular, for headcount poverty, adding household utility consumption expenditures (including electricity, water, and garbage) to a basic imputation model that includes household demographic and employment characteristics (Model 9) performs the best. Compared to a reference imputation model with basic demographic and employment variables, it raises the probability of imputation accuracy by 0.4. Model 9 also works for most of the remaining indicators but to varying degrees, raising the probability of imputation accuracy for near-poverty, extreme poverty, and poverty gap by around 0.1-0.2. Further adding agricultural soil quality information to Model 9 increases the probability of imputation accuracy by 0.5 for headcount poverty. Alternatively, adding food consumption expenditures to an imputation model that includes household demographics, employment, assets, and house characteristics (Model 3) works for headcount poverty, extreme poverty, poverty gap, and consumption mean, raising the probability of accurate imputation for these indicators by around 0.3. The results are strongly statistically different at the five percent level or less. We also find that the proposed imputation method works better than some common machine learning techniques.

Further testing the imputation models with meta-analysis, we find that certain, but not all, model specifications work. In particular, Model 9 works better for mean consumption levels under a model specification (Specification 2) that includes model characteristics rather than country dummy variables (Specification 1). For the USAID poverty gap, several models work under Specification 2 but do not work in Specification 1. These include Models 3, 4, 5, 6, 7, and 8, with varying degrees of statistical significance. While these results provide some tentative evidence that these models may be used to obtain poverty estimates, they also need further study for more robustness.

The estimation results using the estimated model parameters (Specifications 2) also indicate that a larger time interval between the base survey and the target survey is generally associated with lower imputation accuracy for headcount poverty, extreme poverty, and poverty gap. On the other hand, more (higher) extreme poverty, poverty gap and consumption mean are positively associated with imputation accuracy, as is more survey rounds. A higher model goodness-of-fit (R^2) does not necessarily help with raising across year imputation accuracy. However, as discussed earlier, the relationship between the estimated model parameters and the imputation accuracy is, at best, correlational, so these results should be regarded as indicative and should be further investigated.

These results are broadly consistent with earlier studies and offer useful inputs for future survey design. Collecting data on utilities expenditures or food expenditures clearly requires fewer resources and less time than implementing a fully-fledged household consumption survey, thus employing imputation methods in combination with these data to provide updated poverty estimates presents a cost-effective option.[23] Furthermore, in contexts where relatively less intensive survey efforts can be spent on collecting such data (e.g., especially where receipts for such expenditures are strongly digitalized), these advantages appear even stronger. Given the increasingly popular digitalization of payment transactions around the world, these contexts may become much more available in the near future.

[23] Dang *et al.* (2024) offer experimental evidence from Tanzania that suggests collecting data on either reduced or more aggregated food consumption categories could help significantly improve the imputation accuracy of poverty estimates.

References

Altındağ, O., O'Connell, S. D., Şaşmaz, A., Balcıoğlu, Z., Cadoni, P., Jerneck, M., & Foong, A. K. (2021). Targeting humanitarian aid using administrative data: model design and validation. *Journal of Development Economics*, *148*, 102564.

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*, 685-725.

Baulch, Bob. (Ed.) (2011). *Why poverty persists: Poverty dynamics in Asia and Africa*. Cheltenham, UK: Edward Elgar Publishing.

Beegle, K., Christiaensen, L., Dabalen, A., & Gaddis, I. (2016). *Poverty in a rising Africa*. Washington, DC: The World Bank.

Beltramo, Theresa, Hai-Anh Dang, Ibrahima Sarr and Paolo Verme. (2024). "Estimating Poverty among Refugee Populations: A Cross-Survey Imputation Exercise for Chad". *Oxford Development Studies*, 52(1): 94-113.

Christiaensen, Luc, Ethan Ligon, and Thomas P. Sohnesen. (2022). "Should Consumption Sub-Aggregates be Used to Measure Poverty?" *World Bank Economic Review*, 36(2): 413–432.

Christiaensen, Luc, Peter Lanjouw, Jill Luoto, and David Stifel. (2012). "Small Area Estimation-based Prediction Models to Track Poverty: Validation and Applications." *Journal of Economic Inequality,* 10(2): 267-297.

Cuesta, J., & Ibarra, G. L. (2017). Comparing cross-survey micro imputation and macro projection techniques: Poverty in post revolution Tunisia. *Journal of Income Distribution*, *25*(1), 1-30.

Dang, Hai-Anh and Peter Lanjouw. (2023). "Regression-based Imputation for Poverty Measurement in Data Scarce Settings". In Jacques Silber. (Eds.). *Handbook of Research on Measuring Poverty and Deprivation*. Edward Elgar Press.

Dang, Hai-Anh and Paolo Verme. (2023). "Estimating Poverty for Refugees in Data-scarce Contexts: An Application of Cross-Survey Imputation." *Journal of Population Economics*, 36(2), 653 – 679.

Dang, Hai-Anh, Dean Jolliffe, and Calogero Carletto. (2019). "Data Gaps, Data Incomparability, and Data Imputation: A Review of Poverty Measurement Methods for Data-Scarce Environments". *Journal of Economic Surveys*, 33(3): 757-797.

Dang, Hai-Anh, Peter Lanjouw, Umar Serajuddin. (2017). "Updating Poverty Estimates at Frequent Intervals in the Absence of Consumption Data: Methods and Illustration with Reference to a Middle-Income Country." *Oxford Economic Papers*, 69(4): 939-962.

Dang, Hai-Anh, Talip Kilic, Calogero Carletto, Kseniya Abanokova. (forthcoming). "Poverty Imputation in Contexts without Consumption Data: A Revisit with Further Refinements." *Review of Income and Wealth*.

Dang, Hai-Anh, Talip Kilic, Vladimir Hlasny, Kseniya Abanokova, and Calogero Carletto. (2024). “Using Survey-to-Survey Imputation to Fill Poverty Data Gaps at a Low Cost: Evidence from a Randomized Survey Experiment”. IZA Discussion Paper no. 16792.

De Luca, Giuseppe, Jan R. Magnus, and Franco Peracchi. (2018). "Balanced variable addition in linear models." *Journal of Economic Surveys*, 32(4): 1183-1200.

Deaton, Angus and Valerie Kozel. (2005). *The Great Indian Poverty Debate*. New Delhi: Macmillan.

Deaton, Angus, and John Muellbauer. (1980). *Economics and Consumer Behavior*. Cambridge, UK: Cambridge University Press.

Deaton, Angus and Salman Zaidi. (2002). *Guidelines for Constructing Consumption Aggregates for Welfare Analysis* (Vol. 135). Washington, DC: World Bank Publications.

Devarajan, Shantayanan. (2013). "Africa's statistical tragedy." *Review of Income and Wealth*, 59: S9-S15.

Douidich, Mohamed, Abdeljaouad Ezzrari, Roy van der Weide, and Paolo Verme. (2016). “Estimating Quarterly Poverty Rates Using Labor Force Surveys: A Primer.” *World Bank Economic Review*, 30(3): 475-500.

Elbers, Chris, Jean O. Lanjouw, and Peter Lanjouw. (2003). “Micro-Level Estimation of Poverty and Inequality.” *Econometrica*, 71(1): 355-364.

Fernandez, J., Olivieri, S., & Wambile, A. (2024). “Reconstructing 2010–2022 Poverty and Inequality Trends in Bangladesh: A Statistical Matching Approach”. Policy Research Working Paper 10749. Washington, DC: World Bank.

Foster, J., Greer, J., & Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52(3): 761-766.

---. (2010). The Foster–Greer–Thorbecke (FGT) poverty measures: 25 years later. *The Journal of Economic Inequality*, *8*, 491-524.

Jerven, Morten. (2019). "The Problems of Economic Data in Africa." In *Oxford Research Encyclopedia of Politics*.

Kilic, T., and Sohnesen, T. (2019). “Same question but different answer: experimental evidence on questionnaire design’s impact on poverty measured by proxies.” *Review of Income and Wealth*, 65.1, pp. 144-165.

Lain, J. and Vishwanath, T. (2022). *A Better Future for All Nigerians: Nigeria Poverty Assessment 2022*. Washington, D.C.: World Bank Group.

Letta, Marco, Pierluigi Montalbano, and Richard SJ Tol. (2018). "Temperature shocks, short-term growth and poverty thresholds: Evidence from rural Tanzania." *World Development*, 112: 13-32.

Mathiassen, Astrid. (2009). “A Model Based Approach for Predicting Annual Poverty Rates without Expenditure Data”. *Journal of Economic Inequality*, 7:117–135.

Mathiassen, Astrid, and Bjørn K. Getz Wold. (2021). "Predicting poverty trends by survey-to-survey imputation: the challenge of comparability." *Oxford Economic Papers* 73, no. 3: 1153-1174.

Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87-106.

Naudé, W. and Vinuesa, R. (2020). Data, global development, and COVID-19. *WIDER Working Paper 2020/109*.

Ravallion, Martin. (2016). *The economics of poverty: History, measurement, and policy*, New York: Oxford University Press.

Serajuddin, Umar, Hiroki Uematsu, Christina Wieser, Nobuo Yoshida, and Andrew Dabalen. (2015). "Data deprivation: another deprivation to end." *World Bank Policy Research Paper no. 7252*, World Bank, Washington, DC.

Sheahan, Megan, and Christopher B. Barrett. (2017). "Ten striking facts about agricultural input use in Sub-Saharan Africa." *Food Policy*, 67: 12-25.

Snijders, Tom AB, and Roel J. Bosker. (1994). "Modeled variance in two-level models." *Sociological Methods & Research*, 22(3): 342-363.

Stifel, D. and Christiaensen, L. (2007) “Tracking Poverty over Time in the Absence of Comparable Consumption Data”. *World Bank Economic Review*, 21, 317-341.

Tanzania’s National Bureau of Statistics. (2011). *Basic Information Document—National Panel Survey 2010-11*.

Tarozzi, Alessandro. (2007). “Calculating Comparable Statistics from Incomparable Surveys, With an Application to Poverty in India”. *Journal of Business and Economic Statistics* 25, no. 3:314-336.

UNESCO-UIS/OECD/EUROSTAT. (UOE). (2020). *Data collection on formal education—Manual on concepts, definitions and classifications*. Montreal/ Paris/ Luxembourg.

United States Census Bureau. (2017). Current Population Survey, Imputation of Unreported Data Items. Accessed on the Internet on May 24, 2021 at https://www.census.gov/programs-surveys/cps/technical-documentation/methodology/imputation-of-unreported-data-items.html

World Bank. (2021). *World Development Report 2021: Data for Better Lives*. Washington, DC: World Bank.

Table 1. Predicted Poverty Rates Based on Imputation (percentage)

Indicators	Second round									True rates
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	
Panel A: Bangladesh 2015-2018/19										
Headcount poverty rate	13.0	**7.7***	8.5	**7.0***	**8.0**	**7.0***	**7.8***	N/A	N/A	7.3
	(0.7)	(0.6)	(0.8)	(0.7)	(0.7)	(0.6)	(0.7)			(0.6)
Near-poverty rate	**12.2***	9.5	**11.3**	10.2	9.9	9.2	9.5	N/A	N/A	12.2
	(0.5)	(0.6)	(0.7)	(0.6)	(0.6)	(0.6)	(0.6)			(0.6)
Extreme poverty rate	2.3	1.0	**0.9**	**0.6***	0.9	**0.8**	1.0	N/A	N/A	0.6
	(0.3)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)			(0.2)
Poverty gap	2.6	1.4	1.4	**1.1***	1.4	**1.2***	1.4	N/A	N/A	1.1
	(0.2)	(0.2)	(0.2)	(0.1)	(0.2)	(0.1)	(0.2)			(0.1)
USAID Poverty gap	19.9	17.5	**16.5**	**15.9***	17.2	17.0	17.6	N/A	N/A	15.2
	(0.7)	(1.0)	(0.9)	(0.9)	(0.9)	(1.0)	(1.0)			(0.8)
N	5,604	5,604	5,604	5,604	5,604	5,604	5,604			5,604
Panel B: Malawi 2016/17-2019/20										
Headcount poverty rate	53.2	**52.7**	60.5	**52.7**	53.6	**52.6**	**52.6**	**52.2**	**52.3**	51.1
	(0.9)	(1.0)	(1.2)	(1.2)	(1.0)	(1.0)	(1.0)	(1.0)	(0.9)	(0.9)
Near-poverty rate	**14.0***	**14.3***	12.4	**14.2***	**14.4***	**14.4***	**14.3***	**14.5**	**14.4***	14.1
	(0.4)	(0.4)	(0.5)	(0.5)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)
Extreme poverty rate	22.8	**21.7**	29.5	**21.7**	**22.0**	**21.5**	**21.8**	**21.2***	**21.8**	20.6
	(0.7)	(0.7)	(1.1)	(0.9)	(0.8)	(0.7)	(0.8)	(0.8)	(0.7)	(0.8)
Poverty gap	18.2	**17.6**	22.5	**17.6**	**17.9**	**17.5***	**17.7**	**17.3***	**17.7**	17.1
	(0.4)	(0.5)	(0.7)	(0.6)	(0.5)	(0.5)	(0.5)	(0.5)	(0.4)	(0.4)
USAID Poverty gap	**34.3**	**33.5***	37.2	**33.4***	**33.4***	**33.3***	**33.6***	**33.3***	**33.8***	33.5
	(0.4)	(0.4)	(0.6)	(0.5)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)	(0.5)
N	11,432	11,432	11,432	11,432	11,432	11,432	11,432	11,432	11,432	11,432
Panel C: Nigeria 2012/13-2018/19										
Headcount poverty rate	33.4	33.3	50.8	27.2	24.7	29.4	33.2	34.1	34.5	46.4
	(1.9)	(2.0)	(2.6)	(2.1)	(1.9)	(1.9)	(2.1)	(2.1)	(2.0)	(1.9)
Near-poverty rate	**12.2**	**12.7**	**13.6***	**13.4***	11.3	**12.6**	**12.5**	**12.6**	**12.2**	13.7
	(0.9)	(0.9)	(1.0)	(1.0)	(0.9)	(0.9)	(0.9)	(0.9)	(0.9)	(1.0)
Extreme poverty rate	15.1	14.8	26.4	9.2	9.9	12.3	14.9	15.5	15.8	22.2
	(1.4)	(1.5)	(2.2)	(1.3)	(1.2)	(1.3)	(1.5)	(1.5)	(1.4)	(1.4)
Poverty gap	10.9	10.9	17.9	7.4	7.5	9.2	10.9	11.3	11.4	15.3
	(0.8)	(0.9)	(1.3)	(0.8)	(0.8)	(0.8)	(0.9)	(0.9)	(0.9)	(0.8)
USAID poverty gap	**32.7***	**32.7***	35.3	27.2	30.4	31.3	**32.9***	**33.0***	**33.1***	33.0
	(1.2)	(1.3)	(1.1)	(1.2)	(1.5)	(1.4)	(1.3)	(1.3)	(1.1)	(0.8)
N	4,976	4,976	4,976	4,976	4,976	4,976	4,976	4,976	4,976	4,976
Panel D: Tanzania 2019/20-2020/21										
Headcount poverty rate	**17.4***	**17.5***	**19.4**	**16.9***	**17.3***	**16.7**	**17.5***	**17.8***	**18.2***	17.8
	(1.0)	(1.3)	(1.4)	(1.3)	(1.2)	(1.3)	(1.3)	(1.3)	(1.1)	(1.1)
Near-poverty rate	**10.5***	**10.9**	**10.6***	**10.9**	**10.9***	**10.3***	**10.8***	**11.0**	**10.4***	10.2
	(0.7)	(0.7)	(0.8)	(0.8)	(0.8)	(0.7)	(0.7)	(0.8)	(0.7)	(0.7)
Extreme poverty rate	**9.7***	**9.4***	**11.1**	**9.1***	**9.3***	**9.0***	**9.5***	**9.7***	**10.3***	9.8
	(0.8)	(0.9)	(1.1)	(1.0)	(0.9)	(1.0)	(1.0)	(1.0)	(0.9)	(0.8)

Poverty gap	**4.6***	**4.4***	5.3	**4.2**	**4.3***	**4.2**	**4.4***	**4.5***	**4.9***	4.6
	(0.4)	(0.4)	(0.5)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)	(0.3)
USAID Poverty gap	**26.3***	**25.3***	**27.5**	**24.9***	**25.1***	**25.2***	**25.4***	**25.5***	**26.9***	25.9
	(1.1)	(1.1)	(1.3)	(1.1)	(1.1)	(1.2)	(1.1)	(1.1)	(1.1)	(1.1)
N	4,644	4,644	4,644	4,644	4,644	4,644	4,644	4,644	4,644	4,644
Panel E: Vietnam 2014-2016										
Headcount poverty rate	15.0	13.3	6.1	8.4	10.6	12.5	13.3	11.4	**9.6***	9.6
	(0.5)	(0.6)	(0.5)	(0.5)	(0.5)	(0.6)	(0.6)	(0.6)	(0.5)	(0.4)
Near-poverty rate	9.3	8.5	4.9	6.0	**7.1***	8.0	8.5	7.7	**6.8***	6.9
	(0.4)	(0.4)	(0.3)	(0.4)	(0.3)	(0.4)	(0.4)	(0.4)	(0.3)	(0.3)
Extreme poverty rate	2.0	2.0	0.7	**1.2***	**1.5**	2.1	2.1	2.0	2.0	1.2
	(0.2)	(0.3)	(0.1)	(0.2)	(0.2)	(0.3)	(0.3)	(0.3)	(0.3)	(0.2)
Poverty gap	4.0	3.7	1.5	**2.3**	2.9	3.5	3.7	3.2	2.9	2.5
	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.3)	(0.2)
USAID Poverty gap	**26.7**	**27.6**	**24.9**	**27.4**	**27.1**	28.1	**27.8**	28.3	30.1	26.1
	(0.7)	(0.9)	(1.3)	(1.2)	(1.0)	(1.0)	(0.9)	(1.2)	(1.7)	(0.9)
N	9,347	9,347	9,347	9,347	9,347	9,347	9,347	9,347	9,347	9,347
Control variables										
Food expenditures			Y							
Non-food expenditures				Y						
Furnishings and durable household expenses					Y					
Health expenditures						Y				
Education expenditures							Y			
Utilities expenditures								Y	Y	
Household assets & house characteristics		Y	Y	Y	Y	Y	Y	Y		
Demographics & employment	Y	Y	Y	Y	Y	Y	Y	Y	Y	

Note: Normal linear regression with bootstrapped standard errors is used. The standard errors are calculated using 100 bootstrap replications and are adjusted for complex survey design. All estimates are obtained with population weights. The normal linear regression model with the theoretical distribution of the error terms employs cluster random effects. 'Near poor' status is defined as living on an income between 100 and 125% of the poverty line. All indicators are expressed in percentage. The true rate is the estimate directly obtained from the survey data. Estimates shown in boldface or with a "*" respectively fall within the 95% confidence interval or one standard error of the true rate.

Table 2. Predicted Mean (Log) Consumption Based on Imputation

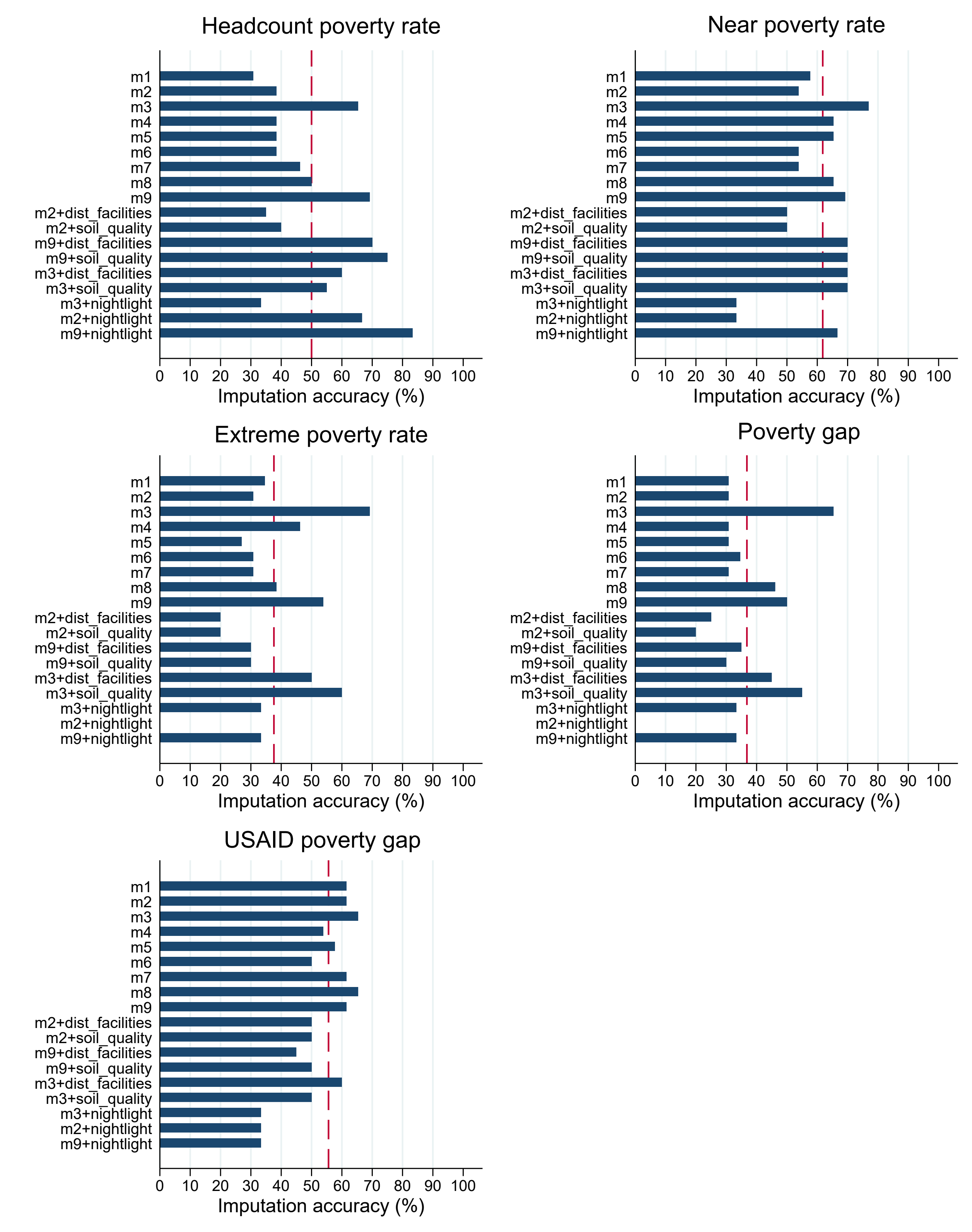
Indicators	Second round									True rates
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	
Bangladesh, from 2015 to 2018/19	10.7 (0.0)	10.8 (0.0)	**10.8** (0.0)	10.8 (0.0)	10.8 (0.0)	10.9 (0.0)	10.8 (0.0)	N/A	N/A	10.8 (0.0)
Malawi, from 2016/17 to 2019/20	12.1 (0.0)	12.1 (0.0)	**12.0** (0.0)	12.1 (0.0)	12.1 (0.0)	12.1 (0.0)	12.1 (0.0)	12.1 (0.0)	12.1 (0.0)	12.1 (0.0)
Nigeria, from 2012/13 to 2018/19	11.3 (0.0)	11.3 (0.0)	11.0 (0.0)	11.4 (0.0)	11.4 (0.0)	11.3 (0.0)	11.3 (0.0)	11.3 (0.0)	11.3 (0.0)	11.0 (0.0)
Tanzania, from 2019/20 to 2020/21	**13.7** (0.0)	13.7 (0.0)	13.7 (0.0)	**13.7** (0.0)	**13.7** (0.0)	**13.7*** (0.0)	**13.7** (0.0)	13.7 (0.0)	13.7 (0.0)	13.7 (0.0)
Vietnam, from 2014 to 2016	9.6 (0.0)	9.7 (0.0)	10.0 (0.0)	9.8 (0.0)	**9.8*** (0.0)	9.7 (0.0)	9.7 (0.0)	9.7 (0.0)	**9.8*** (0.0)	9.8 (0.0)
Food expenditures			Y							
Non-food expenditures				Y						
Furnishings and household expenses					Y					
Health expenditures						Y				
Education expenditures							Y			
Utilities expenditures								Y	Y	
Household assets & house characteristics		Y	Y	Y	Y	Y	Y	Y		
Demographics & employment	Y	Y	Y	Y	Y	Y	Y	Y	Y	

Note: Normal linear regression with bootstrapped standard errors is used. The standard errors are calculated using 100 bootstrap replications and are adjusted for complex survey design. All estimates are obtained with population weights. The normal linear regression model with the theoretical distribution of the error terms employs cluster random effects. Imputed consumption per capita for the second round uses the estimated parameters based on the data from the first round. 100 simulations are implemented. True consumption per capita is the estimate directly obtained from the survey data. Estimates shown in boldface or with a "*" respectively fall within the 95% confidence interval or one standard error of the true rate.

Table 3. Meta-analysis of Imputation Models and Their Parameters, Marginal Effects from Logit Regressions

	Headcount poverty rate		Near poverty rate		Extreme poverty rate		Poverty gap		USAID Poverty gap		Consumption mean	
	Spec.1	*Spec.2*	*Spec.1*	*Spec.2*	*Spec.1*	*Spec.2*	*Spec.1*	*Spec.2*	*Spec.1*	*Spec.2*	*Spec.1*	*Spec.2*
Model 2: Demographics, employment,	0.078	-0.017	-0.036	-0.345	-0.037	0.185	-0.000	0.203	-0.000	0.530*	-0.000	0.263
assets, house characteristics	(0.26)	(0.29)	(0.03)	(0.27)	(0.19)	(0.30)	(0.22)	(0.35)	(0.13)	(0.27)	(0.12)	(0.26)
Model 3	0.331***	0.048	0.196	-0.735	0.314***	0.890*	0.319**	0.855	0.040	1.529***	0.272***	1.227**
(adds food exp. to Model 2)	(0.06)	(0.44)	(0.16)	(0.83)	(0.12)	(0.49)	(0.15)	(0.53)	(0.04)	(0.54)	(0.05)	(0.58)
Model 4	0.078	-0.128	0.074	-0.626	0.104	0.591	-0.000	0.419	-0.077	1.082**	-0.000	0.583
(adds nonfood exp. to Model 2)	(0.08)	(0.34)	(0.14)	(0.48)	(0.13)	(0.44)	(0.17)	(0.43)	(0.07)	(0.47)	(0.00)	(0.43)
Model 5	0.078	-0.032	0.036	-0.327	-0.077	0.178	-0.000	0.234	-0.039	0.583**	0.039	0.335
(adds durables exp. to Model 2)	(0.23)	(0.30)	(0.19)	(0.45)	(0.18)	(0.32)	(0.26)	(0.41)	(0.10)	(0.30)	(0.11)	(0.30)
Model 6	0.078	-0.026	-0.036	-0.378	-0.037	0.209	0.038	0.264	-0.115*	0.467*	-0.000	0.289
(adds health exp. to Model 2)	(0.17)	(0.26)	(0.03)	(0.30)	(0.19)	(0.31)	(0.19)	(0.34)	(0.06)	(0.24)	(0.14)	(0.25)
Model 7	0.150	0.053	-0.036	-0.348	-0.037	0.186	-0.000	0.205	-0.000	0.538*	0.039	0.304
(adds education exp. to Model 2)	(0.26)	(0.31)	(0.03)	(0.27)	(0.19)	(0.30)	(0.22)	(0.35)	(0.13)	(0.28)	(0.11)	(0.27)
Model 8	0.186	0.083	0.074	-0.273	0.036	0.281	0.144	0.375	0.040	0.617**	0.039	0.329
(adds utilities exp. to Model 2)	(0.17)	(0.27)	(0.08)	(0.29)	(0.15)	(0.28)	(0.16)	(0.30)	(0.11)	(0.24)	(0.11)	(0.29)
Model 9	0.371**	0.357**	0.112*	-0.000	0.171***	0.258***	0.178*	0.257**	-0.000	0.176	0.073	0.150***
(adds utilities exp. to demographic & employment)	(0.17)	(0.15)	(0.06)	(0.17)	(0.06)	(0.08)	(0.10)	(0.13)	(0.13)	(0.18)	(0.05)	(0.01)
Model 10	0.056	-0.046	-0.114	-0.397	-0.109	0.098	-0.010	0.160	-0.048	0.488	-0.063	0.206
(adds distance to facilities to Model 2)	(0.31)	(0.33)	(0.11)	(0.29)	(0.28)	(0.36)	(0.31)	(0.42)	(0.18)	(0.34)	(0.21)	(0.32)
Model 11	0.105	0.008	-0.114	-0.396	-0.109	0.129	-0.071	0.134	-0.048	0.481	0.004	0.255
(adds agricultural soil quality to Model 2)	(0.30)	(0.32)	(0.11)	(0.29)	(0.28)	(0.36)	(0.32)	(0.43)	(0.18)	(0.33)	(0.14)	(0.28)
Model 12	0.392	0.351	0.090	-0.009	0.004	0.064	0.093	0.146	-0.096	0.101	0.056	0.157
(adds distance to facilities to Model 9)	(0.25)	(0.22)	(0.09)	(0.18)	(0.12)	(0.17)	(0.23)	(0.26)	(0.15)	(0.22)	(0.17)	(0.14)
Model 13	0.450***	0.424***	0.090	0.001	0.004	0.081	0.044	0.117	-0.048	0.125	0.056	0.136***
(adds agricultural soil quality to Model 9)	(0.14)	(0.12)	(0.09)	(0.17)	(0.12)	(0.17)	(0.20)	(0.24)	(0.18)	(0.23)	(0.05)	(0.02)
True estimates												
Headcount poverty rate		0.006										
		(0.01)										
Near poverty rate				0.069*								
				(0.04)								
Extreme poverty rate						0.028***						
						(0.00)						
Poverty gap								0.036***				
								(0.00)				
USAID poverty gap										0.010		
										(0.02)		
Consumption mean												0.208**
												(0.10)
Other model parameters												
Log of sample size of base survey		-0.157***		-0.590***		0.071		0.005		-0.011		-0.076
		(0.04)		(0.13)		(0.10)		(0.10)		(0.14)		(0.05)
Interval length between base & target surveys		-0.160***		-0.015		-0.174***		-0.132**		-0.052		-0.041
		(0.05)		(0.08)		(0.04)		(0.06)		(0.05)		(0.05)
Normal linear regression model		0.100**		0.009*		0.022		0.004		-0.008		0.019
		(0.05)		(0.01)		(0.02)		(0.02)		(0.01)		(0.04)
Number of rounds used		-0.015		-0.143*		0.064**		0.103***		0.029		-0.230**
		(0.05)		(0.08)		(0.03)		(0.04)		(0.05)		(0.10)
R squared		0.626		2.079		-1.295		-1.189		-3.261***		-1.962
		(0.86)		(1.90)		(0.80)		(0.85)		(1.21)		(1.33)
Country FE												
Tanzania	0.008		0.044***		0.362***		0.366***		0.413***		2.410***	
	(0.02)		(0.01)		(0.01)		(0.02)		(0.02)		(0.19)	
Malawi	-0.090***		0.242***		0.180***		0.191***		0.022***		2.435***	
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.19)	
Nigeria	-0.103***		0.693***		0.205***		0.166***		0.393***		2.462***	
	(0.00)		(0.01)		(0.00)		(0.00)		(0.00)		(0.19)	
N	314	314	314	314	314	314	314	314	314	314	314	314

Figure 1. Imputation Accuracy for Different Imputation Models, Poverty Indicators



Note: Imputation accuracy is the share of the estimates that are statistically insignificantly different from the true poverty rates for all countries and years. Red dashed line indicates mean accuracy.

Figure 2. Imputation Accuracy for Different Imputation Models

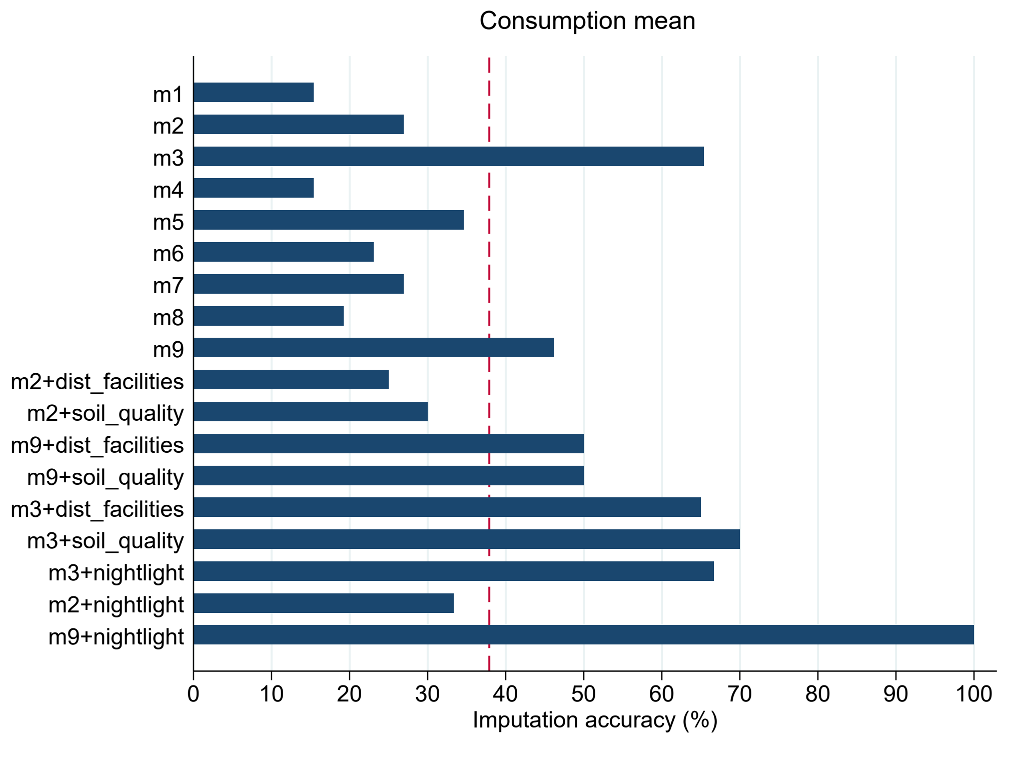


Figure 3. Distribution of imputed log of consumption

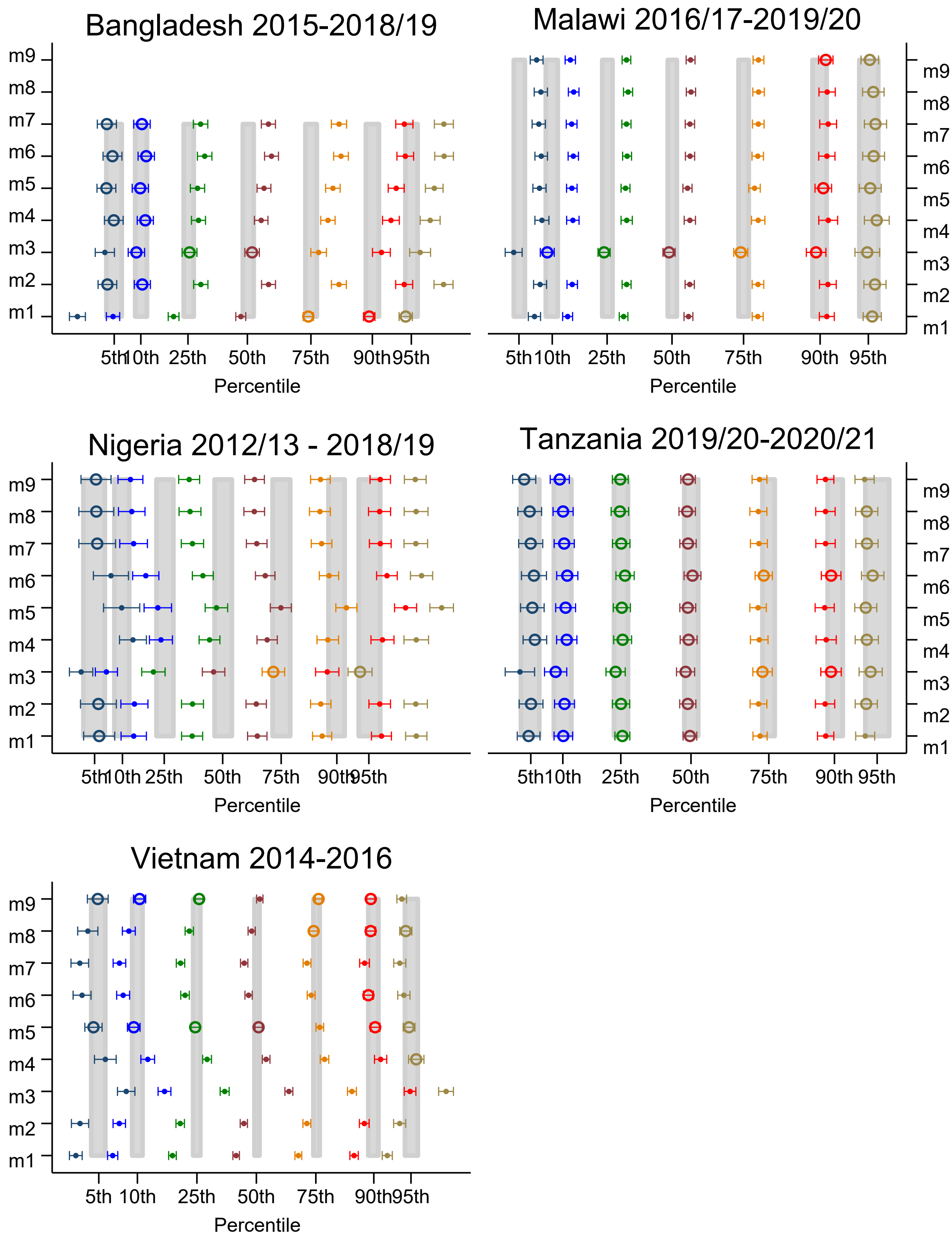


Figure 4. Predicted consumption distribution, Model 3

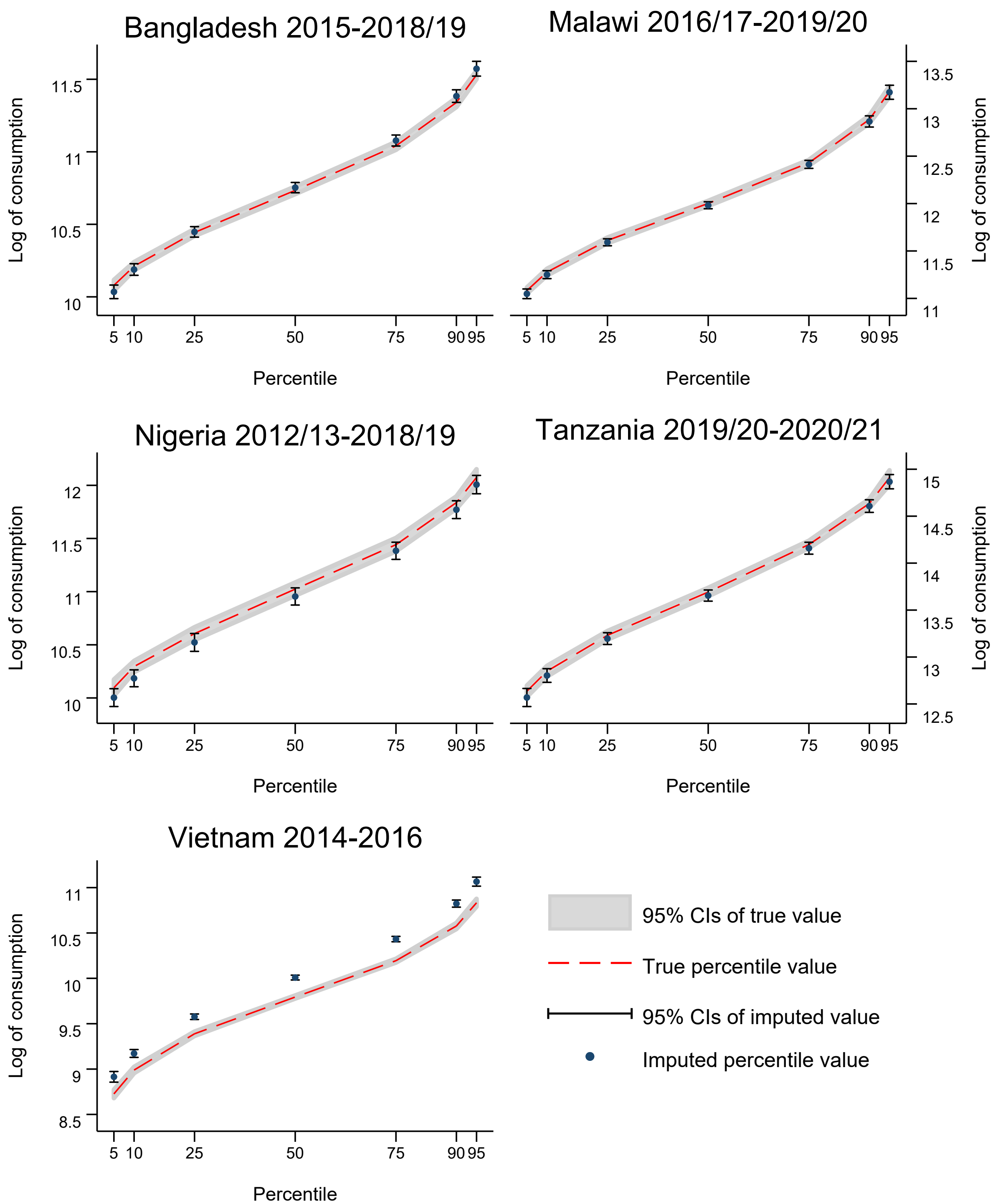
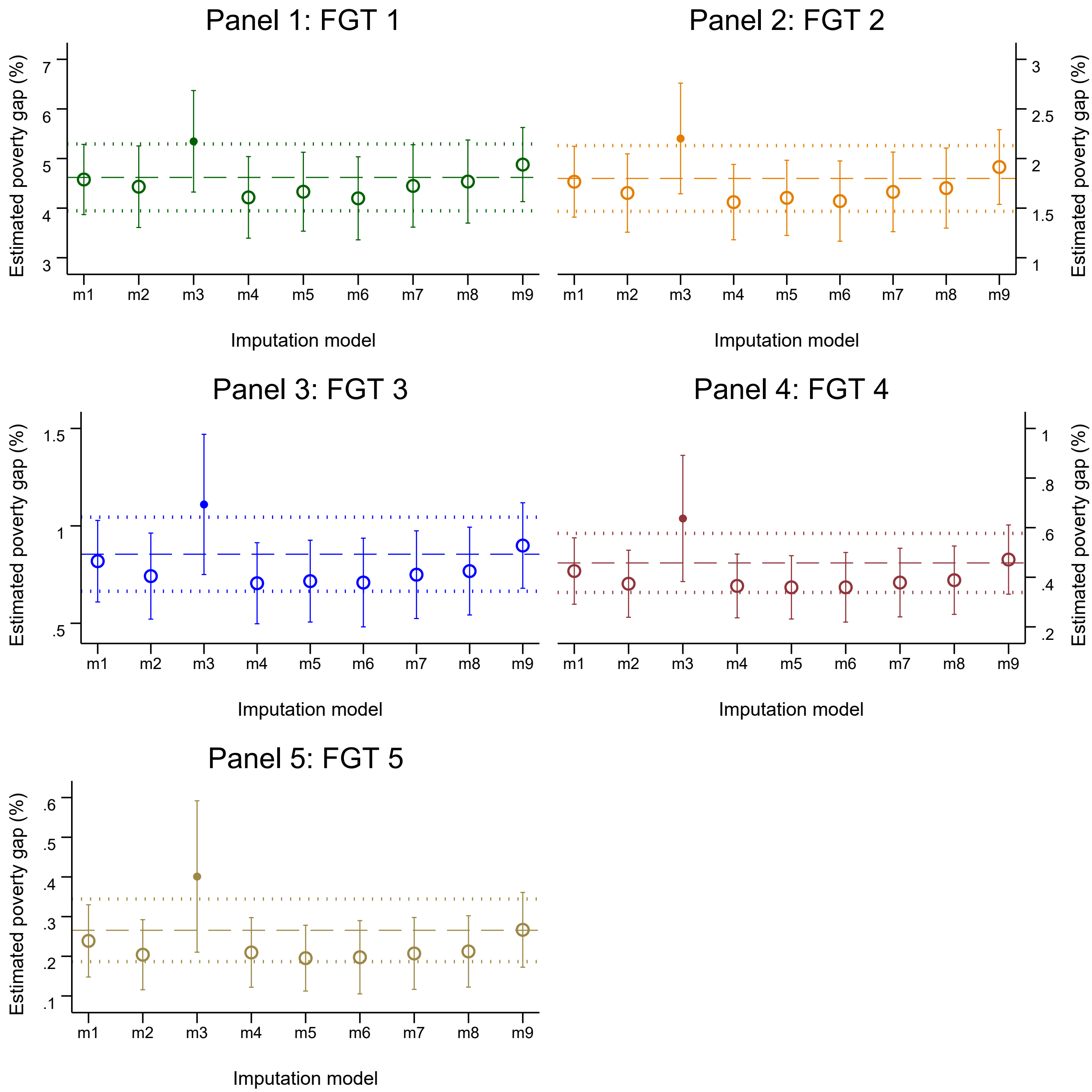
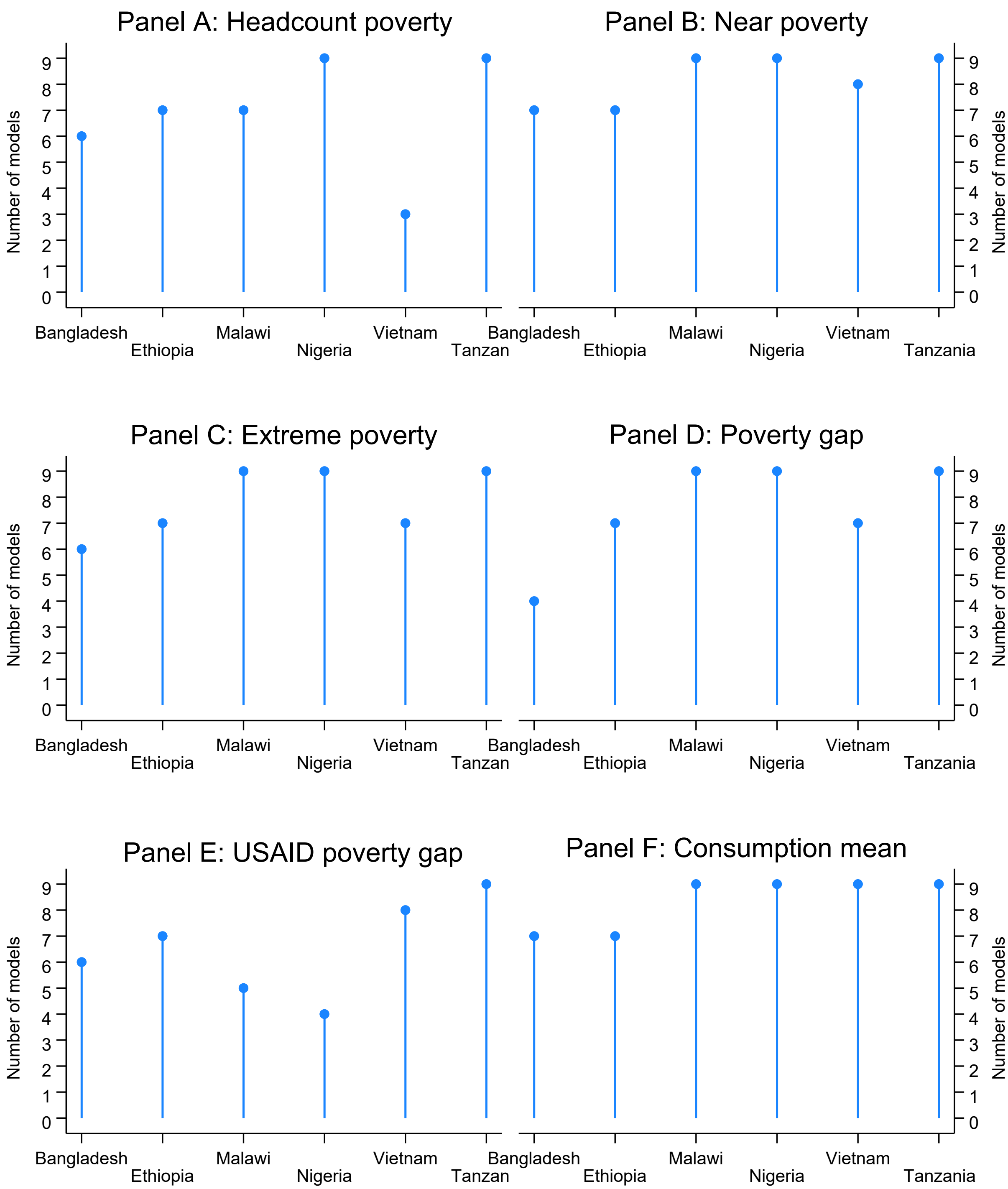


Figure 5. Predicted FGT indexes, Tanzania 2019/20- 2020/21



Note: Estimates are obtained by imputing from sample 1 into sample 2. 100 simulations are implemented. The standard errors are calculated using 100 bootstrap replications and are adjusted for complex survey design. Larger hollow symbols indicates that the estimates are statistically insignificantly different from the true poverty gap. Dashed lines represent the true poverty gap. Dotted lines represent confidence intervals of the true poverty gap. Estimates are obtained using the normal linear regression models.

Figure 6. Number of Models with Predicted Estimates That Are Statistically Insignificantly Different from the True Poverty Estimates, Within-Year Imputation



Note: Estimates are obtained by imputing from sample 1 into sample 2. 100 simulations are implemented. Estimates are obtained using the normal linear regression models.

Appendix A: Additional Tables and Figures

Table A.1. Household consumption model, Bangladesh 2015

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Household size	-0.054***	-0.067***	-0.024***	-0.031***	-0.055***	-0.062***	-0.072***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age of HH Head	0.002***	0.001**	0.000	0.000	0.002***	0.001**	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
HH Head is Female	0.058***	-0.005	0.033***	-0.023**	-0.014	0.003	-0.020
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Head has less than 5 years of schooling	0.091***	0.039***	0.022***	0.005	0.029**	0.031**	0.035**
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Head has 5-9 years of schooling	0.204***	0.078***	0.040***	0.015*	0.051***	0.072***	0.074***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Head has 10 or more years of schooling	0.451***	0.151***	0.079***	0.033**	0.115***	0.145***	0.134***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Share of HH members in 0-14	-0.742***	-0.451***	-0.184***	-0.202***	-0.348***	-0.386***	-0.533***
	(0.04)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Share of HH members in 15-24	-0.187***	-0.116***	-0.027	-0.092***	-0.139***	-0.068**	-0.135***
	(0.04)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Share of HH members in 60 and older	-0.322***	-0.187***	-0.086***	-0.047**	-0.128***	-0.210***	-0.129***
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
HH Head did wage/salary work during the last 7 days	-0.181***	-0.094***	-0.046***	-0.029***	-0.060***	-0.076***	-0.095***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
HH Head was self-employed during the last 7 days	0.027**	0.022*	-0.014**	0.025***	0.015	0.037***	0.013
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Log of food expenditures			0.698***				
			(0.01)				
Log of nonfood expenditures				0.504***			
				(0.01)			
Log of durable expenditures					0.182***		
					(0.01)		
Log of health expenditures						0.094***	
						(0.00)	
Log of education expenditures							0.017***
							(0.00)
Household owns a radio		0.021	0.000	0.025	0.013	0.016	0.028
		(0.04)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)
Household owns a television		0.134***	0.055***	0.072***	0.037***	0.120***	0.131***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Household owns a audio cassette/cd player		0.028	0.024*	0.006	0.001	0.028	0.030
		(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Household owns a sewing machine		-0.007	0.019*	-0.013	-0.039**	-0.005	-0.008
		(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Household owns a stove / gas burner		0.164***	0.091***	0.062***	0.104***	0.145***	0.166***
		(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Household owns a bicycle		0.040***	0.029***	-0.001	0.001	0.034***	0.026**
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Household owns a motor vehicles		0.242***	0.154***	0.088***	0.087***	0.227***	0.250***
		(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Household owns a mobile phone		0.163***	0.084***	0.043***	0.029**	0.143***	0.154***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Household owns an iron		0.114***	0.076***	0.019	0.088***	0.091***	0.110***
		(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Household owns an electric fan		0.127***	0.064***	0.048***	0.051***	0.119***	0.122***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Log of total floor area of the dwelling		0.143***	0.075***	0.058***	0.093***	0.130***	0.141***
		(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Household dwelling wall materials		0.015	0.023***	0.010	-0.015	0.017	0.012
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Household dwelling roof materials		0.014	0.032*	-0.015	-0.018	0.003	0.014
		(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Household dwelling floor materials		0.138***	0.084***	0.062***	0.102***	0.132***	0.135***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Household dwelling water access		0.078***	0.037***	0.022*	0.064***	0.066***	0.076***
		(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Household toilet is water sealed		0.036***	0.025***	0.008	0.015	0.031***	0.034***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Household toilet is other types		-0.021	-0.002	-0.015	-0.016	-0.016	-0.019
		(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
_cons	11.110***	10.085***	3.188***	5.777***	9.212***	9.495***	10.131***
	(0.04)	(0.06)	(0.07)	(0.08)	(0.06)	(0.06)	(0.06)
sigma_e	0.38	0.32	0.17	0.24	0.29	0.30	0.32
sigma_u	0.14	0.11	0.05	0.07	0.10	0.09	0.10
rho	0.13	0.10	0.06	0.09	0.10	0.09	0.10
r2_o	0.33	0.54	0.87	0.75	0.61	0.60	0.55
N	5447	5447	5445	5447	5441	5447	5447

Table A.2. Household consumption model, Ethiopia 2018/19

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Household size	-0.103**	-0.002	-0.083***	-0.107**	-0.105***	-0.100***	-0.098***
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Head's age	-0.003**	0.001**	-0.003***	-0.003**	-0.003***	-0.004***	-0.004***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Head is female	-0.037*	0.015**	-0.032	-0.044**	-0.040*	-0.036*	-0.044**
	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Head has primary education	0.120**	0.034**	0.061***	0.117**	0.086***	0.050**	0.045*
	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Head has secondary education	0.259**	0.095**	0.157***	0.249**	0.205***	0.124***	0.117***
	(0.03)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Head has higher education	0.448**	0.147**	0.270***	0.436**	0.386***	0.256***	0.249***
	(0.04)	(0.01)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Share of household members age 0-14	-0.261**	-0.010	-0.236***	-0.286**	-0.205***	-0.093	-0.090
	(0.06)	(0.02)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Share of household members age 15-24	-0.122**	0.043**	-0.111***	-0.143**	-0.118***	-0.053	-0.058
	(0.04)	(0.01)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Share of household members age 60 and older	0.098	-0.063**	0.184***	0.115*	0.162**	0.172***	0.193***
	(0.07)	(0.02)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
Head is working	0.037*	0.004	-0.005	0.034	0.032	0.036*	0.033
	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Rural	-0.344**	-0.149**	-0.141***	-0.333**	-0.171***	-0.130***	-0.063
	(0.04)	(0.01)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Log of food consumption peraeq		0.888**					
		(0.01)					
Log of non-food consumption peraeq			0.273***				
			(0.01)				
Log of education expenditures peraeq				0.009***			
				(0.00)			
Log of utilities peraeq					0.058***		0.042***
					(0.00)		(0.00)
Household owns a television						0.083***	0.074**
						(0.03)	(0.03)
Household owns an CD / DVD player						0.058*	0.051
						(0.03)	(0.03)
Household owns a refrigerator						0.153***	0.144***
						(0.03)	(0.03)
Household owns a bicycle						0.069	0.060
						(0.07)	(0.07)
Household owns a motorcycle						0.027	0.018
						(0.09)	(0.09)
Household owns a car						0.247***	0.230***
						(0.06)	(0.06)
Household owns a desk phone						0.020	0.002
						(0.03)	(0.03)
Household owns a mobile phone						0.128***	0.041
						(0.02)	(0.02)
Household owns a radio						0.091***	0.081***
						(0.02)	(0.02)
Household owns elecrtic stove						-0.001	-0.010
						(0.03)	(0.03)
Log of residential area per capita						0.423***	0.391***
						(0.05)	(0.05)
Roof is made of concrete/metal sheets						0.054*	0.037
						(0.03)	(0.03)
Roof is made of other materials						0.089**	0.086**
						(0.04)	(0.04)
Wall is made of stones/stones and other						0.063*	0.055
						(0.04)	(0.04)
Wall is made of blocks/bricks						0.069*	0.054
						(0.04)	(0.04)
Wall is made of other						0.110*	0.089
						(0.06)	(0.06)
Improved water						0.014	-0.012
						(0.03)	(0.03)
Improved toilet						0.075***	0.072***
						(0.03)	(0.02)
Floor is made of cement						0.104***	0.096***
						(0.03)	(0.03)
Floor is made of other material						0.171***	0.175***
						(0.04)	(0.04)
_cons	10.485**	1.438**	8.228***	10.500**	10.122***	9.789***	9.688***
	(0.08)	(0.06)	(0.10)	(0.08)	(0.08)	(0.09)	(0.09)
sigma_e	0.47	0.16	0.43	0.47	0.46	0.45	0.44
sigma_u	0.35	0.08	0.29	0.35	0.33	0.33	0.32
rho	0.36	0.18	0.32	0.36	0.34	0.35	0.34
r2_o	0.42	0.95	0.57	0.43	0.47	0.49	0.51
N	3377	3377	3377	3377	3377	3377	3377

Table A.3. Household consumption model, Malawi 2016/17

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Household size	-0.084***	-0.096***	-0.031***	-0.027***	-0.084***	-0.098***	-0.103***	-0.086***	-0.070***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Head's age	-0.002***	-0.002***	-0.002***	0.000	-0.002***	-0.002***	-0.003***	-0.002***	-0.002***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Head is female	-0.016*	-0.009	0.007	-0.011*	-0.003	-0.006	-0.018**	-0.007	-0.016*
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Head has primary education	0.149***	0.072***	0.020***	0.026***	0.055***	0.068***	0.070***	0.071***	0.143***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Head has junior secondary education	0.210***	0.094***	0.024***	0.032***	0.082***	0.094***	0.089***	0.089***	0.196***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Head has secondary education	0.359***	0.139***	0.032***	0.056***	0.125***	0.138***	0.134***	0.139***	0.344***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Head has diploma/degree	0.824***	0.285***	0.128***	0.083***	0.244***	0.292***	0.276***	0.288***	0.791***
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Share of household members age 0-14	-0.721***	-0.458***	-0.216***	-0.087***	-0.417***	-0.478***	-0.494***	-0.434***	-0.669***
	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
Share of household members age 15-24	-0.232***	-0.116***	-0.051***	-0.037***	-0.108***	-0.126***	-0.131***	-0.111***	-0.216***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Share of household members age 60 and older	-0.215***	-0.099***	-0.046***	-0.052***	-0.073***	-0.111***	-0.063***	-0.097***	-0.205***
	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Head is employed for a wage/salary/commission in the last 12 months	0.038***	0.040***	0.013**	0.013*	0.041***	0.041***	0.039***	0.040***	0.038***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Head engaged in casual/ganyu labor in the last 12 months	-0.202***	-0.094***	-0.017***	-0.055***	-0.091***	-0.097***	-0.094***	-0.100***	-0.202***
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Urban	-0.441***	-0.226***	-0.109***	-0.011	-0.226***	-0.205***	-0.223***	-0.203***	-0.391***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Log of food consumption per capita			0.736***						
			(0.00)						
Log of non-food consumption per capita				0.623***					
				(0.01)					
Log of furnishings expenses per capita					0.109***				
					(0.00)				
Log of health expenditures per capita						0.019***			
						(0.00)			
Log of education expenditures per capita							0.010***		
							(0.00)		
Lof of utilities per capita								0.068***	0.091***
								(0.00)	(0.00)
Household owns a car		0.427***	0.339***	0.054***	0.376***	0.431***	0.427***	0.412***	
		(0.03)	(0.01)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	
Household owns a motorcycle		0.207***	0.094***	0.022	0.168***	0.206***	0.208***	0.198***	
		(0.03)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	
Household owns a bicycle		0.088***	0.018***	0.028***	0.064***	0.083***	0.086***	0.087***	
		(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Household owns a mobile phone		0.204***	0.103***	0.011*	0.154***	0.198***	0.199***	0.196***	
		(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Household owns an CD / DVD player		0.086***	0.038***	0.016	0.083***	0.087***	0.082***	0.081***	
		(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	
Household owns a television		0.148***	0.054***	0.034***	0.132***	0.147***	0.146***	0.133***	
		(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	
Household owns a computer		0.181***	0.111***	0.031*	0.146***	0.185***	0.174***	0.173***	
		(0.03)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	
Household owns a refrigerator		0.130***	0.055***	0.016	0.052**	0.129***	0.124***	0.116***	
		(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Household owns a air conditioner		-0.002	-0.012	-0.003	0.011	-0.001	-0.007	0.095	
		(0.11)	(0.06)	(0.08)	(0.11)	(0.11)	(0.11)	(0.11)	
Household owns a fan		0.078***	0.044***	0.023	0.059***	0.071***	0.077***	0.076***	
		(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Household owns a washing machine		0.289***	0.248***	0.091	0.200**	0.336***	0.301***	0.282***	
		(0.09)	(0.05)	(0.06)	(0.09)	(0.09)	(0.09)	(0.09)	
Log of residential area per capita		0.305***	0.210***	-0.021	0.253***	0.314***	0.307***	0.279***	
		(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	
Household dwelling has improved walls		0.029***	0.007	0.004	0.020**	0.025***	0.027***	0.029***	
		(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Household dwelling has improved roof		0.068***	0.028***	0.001	0.059***	0.066***	0.064***	0.064***	
		(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Household dwelling has improved floor		0.080***	0.049***	0.003	0.067***	0.086***	0.079***	0.075***	
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Household water source is improved		0.033***	0.005	0.014*	0.024**	0.037***	0.032***	0.032***	
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Household toilet facility is improved		0.105***	0.035***	0.033***	0.098***	0.097***	0.103***	0.098***	
		(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	
Household has mosquito nets		0.032***	-0.004	0.024***	0.024***	0.030***	0.034***	0.032***	
		(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	
_cons	13.944***	13.045***	3.923***	5.414***	12.131***	12.932***	13.079***	12.299***	12.872***
	(0.04)	(0.04)	(0.05)	(0.07)	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)
sigma_e	0.41	0.36	0.18	0.25	0.34	0.36	0.36	0.35	0.40
sigma_u	0.16	0.14	0.05	0.08	0.12	0.12	0.14	0.13	0.15
rho	0.14	0.13	0.06	0.09	0.11	0.11	0.13	0.12	0.12
r2_o	0.53	0.64	0.92	0.84	0.69	0.66	0.65	0.66	0.57
N	12446	12446	12446	12446	12446	12446	12446	12446	12446

Table A.4. Household consumption model, Nigeria 2012/13

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Household size	-0.069***	-0.089***	-0.015***	-0.038***	-0.089***	-0.088***	-0.091***	-0.088***	-0.068***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Head's age	-0.001**	-0.003***	-0.000	-0.002***	-0.003***	-0.003***	-0.004***	-0.003***	-0.002**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Head is female	-0.052**	-0.002	0.007	-0.015	0.005	-0.001	-0.007	-0.004	-0.053**
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Head has primary education	0.108***	0.039**	0.016**	-0.011	0.043**	0.035*	0.027	0.038**	0.106***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Head has secondary education	0.243***	0.118***	0.033***	0.014	0.117***	0.115***	0.099***	0.121***	0.245***
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Head has secondary vocational education and higher	0.558***	0.250***	0.110***	0.026	0.237***	0.243***	0.227***	0.246***	0.551***
	(0.03)	(0.03)	(0.01)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Share of household members in 0-14	-0.439***	-0.321***	-0.131***	-0.041	-0.325***	-0.322***	-0.310***	-0.320***	-0.435***
	(0.04)	(0.04)	(0.01)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Share of household members in 25-59	0.203***	0.181***	-0.027*	0.163***	0.174***	0.177***	0.282***	0.180***	0.202***
	(0.05)	(0.04)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Share of household members in 60 and older	0.076	0.175***	-0.006	0.134***	0.173***	0.165***	0.312***	0.172***	0.074
	(0.05)	(0.05)	(0.02)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Head did any work in last 7 days	0.157***	0.113***	0.009	0.070***	0.109***	0.114***	0.112***	0.111***	0.153***
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Urban	-0.337***	-0.130***	-0.037***	-0.029	-0.126***	-0.137***	-0.126***	-0.123***	-0.324***
	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Log of perca food consumption			0.855***						
			(0.00)						
Log of perca non-food consumption				0.522***					
				(0.01)					
Log of perca infrequent non-food consumption					0.029***				
					(0.00)				
Lof of perca health expenditures						0.025***			
						(0.00)			
Log of perca education expenditures							0.015***		
							(0.00)		
Log of perca utilities								0.013***	0.018***
								(0.00)	(0.00)
Household owns a motorcycle		0.086***	0.027***	0.023	0.081***	0.078***	0.079***	0.088***	
		(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Household owns a bicycle		0.167***	0.082***	0.047	0.175***	0.167***	0.163***	0.172***	
		(0.06)	(0.02)	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)	
Household owns a mobile phone		0.086	-0.062*	0.059	0.081	0.091	0.083	0.099	
		(0.10)	(0.04)	(0.08)	(0.10)	(0.10)	(0.10)	(0.10)	
Household owns a DVD player		-0.013	-0.011*	0.003	-0.016	-0.017	-0.014	-0.013	
		(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	
Household owns a television		0.065***	0.013**	0.001	0.062***	0.062***	0.065***	0.064***	
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Household owns a computer		0.214***	0.091***	0.038*	0.201***	0.212***	0.211***	0.214***	
		(0.03)	(0.01)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	
Household owns a refrigerator		0.109***	0.024***	0.037**	0.111***	0.106***	0.104***	0.106***	
		(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Household owns an air conditioner		0.063***	0.021**	0.015	0.062***	0.061***	0.063***	0.061**	
		(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Household owns a washing machine		0.136***	0.087***	0.024	0.127***	0.121***	0.127***	0.133***	
		(0.04)	(0.01)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	
Household owns a car		0.091***	0.011	0.040**	0.093***	0.083***	0.086***	0.093***	
		(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Household owns a fan		0.133***	0.046***	-0.027**	0.132***	0.126***	0.124***	0.132***	
		(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	
Household owns a satellite		0.054*	-0.022*	0.046*	0.056*	0.058*	0.053	0.054*	
		(0.03)	(0.01)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	
Log of residential area		0.145***	0.057***	0.011	0.131***	0.150***	0.141***	0.142***	
		(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	
Roof is made of concrete/metal sheets/tiles		0.090***	0.025***	0.012	0.098***	0.085***	0.087***	0.089***	
		(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	
Wall is made of burnt bricks/concrete/metal sheets		0.048**	0.016**	-0.004	0.050***	0.050***	0.044**	0.045**	
		(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	
Piped water/Truck		0.022	-0.005	0.012	0.021	0.029	0.022	0.020	
		(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Any well water		0.016	-0.001	0.019	0.008	0.015	0.021	0.017	
		(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	
On water/Flush/VIP toilet		0.152***	0.071***	-0.026	0.156***	0.133***	0.151***	0.156***	
		(0.03)	(0.01)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	
Other toilet		0.050**	0.017**	-0.036**	0.052**	0.036*	0.048**	0.051**	
		(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Toilet is not shared		0.063***	0.035***	0.021	0.059***	0.077***	0.064***	0.062***	
		(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	
_cons	12.270***	11.544***	1.856***	6.524***	11.554***	11.480***	11.491***	11.434***	12.109***
	(0.08)	(0.08)	(0.06)	(0.11)	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)
sigma_e	0.44	0.40	0.14	0.31	0.39	0.40	0.40	0.40	0.44
sigma_u	0.23	0.16	0.05	0.12	0.16	0.16	0.16	0.16	0.22
rho	0.21	0.14	0.12	0.12	0.15	0.14	0.15	0.14	0.20
r2_o	0.45	0.58	0.95	0.76	0.59	0.59	0.58	0.58	0.46
N	4405	4405	4405	4405	4405	4405	4405	4405	4405

Note: ***p<0.01, **p<0.05, *p<0.1. Standard errors are in parentheses. All estimations employs cluster random effects models and controls for the regional dummy variables.

Table A.5. Household consumption model, Tanzania 2019/20

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Household size	-0.031***	-0.029***	-0.006*	-0.018***	-0.031***	-0.031***	-0.030***	-0.027***	-0.026***
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Head's age	-0.003***	-0.005***	-0.000	-0.004***	-0.005***	-0.005***	-0.005***	-0.005***	-0.004***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Head is female	0.089**	0.073**	0.013	0.029	0.065*	0.068*	0.072*	0.072**	0.080**
	(0.04)	(0.04)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Head has primary education	0.063	-0.026	0.013	-0.024	-0.020	-0.033	-0.026	-0.032	0.032
	(0.05)	(0.04)	(0.02)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Head has secondary ordinary education	0.355***	0.114**	0.039	0.058	0.108**	0.093*	0.117**	0.110**	0.309***
	(0.06)	(0.05)	(0.03)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)
Head has secondary advanced education and higher	0.649***	0.176**	0.143***	0.041	0.169**	0.138*	0.180**	0.161*	0.567***
	(0.08)	(0.09)	(0.04)	(0.07)	(0.08)	(0.08)	(0.09)	(0.08)	(0.08)
Share of household members age 0-14	-0.736***	-0.498***	-0.135***	-0.350***	-0.492***	-0.541***	-0.530***	-0.500***	-0.729***
	(0.09)	(0.09)	(0.04)	(0.08)	(0.09)	(0.09)	(0.10)	(0.09)	(0.09)
Share of household members age 15-24	-0.297***	-0.224***	-0.032	-0.141**	-0.233***	-0.211***	-0.235***	-0.206***	-0.238***
	(0.08)	(0.07)	(0.03)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
Share of household members age 60 and older	0.104	0.100	-0.024	0.164*	0.098	0.083	0.113	0.093	0.109
	(0.11)	(0.10)	(0.05)	(0.09)	(0.10)	(0.10)	(0.10)	(0.10)	(0.11)
Head was working for wage/salary last 7 days	0.143***	0.099***	0.026	0.064**	0.086**	0.104***	0.098***	0.099***	0.135***
	(0.04)	(0.04)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Head was self-employed (non-farm) last 7 days	0.165***	0.049	0.022	0.010	0.049	0.056	0.048	0.040	0.125***
	(0.04)	(0.04)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Dar es Salam			0.831***						
			(0.01)						
Rest of urban				0.274***					
				(0.01)					
Zanzibar					0.022***				
					(0.00)				
Log of food consumption peraeq						0.020***			
						(0.00)			
Log of non-food consumption peraeq							0.003		
							(0.00)		
Log of furnishings and household expenses peraeq								0.017***	0.042***
								(0.01)	(0.01)
Log of health expenditures peraeq		0.219***	0.138***	0.099**	0.204***	0.215***	0.218***	0.218***	
		(0.05)	(0.02)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	
Log of education expenditures peraeq		0.029	-0.027*	0.050*	0.020	0.039	0.029	0.031	
		(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Lof of utilities peraeq		0.048	0.011	-0.176	-0.001	-0.126	0.059	0.156	
		(0.48)	(0.23)	(0.41)	(0.48)	(0.47)	(0.48)	(0.48)	
Household owns a motorcycle		0.141***	0.068***	-0.000	0.121**	0.131***	0.138***	0.141***	
		(0.05)	(0.02)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	
Household owns a bicycle		0.128**	0.017	0.094**	0.110**	0.126**	0.128**	0.120**	
		(0.05)	(0.02)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	
Household owns a desk phone		0.061	0.038	0.019	0.061	0.064	0.059	0.054	
		(0.05)	(0.02)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	
Household owns a mobile phone		0.183**	0.146***	0.078	0.176**	0.221***	0.179**	0.185**	
		(0.08)	(0.04)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	
Household owns an CD / DVD player		0.130**	0.085***	0.065	0.136***	0.130**	0.124**	0.121**	
		(0.05)	(0.02)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	
Household owns a television		0.200***	0.067***	0.107**	0.176***	0.178***	0.200***	0.196***	
		(0.05)	(0.02)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	
Household owns a compute		0.027	0.001	0.004	0.022	0.030	0.027	0.028	
		(0.03)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Household owns a refrigerator		0.082**	0.016	0.048	0.084**	0.069*	0.084**	0.081**	
		(0.04)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	
Household owns a air conditioner/fan		0.424***	0.096***	0.357***	0.418***	0.417***	0.423***	0.420***	
		(0.07)	(0.03)	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)	
Household owns a radio		0.095*	0.007	0.045	0.100**	0.097**	0.094*	0.079	
		(0.05)	(0.02)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	
Household owns a mosquito net		-0.138*	-0.044	-0.063	-0.162**	-0.122	-0.139*	-0.127	
		(0.08)	(0.04)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	
Log of residential area per capita		-0.167**	-0.079**	-0.083	-0.180**	-0.157**	-0.166**	-0.157**	
		(0.08)	(0.04)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	
Roof is made of concrete/metal sheets/tiles		0.056	0.057***	-0.041	0.055	0.054	0.054	0.043	
		(0.04)	(0.02)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	
Wall is made of burnt bricks/stones		0.161***	0.024	0.126***	0.168***	0.181***	0.161***	0.144***	
		(0.05)	(0.02)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	
Wall is made of mud bricks/mud stones		0.026	-0.014	0.030	0.024	0.042	0.026	0.032	
		(0.05)	(0.02)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	
Floor is made of concrete/cement/tiles		0.127***	0.029	0.094***	0.129***	0.119***	0.126***	0.115***	
		(0.04)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	
Piped water	0.416***	0.094*	0.079***	0.006	0.122**	0.106**	0.095*	0.076	0.294***
	(0.06)	(0.06)	(0.03)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
Any well water	0.280***	0.091**	0.054***	0.011	0.101***	0.082**	0.089**	0.078**	0.205***
	(0.04)	(0.04)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Flush/VIP toilet	0.172**	-0.098*	-0.066**	-0.030	-0.009	-0.106*	-0.096	-0.113*	0.079
	(0.07)	(0.06)	(0.03)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
_cons	14.135***	13.692***	2.505***	10.536***	13.539***	13.611***	13.713***	13.566***	13.809***
	(0.10)	(0.13)	(0.19)	(0.19)	(0.13)	(0.13)	(0.13)	(0.14)	(0.10)
sigma_e	0.51	0.46	0.22	0.40	0.46	0.45	0.46	0.46	0.50
sigma_u	0.19	0.06	0.00	0.04	0.06	0.03	0.06	0.07	0.15
rho	0.12	0.02	0.00	0.01	0.02	0.01	0.02	0.02	0.08
r2_o	0.40	0.55	0.90	0.67	0.56	0.57	0.55	0.55	0.43
N	1179	1179	1179	1179	1179	1179	1179	1179	1179

Table A.6. Household consumption model, Vietnam 2014

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Household size	-0.065***	-0.132***	-0.068***	-0.014***	-0.101***	-0.128***	-0.134***	-0.115***	-0.046***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Head's age	0.002***	-0.001	-0.000	-0.000**	0.000	-0.001**	-0.001	-0.001	0.001*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Head is female	0.019	0.037***	0.036***	-0.011***	0.030***	0.031***	0.039***	0.030***	0.006
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Head belongs to ethnic minority group	-0.419***	-0.187***	-0.092***	-0.015**	-0.174***	-0.135***	-0.173***	-0.146***	-0.189***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Head completed primary school	0.177***	0.046***	0.033***	-0.003	0.037***	0.037***	0.043***	0.039***	0.102***
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Head completed lower secondary school	0.281***	0.074***	0.047***	-0.003	0.058***	0.060***	0.071***	0.064***	0.168***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Head completed upper secondary school	0.468***	0.143***	0.086***	0.008	0.120***	0.128***	0.135***	0.126***	0.305***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Head has (some) college education	0.729***	0.216***	0.090***	0.038***	0.172***	0.209***	0.206***	0.200***	0.515***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Share of household members age 0-14	-0.286***	-0.324***	-0.145***	-0.059***	-0.239***	-0.267***	-0.505***	-0.280***	-0.193***
	(0.04)	(0.03)	(0.02)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Share of household members age 15-24	0.113***	0.029	0.026	0.007	0.061**	0.082***	-0.099***	0.064**	0.182***
	(0.04)	(0.03)	(0.02)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Share of household members age 15-24	0.275***	0.057***	0.003	0.044***	0.042**	0.084***	0.046**	0.058***	0.207***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Head worked in the last 12 months	-0.020	-0.013	-0.027***	0.007	-0.020*	0.004	-0.016	-0.011	-0.015
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Log of food consumption per capita			0.673***						
			(0.01)						
Log of non-food consumption per capita				0.748***					
				(0.00)					
Log of durables consumption per capita					0.183***				
					(0.00)				
Log of health expenditures per capita						0.070***			
						(0.00)			
Log of education expenditures per capita							0.019***		
							(0.00)		
Log of electricity, water & garabage expenditures per capita								0.103***	0.243***
								(0.01)	(0.01)
Household owns a car		0.609***	0.540***	0.030**	0.274***	0.620***	0.601***	0.590***	
		(0.03)	(0.02)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	
Household owns a motorbike		0.167***	0.108***	-0.029***	-0.002	0.161***	0.164***	0.164***	
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Household owns a bicycle		-0.041***	-0.010*	-0.017***	-0.033***	-0.044***	-0.058***	-0.043***	
		(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	
Household owns a desk phone		0.067***	0.042***	0.009*	0.070***	0.062***	0.067***	0.060***	
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Household owns a cell phone		0.128***	0.065***	0.001	0.053***	0.111***	0.124***	0.113***	
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Household owns a DVD player		0.051***	0.021***	0.007**	0.015**	0.050***	0.052***	0.045***	
		(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	
Household owns a television		0.065***	0.036***	-0.009	-0.009	0.054***	0.058***	-0.003	
		(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	
Household owns a computer		0.168***	0.105***	0.007	0.090***	0.163***	0.149***	0.156***	
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Household owns a refrigerator		0.160***	0.094***	0.001	0.074***	0.147***	0.155***	0.110***	
		(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	
Household owns an airconditioner		0.207***	0.121***	0.034***	0.139***	0.201***	0.213***	0.176***	
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Household owns a washing machine		0.109***	0.055***	0.013***	0.052***	0.097***	0.102***	0.095***	
		(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	
Household owns an electric fan		0.068***	0.034***	0.002	0.039***	0.049***	0.065***	0.030**	
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Log of residential area		0.187***	0.141***	-0.006*	0.151***	0.183***	0.184***	0.178***	
		(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	
House wall materials		0.026***	0.017***	-0.006***	0.020***	0.024***	0.025***	0.020***	
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
Access to drinking water		0.007***	0.004***	0.000	0.008***	0.007***	0.007***	0.003	
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
Type of toilet		0.040***	0.016***	0.006***	0.031***	0.038***	0.039***	0.032***	
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
Urban	0.269***	0.063***	0.042***	0.004	0.103***	0.076***	0.059***	0.047***	0.139***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	9.475***	8.422***	2.949***	2.701***	7.611***	8.064***	8.470***	8.029***	8.088***
	(0.05)	(0.05)	(0.06)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
σ_ε	0.39	0.30	0.21	0.14	0.28	0.29	0.30	0.29	0.35
σ_u	0.25	0.19	0.11	0.08	0.18	0.18	0.19	0.19	0.23
ρ	0.29	0.28	0.21	0.24	0.29	0.28	0.29	0.29	0.31
R^2	0.46	0.69	0.86	0.94	0.73	0.71	0.69	0.70	0.56
N	9300	9300	9300	9300	9300	9300	9300	9300	9300

Note: * p<0.10, ** p<0.05 *** p<0.01. Standard errors are in parentheses. All estimation employs commune random effects models and control for regional dummy variables. House wall material is assigned numerical values using the following categories: 6 "cement", 5 "brick", 4 "iron/wood", 3 "earth/straw", 2 "bamboo/board", and 1 "others". The types of toilet are assigned numerical values using the following categories: 6 "septic", 5 "suilabh", 4 "double septic", 3 "fish bridge", 2 "others", and 1 "none".

Table A.7. Predicted Poverty Rates Based on Imputation (percentage)

Indicators	Second period									True rates
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	
Panel A: Bangladesh 2015-2018/19										
Headcount poverty rate	12.5	**7.5***	**8.2**	**6.7**	**7.7***	**6.8***	**7.6***	N/A	N/A	7.3
	(0.7)	(0.6)	(0.8)	(0.7)	(0.6)	(0.6)	(0.7)			(0.6)
Near-poverty rate	13.1	9.7	**11.8***	10.5	10.1	9.4	9.7	13.1	9.7	12.2
	(0.6)	(0.6)	(0.7)	(0.6)	(0.6)	(0.6)	(0.6)	(0.6)	(0.6)	(0.6)
Extreme poverty rate	1.9	**0.9**	**0.8**	**0.5***	**0.8**	**0.7***	**0.9**	1.9	**0.9**	0.6
	(0.3)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.3)	(0.2)	(0.2)
Poverty gap	2.4	**1.3**	**1.3**	**1.0***	**1.3**	**1.1***	**1.3**	2.4	**1.3**	1.1
	(0.2)	(0.1)	(0.2)	(0.1)	(0.1)	(0.1)	(0.1)	(0.2)	(0.1)	(0.1)
USAID Poverty gap	18.8	17.0	**16.1**	**15.1***	**16.5**	**16.3**	17.0	18.8	17.0	15.2
	(0.7)	(0.9)	(0.9)	(0.9)	(0.9)	(0.9)	(0.9)	(0.7)	(0.9)	(0.8)
N										
Panel B: Malawi 2016/17-2019/20										
Headcount poverty rate	53.5	**52.8**	60.8	52.8	53.9	**52.7**	**52.7**	**52.3**	**52.7**	51.1
	(0.9)	(1.0)	(1.1)	(1.2)	(1.0)	(1.0)	(1.0)	(1.0)	(0.9)	(0.9)
Near-poverty rate	**14.0***	**14.2***	12.3	**14.3***	**14.3***	**14.4***	**14.2***	**14.4***	**14.3***	14.1
	(0.4)	(0.4)	(0.5)	(0.5)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)
Extreme poverty rate	22.7	**21.5**	29.6	**21.6**	**21.9**	**21.5**	**21.7**	**21.1***	**21.8**	20.6
	(0.7)	(0.8)	(1.1)	(0.9)	(0.8)	(0.8)	(0.8)	(0.8)	(0.7)	(0.8)
Poverty gap	18.2	**17.6**	22.6	**17.5**	18.0	**17.5***	**17.7**	**17.3***	**17.7**	17.1
	(0.4)	(0.5)	(0.6)	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)	(0.4)	(0.4)
USAID Poverty gap	**34.1**	**33.4***	37.2	**33.2***	**33.3***	**33.3***	**33.6***	**33.2***	**33.6***	33.5
	(0.4)	(0.4)	(0.6)	(0.5)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)	(0.5)
N	11,432	11,432	11,432	11,432	11,432	11,432	11,432	11,432	11,432	11,432
Panel C: Nigeria 2012/13-2018/19										
Headcount poverty rate	33.4	33.5	51.3	27.6	24.7	29.7	33.4	34.4	34.6	46.4
	(2.0)	(2.0)	(2.4)	(1.9)	(1.8)	(1.9)	(2.0)	(2.0)	(2.0)	(1.9)
Near-poverty rate	**12.5**	**12.7***	**13.4***	**13.3***	11.5	**12.6**	**12.6**	**12.6**	**12.4**	13.7
	(0.9)	(0.9)	(1.0)	(1.0)	(0.9)	(0.9)	(0.9)	(0.9)	(0.9)	(1.0)
Extreme poverty rate	14.7	14.8	26.4	9.0	10.0	12.2	14.9	15.4	15.5	22.2
	(1.4)	(1.5)	(2.1)	(1.1)	(1.2)	(1.3)	(1.5)	(1.5)	(1.4)	(1.4)
Poverty gap	10.9	10.9	18.0	7.4	7.6	9.3	11.0	11.3	11.4	15.3
	(0.9)	(0.9)	(1.2)	(0.7)	(0.7)	(0.8)	(0.9)	(0.9)	(0.9)	(0.8)
USAID poverty gap	**32.6***	**32.7***	35.1	26.9	30.7	31.2	**33.0***	**33.0***	**33.0***	33.0
	(1.1)	(1.3)	(1.1)	(1.1)	(1.5)	(1.3)	(1.3)	(1.2)	(1.1)	(0.8)
N	4,976	4,976	4,976	4,976	4,976	4,976	4,976	4,976	4,976	4,976
Panel D: Tanzania 2019/20-2020/21										

Headcount poverty rate	**17.1***	**17.3***	**19.4**	**16.4**	**17.1***	**16.5**	**17.3***	**17.6***	**18.2***	17.8
	(1.1)	(1.3)	(1.5)	(1.3)	(1.2)	(1.3)	(1.3)	(1.3)	(1.1)	(1.1)
Near-poverty rate	**10.5***	**10.9***	**10.7***	**11.3**	**10.8***	**10.3***	**10.8***	**10.8***	**10.5***	10.2
	(0.7)	(0.8)	(0.8)	(0.8)	(0.8)	(0.8)	(0.8)	(0.8)	(0.7)	(0.7)
Extreme poverty rate	**9.8***	**9.4***	**10.9**	**8.4**	**9.3***	**9.0**	**9.5***	**9.6***	**10.4***	9.8
	(0.8)	(1.0)	(1.1)	(0.9)	(0.9)	(1.0)	(1.0)	(1.0)	(0.9)	(0.8)
Poverty gap	**4.6***	**4.5***	**5.2**	**3.9**	**4.4***	**4.2**	**4.5***	**4.5***	**4.9***	4.6
	(0.4)	(0.4)	(0.5)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)	(0.3)
USAID Poverty gap	**26.9***	**25.8***	**27.0***	**24.0**	**25.6***	**25.5***	**25.8***	**25.7***	**26.8***	25.9
	(1.1)	(1.1)	(1.4)	(1.1)	(1.1)	(1.2)	(1.1)	(1.1)	(1.1)	(1.1)
N	4,644	4,644	4,644	4,644	4,644	4,644	4,644	4,644	4,644	4,644
Panel E: Vietnam 2014-2016										
Headcount poverty rate	14.7	13.1	6.0	8.3	**10.3**	12.3	13.1	11.2	**9.0**	9.6
	(0.5)	(0.6)	(0.5)	(0.6)	(0.6)	(0.6)	(0.6)	(0.6)	(0.5)	(0.4)
Near-poverty rate	9.4	8.6	4.9	6.1	**7.2***	8.0	8.6	7.8	**6.9***	6.9
	(0.4)	(0.4)	(0.3)	(0.4)	(0.3)	(0.4)	(0.4)	(0.4)	(0.3)	(0.3)
Extreme poverty rate	1.9	1.9	0.6	**1.2***	**1.5**	2.0	2.0	1.9	1.9	1.2
	(0.2)	(0.3)	(0.1)	(0.2)	(0.2)	(0.3)	(0.3)	(0.3)	(0.4)	(0.2)
Poverty gap	3.9	3.6	1.5	**2.3**	**2.8**	3.4	3.6	3.1	**2.7**	2.5
	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.3)	(0.3)	(0.2)
USAID Poverty gap	**26.5***	**27.2**	**24.5**	**27.4**	**26.9***	27.9	**27.4**	28.0	30.0	26.1
	(0.7)	(1.0)	(1.3)	(1.3)	(1.0)	(1.0)	(1.0)	(1.4)	(1.9)	(0.9)
N	9,347	9,347	9,347	9,347	9,347	9,347	9,347	9,347	9,347	
Control variables										
Food expenditures			Y							
Non-food expenditures				Y						
Furnishings and household expenses					Y					
Health expenditures						Y				
Education expenditures							Y			
Utilities expenditures								Y	Y	
Household assets & house characteristics		Y	Y	Y	Y	Y	Y	Y		
Demographics & employment	Y	Y	Y	Y	Y	Y	Y	Y	Y	

Note: Empirical distribution of the error terms model with bootstrapped SEs is used. The standard errors are calculated using 100 bootstrap replications and are adjusted for complex survey design. All estimates are obtained with population weights. The normal linear regression model with the theoretical distribution of the error terms employs cluster random effects. Imputed poverty rates for 2019/20 use the estimated parameters based on the 2016/17 data. 100 simulations are implemented. 'Near poor' status is defined as living on an income between 100 and 125% of the poverty line. All indicators are expressed in %. True rate is the estimate directly obtained from the survey data. Estimates shown in boldface or with a "*" respectively fall within the 95% confidence interval or one standard error of the true rate.

Table A.8. Predicted Mean (Log) Consumption Based on Imputation

Indicators	Second round									
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**	**Model 9**	**True rates**
Bangladesh, from 2015 to 2018/19	10.7 (0.0)	10.8 (0.0)	**10.8** (0.0)	10.8 (0.0)	10.8 (0.0)	10.9 (0.0)	10.8 (0.0)	N/A	N/A	10.8 (0.0)
Malawi, from 2016/17 to 2018/19	12.1 (0.0)	12.1 (0.0)	**12.0** (0.0)	12.1 (0.0)	12.1 (0.0)	12.1 (0.0)	12.1 (0.0)	12.2 (0.0)	12.1 (0.0)	12.1 (0.0)
Nigeria, from 2012/13 to 2018/19	11.3 (0.0)	11.3 (0.0)	11.0 (0.0)	11.4 (0.0)	11.4 (0.0)	11.3 (0.0)	11.3 (0.0)	11.2 (0.0)	11.2 (0.0)	11.0 (0.0)
Tanzania, from 2019/20 to 2020/21	13.7 (0.0)	13.7 (0.0)	13.7 (0.0)	**13.7** (0.0)	**13.7** (0.0)	**13.7*** (0.0)	**13.7** (0.0)	13.7 (0.0)	13.7 (0.0)	13.7 (0.0)
Vietnam, from 2014 to 2016	9.6 (0.0)	9.7 (0.0)	10.0 (0.0)	9.8 (0.0)	**9.8*** (0.0)	9.7 (0.0)	9.7 (0.0)	9.7 (0.0)	**9.8*** (0.0)	9.6 (0.0)
Food expenditures			Y							
Non-food expenditures				Y						
Furnishings and household expenses					Y					
Health expenditures						Y				
Education expenditures							Y			
Utilities: water, kerosene, lighting								Y	Y	
Household assets & house characteristics		Y	Y	Y	Y	Y	Y	Y		
Demographics & employment	Y	Y	Y	Y	Y	Y	Y	Y	Y	

Note: Empirical distribution of the error terms model with bootstrapped SEs is used. The standard errors are calculated using 100 bootstrap replications and are adjusted for complex survey design. All estimates are obtained with population weights. The normal linear regression model with the theoretical distribution of the error terms employs cluster random effects. Imputed consumption per capita for the second round uses the estimated parameters based on the data from the first round. 100 simulations are implemented. True consumption per capita is the estimate directly obtained from the survey data. Estimates shown in boldface or with a "*" respectively fall within the 95% confidence interval or one standard error of the true rate.

Table A.9. Meta-analysis of Imputation Models and Their Parameters, Logit Regressions

	Headcount poverty rate		Near poverty rate		Extreme poverty rate		Poverty gap		USAID Poverty gap		Consumption mean	
	Spec.1	*Spec.2*	*Spec.1*	*Spec.2*	*Spec.1*	*Spec.2*	*Spec.1*	*Spec.2*	*Spec.1*	*Spec.2*	*Spec.1*	*Spec.2*
Model 2: Demographics, employment, assets, house	0.344	-0.102	-0.187	-2.435	-0.192	1.264	-0.000	1.185	-0.000	2.836*	-0.000	2.135
characteristics	(1.14)	(1.70)	(0.14)	(2.37)	(0.97)	(2.05)	(1.12)	(2.07)	(0.68)	(1.60)	(0.89)	(2.06)
Model 3	1.463***	0.284	1.014	-5.185	1.621***	6.087*	1.620**	4.993	0.212	8.177**	2.016***	9.970**
(adds food exp. to Model 2)	(0.19)	(2.60)	(0.90)	(6.88)	(0.56)	(3.45)	(0.71)	(3.28)	(0.20)	(3.38)	(0.53)	(4.28)
Model 4	0.344	-0.756	0.380	-4.418	0.537	4.042	-0.000	2.446	-0.410	5.785**	-0.000	4.735
(adds nonfood exp. to Model 2)	(0.36)	(2.03)	(0.72)	(4.29)	(0.66)	(3.01)	(0.86)	(2.63)	(0.38)	(2.77)	(0.00)	(3.24)
Model 5	0.344	-0.190	0.188	-2.307	-0.397	1.215	-0.000	1.369	-0.207	3.120*	0.289	2.722
(adds durables exp. to Model 2)	(1.00)	(1.74)	(0.98)	(3.66)	(0.94)	(2.20)	(1.34)	(2.46)	(0.51)	(1.70)	(0.77)	(2.29)
Model 6	0.344	-0.155	-0.187	-2.668	-0.192	1.429	0.193	1.539	-0.611*	2.497*	-0.000	2.353
(adds health exp. to Model 2)	(0.76)	(1.53)	(0.14)	(2.65)	(0.97)	(2.16)	(0.98)	(2.02)	(0.33)	(1.42)	(1.02)	(1.95)
Model 7	0.664	0.311	-0.187	-2.458	-0.192	1.274	-0.000	1.196	-0.000	2.876*	0.289	2.472
(adds education exp. to Model 2)	(1.13)	(1.80)	(0.14)	(2.38)	(0.97)	(2.07)	(1.12)	(2.08)	(0.68)	(1.64)	(0.77)	(2.12)
Model 8	0.820	0.486	0.380	-1.929	0.184	1.925	0.731	2.186	0.212	3.298**	0.289	2.675
(adds utilities exp. to Model 2)	(0.72)	(1.59)	(0.43)	(2.44)	(0.78)	(1.91)	(0.77)	(1.82)	(0.57)	(1.51)	(0.77)	(2.23)
Model 9	1.640**	2.099**	0.579*	-0.002	0.883**	1.764***	0.904*	1.501*	-0.000	0.939	0.541	1.223***
(adds utilities exp. to demographic & employment)	(0.75)	(1.01)	(0.31)	(1.18)	(0.35)	(0.64)	(0.49)	(0.83)	(0.68)	(0.98)	(0.36)	(0.14)
Model 10	0.249	-0.270	-0.588	-2.802	-0.561	0.668	-0.051	0.937	-0.255	2.611	-0.463	1.673
(adds distance to facilities to Model 2)	(1.39)	(1.92)	(0.60)	(2.61)	(1.48)	(2.44)	(1.60)	(2.46)	(0.96)	(1.92)	(1.58)	(2.53)
Model 11	0.465	0.049	-0.588	-2.797	-0.561	0.885	-0.361	0.779	-0.255	2.574	0.031	2.074
(adds agricultural soil quality to Model 2)	(1.30)	(1.86)	(0.60)	(2.62)	(1.48)	(2.46)	(1.67)	(2.55)	(0.96)	(1.87)	(1.05)	(2.22)
Model 12	1.732	2.065	0.465	-0.064	0.020	0.437	0.471	0.854	-0.513	0.541	0.415	1.280
(adds distance to facilities to Model 9)	(1.12)	(1.46)	(0.48)	(1.27)	(0.64)	(1.13)	(1.17)	(1.54)	(0.83)	(1.20)	(1.22)	(1.12)
Model 13	1.986***	2.495***	0.465	0.004	0.020	0.557	0.221	0.681	-0.255	0.670	0.415	1.103***
(adds agricultural soil quality to Model 9)	(0.63)	(0.79)	(0.48)	(1.23)	(0.64)	(1.16)	(1.04)	(1.43)	(0.96)	(1.26)	(0.41)	(0.17)
True estimates												
Headcount poverty rate		0.033										
		(0.04)										
Near poverty rate				0.484								
				(0.37)								
Extreme poverty rate						0.194***						
						(0.03)						
Poverty gap								0.213***				
								(0.01)				
USAID poverty gap										0.051		
										(0.08)		
Consumption mean												
												1.690**
Other model parameters												(0.75)
Log of sample size of base survey		-0.927***		-4.164**		0.488		0.029		-0.058		-0.622
		(0.30)		(1.68)		(0.68)		(0.60)		(0.74)		(0.39)
Interval length between base & target surveys		-0.942**		-0.105		-1.188***		-0.768**		-0.279		-0.337
		(0.39)		(0.53)		(0.25)		(0.37)		(0.25)		(0.40)
Normal linear regression model		0.589**		0.065		0.152		0.026		-0.042		0.157
		(0.24)		(0.05)		(0.15)		(0.09)		(0.07)		(0.37)
Number of rounds used		-0.089		-1.010***		0.440**		0.601***		0.157		-1.873***
		(0.26)		(0.35)		(0.22)		(0.21)		(0.27)		(0.71)
R squared		3.682		14.670		-8.859		-6.941		-17.439**		-15.951
		(5.13)		(16.35)		(5.49)		(5.15)		(7.51)		(10.09)
Country FE												
Tanzania	0.037		0.228***		1.866***		1.861***		2.205***		17.845***	
	(0.08)		(0.06)		(0.11)		(0.07)		(0.07)		(1.44)	
Malawi	-0.397***		1.249***		0.929***		0.972***		0.117***		18.030***	
	(0.02)		(0.06)		(0.07)		(0.04)		(0.00)		(1.48)	
Nigeria	-0.455***		3.583***		1.054***		0.846***		2.096***		18.233***	
	(0.02)		(0.11)		(0.08)		(0.04)		(0.04)		(1.49)	
Constant	-0.673	7.052***	-0.404***	29.451***	-1.814***	-1.691	-1.982**	-0.195	-0.602	7.676**	-19.371***	-1.538
	(0.75)	(1.80)	(0.15)	(5.36)	(0.61)	(4.14)	(0.87)	(3.90)	(0.53)	(3.83)	(1.27)	(3.04)
Log likelihood	-202.33	-159.22	-175.90	-133.36	-179.25	-140.98	-181.26	-161.47	-175.55	-174.20	-128.22	-118.82
Pseudo R2	0.07	0.27	0.16	0.36	0.13	0.31	0.12	0.21	0.18	0.19	0.19	0.25
N	314	314	314	314	314	314	314	314	314	314	314	314

Table A.10. Meta-analysis of Imputation Models and Their Parameters, Ordered Logit Regressions

	Headcount poverty rate		Near poverty rate		Extreme poverty rate		Poverty gap		USAID Poverty gap		Consumption mean	
	Spec. 1	*Spec. 2*	*Spec. 1*	*Spec. 2*	*Spec. 1*	*Spec. 2*	*Spec. 1*	*Spec. 2*	*Spec. 1*	*Spec. 2*	*Spec. 1*	*Spec. 2*
Model 2: Demographics, employment, assets, house	0.231	0.130	-0.030	-0.622	-0.299	1.123	-0.161	1.263	-0.027	1.735*	0.136	2.275
characteristics	(1.32)	(2.05)	(0.34)	(1.66)	(1.01)	(2.15)	(1.20)	(2.31)	(0.34)	(0.89)	(0.94)	(2.36)
Model 3	1.044*	0.699	1.244**	-0.654	0.985	5.446	0.926	5.497	0.203	5.177***	2.086***	9.502**
(adds food exp. to Model 2)	(0.61)	(2.82)	(0.59)	(3.86)	(0.69)	(3.86)	(1.02)	(3.86)	(0.12)	(1.93)	(0.53)	(4.34)
Model 4	-0.054	-0.369	0.335	-1.050	-0.100	2.867	-0.325	2.514	-0.636	3.076	0.000	4.558
(adds nonfood exp. to Model 2)	(0.46)	(2.14)	(0.69)	(2.71)	(0.87)	(3.11)	(0.97)	(3.08)	(0.59)	(2.08)	(.)	(3.28)
Model 5	0.379	0.321	0.691	0.044	-0.507	1.126	-0.100	1.606	-0.167	1.877	0.389	2.788
(adds durables exp. to Model 2)	(1.29)	(2.18)	(0.87)	(2.30)	(1.05)	(2.42)	(1.40)	(2.75)	(0.48)	(1.18)	(0.81)	(2.50)
Model 6	0.034	-0.174	0.099	-0.479	-0.389	1.125	-0.108	1.467	-0.354	1.564*	0.257	2.683
(adds health exp. to Model 2)	(0.96)	(1.92)	(0.26)	(1.62)	(1.05)	(2.36)	(1.14)	(2.39)	(0.33)	(0.92)	(1.05)	(2.33)
Model 7	0.423	0.331	-0.030	-0.617	-0.412	0.928	-0.260	1.109	-0.027	1.754*	0.389	2.529
(adds education exp. to Model 2)	(1.21)	(2.00)	(0.34)	(1.67)	(0.91)	(1.92)	(1.11)	(2.11)	(0.34)	(0.92)	(0.81)	(2.33)
Model 8	0.700	0.745	0.167	-0.455	0.056	1.802	0.651	2.444	0.263	2.154**	0.456	2.843
(adds utilities exp. to Model 2)	(0.90)	(1.84)	(0.26)	(1.67)	(0.94)	(2.30)	(0.94)	(2.22)	(0.54)	(1.07)	(0.81)	(2.45)
Model 9	1.406*	1.490	0.436	0.280	0.573**	1.261**	0.553	1.096	-0.098	0.409	0.660**	1.376***
(adds utilities exp. to demographic & employment)	(0.76)	(0.95)	(0.30)	(0.75)	(0.27)	(0.60)	(0.68)	(1.13)	(0.45)	(0.70)	(0.26)	(0.16)
Model 10	0.120	-0.010	-0.153	-0.606	-0.736	0.508	-0.238	1.054	-0.562	1.240	-0.257	1.956
(adds distance to facilities to Model 2)	(1.58)	(2.36)	(0.70)	(1.83)	(1.58)	(2.69)	(1.69)	(2.81)	(0.71)	(1.24)	(1.63)	(2.91)
Model 11	0.307	0.187	-0.153	-0.630	-0.736	0.656	-0.528	0.855	-0.288	1.466	0.197	2.293
(adds agricultural soil quality to Model 2)	(1.48)	(2.25)	(0.70)	(1.81)	(1.58)	(2.68)	(1.76)	(2.88)	(0.83)	(1.34)	(1.11)	(2.50)
Model 12	1.166	1.225	0.304	0.180	-0.088	0.252	0.350	0.771	-0.611	0.052	0.315	1.155
(adds distance to facilities to Model 9)	(1.03)	(1.25)	(0.53)	(1.02)	(0.59)	(1.04)	(1.11)	(1.57)	(0.61)	(0.88)	(0.99)	(0.85)
Model 13	1.195	1.184	0.304	0.199	-0.088	0.330	0.132	0.582	-0.211	0.319	0.705	1.425***
(adds agricultural soil quality to Model 9)	(0.78)	(0.99)	(0.53)	(0.98)	(0.59)	(1.03)	(0.99)	(1.44)	(0.61)	(0.80)	(0.56)	(0.30)
True estimates												
Headcount poverty rate		0.021										
		(0.03)										
Near poverty rate				0.393								
				(0.28)								
Extreme poverty rate						0.154***						
						(0.03)						
Poverty gap								0.171***				
								(0.02)				
USAID poverty gap										0.073		
										(0.06)		
Consumption mean												1.374***
												(0.53)
Other model parameters												
Log of sample size of base survey		-0.613***		-0.979		0.271		-0.003		-0.140		-0.422
		(0.24)		(0.82)		(0.42)		(0.51)		(0.65)		(0.45)
Interval length between base & target surveys		-0.897***		-0.154		-1.001***		-0.741**		-0.102		-0.323
		(0.30)		(0.39)		(0.23)		(0.31)		(0.27)		(0.47)
Normal linear regression model		-0.093		-0.193		0.335		0.416*		0.314		-1.582***
		(0.25)		(0.41)		(0.28)		(0.25)		(0.19)		(0.61)
Number of rounds used		0.391**		0.020		0.153		0.014		-0.202		0.237
		(0.17)		(0.13)		(0.14)		(0.05)		(0.13)		(0.28)
R squared		0.632		4.207		-9.593		-9.511		-11.049***		-15.228
		(5.40)		(8.54)		(6.41)		(6.07)		(4.13)		(10.52)
Country FE												
Tanzania	0.022		-0.151*		1.769***		1.668***		2.028***		16.806***	
	(0.10)		(0.08)		(0.13)		(0.10)		(0.13)		(1.17)	
Malawi	-0.650***		0.973***		0.722***		0.706***		0.277***		16.984***	
	(0.05)		(0.09)		(0.05)		(0.04)		(0.02)		(1.19)	
Nigeria	-0.415***		1.463***		1.022***		0.845***		1.723***		17.394***	
	(0.08)		(0.11)		(0.05)		(0.04)		(0.16)		(1.19)	
cut1	0.375	-6.531***	0.197	-3.680	1.502**	-1.284	1.633	-2.676	0.564	-3.748	18.478***	1.020
	(0.95)	(2.32)	(0.33)	(6.63)	(0.71)	(2.80)	(1.00)	(2.82)	(0.42)	(3.13)	(1.17)	(2.29)
cut2	1.628*	-5.007**	0.924***	-2.827	2.512***	-0.047	2.746***	-1.405	1.323***	-3.005	19.389***	1.978
	(0.90)	(2.20)	(0.27)	(6.74)	(0.83)	(3.03)	(0.98)	(3.11)	(0.33)	(2.93)	(1.18)	(2.30)
Log likelihood	-309.21	-267.97	-298.62	-271.39	-260.02	-223.32	-262.39	-238.22	-283.31	-285.77	-170.82	-164.38
Pseudo R2	0.03	0.15	0.09	0.15	0.08	0.14	0.07	0.12	0.10	0.07	0.15	0.17
N	314	314	314	314	314	314	314	314	314	314	314	314

Table A.11. Comparison of utility expenditure variables in the BIHSs 2010/11 to 2018/19

Item name	2010/11 BIHS	2015 BIHS	2018/19 BIHS
Module P1 (monthly recall)			
1. Firewood	+	+	+
2. Cow dung/cakes/bhushi/wood-powder	+	+	+
3. Jute stick	+	+	+
4. Kerosene	+	+	+
5. Agriculture by-products used for fuel: paddy, hag, pressed sugarcane, and dried com plants	+	+	+
6. Gas (natural, bio-gas) or liquified petroleum gas (LPG)	+	+	+
7. Electricity	+	+	
8. Pit coal, char coal, wood coal	+	+	+
9. Other fuels and light (e.g., matches and candles)	+	+	+
10. Electricity (national grid)			+
11. Electricity (generator)		+	+
12. Electricity (solar)		+	+

Notes: The inconsistent variables are marked in red. The sign "+" indicates the variable is included in the specific survey round. For item 8, very few households (8 households in 2015 BIHS and 1 household in 2018/19 BIHS) reported use of "Pit coal, char coal, wood coal".

Figure A.1. Number of Models with Predicted Estimates of (Log) of Consumption That Are Statistically Insignificantly Different from the True Estimates

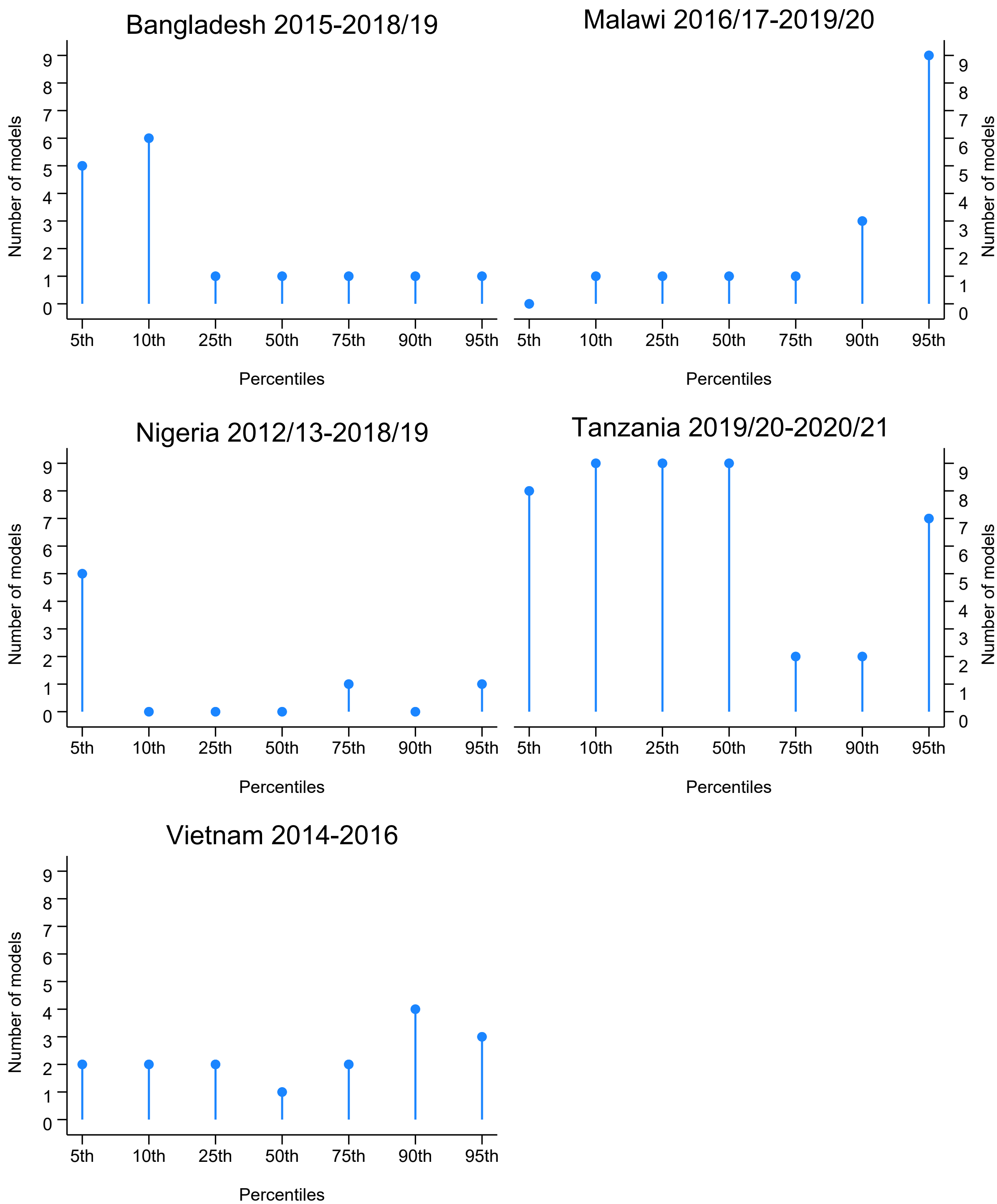


Figure A.2. Predicted consumption distribution, Model 9

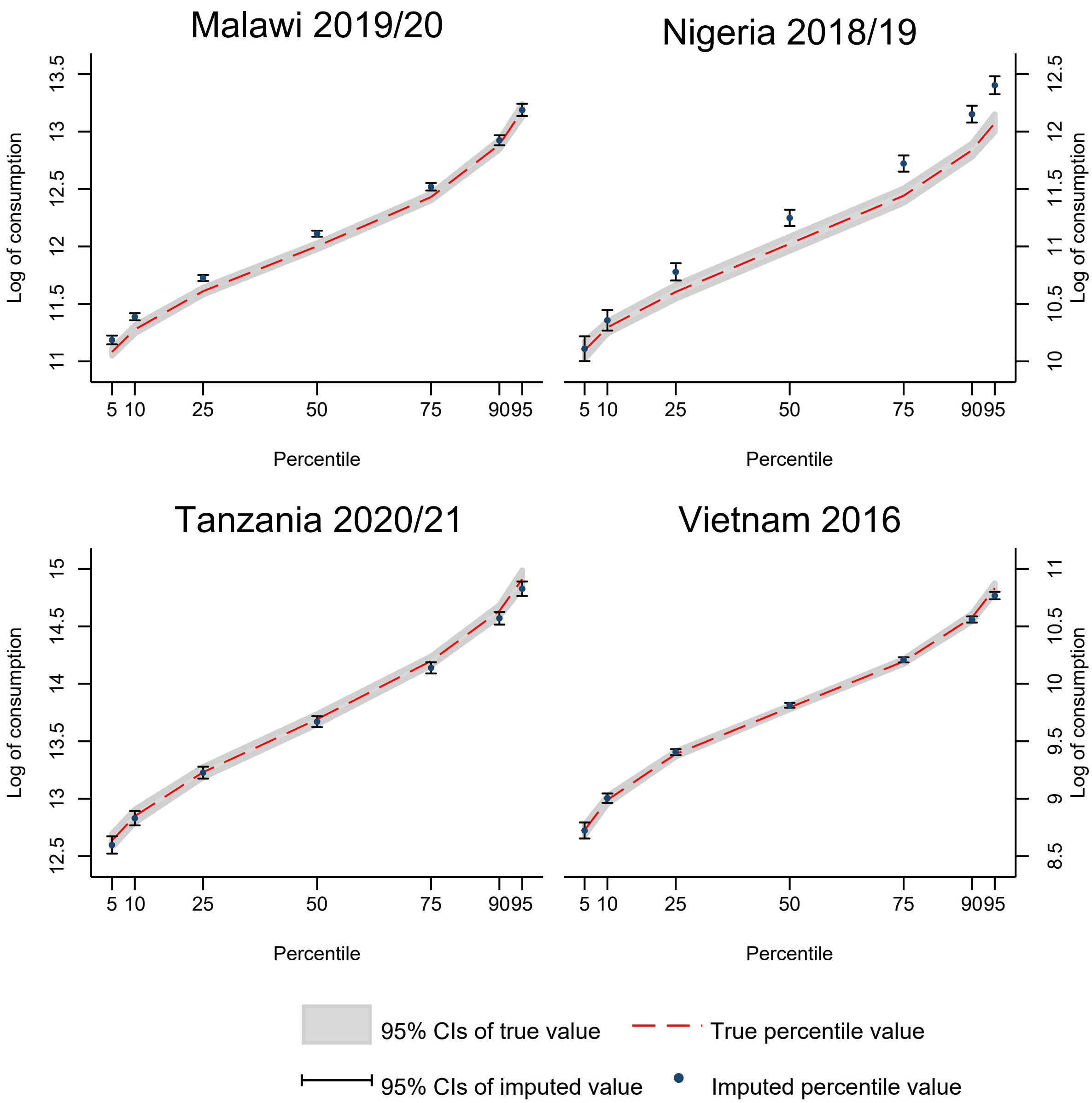
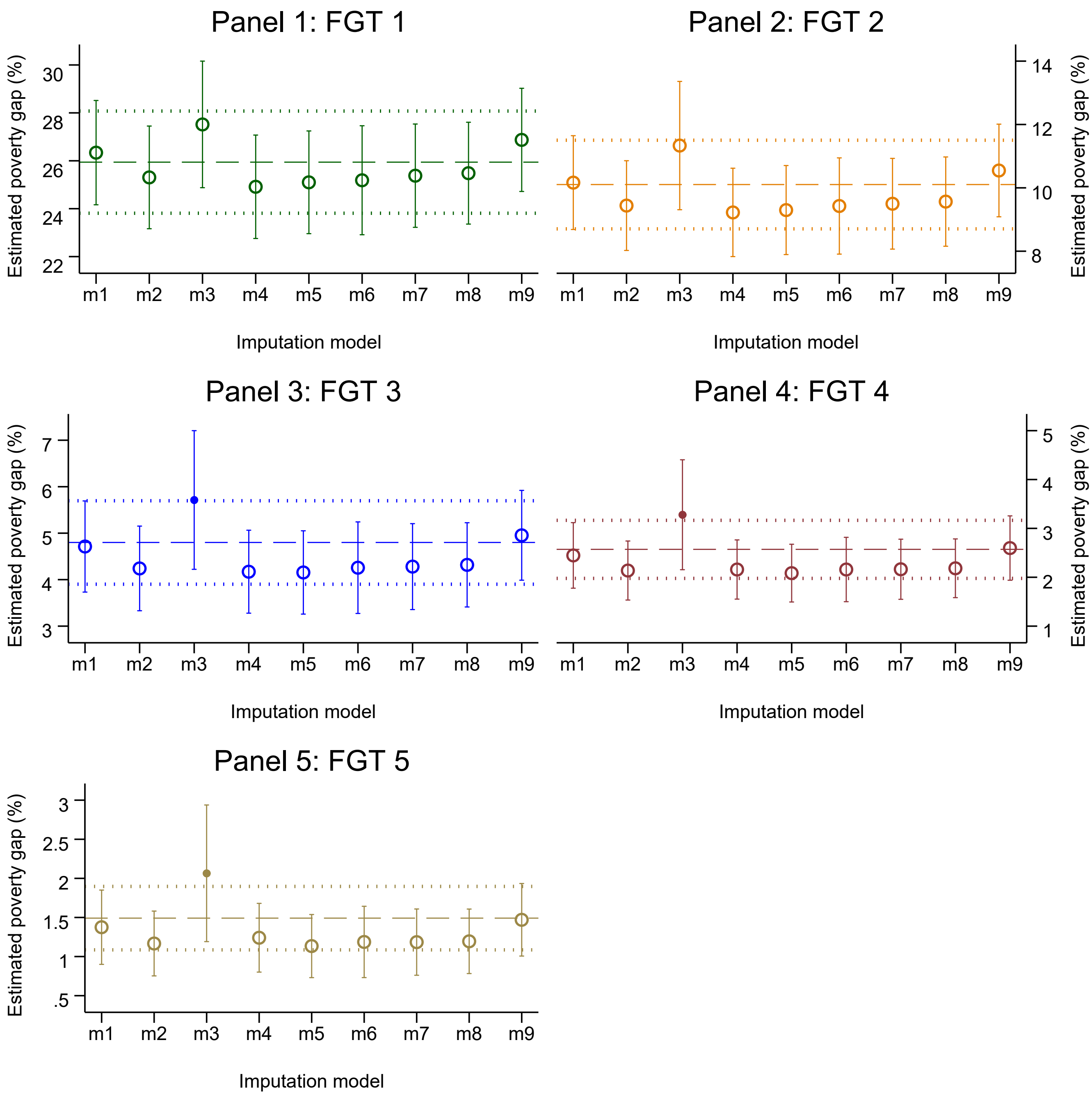
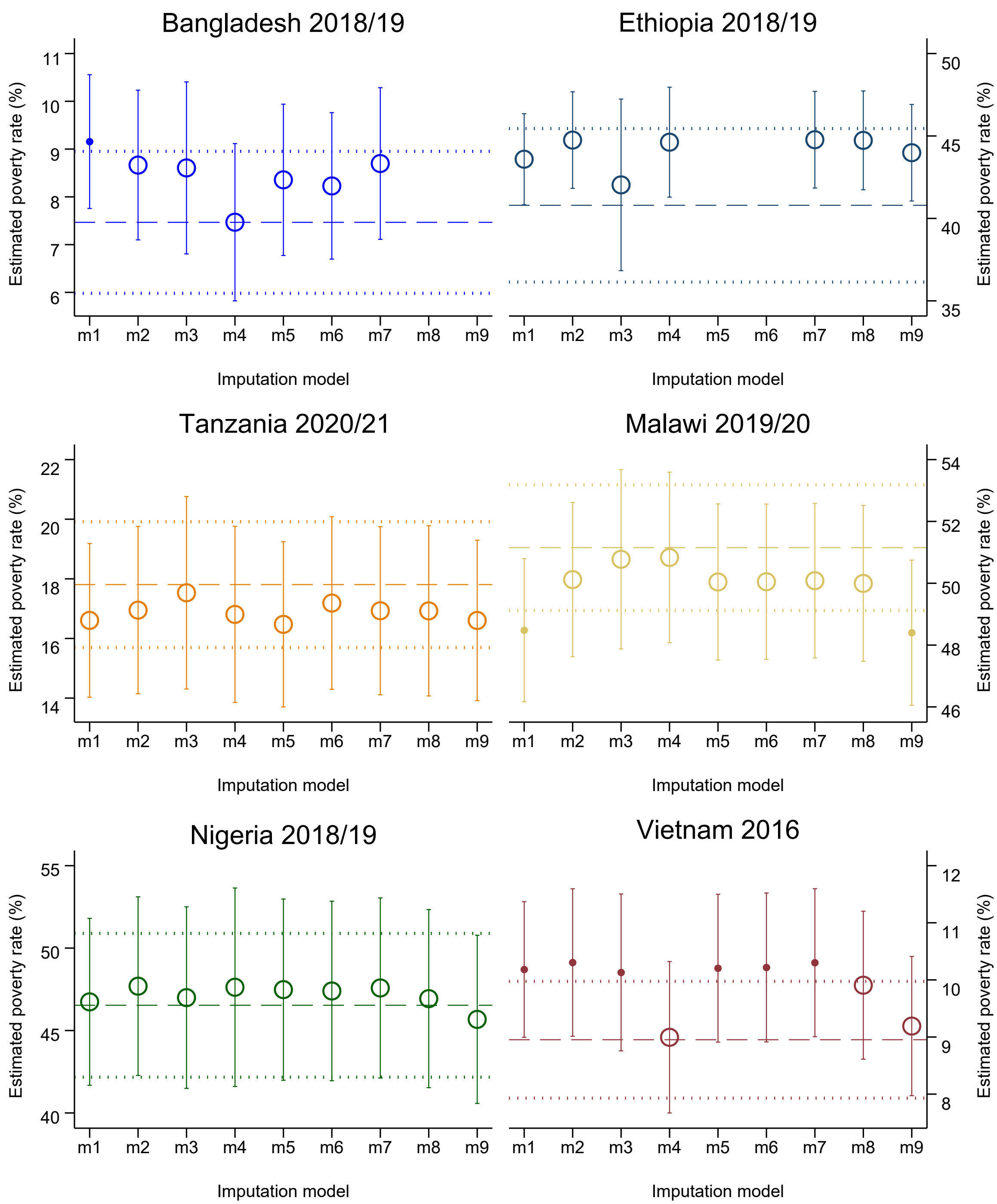


Figure A.3. Predicted USAID FGT indexes, Tanzania 2019/20- 2020/21



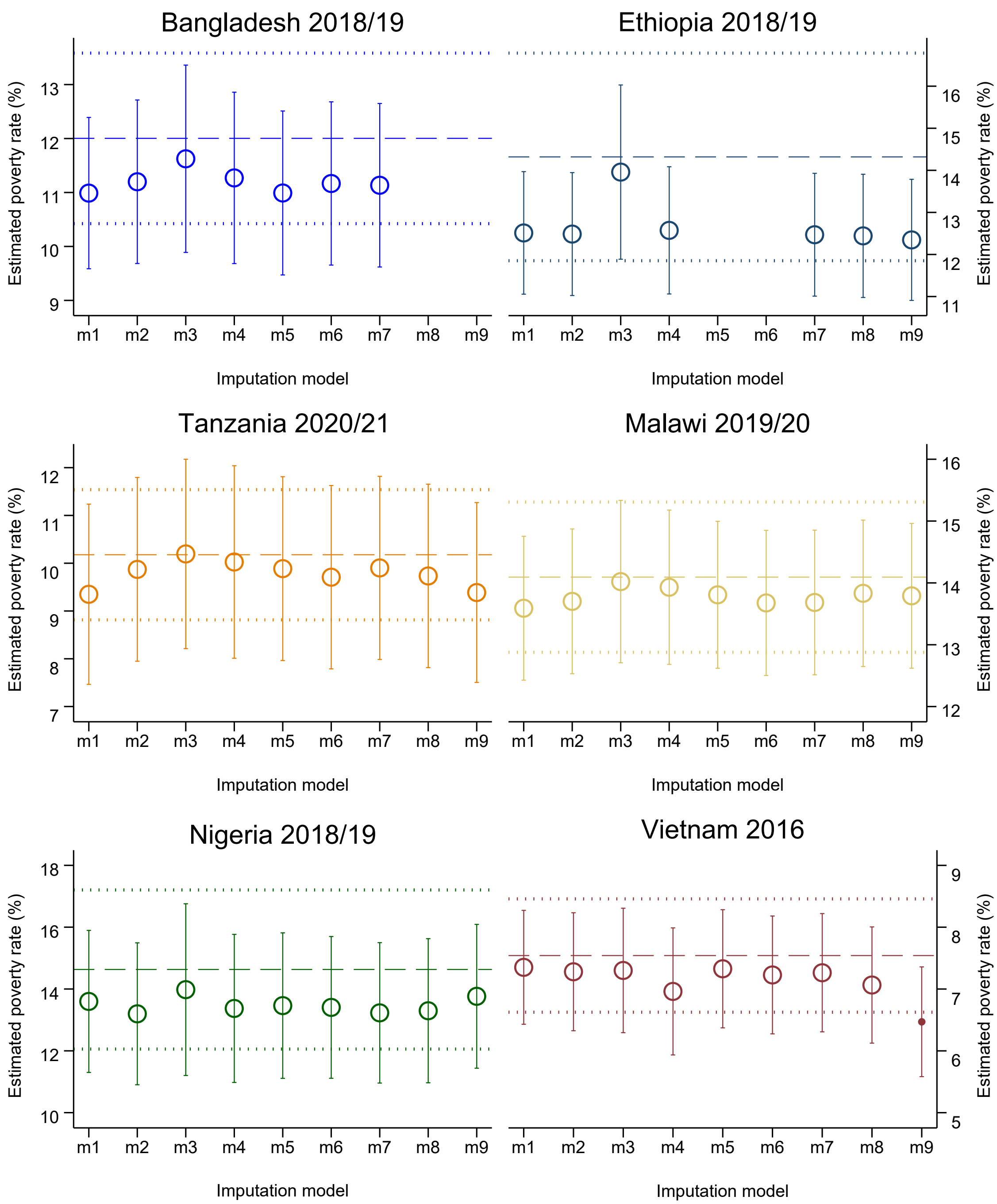
Note: Estimates are obtained by imputing from sample 1 into sample 2. 100 simulations are implemented. The standard errors are calculated using 100 bootstrap replications and are adjusted for complex survey design. Larger hollow symbols indicates that the estimates are statistically insignificantly different from the true poverty gap. Dashed lines represent the true poverty gap. Dotted lines represent confidence intervals of the true poverty gap. Estimates are obtained using the normal linear regression models.

Figure A.4. Predicted Headcount Poverty Based on Within-Year Imputation



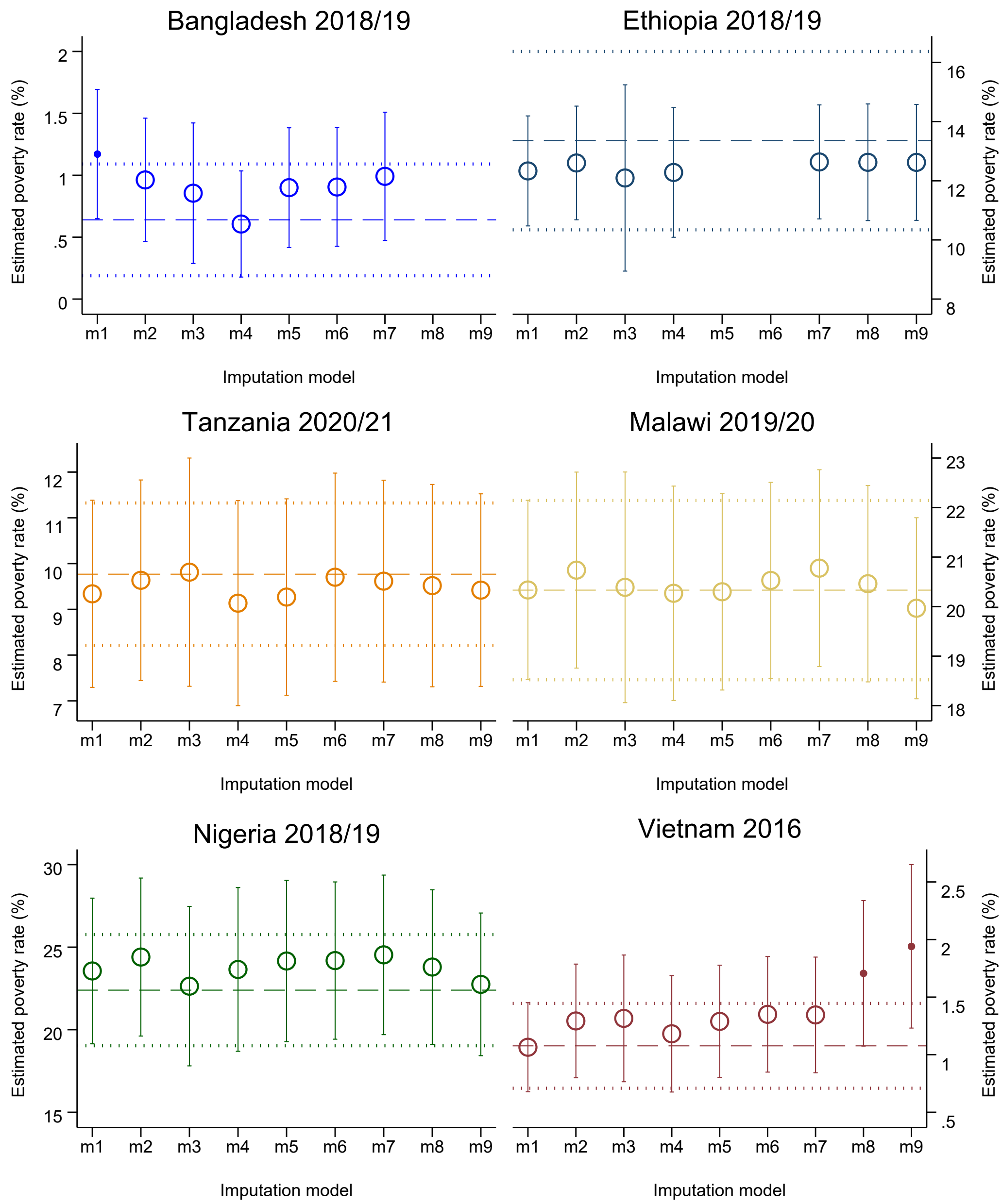
Note: Estimates are obtained by imputing from sample 1 into sample 2. 100 simulations are implemented. The standard errors are calculated using 100 bootstrap replications and are adjusted for complex survey design. Larger hollow symbols indicates that the estimates are statistically insignificantly different from the true poverty rates. Dashed lines represent the true poverty rates. Dotted lines represent confidence intervals of the true poverty rates. Estimates are obtained using the normal linear regression models.

Figure A.5. Predicted Near - Poverty Rates Based on Within-Year Imputation



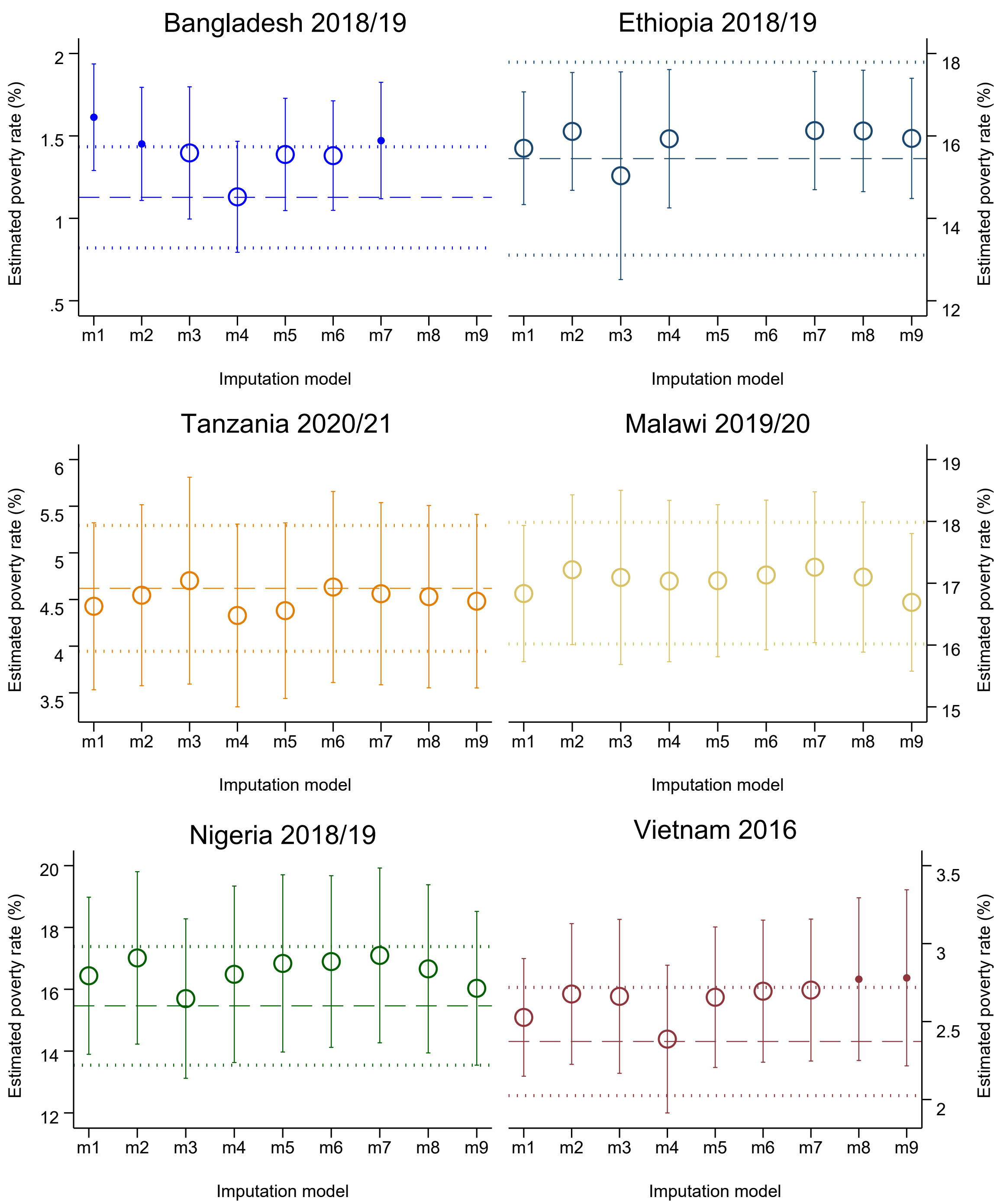
Note: Estimates are obtained by imputing from sample 1 into sample 2. 100 simulations are implemented. The standard errors are calculated using 100 bootstrap replications and are adjusted for complex survey design. Larger hollow symbols indicates that the estimates are statistically insignificantly different from the true near- poverty rates. Dashed lines represent the true near-poverty rates defined as living on an income between 100 and 125% of the poverty line. Dotted lines represent confidence intervals of the true near-poverty rates. Estimates are obtained using the normal linear regression models.

Figure A.6. Predicted Extreme Poverty Rates Based on Within-Year Imputation



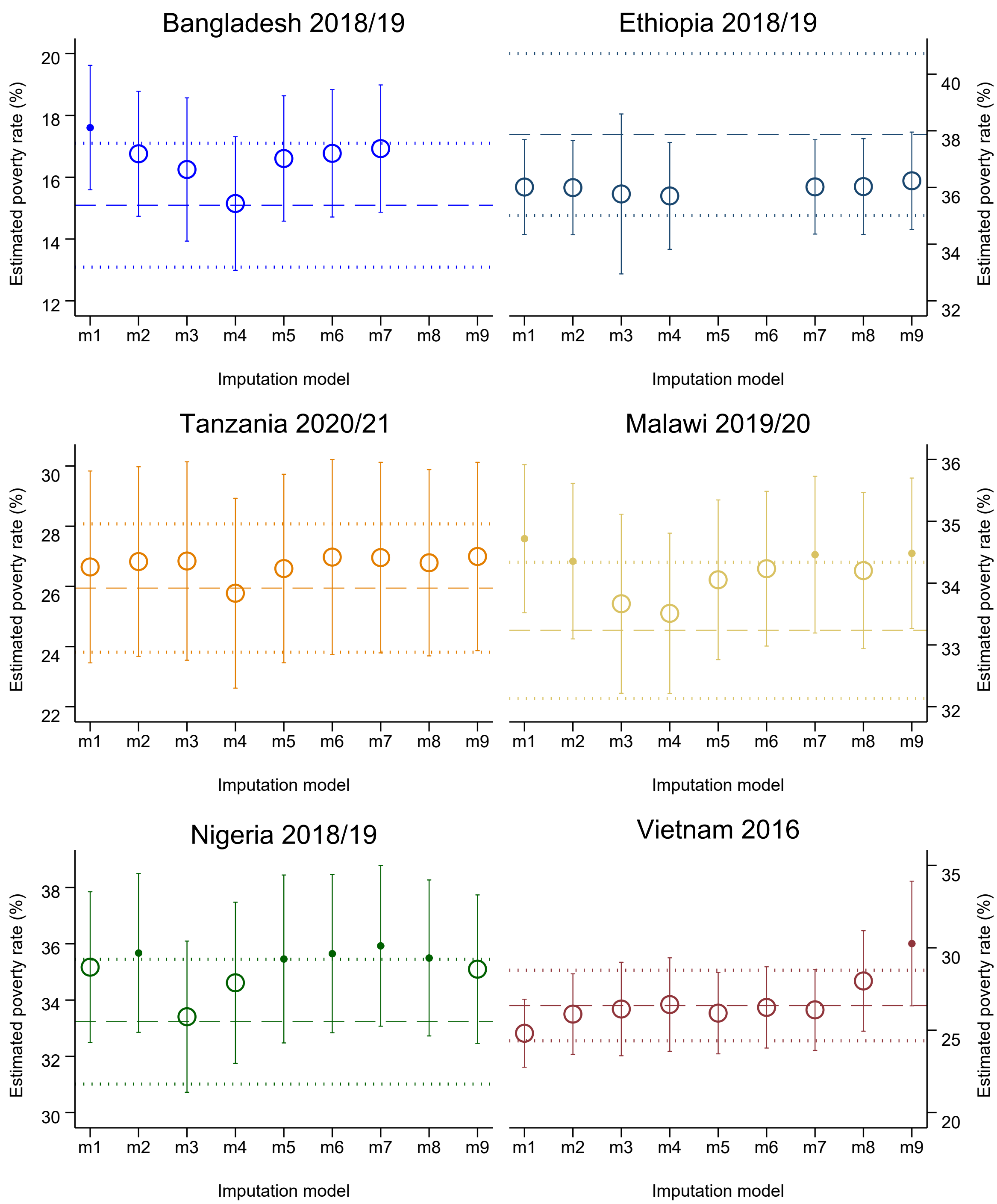
Note: Estimates are obtained by imputing from sample 1 into sample 2. 100 simulations are implemented. The standard errors are calculated using 100 bootstrap replications and are adjusted for complex survey design. Larger hollow symbols indicates that the estimates are statistically insignificantly different from the true extreme poverty rates. Dashed lines represent the true extreme poverty rates. Extreme poverty line is defined as US$1.25 (2011 PPP) per day per capita in Bangladesh and Nigeria and as half of the national poverty line in Vietnam and Ethiopia. National food poverty lines are used as extreme poverty line in Malawi and Tanzania. Dotted lines represent confidence intervals of the true extreme poverty rates. Estimates are obtained using the normal linear regression models.

Figure A.7. Predicted Poverty Gap Based on Within-Year Imputation



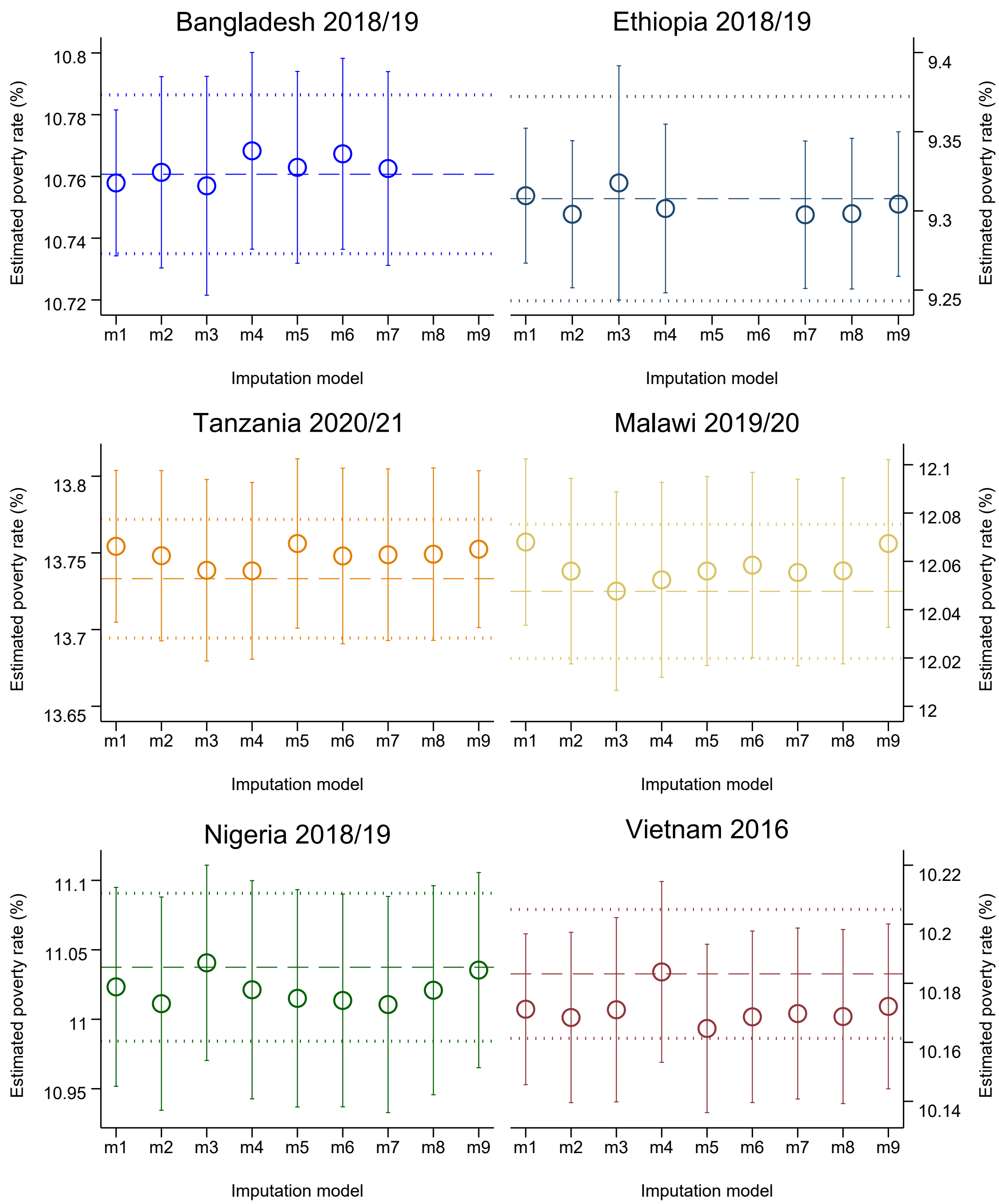
Note: Estimates are obtained by imputing from sample 1 into sample 2. 100 simulations are implemented. The standard errors are calculated using 100 bootstrap replications and are adjusted for complex survey design. Larger hollow symbols indicate that the estimates are statistically insignificantly different from the true poverty gap. Dashed lines represent the true poverty gap. Dotted lines represent confidence intervals of the true poverty gap. Estimates are obtained using the normal linear regression models.

Figure A.8. Predicted USAID Poverty Gap Based on Within-Year Imputation



Note: Estimates are obtained by imputing from sample 1 into sample 2. 100 simulations are implemented. The standard errors are calculated using 100 bootstrap replications and are adjusted for complex survey design. Larger hollow symbols indicate that the estimates are statistically insignificantly different from the true poverty gap. Dashed lines represent the true poverty gap. Dotted lines represent confidence intervals of the true poverty gap. Estimates are obtained using the normal linear regression models.

Figure A.9. Predicted (Log) Consumption Based on Within-Year Imputation



Note: Estimates are obtained by imputing from sample 1 into sample 2. 100 simulations are implemented. The standard errors are calculated using 100 bootstrap replications and are adjusted for complex survey design. Larger hollow symbols indicate that the estimates are statistically insignificantly different from the true consumption mean. Dashed lines represent the true consumption mean. Dotted lines represent confidence intervals of the true consumption mean. Estimates are obtained using the normal linear regression models.

Appendix B: Machine Learning Results

Table B.1. ML Estimates, Tanzania 2020/21

Indicators	LASSO	Elastic Net	Random Forest	True rates
Headcount poverty rate	9.8	9.7	5.8	17.8
	(0.9)	(0.9)	(0.7)	(1.1)
Near-poverty rate	12.0	12.0	16.6	10.2
	(0.8)	(0.8)	(1.0)	(0.7)
Extreme poverty rate	3.8	3.8	0.1	9.8
	(0.6)	(0.6)	(0.0)	(0.8)
Poverty gap	1.9	1.8	0.4	4.6
	(0.3)	(0.3)	(0.4)	(0.3)
USAID Poverty gap	18.9	19.0	6.6	25.9
	(1.3)	(1.3)	(0.5)	(1.1)
Log of consumption mean	**13.7***	**13.7***	**13.7***	13.7
	(0.0)	(0.0)	(0.0)	(0.0)
N	4,644			

Table B.2. ML Estimates, Vitenam 2016

Indicators	LASSO	Elastic Net	Random Forest	True rates
Headcount poverty rate	25.2	25.2	21.5	9.5
	(0.6)	(0.6)	(0.6)	(0.4)
Near-poverty rate	14.8	14.8	16.3	6.9
	(0.4)	(0.4)	(0.5)	(0.3)
Extreme poverty rate	4.7	4.7	3.5	1.2
	(0.4)	(0.4)	(0.3)	(0.2)
Poverty gap	7.3	7.3	5.9	2.5
	(0.3)	(0.3)	(0.2)	(0.2)
USAID Poverty gap	29.1	29.1	27.3	26.1
	(0.7)	(0.7)	(0.7)	(0.9)
Log of consumption mean	9.7	9.7	9.7	9.8
	(0.0)	(0.0)	(0.0)	(0.0)
N	4,644			

Table B.3. The list of selected variables in Lasso and Elastic Net models, Tanzania

	LASSO		Elastic Net	
	Non-standardized	Standardized	Non-standardized	Standardized
Head has primary education	-0.050	-0.025	-0.049	-0.024
Head has secondary ordinary education	0.024	0.010	0.025	0.010
Head has secondary advanced education and higher	0.003	0.001	0.004	0.001
Household size	-0.051	-0.167	-0.050	-0.164
Dependency Ratio	-0.023	-0.021	-0.023	-0.022
Gender Ratio	-0.046	-0.051	-0.046	-0.050
Household Head worked in unpaid apprentice in the last 12 months	0.048	0.003	0.045	0.003
Household Head worked in farm in the last 12 months	-0.008	-0.004	-0.008	-0.004
Household Head employed in mining, manufacturing, construction	-0.014	-0.003	-0.012	-0.003
Household Head employed in retail, transportation	0.105	0.028	0.103	0.027
Household Head employed in information and communication	0.209	0.020	0.208	0.020
Household Head employed in technical, administrative, education	0.084	0.019	0.083	0.019
Proportion of adult males that worked for a wage, salary, or commission	-0.053	-0.025	-0.051	-0.024
Proportion of adult males that engaged in casual/ganyu labor in the last 12 months	-0.003	-0.001	-0.000	0.000
Proportion of adult males that worked in unpaid apprentice in the last 12 months	0.091	0.005	0.083	0.005
Proportion of adult males that worked in farm in the last 12 months	-0.004	-0.002	-0.005	-0.002
Proportion of adult males that worked in electricity or water supply	0.000	0.000		
Proportion of adult males that worked in information and communication	0.075	0.006	0.072	0.006
Overcrowded	-0.032	-0.043	-0.033	-0.045
Household dwelling roof materials	-0.092	-0.030	-0.091	-0.030
Household dwelling floor materials	-0.042	-0.021	-0.043	-0.021
Burnt bricks/Concrete walls	0.007	0.003	0.007	0.003
Piped water/Truck water	0.075	0.038	0.075	0.038
Flush/VIP toilet	0.061	0.031	0.062	0.031
Electricity for lighting	0.072	0.035	0.072	0.035
Household owns a chair or sofa	-0.037	-0.018	-0.037	-0.018
Household owns a sewing machine	-0.072	-0.019	-0.070	-0.019
Household owns an electric/gas stove	0.174	0.068	0.173	0.068
Household owns a refrigerator/freezer	0.051	0.018	0.050	0.018
Household owns a bicycle	-0.014	-0.006	-0.013	-0.006
Household owns a motor vehicle	0.030	0.010	0.031	0.010
Household owns a computer	0.166	0.032	0.164	0.032
Household owns a mobile phone	0.042	0.014	0.040	0.013
Household owns an iron	0.010	0.005	0.010	0.005
Household owns an air c/fans	0.120	0.041	0.119	0.041
Anyone in the household owns livestock	-0.039	-0.019	-0.038	-0.019
Household owns cows	0.033	0.010	0.030	0.009
Household consumed spaghetti, macaroni	0.042	0.012	0.042	0.012
Household consumed onions, tomatoes, carrots	-0.138	-0.044	-0.135	-0.043
Household consumed sweets	-0.035	-0.008	-0.031	-0.007

Household consumed biscuits, buns, scones	0.047	0.023	0.047	0.023
Household consumed potato	0.080	0.040	0.079	0.040
Household consumed beef	0.130	0.062	0.129	0.062
Household consumed eggs	0.105	0.040	0.104	0.040
Household purchased cigarettes or other tobacco	0.005	0.001	0.001	0.000
Household purchased matches	-0.024	-0.010	-0.023	-0.010
Household purchased toothpaste, toothbrush	0.131	0.063	0.130	0.063
Household purchased personal products	0.058	0.029	0.058	0.029
Household purchased petrol or diesel	0.167	0.047	0.164	0.046
Household purchased cleaning products	0.162	0.042	0.161	0.042
Household spent on taxes	0.032	0.008	0.032	0.008
Household spent on wedding	0.076	0.035	0.075	0.035
Household purchased education	0.078	0.012	0.074	0.011
Household purchased schoolbooks	-0.044	-0.017	-0.043	-0.017
Cereals, Grains, and Cereal Products	-0.089	-0.022	-0.088	-0.022
Fruits	0.077	0.038	0.077	0.038
Meat, Fish and Animal Products	0.041	0.014	0.039	0.013
Milk/Milk Products	0.124	0.056	0.123	0.056
Nuts and Pulses	0.060	0.024	0.059	0.023
Root, Tubers, and Plantains	0.045	0.020	0.044	0.019
Spices/Condiments	-0.290	-0.066	-0.282	-0.064
Sugar/Sugar Products/Honey	0.080	0.032	0.080	0.032
Vegetables	-0.000	-0.000	-0.005	-0.001
_cons	14.281	0.000	14.277	0.000
MSE	0.18		0.18	
R squared	0.63		0.63	
N	1182		1,182	

Table B.4. The list of selected variables in Lasso and Elastic Net models, Vietnam

	LASSO		Elastic Net	
	Non-standardized	Standardized	Non-standardized	Standardized
Head`s age	0.003	0.039	0.003	0.039
Head`s ethnicity	-0.125	-0.046	-0.125	-0.046
Primary education	0.037	0.016	0.036	0.016
Lower secondary education	0.051	0.023	0.050	0.023
Upper secondary education	0.121	0.043	0.120	0.043
College	0.220	0.055	0.219	0.055
Household size	-0.151	-0.236	-0.150	-0.236
Dependency Ratio	-0.087	-0.060	-0.087	-0.060
Gender Ratio	-0.015	-0.013	-0.015	-0.013
Household Head worked for a wage, salary, or commission in the last 12 months	-0.029	-0.010	-0.029	-0.010
Household Head engaged in casual/ganyu labor in the last 12 months	-0.001	0.000	-0.001	-0.000
Household Head employed in industry 1	-0.018	-0.007	-0.018	-0.007
Household Head employed in industry 2	0.036	0.012	0.035	0.012
Household Head employed in industry 3	0.028	0.004	0.028	0.004
Household Head employed in industry 5	0.033	0.003	0.032	0.003
Proportion of adult males that engaged in casual/ganyu labor in the last 12 months	-0.007	-0.004	-0.007	-0.004
Proportion of adult males that worked in farm in the last 12 months	-0.019	-0.009	-0.019	-0.009
Proportion of adult males that worked in industry 1	-0.008	-0.003	-0.008	-0.003
Proportion of adult males that worked in industry 4	0.012	0.003	0.012	0.003
Proportion of adult males that worked in industry 5	0.019	0.002	0.019	0.001
log of residential area	0.196	0.113	0.196	0.112
Roof: cement	0.017	0.007	0.017	0.007
Wall: bricks	0.029	0.012	0.029	0.012
Wall:cement	0.051	0.007	0.050	0.007
Improved water source	0.006	0.017	0.006	0.017
Improved toilet source	0.029	0.049	0.029	0.049
Lighting source - Electricity	-0.108	-0.015	-0.107	-0.015
Dwelling_cookfuel	0.016	0.006	0.016	0.006
Household owns a car	0.625	0.080	0.624	0.080
Household owns a motorbike	0.094	0.035	0.094	0.035
Household owns a bicycle	-0.048	-0.024	-0.048	-0.024
Household owns a DVD	0.060	0.030	0.060	0.030
Household owns a TV	0.054	0.014	0.054	0.014
Household owns a computer	0.182	0.072	0.182	0.072
Household owns a refrigerator	0.124	0.061	0.125	0.061
Household owns a sewing machine	0.063	0.013	0.063	0.013
Household owns an electric/gas cooker	0.122	0.042	0.122	0.042
Anyone in the household cultivate any plot	-0.072	-0.036	-0.072	-0.036
Anyone in the household earn revenues from husbandry, hunting, trapping and dome	-0.023	-0.012	-0.023	-0.012
Household obtains goat/sheep	0.027	0.002	0.026	0.002

Household obtains chickens	-0.023	-0.011	-0.022	-0.011
Household consumed Noodle last 30 days	0.026	0.009	0.026	0.009
Household consumed Peas, beans last 30 days	0.047	0.024	0.047	0.024
Household consumed tomatoes last 30 days	0.064	0.028	0.064	0.028
Household consumed tea, coffee last 30 days	0.022	0.010	0.064	0.024
Household consumed potatoes last 30 days	0.033	0.015	0.022	0.010
Household consumed beef last 30 days	0.117	0.056	0.033	0.015
Household consumed ice cream & yogurt last 30 days	0.045	0.019	0.117	0.056
Household consumed chicken last 30 days	0.078	0.038	0.045	0.019
Household purchased matches	-0.040	-0.014	0.078	0.038
Household purchased petrol or diesel	0.068	0.026	-0.040	-0.014
Household purchased cleaning products	0.059	0.014	0.068	0.026
Household purchased soap	0.055	0.023	0.059	0.014
Household spent on wedding last 12 months	0.252	0.040	0.055	0.023
Household spent on building housing accommodation	-0.039	-0.015	0.251	0.039
Household spent on house repair and maintenance over the past 12 months	0.038	0.011	-0.039	-0.014
Household purchased tuition fee	0.080	0.039	0.038	0.011
Household purchased school uniform	-0.011	-0.006	0.080	0.039
Consumption category last 30 days: Fruits	0.064	0.024	-0.011	-0.006
Consumption category last 30 days: Milk/Milk Products	0.042	0.021	0.042	0.021
Consumption category last 30 days: Peanuts & sesame	0.014	0.006	0.014	0.006
Consumption category last 30 days: Sugar/confectionery/ molasses	0.053	0.019	0.052	0.019
Consumption category last 30 days: Vegetables	-0.052	-0.006	-0.052	-0.006
Household has living conditions improved in 5 years	0.002	0.001	0.002	0.001
_cons	8.494	0.000	8.494	0.000
MSE	0.12		0.12	
R squared	0.71		0.71	
N	9,296		9,296	

Table B.5. Variable importance scores in Random Forest, Tanzania

Variable Importance	
Head`s age	0.0546
Head is literate	0.0409
Head has primary education	0.0462
Head has secondary ordinary education	0.0450
Head has secondary advanced education and higher	0.0640
Household size	0.1974
Dependency Ratio	0.1003
Gender Ratio	0.0620
Household Head worked as an employee in the last 12 months	0.0396
Household Head worked as self-employed in the last 12 months	0.0388
Household Head worked in unpaid apprentice in the last 12 months	0.0318
Household Head worked in farm in the last 12 months	0.1056
Household Head employed in mining, manufacturing, construction	0.0492
Household Head employed in retail, transportation	0.0452
Household Head employed in electricity or water supply	0.0371
Household Head employed in information and communication	0.0474
Household Head employed in technical, administrative, education	0.0768
Proportion of adult males that worked for a wage, salary, or commission	0.0412
Proportion of adult males that self-employed	0.0397
Proportion of adult males that worked in unpaid apprentice	0.0524
Proportion of adult males that worked in farming	0.0443
Proportion of adult males that worked in mining, manufacturing, construction	0.0516
Proportion of adult males that worked in retail, transportation	0.0445
Proportion of adult males that worked in electricity or water supply	0.0348
Proportion of adult males that worked in information and communication	0.0495
Proportion of adult males that worked in technical, administrative, education	0.0708
Number of rooms	0.0549
Overcrowded	0.1625
Household dwelling roof materials	0.0631
Household dwelling floor materials	0.3701
Burnt bricks/Concrete walls	0.0521
Piped water/Truck water	0.2013
Flush/VIP toilet	0.6522
Charcoal for cooking	0.1125
Electricity for lighting	1.0000
Household owns a chair or sofa	0.0440
Household owns a radio	0.0433

Household owns a tv	0.4729
Household owns a DVD	0.1763
Household owns a sewing machine	0.0475
Household owns an electric/gas stove	0.9191
Household owns a refrigerator/freezer	0.1640
Household owns a bicycle	0.0443
Household owns a motor vehicle	0.0803
Household owns a computer	0.1311
Household owns a mobile phone	0.0618
Household owns an iron	0.1057
Household owns an air c/fan	0.1585
Household owns decoder	0.0770
Anyone in the household owns livestock	0.0521
Household owns goat	0.0542
Household owns chicken	0.0455
Household owns cows	0.0532
Household consumed spaghetti, macaroni	0.0774
Household consumed beans	0.0696
Household consumed onions, tomatoes, carrots	0.0782
Household consumed fruits	0.0938
Household consumed sugar	0.1569
Household consumed sweets	0.0553
Household consumed tea	0.0209
Household consumed biscuits, buns, scones	0.1033
Household consumed potato	0.0627
Household consumed beef	0.1249
Household consumed yogurt	0.0600
Household consumed chicken	0.0757
Household consumed eggs	0.1615
Household purchased cigarettes or other tobacco	0.0459
Household purchased matches	0.0514
Household purchased toothpaste, toothbrush	0.1203
Household purchased personal products	0.0659
Household purchased petrol or diesel	0.0998
Household purchased cleaning products	0.2526
Household purchased soap	0.0623
Household spent on taxes	0.0842
Household spent on construction	0.0480
Household spent on wedding	0.0548
Household spent on repair	0.0397

Household purchased education	0.0623
Household purchased schoolbooks	0.0541
Household purchased uniform	0.0545
Cereals, Grains, and Cereal Products	0.0920
Oil/fats	0.0643
Fruits	0.1559
Meat, Fish and Animal Products	0.0886
Milk/Milk Products	0.0677
Nuts and Pulses	0.0906
Root, Tubers, and Plantains	0.0705
Spices/Condiments	0.1161
Sugar/Sugar Products/Honey	0.1868
Vegetables	0.1040

Note: The values are scaled proportional to the largest value in the set.

Table B.6. Variable importance scores in Random Forest, Vietnam

Variable	Importance
Head`s age	0.0086
Head`s ethnicity	0.2932
Primary education	0.0069
Lower secondary education	0.0072
Upper secondary education	0.0102
College	0.0309
Household size	0.0493
Dependency Ratio	0.0218
Gender Ratio	0.0099
Household Head worked for a wage, salary, or commission in the last 12 months	0.0094
Household Head worked as self-employed in the last 12 months	0.0083
Household Head engaged in casual/ganyu labor in the last 12 months	0.0086
Household Head employed in industry 1	0.0080
Household Head employed in industry 2	0.0110
Household Head employed in industry 4	0.0119
Household Head employed in industry 3	0.0115
Household Head employed in industry 5	0.0137
Proportion of adult males worked for a wage, salary, or commission	0.0136
Proportion of adult males that engaged in casual/ganyu labor in the last 12 months	0.0103
Proportion of adult males that worked in farm in the last 12 months	0.0121
Proportion of adult males that worked in industry 1	0.0094
Proportion of adult males that worked in industry 2	0.0107
Proportion of adult males that worked in industry 3	0.0120
Proportion of adult males that worked in industry 4	0.0126
Proportion of adult males that worked in industry 5	0.0138
log of residential area	0.0212
Roof: cement	0.0087
Roof: cement	0.0140
Wall: bricks	0.0144
Wall:cement	0.0161
Improved water source	0.0208
Improved toilet source	0.1869
Lighting source- Electricity	0.0115
Dwelling_cookfuel	0.0089
Household owns a car	0.1665
Household owns a motorbike	0.0253
Household owns a bicycle	0.0109

Household owns a DVD	0.0117
Household owns a TV	0.0183
Household owns a computer	0.4517
Household owns a refrigerator	1.0000
Household owns a sewing machine	0.0100
Household owns an electric/gas cooker	0.2177
Anyone in the household cultivate any plot	0.0560
Anyone in the household earn revenues from husbandry, hunting, trapping and dome	0.0169
Household obtains goat/sheep	0.0112
Household obtains chickens	0.0120
Household obtains pigs	0.0101
Household consumed Noodle last 30 days	0.0118
Household consumed Peas, beans last 30 days	0.0134
Household consumed tomatoes last 30 days	0.0166
Household consumed fruits last 30 days	0.0239
Household consumed sugar last 30 days	0.0125
Household consumed tea, coffee last 30 days	0.0164
Household consumed potatoes last 30 days	0.0114
Household consumed beef last 30 days	0.2596
Household consumed ice cream & yogurt last 30 days	0.0147
Household consumed eggs last 30 days	0.0177
Household consumed chicken last 30 days	0.0117
Household purchased matches	0.0111
Household purchased petrol or diesel	0.0291
Household purchased cleaning products	0.0190
Household purchased soap	0.0131
Household spent on wedding last 12 months	0.0220
Household spent on building housing accommodation	0.0125
Household spent on house repair and maintenance over the past 12 months	0.0104
Household purchased tuition fee	0.0143
Household purchased schoolbooks	0.0108
Household purchased school uniform	0.0109
Consumption category last 30 days: Fruits	0.0321
Consumption category last 30 days: Milk/Milk Products	0.0131
Consumption category last 30 days: Peanuts & sesame	0.0107
Consumption category last 30 days: Sugar/confectionery/ molasses	0.0138
Consumption category last 30 days: Vegetables	0.0169
Household has living conditions improved in 5 years	0.0117