ASSET-BASED MEASURES FOR MACHINE-LEARNING POVERTY MAPS

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Peizan Sheng

August 2022

ABSTRACT

This paper develops a machine learning approach to estimate internationally-and-intertemporally comparable, decomposable, structural asset poverty measures. These measures are founded in theory, link directly to official poverty lines, and are amenable to ML-based prediction using Earth Observation data. Using household survey data from Tanzania, Uganda, and Malawi, we model the relationship between household consumption expenditures and productive assets, directly linking flow-based poverty measures with asset-based structural poverty measures. The poverty measures we construct can serve as new, improved dependent variables for ML poverty prediction. We also assess whether our poverty estimates vary from readily available poverty estimates and whether this difference in poverty measures matters.

BIOGRAPHICAL SKETCH

Peizan Sheng was born in Jiangsu, China. He started his academic journey from Renmin University of China, where he obtained Bachelor of Science in Mathematics and received rigorous math training. To further his research in economics, Peizan entered the Master of Science program in Applied Economics and Management at Cornell University. After graduating from Cornell University, Peizan will join the Ph.D. program in Public Policy at University of Chicago in Fall 2022.

This paper is dedicated to my close family, my love, and my committee members for motivating me throughout my academic journey.

ACKNOWLEDGEMENTS

I want to express great gratitude to Professor Christopher B. Barrett for his guidance, support, and encouragement in my academic pursuit. The weekly meeting with him was the most exciting and fruitful hour of the week. I learned how to structure research projects logically, how to present results neatly, and how to apply mathematics to answer development economics questions. From mathematics to development economics, from theory to empirical research, I eventually found my interests. During this transition, I feel extremely fortunate to work with such an exceptional advisor, instructor, and human being in all aspects.

I also want to thank my chair, Professor Brian Dillon, for making a lot of time to instruct me and give advice on this project. The econometrics and development economics courses he instructed laid a solid foundation for my future economic research. I extremely appreciate his utmost support during my PhD application. Professor Dillon is also a great example of moral integrity and exceptional professionality. Having his advice through two years has been incredibly useful.

I am also indebted to Elizabeth Tennant and Yating Ru, both of whom have rich experience in regional science and machine learning. Their technical guidance allowed me to work faster and think better on this project. They are great persons to work with, always ready to help and sincerely preoccupied with the well-being of the ones that surround them.

Last, I would like to thank Liwen Gu and my parents for their unconditional love and support, and for always believing in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Accurate poverty measurement offers valuable information to track progress towards improving human wellbeing and to shape decisions about distributing scarce resources. Although the quantity and quality of economic data available in developing economies have improved in recent years, data on key measures of economic well-being remain limited for much of the developing world, especially in the poorest countries. Furthermore, poverty estimates available in low-and middle-income countries are spatially coarse, typically at no finer resolution than the first sub-national administrative unit level (i.e., province, region or state). The resulting data gap of reliable poverty estimates of finer spatial resolution hampers efforts to identify cohorts of greatest need and to geographically target interventions effectively.

Data gaps of poverty estimates are mainly attributed to the scarcity of nationally representative household surveys. Poverty estimates have historically relied on data from nationally-representative, large-scale household surveys. However, such surveys are expensive, time-consuming, and hard to obtain across broader geographies. Considering the monetary and time cost, nationwide surveys or censuses typically occur only every few years, at best, which does not satisfy practitioners' need for higher-frequency poverty estimates (Yeh et al., 2020). Furthermore, face-to-face interviews are sometimes impractical due to infectious disease outbreaks, civil conflict, or other reasons. Based on World Bank data, 39 of 59 African countries fielded fewer than two nationally representative consumption surveys between 2000 and 2010 (Jean et al., 2016). Closing the data gap through more frequent nationally-representative household surveys is unlikely due to cost, risks, and institutional capacity constraints.

Fortunately, recent advances in machine learning (ML)-based methods and remote sensing (RS) data have enabled researchers to generate poverty estimates at higher spatial and temporal resolutions (Jean et al., 2016; Pokhriyal and Jacques, 2017; Yeh et al., 2020; Chi et al., 2022; Browne et al., 2021; Tang et al., 2022; Ayush et al., 2020). These studies train ML models on widely available, high spatial resolution Earth Observation (EO) data such as Google maps, satellite imagery, mobile phone networks, Normalized Difference Vegetation Index (NDVI) to predict survey-based measures of people's wellbeing in places not covered by surveys, thereby generating poverty maps without requiring regular and expensive census data. In these studies, a variety of ML methods — including convolutional neural networks (CNN), transfer learning, deep learning, and random forests — have been used to improve model prediction performance. The lower cost, wider and more frequent availability of EO data sources, and easier extension to include places not recently sampled by household surveys have sparked great interest among development economists and policymakers in ML-based poverty mapping methods.

A closer look at the ML-based poverty prediction literature, however, reveals a number of gaps and shortcomings. First, the current ML-based poverty prediction literature has focused mainly on improving the out-of-sample accuracy of ML models through advances in ML algorithms or in the feature sets (i.e., explanatory variables) used. But the usefulness of such methods depends fundamentally on the quality of the left-hand side variable, i.e., on the poverty measures predicted by those algorithms and feature sets. Consider the basic poverty prediction model as $y = f(x)$, where y is the poverty estimate, f's are machine learning algorithms, and x's are input features. The ML poverty prediction literature has tried different kinds of algorithms f's (including CNN, deep learning, transfer learning, random forests) and a variety of feature sets x's (including Google maps, satellite imagery, mobile phone networks, NDVI, land cover and land use change) to fit models. The objective of these studies is to achieve better predictive accuracy, i.e., higher R-squared, higher accuracy rate, and lower RMSE. Few studies have focused on the quality and theoretical properties of the left-hand variable, i.e., on its correspondence with official poverty measures. Few use rank correlation coefficients as performance tests, although the ordering of observations is arguably the most policy-relevant diagnostic.

To date, the ML poverty prediction literature has tended to use readily available poverty proxies that have little grounding in theory nor correspondence with the national or global poverty lines used in policy-making. The most widely used poverty proxies are standardized asset wealth measures from the Demographic and Health Surveys (DHS) as used by, for example, Jean et al. (2016); Yeh et al. (2020); Chi et al. (2022); Lee and Braithwaite (2020); Browne et al. (2021). The DHS team collects detailed data on specific assets, and computes and releases a Relative Wealth Index (RWI) using a standardized methodology (Rutstein and Johnson, 2004). The ML poverty estimation literature traditionally uses the DHS RWI to proxy for poverty. In particular, households are assigned to "poorest", "poorer", "middle", "richer" and "richest" categories based on cut points established for the DHS RWI.

Unfortunately, this commonly used outcome measure, DHS RWI, suffers from key weaknesses. First, the cut points used by RWI have no direct correspondence to the official poverty lines declared and used by governments nor by multilateral development banks (e.g., the World Bank) and other international organizations. RWI does not map to expenditure- or income-based measures because DHS does not collect such data. However, the official poverty lines, including the $1.90/day per capita (2011 USD purchasing power parity (PPP)-adjusted) global poverty line which underpins Sustainable Development Goal 1 poverty commitments, are based on expenditure or income flow measures, not asset stock measures.

Second, the RWI is a survey-specific measure that only captures the distribution of asset wealth for a particular country at that moment in time. Therefore, it is not suitable for intertemporal or cross-country absolute wealth or poverty comparisons. For example, in an extremely poor country, a household identified as the highest wealth quantiles is not necessarily well-off in absolute terms compared to later years in that same country or to households in other, wealthier countries. As a result, it is somewhat problematic to apply DHS RWI as the left-hand side variable in poverty estimation studies, especially in multi-country or intertemporal poverty prediction.

In this paper, we develop internationally-and-intertemporally comparable, asset-based structural poverty estimates that correspond directly to official income- or expenditure-based poverty lines. We pilot this approach for these contiguous countries in Africa to create national and pooled asset poverty estimates. This paper ties the asset-based measures directly to consumption-based poverty measures (e.g., the World Bank's $1.90/day per capita international poverty line). The estimates we construct are consistent with Foster–Greer–Thorbecke (FGT) principles and represent decomposable expected or structural poverty and lend themselves to measuring the depth as well as prevalence of poverty (Carter and Barrett, 2006). Compared with poverty indices currently employed in ML-based poverty mapping literature, our poverty estimates satisfy key poverty measurement axioms, apply directly to official poverty lines, and enable greater interpretability and flexibility in application. The poverty measures we construct can serve as an improved dependent variables for ML-based prediction using remote sensing EO data. We predict not only the prevalence of poverty in a location (i.e., how many people fall below the poverty line) but also its depth or gap (i.e., how far beneath the poverty line people fall). Such improvements can enhance the usefulness of ML-based advances in poverty estimation for both science and society.

This paper contributes to the growing literature on ML poverty prediction in three ways. First, it constructs an internationally-and-intertemporally comparable structural poverty measure, giving practitioners improved poverty measures for ML-based poverty prediction. Second, we harness the strengths of modern poverty measurement, estimating not only prevalence, but also poverty depth or gap, and link expressly to the official national poverty or dominant international poverty line. To our knowledge, this paper is the first ML study constructing comparative asset-based, structural poverty measures and the associated poverty gap and squared poverty gap estimates tied directly to poverty lines.

Third, we assess whether this difference in poverty estimation method matters. This paper compares the predictive performance of ML methods with poverty prediction literature, i.e., does the evaluation criteria (RMSE, R squared, and rank correlation coefficient) vary from that in ML poverty prediction literature when using the same data set, same method, but a different LHS variable? Following Browne et al. (2021), this paper uses very similar feature sets and ML algorithms, but applies asset-based structural poverty estimates as the LHS variable. Intuitively, we hope the predictive performance is comparable to ML poverty prediction literature using alternative poverty estimates (e.g., RWI), which will support the claim that our asset-based structural poverty estimates can serve as new, improved dependent variable in ML poverty prediction task. Also, we compare the FGT0 measure with FGT1, 2 measures through rank correlation coefficient to see whether targeting based on poverty prevalence estimates would differ from those based on more distributionally-sensitive poverty gap or squared poverty gap measures.

The rest of the paper is organized as follows. Section 2 presents the concept of poverty measurement. Section 3 provides an overview of a variety of alternative poverty measures used in the literature. Section 4 gives the theoretical framework and the methodology of constructing internationally-and-intertemporally comparable asset measures. Section 5 describes the data and variables used. Section 6 provides the results of both country-specific estimation and cross-country estimation. Section 7 concludes and discusses the prospects for future applications and limitations.

CHAPTER 2

CONCEPT OF POVERTY MEASUREMENT

What do we want to know about poverty? Scholars want to estimate 1) what share of population are poor (poverty headcount ratio), 2) how poor are they (poverty gap or depth), and 3) where they are (poverty mapping or poverty spatial distribution). The main method poverty scholars use is based on the Foster-Greer-Thorbecke (Foster et al., 1984) class of poverty measures, which satisfy several desirable axiomatic properties of poverty measurement. Applications of FGT measures have included over four generations of poverty measurement, from static expenditure poverty to dynamic asset poverty (Carter and Barrett, 2006).

2.1 Foster-Greer-Thorbecke poverty measures

The Foster-Greer-Thorbecke class of poverty measures (Foster et al., 1984), which have become the standard for international evaluations of poverty, take the form:

$$P_{\alpha} = \frac{1}{n}\sum_{i=1}^{q}(\frac{z-y_i}{z})^{\alpha} \tag{2.1}$$

where z is the poverty line; n is the total population; q is the population with wellbeing below the poverty line; and y_i is the i^{th} wellbeing rank (ordered from lowest to highest) in the population. The poverty aversion parameter α takes non-negative values. As α increases from zero, P_{α} becomes more sensitive to the severity of poverty experienced by the poor, represented by the normalized poverty gap for individual i:$(z - y_i)/z$. We compute the three most commonly used FGT poverty measures: $P_{\alpha=0}$, P_1, and P_2. The P_0 measure is the poverty headcount ratio or prevalence, while P_1 and P_2 are frequently referred to as the "poverty gap" and "squared poverty gap" measure, respectively.

The FGT class of poverty measures satisfy desirable axiomatic properties. Sen (1976) argues that: (axiom i) a measure of poverty should increase if the income of a poor person falls; (axiom ii) a measure of poverty should increase if income is transferred from a very poor person to a less poor person. Foster et al. (1984, 2010) provide a retrospective view of Sen's paper and add (axiom iii) additive decomposability: total poverty is a weighted average of the subgroups; and (axiom iv) subgroup consistency: an overall measure of poverty should decrease when poverty is reduced in one subgroup and the other subgroups are unchanged. The P_0 measure satisfies axiom iii and iv; P_1 additionally satisfies axiom i; P_2 satisfies all four axioms. So it is meaningful and necessary to generate not only headcount ratio but also poverty gap and squared poverty gap measures. In addition to its sound theoretical properties, FGT measures' simple structure – based on powers of normalized shortfalls – also facilitates communication with practitioners. For these reasons, FGT measures have become the workhorse poverty measures globally over the last 30-plus years.

2.2 Four generations of poverty measurements

Economists have developed four generations of poverty measurements (Carter and Barrett, 2006) from static expenditure poverty to dynamic asset poverty (Table 2.1). The first generation poverty measurement relied on household expenditure or income data at a single point of time. Once a money metric poverty line is defined (e.g., \$1.90 PPP-adjusted per capita per day consumption-based poverty line), the population can be divided into poor and non-poor categories, and then FGT measures can be used to calculate the prevalence and depth of poverty within an economy.

Table 2.1: Alternative approaches to poverty measurements

Generation	Poverty Measurement	Categories
1st Generation	static expenditure poverty	poor, non-poor
2nd Generation	dynamic expenditure poverty	chronically poor, transitorily poor, never poor
3rd Generation	static asset poverty	structural poor, stochastic poor, never poor
4th Generation	dynamic asset poverty	persistently poor, dynamically mobile (long term)

The first-generation poverty measures cannot, however, distinguish between two significantly different poverty process. Say, one estimates a 20 percent poverty headcount ratio under the first-generation framework. This could reflect the same 20 percent of individuals are persistently poor each period, or that over time all households are poor one-fifth of the time. The former tends to be a more polarized society – a more hopeless large subpopulation of chronically poor individuals, whereas in the latter economy people move in and out of poverty or only occasionally fall below the poverty line for a short period of time (Carter and Barrett, 2006).

Interest in distinguishing between these two different situations inspired the second generation of poverty analysis based on panel data, i.e., longitudinal observations of the same households over time. The panel structure enables researchers to categorize households into three categories: the chronically poor, the transitorily poor and the never poor (Baulch and Hoddinott, 2000). In a society typified by chronic poverty, a large subpopulation are persistently poor period after period and vulnerable to despair and hopelessness, whereas in transitorily poor society, people share the burden of poverty almost equally for a minority of time on the basis of random outcomes.

The second-generation measurement cannot, however, distinguish structural poverty transitions from stochastic transitions. Some people may be originally poor just due to bad luck, and their subsequent transition to non-poor status just reflects a return to their expected living standard. This is so-called "stochastic poverty", i.e., one is temporally poor due to bad lack. For others, the transition may be structural, because they accumulate new assets, or due to increased returns on the assets that they already own, leading to a rise in their expected standard of living. Previously, they were expected to be "structurally" poor, but now they are expected to be non-poor.

To address concerns about second generation poverty measures and distinguish between the stochastic and structural poor, Carter and May (1999, 2001) developed a poverty measurement method that mapped expenditure or income flows into asset space. This is the third generation poverty measurement. Productive assets are the structural source of income generation, which underpins expenditures over time (Friedman, 1957; Barrett et al., 2016). The term "asset" extensively includes traditional, privately held productive assets (e.g., land, livestock, equipment, human capital) and financial capital, as well as social, geographical and market access status that confers economic advantage. Carter and May identify a capital stock-based poverty line, which can be applied to differentiate stochastic from structural transitions, as an extension of flow-based expenditure poverty line.

The long-term persistence of structural poverty remains unsettled in the third generation measurement, however. It is unclear whether the currently structurally poor are likely to remain poor over the longer term, to fall into a poverty trap, or if they are on a predictable, if as-yet-unrealized, climb out of poverty. The fourth generation poverty measures permit a forward-looking approach based on estimation of poverty dynamics (Carter and Barrett, 2006; Barrett and Carter, 2013).

This paper's contribution is to bring the third generation poverty measures, based on static asset poverty lines, to the ML poverty mapping space. We bring together the gains of these two workstreams – poverty measurement that satisfies important axiomatic properties for identifying people experiencing severe and enduring poverty, and advances in our ability to estimate poverty in places and times not sampled by household surveys, especially stock-based measures like asset poverty.

To construct a 3rd generation, asset-based poverty measure, we use household consumption expenditure as the left-hand side variable, and productive assets as independent variables (or input features in ML language). The flow-based poverty measure, consumption expenditure, is linked to the asset-based measures to estimate structural poverty following the third generation poverty measures (Carter and May, 1999), expressed in FGT form as P_0, P_1, P_2 estimates. The poverty measure we construct can serve as a dependent variable in the poverty estimation studies using supervised learning algorithms and remote sensing data products as a feature set. Our poverty measurement satisfies key poverty measurement axioms that asset indices currently employed in the ML poverty estimation literature do not satisfy and enables greater interpretability and flexibility in application with respect to policy-relevant poverty lines. The mechanics of constructing asset-based structural poverty measures are given in Section 4.1 and 4.2.

CHAPTER 3

ALTERNATIVE POVERTY MEASURES

There are multiple approaches to make comparable wealth or poverty indices from household survey data. Section 3.1 - 3.4 explain some of the frequent measures' construction, where it has been applied to date, and the strengths and weaknesses for each alternative index. The poverty measures most often used in the current ML-based poverty prediction literature, however, have little grounding in theory nor correspondence to official poverty lines used by policymakers. None of them harness the strengths of modern poverty measurement, typically reflecting only prevalence, not depth, nor worrying about international and intertemporal comparability, nor correspondence with official poverty lines. In other words, the community of ML poverty prediction has been using somewhat problematic poverty measurements as dependent variables in prediction.

3.1 Comparative Wealth Index (CWI)

Rutstein and Staveteig (2014) proposed the Comparative Wealth Index (CWI), an experimental methodology to calculate a wealth index that is meant to be comparable across countries and time. 172 standard DHS surveys conducted between 1990 and 2012 in 69 countries were used to estimate the CWI. The construction of CWI takes four steps. First, select a baseline survey, an idea similar to selecting the base year used for price indexes. Rutstein and Staveteig (2014) chose the 2002 Vietnam survey's wealth as the baseline because it was in the middle of the time period of DHS surveys and in the middle income per capita of countries with DHS surveys. Second, use unsatisfied basic needs (UBN) [1]

[1] Unsatisfied Basic Needs framework is developed by the U.N. Economic Commission for

and other variables common to most DHS surveys as “anchoring” points. The UBN-based framework assigns points based on four dimensions: inadequate housing, overcrowding, inadequate sanitation, and high economic dependency. Wealth scores were calculated for the percentage of households that had all four unsatisfied basic needs (4 points), three or more unsatisfied basic needs (3 points), two or more unsatisfied basic needs (2 points), and one or more unsatisfied basic need (1 point). These scores are then used as anchoring points for the Relative Wealth Index (RWI). Third, use proportions of households at given levels of the “anchors” to determine cut-points. Finally, transform and adjust the survey-specific DHS RWI through linear regression on anchor cut-points of the baseline wealth index. The CWI generates results in the form of rankings of countries and regions as well as trends and regional averages that generally accord with per capita income measures.

The CWI is useful in cross-country and trend analysis of demographic and health outcomes. In the illustrative applications for indicators of child mortality, fertility, maternal health care, and child nutritional status, the CWI performed well, indicating the value of absolute wealth measures in comparing national survey data – as important and usually more important than the RWI (Rutstein and Staveteig, 2014). Applying the CWI in trend analysis within and across countries helps to figure out the effects of anti-poverty health programs versus the effects of changes in the economic status of the population. In short, the CWI provides an important option for analysis of health disparities across regions and over time.

Latin America (ECLAC) to measure nonmonetary dimensions of poverty (Feres and Mancero, 2001a). Six key indicators of UBNs in Latin America are: 1) overcrowding, 2) inadequate housing, 3) inadequate source of water, 4) lack of or unsuitability of toilet facilities, 5) children not attending school, and 6) economic capacity.

However, limitations remain in the anchoring point method and results should be interpreted with caution. In the construction of the CWI, selection of the anchoring point criteria is limited for purposes of illustration, and the results could vary if other criteria are used. Indeed, these criteria are fairly arbitrary chosen automatically and not all DHS surveys include all selected criteria. For example, information on the number of sleeping rooms in the home, sharing of toilet facilities with other households, or possession of a fixed telephone were remarkably absent in DHS surveys, but they are important for calculating Unsatisfied Basic Needs (UBN) points. More sensitivity analysis needs to be done to determine 1) the effects of using fewer anchoring points; 2) the effects of choosing the baseline survey; 3) whether a specific functional form (e.g., non-linear regression) would work better (Rutstein and Staveteig, 2014). Plus, the CWI is not tied to any official international or national poverty lines. The magnitude of deviation from cut points does not tell us anything but prevalence relative to a poverty line not anchored to official poverty lines. All these limitations inhibit the wide adoption of CWI by development economists and practitioners.

3.2 Harmonized Wealth Index (HWI)

To extend, test and complement the CWI approach, Staveteig and Mallick (2014) constructed the Harmonized Wealth Index (HWI) using pooled data in eight focal countries (Bangladesh, Bolivia, Cameroon, Egypt, Ghana, Indonesia, Nepal, and Zimbabwe) from the mid-1990s to present. Households in any specific survey having the same basic assets will, by definition, score exactly the same. For the purpose of harmonization, the disaggregated asset categories in later years naturally had to be collapsed back into a single group. Categories that appeared in later years only, such as wooden planks, were impossible to harmonize from earlier surveys and were necessarily collapsed into the 'other' category.

Unlike the CWI, HWI is not a universal metric. Rather, it is a method of computing a pooled asset index for a small set of surveys. The HWI is particularly useful for policymakers and researchers to conduct trend analysis in a small set of countries. On the other hand, its reliability declines substantially as the number of harmonized assets is reduced. Considering the need of common set of assets in the analysis, it should be computed within a single or small number of countries, which indicates low cross-country comparability. Harmonizing assets obviously leads to a loss of information that can differentiate households from one another.

3.3 International Wealth Index (IWI)

To create a metric that can be used for all low- and middle-income countries, Smits and Steendijk (2015) proposed the International Wealth Index (IWI), a comparable wealth index based on data from 165 household surveys between 1996 to 2011, primarily DHS surveys. The authors pooled the data and used principal component analysis (PCA) to compute an index for a common set of assets. The factors were refined into a set of generalized weights that scale households between 0 and 100. Households with an IWI value of 0 have none of the durables, lowest housing quality, and no access to public services, while the value of 100 indicates having all consumer durables and highest quality services and housing characteristics.

A major advantage of IWI is that it creates a stable set of asset weights that can be applied to successive surveys without additional computation. Also, the IWI has the good property of easy reproducibility and usability: a comparable score can be easily produced for any individual household with the requisite information.

However, there exist two key built-in disadvantages for this pooled data method. First, constructing IWI loses a great amount of useful information. Finding a small set of assets common to all 165 surveys requires discarding much useful asset information collected in some but not all surveys. In particular, the index suffers from the problem of missing and reduced number of asset questions in earlier surveys compared with later surveys. Another limitation is the low comparability when adding additional surveys. As the construction of IWI was done at a single time period, adding a new survey in the future requires re-pooling the data each time and then generating different weights from PCA. As the socioeconomic status of the population changes over time, the original weights become increasingly less applicable for further practitioners. With one or two additional surveys, there might be negligible differences, but adding a bunch of additional surveys will increasingly deviate from the original weights. Similar to CWI and HWI, IWI faces the same criticism as it selects an arbitrary set of assets and is not anchored to any official poverty line.

3.4 Absolute Wealth Estimates (AWE)

Hruschka et al. (2015) proposed the absolute wealth estimates (AWE) based on data from 156 DHS surveys from 66 countries. The core method relies on two main inputs for each country-year combination: 1) the relative rank of each surveyed household based on its assets holding; 2) the assumed shape of the resulting wealth distribution. The first input, the relative rank of household, is given by wealth factor score in the DHS surveys. The shape of wealth distribution is determined by three parameters: the mean wealth per capita, the Gini coefficient (to measure variance), and the best combination of the Pareto and log-normal distributions (to estimate skewness). With these inputs, the authors identify the specific rank of each household in the wealth distribution and generate the absolute wealth estimates.

As a comparable wealth measure, AWE makes it possible to examine the relationship between absolute economic status and health, growth, and other outcomes of interest. Unlike other methods to produce comparable wealth measures (e.g., CWI), AWE only depends on two inputs – the relative household ranking and the assumed shape of wealth distribution – but does not rely on countries selection as well as selected baseline surveys.

Hruschka et al. (2015) also acknowledge several limitations. First, AWE relies on a fairly arbitrary distributional assumptions and assumes that all country-years follow the same wealth distribution. Second, the estimates for the wealthiest households will be particularly uncertain because of the skewed distribution of wealth and the limited number of very rich households included in DHS. In addition, the main purpose of AWE is for aggregate analysis; hence, it is likely to be highly uncertain at the level of individual households.

In short, these alternative measures of economic status are somewhat problematic. The major limitation is that they are not anchored to established axiomatic poverty measures, nor to the official poverty lines used by governments and international agencies. Additionally, they either do not satisfy the intertemporal and cross-country comparability or lose lots of information during the construction procedure. Table 3.1 summarizes the poverty estimates used in the existing literature. We find that DHS Relative Wealth Index and other DHS-based wealth index are commonly used outcome measures in ML poverty prediction literature, and few studies use measures tied to expenditure or income-based FGT measures. Therefore, generating internationally-and-intertemporally comparable poverty estimates based on axiomatic measurement principles and official poverty lines can potentially improve the poverty prediction; here, we pilot a new approach to address these gaps.

Table 3.1: Poverty measures used in the existing literature

Literature	Outcome variable	R-squared
Jean et al. (2016)	DHS (asset index), LSMS (expenditure)	0.37-0.75
Ayush et al. (2020)	LSMS (expenditure)	0.46-0.54
Browne et al. (2021)	DHS (malnutrition, asset poverty)	-0.21-0.31
Lee and Braithwaite (2020)	DHS (International Wealth Index)	0.78-0.99
Tang et al. (2022)	DHS (asset index), LSMS (expenditure)	0.34-0.75
Pokhriyal and Jacques (2017)	Global Multidimensional Poverty Index	NA
Yeh et al. (2020)	DHS (Relative Wealth Index)	0.50-0.70
Chi et al. (2022)	DHS (Relative Wealth Index)	0.56-0.70

Notes: NA indicates that the literature does not report R-squared.

CHAPTER 4

METHODOLOGY

To execute ML-based structural poverty estimation, we first apply ML methods to model the relationship between household-level expenditure and assets, and use an official poverty line tied to household expenditure to map to asset structural poverty. Then we predict the FGT measures of structural poverty for locations (e.g., enumeration areas). Then we use ML methods and Earth Observation feature sets to predict FGT measures of structural poverty so that practitioners can project them onto entire national poverty map.

4.1 Construct national asset poverty measures

First, we execute the country-specific estimation and construct single-country asset-based structural poverty measures that correspond to country-specific official poverty lines based on household-level expenditure.

The country-specific estimation procedures are given in Table 4.1. First, employing nationally representative household survey data, we model the relationship between household-level consumption expenditures and the household's asset holdings as $E_i = f(A_i) + \epsilon_i$, where E_i is household-level consumption expenditure, A_i are assets of household i. Second, we generate household-level structural expenditure estimates $\hat{E}_i$, i.e., a household's expected expenditure given its asset holdings. The predicted consumption expenditure is directly anchored to both the official national poverty line and, when converted into 2011 PPP-adjusted US dollar terms, to the World Bank \$1.90/day and \$3.20/day per capita global poverty line. The third step is to estimate enumeration area (EA)-level structural poverty prevalence, poverty gap, and squared poverty gap (P_0, P_1, P_2) based on the FGT framework and the $\hat{E}_i$ estimates in each enumeration area. We generate $P_{\alpha,s}$ = FGT($\hat{E}_i$| poverty line), where $\alpha \in \{0, 1, 2\}$, and s is the enumeration area (EA). Fourth, we predict EA-level P_0, P_1, P_2 measures based on Earth Observation features and ML methods, i.e., $P_{\alpha,s} = g(Z_s)+\xi$, where Z_s is EO feature set, ξ is the error term, $g(\cdot)$ are machine learning algorithms. Finally, the trained algorithms can be projected onto EO data for the whole country region (not just survey EAs) to generate a complete national poverty map.

Table 4.1: Country-specific estimation procedures

	Model	Expected Ouput
Step 1	$E_i = f(A_i) + \epsilon_i$	trained ML algorithm $\hat{f}$
Step 2	$\hat{E}_i = \hat{f}(A_i)$	HH-level structural expenditure estimates $\hat{E}_i$
Step 3	$P_{\alpha,s}$ = FGT($\hat{E}_i$| poverty line), $\forall i \in s$	EA-level poverty estimates $P_{0,s}, P_{1,s}, P_{2,s}$
Step 4	$P_{\alpha,s} = g(Z_s) + \xi$	trained ML algorithms $\hat{g}$
Step 5	$\hat{P} = \hat{g}(Z)$	complete national poverty map

Notes: E_i is household-level consumption expenditure; A_i are productive assets of household i; $P_{\alpha,s}$ is EA-level poverty estimates based on FGT framework, $\alpha \in \{0, 1, 2\}$, s is enumeration area (EA); Z_s is EO feature set in enumeration area s; Z is EO data for the whole country region (not just survey EAs)

Several data preprocessing procedures are necessary before fitting the model. The data are split into a training set (75%) and a test set (25%) by household ID. We also demean all data so that all household-level features x_i have mean equal to zero. That way, second-order flexible functional forms generate exact second-order approximations of the unknown true functional form at the sample mean.

In terms of model selection, we consider a variety of models including ordinary least squares (first-order approximation, second-order approximation, household fixed effects model, *lny* as dependent variable) as well as machine learning models (LASSO and random forest).

Consistent with the poverty prediction literature, linear regression (Eq. 4.1)

$$y = X\beta + \epsilon \tag{4.1}$$

serves as the benchmark when comparing prediction performance with machine learning algorithms. In addition to simple linear regression, we also tried three other OLS specifications - 1) OLS with second order approximation (i.e., include all independent variables, squared terms, and interaction terms in the input features X); 2) lny as dependent variable because expenditures distributions are often positively skewed; 3) controlling for household fixed effects - to check whether the model performance improves.

The Least Absolute Shrinkage and Selection Operator (LASSO) is an extension of ordinary least squares model (Eq. 4.2). Coefficients of the LASSO model are chosen to minimize the sum of squared residuals plus a penalty term that penalizes the size of the model through the sum of absolute values of the coefficients.

$$\hat{\beta}^{LASSO} = \arg\min_{b} \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} x_{ij} b_j)^2 + \lambda \sum_{j=1}^{p} |b_j| \gamma_j \tag{4.2}$$

where n is number of observations, p is number of features, λ is a tuning parameter (higher value results in more shrinkage of b_j), and γ_j is a normalizing factor to address different scales of X's. The first term of the optimization objective function is exactly the same as standard OLS, and the remaining part is the penalty term, which will set many coefficients equal to zero to meet the parsimonious variable selection goal.

Random forests (RF) are among the most commonplace ML algorithms and can be easily implemented using widely available off-the-shelf packages in many programming languages. RF models refer to ensembles of regression trees where a set of T un-pruned regression trees are generated based on bootstrap sampling from the original training data. For each node, the optimal feature for node splitting is selected from a random set of m features from the total N features. The selection of the feature for node splitting from a random set of features decreases the correlation between different trees and thus the average prediction of multiple regression trees is expected to have lower variance than individual regression trees (Hastie et al., 2009).

The Random Forest is one of the best, most popular and easiest to use out-of-the-box classifiers. There are two reasons for this. First, RF only has two hyper-parameters, number of bagged datasets and number of features we search. The output classifier is extremely insensitive to both of these hyperparameters. Second, decision trees do not require a lot of preprocessing. For example, the features can be of different scale, magnitude, or slope. This can be highly advantageous in scenarios with heterogeneous data, for example the poverty prediction settings where features could be things like size of land, tropical livestock unit, HH head gender, each of which is recorded in completely different variable types and units.

Because the focus of this paper is not to develop a new machine learning algorithms or to improve model performance by using fancier estimators or a better feature set, here we only use the relatively simple specifications like RF following Browne et al. (2021). We hope that practitioners can produce still-better poverty predictions through more user-friendly algorithms without the need of deep learning or transfer learning methods or taking too much storage space or execution time.

4.2 Comparative analysis and construction of a comparative poverty measure

While allocating resources within-country may best be served by own-country structural poverty models relating assets to consumption expenditures, for many applications – especially regional or global research and policy initiatives spanning multiple LMICs – cross-national comparability is critical. We therefore propose to aggregate/pool the national survey data to create a regional dataset of assets and consumption expenditures data for several countries in the study domain (Tanzania, Uganda, and Malawi) and then apply the same estimation approach using pooled data, all put in common, real, PPP-adjusted terms. The resulting, cross-nationally comparable, structural poverty measure remedies the shortcomings of DHS and similar asset indices (see Section 3).

The reason for pooled estimation is that the ambition in ML poverty (or malnutrition) prediction is to be able to project into spaces that have not been sampled. The longer-term goal (beyond this paper) is to generate global estimates that can be matched with, for example, global estimates of the value of ecosystem services or of exposure to natural hazards. Necessarily, that will require projecting estimates into countries that lack nationally representative data that could support the sort of country-specific estimation this paper covers. So the aspiration is to identify enough common correlation structure across countries to identify a *predictive* model (not an inferential model) that performs satisfactorily across multiple countries. This is not analytically different from projecting a nation-specific model based on a sample into unsampled spaces within that same country (i.e., ML-based poverty mapping, as distinct from census-based poverty mapping). There exists well-justified skepticism about pooling across countries, since any such effort relies on a very strong assumption that there exist common patterns across countries that are sufficiently strong to explain intra-national as well as cross-national variation. We need to test that assumption, i.e., to be explicit about how much predictive skill one sacrifices, if any, by pooling.

Comparing these regional results to national models will enable us to better assess how predictive accuracy changes with estimation of a regional model. When one builds a cross-nationally comparable index, how much does this impact the poverty measures generated for a given country and year? We can compare rank correlation coefficients among measures generated using nation-specific versus regional data to test how sensitive predictions are to the geographic breadth of the data included. Put differently, this gets at the prospective tradeoff between regional generalizability and country-specificity.

After the country-specific estimation and comparative analysis, this paper then generates the poverty gap and squared poverty gap measures based on the Foster-Greer-Thorbecke method (see Section 2.1).

4.3 Strengths compared with alternative poverty measures

Compared with the most common outcome measure, the DHS Relative Wealth Index, the approach described in this article enables the production of a more comparable, flexible, interpretable, normatively-anchored, and axiomatically attractive structural poverty estimates (Table 4.2).

Comparability: Because DHS relative wealth index is survey-specific, calibrated to a particular country and moment in time, it is not suitable for intertemporal or cross-country comparison of absolute wealth or poverty. Multiple methodologies have emerged to permit comparison across DHS surveys – including internationally standardized indices (e.g., the Comparative Wealth Index) or study-specific indices generated using principal components analysis (Yeh et al., 2020). Indices that use the same weighting scheme across periods, countries, or both, enhance comparability over space and time, and therefore usefulness for some research and policy applications. However, the relationship between assets and poverty varies over place and time (McBride et al., 2022). For example, the asset stocks associated with a non-poor state likely varies from peri-urban Southeast Asia to rural sub-Saharan Africa to urban Latin America. Thus, computing a study-specific wealth index may have advantages in a regional study of countries with shared asset wealth characteristics. More generally, one needs a clear sense of the purpose of comparisons to build an appropriate, comparable measure.

Interpretability: Most poverty measures - such as the World Bank's global \$1.90/day per person poverty line, which underpins Sustainable Development Goal 1 poverty commitments – rely on income or consumption expenditure measures in monetary units that are readily interpretable. Because the DHS does not collect consumption expenditures nor income data, those asset indices cannot be used directly to generate money-metric poverty measures. Instead, DHS and other asset indices are necessarily unitless measures. This makes them difficult to interpret.

Normative anchoring: An important dimension of comparability and interpretability relates to the normative anchoring of a poverty measure. Poverty lines based on some absolute standard of living typically provide those normative anchors (Ravallion, 1992; Lipton and Ravallion, 1995). But most unitless asset indices lack any explicit connection to a normative anchor, especially to any absolute standard of living. Rather, 'poverty' is typically defined as an arbitrarily defined bottom quantile of the asset index distribution. An asset-based poverty measure may be more useful if it is readily comparable to national or global poverty lines, such as the World Bank's $1.90 a day per person measure, which itself derives from a range of country-specific poverty lines.

Flexibility: In part because of the interpretability and normative anchoring shortcomings of asset index measures, the current literature offers only poverty prevalence predictions. But policymakers often want to know not only where one finds the greatest share of people in poverty, but also how poor they are, i.e., how great is the poverty gap between people's current state and the relevant poverty line. Sometimes policymakers seek even more distributionally-sensitive measures, for example, when trying to target the poorest of the poor. Money metric poverty measures in the Foster-Greer-Thorbecke (Foster et al., 1984) tradition directly permit flexibility in adapting measures to the question at hand. In contrast, the asset indices presently used in ML-based asset poverty prediction do not.

Measures' axiomatic properties: One reason why poverty measures' flexibility is important stems from the degree to which they satisfy desirable axiomatic properties. As Sen (1976) famously showed, a poverty headcount ratio fails to satisfy several important properties. Chief among these are the Monotonicity axiom – a measure of poverty should increase if the income of a poor person falls – and the Transfer axiom – a measure of poverty should increase if income is transferred from a person beneath the poverty line to one above it. The FGT family of poverty measures satisfy these and other axiomatic properties of a poverty measure, including Additive Decomposability and Subgroup Consistency.

Table 4.2: Strengths relative to DHS Relative Wealth Index

Dimension	DHS RWI	Our structural poverty measurement
Comparability	Survey specific	Internationally-and-intertemporally comparable
Interpretability	Does not collect consumption/income data; Cannot be used directly to generate money metric	Rely on consumption expenditure measures in monetary units that are readily interpretable
Normative anchoring	Defined as an arbitrarily defined bottom quantile of the asset index distribution	Readily comparable to national/global poverty lines (e.g., World Bank's $1.90/day per capita)
Flexibility	Not based on FGT class	Based on FGT class; Can generate poverty gap measure
Axiomatic properties	Not satisfy FGT 4 axioms	Satisfy 4 axioms of FGT class

CHAPTER 5

DATA

5.1 Data source

The data used to construct the poverty estimates come from the Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) for Tanzania, Uganda, and Malawi. The LSMS project collaborates with the national statistics offices of its partner countries in Sub-Saharan Africa to design and implement systems of nationally representative household panel surveys. LSMS surveys collect information on a wide range of topics including agricultural production, non-farm income generating activities, consumption expenditures, and a wealth of other socioeconomic characteristics. The survey's main objective is to provide high-quality, household-level data for the government and other stakeholders to monitor poverty and other key characteristics that permit better targeting and impact evaluation of policy initiatives.

Why use LSMS data? The choice of data is driven by the methods we use and their requirements. The ideal data set should include both consumption expenditure as a dependent variable and productive assets as predictors. To the best of our knowledge, LSMS are the best available nationally representative panel data sets satisfying this requirement and reasonably comparable across our study countries. DHS data are most commonly used for ML-based poverty prediction, but they do not include consumption expenditure nor income data and they lack panel structure. In LSMS, the panel structure makes it possible to control for time-invariant household unobservables if one wishes to.

We therefore only consider the LSMS data rounds that include consumption expenditure aggregates, which are key dependent variables when constructing national poverty estimates [1]. The selected datasets are shown in Table 5.1.

Table 5.1: Selected LSMS Datasets

Country	Year	# of observations	Sample size
Tanzania	2008	3,122	16,173
	2010	3,754	
	2012	4,876	
	2014	3,337	
	2019	1,084	
Uganda	2011	1,263	4,011
	2013	2,748	
Malawi	2010	3,098	15,519
	2016	12,421	

Our study area includes multiple countries in eastern Africa (Tanzania, Uganda, and Malawi), one of the poorest regions in the world. Agricultural livelihoods are more readily predictable using the sorts of big Earth Observation data sets commonly used in the ML literature. One can't observe or estimate educational attainment, prior work experience, etc. from remote sensing data. But weather, vegetation status, road conditions, etc. – which are strongly related to agricultural assets and productivity – can be observed. So ML methods based on remote sensing EO data may work best in rural, agrarian settings.

Table 5.2 illustrates the poverty and agricultural dependence in the countries of interest. An estimated 40% (Uganda) to 70% (Malawi) of the populations fall below the international poverty line of $1.90 per day (2011 PPP). In each country, over 70% of the population depends for their livelihoods on agriculture, hunting, fishing, or forestry. Although this paper only focuses on eastern and

[1] After connecting with LSMS team, we are informed that the consumption aggregates for Malawi IHS5 2019/20 and IHPS 2019 are being checked and have not been uploaded. Similarly, consumption aggregates in Uganda 2010/2011, 2015/2016, 2018/2019, 2019/2020 waves are missing because of lack of sufficient documentation of the methodology and syntax files.

southern African countries, our proposed methods can be easily extended to other low- and middle-income countries more broadly for ML poverty prediction if researchers have access to both consumption aggregates and productive assets data.

Table 5.2: National figures of poverty and agricultural dependency in eastern Africa

Country	Poverty headcount ratio (%)	Poverty gap	Human Development Index	Agricultural population (%)	Agricultural value added (% of GDP)
Malawi	69.2	28.7	0.47	72.9	30.5
Tanzania	49.4	15.9	0.51	73.3	28.8
Uganda	41.3	13.1	0.53	73.5	24.7

Definitions: The poverty headcount ratio is defined as the percentage of people living below the international poverty line ($1.90 a day in 2011 PPP). The poverty gap is the mean shortfall in income or consumption from the poverty line of $1.90 a day, expressed as a percentage of the poverty line. The Human Development Index is a summary measure of multiple dimensions of human development (a long and healthy life, knowledge, and a decent standard of living). The agricultural population is the share of the population (including their dependents) with livelihoods in agriculture, hunting, fishing or forestry. The value added in value added of the agricultural sector is the net output after adding up all outputs and subtracting intermediate inputs (not accounting for the depletion and degradation of natural resources). Estimates are for most recent year available, which varies by statistic and country. **Data sources**: The World Development Indicators (World Bank 2021); United Nations Development Program (2016).

5.2 Dependent variables

The outcome of interest in this paper is household-level consumption expenditures, which is the flow-based poverty metric used to compute further poverty statistics per official poverty lines. For example, a household is said to be poor or in poverty if their measured consumption expenditure falls below a defined international poverty line ($1.90, $3.20 per capita per day) or a national poverty line. The current national poverty lines in Uganda, Tanzania and Malawi are UGX 992 (international $0.88-$1.04), TZS 1644 (international $2.16), and MWK

449.84 (international $2.09), respectively.[2] In the more recent poverty estimation literature, studies have leveraged such data to estimate asset poverty thresholds that map directly to poor or non-poor states of expenditure or income (Carter and May, 2001; Carter and Barrett, 2006; Barrett et al., 2006). As Carter and Barrett (2006) explain – and Foster et al. (2010) endorse – the resulting asset-based structural poverty measures have several desirable properties. For example, the asset-based measures are thought to better capture households' longer-run economic status and identify structural poverty separately from stochastic poverty.

Before cleaning consumption data, we first check how consumption aggregates are constructed in each LSMS survey through Basic Information Document (BID) and household questionnaires. The aim of this step is to ensure the regressions we run use aggregates that are created using consistent protocols. The key finding is that in the consumption aggregates of all LSMS surveys, auto-provided food is included, housing is treated in an internally consistent manner (i.e., either impute rents for those who own the residence they occupy and include it in consumption expenditures, as in Uganda, or exclude cash rental payments and don't impute rents for owner-occupied housing, as in Tanzania), and business expenses (e.g., spending on fertilizer or business or farm equipment) and debt repayment are excluded even though they involve monetary outlays. Differences in how housing expenditures are handled will create differences between countries automatically, hence the need for country dummy variables in pooled multi-country estimation. Details regarding the construction of consumption aggregates are given in Appendix A.

To convert different currencies to a common currency and equalize their purchasing power over countries and years, we make purchasing power parity

[2] We convert national poverty lines into 2011 PPP-adjusted US dollar terms.

(PPP) adjustment for consumption aggregate in each country-year combination to generate consistent real values. The PPP conversion factor is defined as the number of units of a country's currency required to buy the same amounts of goods and services in the domestic market as the U.S. dollar would buy in the United States. The PPP conversion factor is a spatial price deflator and currency converter that controls for price level differences as well as currency exchange rates between countries, thereby allowing volume comparisons of gross domestic product (GDP) and its expenditure components. We can generate PPP-adjusted consumption expenditure per capita per day as:

$$\text{PPP-Adjusted real consumption (\$/day/person)} = \frac{\text{Consumption in local currency}}{\text{PPP conversion factor} \times \text{hhsize} \times 365} \tag{5.1}$$

Table 5.3: PPP conversion factor (local currency unit per international $)

Year	Tanzania	Uganda	Malawi
2008	458.34	729.21	65.60
2009	515.83	827.08	71.38
2010	538.97	846.09	75.44
2011	588.79	944.26	78.70
2012	636.55	1017.31	92.54
2013	681.66	1064.90	117.76
2014	706.94	1081.16	146.15
2015	727.54	1127.69	178.78
2016	735.66	1172.63	215.13
2017	754.62	1221.09	241.93
2018	762.37	1223.25	265.49
2019	774.74	1235.95	285.20

Source: International Comparison Program, World Bank | World Development Indicators database, World Bank | Eurostat-OECD PPP Programme

5.3 Independent variables

Why use assets as input features for poverty prediction? Predictive assets (e.g., land, livestock, equipment) are key bases of income generation, therefore assets are strongly correlated with income or expenditures, the dominant flow-based measures of money-metric well-being. But survey data collection of asset data – as compared to expenditures or income data – is far easier, cheaper, and less prone to substantial measurement error (Barrett et al., 2016). Moreover, especially in places subject to multiple market failures that impede consumption expenditure smoothing, productive asset holdings more accurately reflect permanent income, thereby providing a better estimate of expected, structural poverty (Sen, 1995; Carter and May, 2001; Carter and Barrett, 2006; Sullivan et al., 2008). Finally, for analysis ultimately aimed at understanding dynamics – e.g., at changes in living conditions – stock variables (e.g., assets) typically describe the current state of the system better than flow variables (e.g., income) do.

How to choose assets in our regions of interest? Assets vary in importance among households in different regions. In east Africa, the poorest households' livelihood typically depend heavily on smallholder semi-subsistence agricultural production and unskilled agricultural labor markets. Labor power is the main productive asset stock (Barrett et al., 2001). Relatively wealthier households in regions of interest commonly rely more on earnings from land, livestock and skilled employment (e.g., salaried labor or skill-/capital-intensive non-farm enterprises). As a result, we consider a vector of productive assets (livestock, land, labor quality, capital equipment, and so forth) associated with earned income and thus consumption expenditure (Table 5.4).

Table 5.4: Productive assets selected

Category	Variable
Livestock	# of livestock owned, Tropical Livestock Unit (TLU)
Land	size of land, size of irrigated land
Capital equipment: transport	# of cars, trucks, bicycles, motorcycles, boats
Capital equipment: communications	# of mobile phone, computers
Capital equipment: agriculture	# of ploughs, tractors, planters, harvesters, sprayers, tillers
Water and Sanitation (Labor quality)	# of (sleeping) rooms, major material of roof, wall, floor, main source of drinking water, type of toilet facilities
Human Capital	household size, household head gender, age, educational attainment

In the raw LSMS data set, several of these variables – including major material of roof, wall, floor, main source of drinking water, toilet facility type – are recorded as categorical variables. The categories are often survey-specific, i.e., possible responses to survey questions often vary across countries and waves. Take the question "What is the main toilet facilities usually used in this household?" as an example. In the Tanzania 2010 questionnaire, categories include no toilet, flush toilet, pour toilet, VIP, ecosan, unimproved pit latrine, improved pit latrine, other. The Tanzania 2014 survey's possible options are no toilet, pit latrine without slab/open pit, pit latrine with slab (not washable), pit latrine with slab (washable), VIP, pour flush, flush toilet, ecosan, other. As a result, it is necessary to recode the categorical variables to a standardized dummy variable (often 0 indicates "bad/unimproved quality", and 1 indicates "good/improved quality"), with a clear mapping from survey-specific options to the standardized outcomes. In Table 5.5, the transformation we use draws on DHS standards (Croft et al., 2018).

Due to the fact that livestock holdings vary based on agroecological conditions within and among countries, a standardized measure of livestock needs to be constructed. The most commonly used variable is Tropical Livestock Units (TLU), which are livestock numbers converted to a common unit (Jahnke and

Jahnke, 1982). Measuring 250 kg of liveweight, the TLU has been used as the reference point to factor livestock of different species by biomass in LMICs since at least the mid-20th century. The camel, as the largest livestock species in tropical regions at that time, with an average liveweight of 250 kg, was defined as 1 TLU; further conversion factors were established for the remaining species. Cattle were assumed to have an average weight of 175 kg, equating to 0.7 TLU per head, with 1 TLU per head allocated for steer, cows, bulls and heifers, 0.1 for sheep and goats, 0.2 for pigs, 0.8 for horses, 0.7 for mules, 0.5 for asses, and 0.01 for chickens, rabbits, ducks and turkeys (Rothman-Ostrow et al., 2020).

After recoding categorical variable and generating TLUs, we cleaned the input features for fitting the model. The detailed cleaning process is given in Appendix B. Readers should be able to fully reproduce results in this paper from the raw data following the steps outlined below.

5.4 Geographical feature set in predicting poverty

In feature set selection, we focus on publicly available data in order to represent feasible, low-cost features for economists and institutions that are unable to invest much in data collection or procurement. We finalize feature selection based on a range of prior studies (Browne et al., 2021; Elbers et al., 2003; Lang et al., 2013; Christiaensen et al., 2012) to have significant predictive power in explaining geographic patterns of poverty. The settled Earth Observation (EO) features we consider include population, remoteness, night light intensity, elevation, slope, vegetation, climate and weather, and time (Table 5.6).

Population (population count and density): Population data come from World-

Pop (WorldPop and for International Earth Science Information Network , CIESIN), which provides various types of modelled gridded population datasets. The "unconstrained global mosaics of population counts 2000-2020 (1km resolution)" dataset is used in this study and the years are aligned with corresponding survey years. We first calculated the total population count for each EA buffer zone (the radius is 5 km for a rural EA and 2 km for an urban EA) by summing up the values of grids that fall within the buffer boundaries. Then we derived the EA-level population density by dividing the population counts over the buffer areas.

Remoteness (travel time): Travel time to nearest cities represents remoteness of the place and is a good indicator of prosperity. We obtained the data from the Malaria Atlas Project (MAP), which published a global map of travel time to the nearest urban centre (defined as a city with a population larger than 50,000 or a contiguous area with population density larger than 1,500 per square kilometer) for the year 2015 at 1 km resolution (Weiss et al., 2018). We took the average values of the pixels falling within the EA buffer zones. The values of year 2015 are used to approximate other years because there is a lack of annual data and infrastructure change is generally slow in a few years.

Night lights (lights): Nighttime lights emitted from the earth and received by satellites in the outer space have proved to capture economic activities well. We assembled the data from a harmonized global nighttime light dataset 1992-2021 (Li et al., 2020), which integrated Defense Meteorological Satellite Program (DMSP)/Operational Linescan System (OLS) data from 1992-2013 and Visible Infrared Imaging Radiometer Suite (VIIRS) data from 2012-2021. We selected the years that correspond to our survey years and calculated the mean nighttime lights intensity for each EA.

Geography (elevation, slope): Geophysical factors constraint development and poverty in many regions. The relevant variables used in this study include elevation and slope. The elevation data are collected by the NASA Shuttle Radar Topography Mission (SRTM) at 1-arc-second (~30 meters) resolution globally (Nasa, 2013). We computed the slope information from elevation data using a built-in slope function in Google Earth Engine (GEE). Then we calculated the values of slope and elevation for EAs by taking the average of pixels within the EA buffer zones.

Vegetation (NDVI): The Normalized Difference Vegetation Index (NDVI) quantifies vegetation greenness on the ground and can be used to predict agricultural production and rural livelihoods. We sourced NDVI from the NOAA Climate Data Record (CDR) of Advanced Very High-Resolution Radiometer (AVHRR) Surface Reflectance, which has operated from 1981 to present. The original daily data were aggregated to annual average values. And then we calculated the mean for each EA buffer zone.

Climate and Weather (rainfall and temperature): Climatic conditions and episodes of heat and water stress may impact people's wellbeing through multiple avenues, especially via conditions for agriculture and livestock. We use the Climate Hazards group Infrared Precipitation with Stations (CHIRPS) to construct variables for long-term rainfall and its variability, as well as total rainfall and rainfall z-scores (Funk et al., 2014). Binned temperature variables reflecting the hours above 30 degrees Celsius are constructed from the Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2) 2-meter air temperature (Modeling and , GMAO). Rainfall and temperature variables are included for the year of the LSMS survey as well as the preceding year.

Table 5.5: Recode categorical variables to dummies

	Unimproved	Improved
Drinking water source	Unprotected well; Unprotected spring; Surface water (river/ dam/lake/pond/stream/ Canal/irrigation channel); Other Note: If a spring or well was not specified as protected, it was considered unimproved.	Piped into dwelling; Piped to yard/plot; Public tap/standpipe; Piped to neighbor; Tube well or borehole; Protected well; Protected spring; rainwater; Tanker truck, cart with small tank (delivered water); Bottled water
Toilet facility	Flush-to somewhere else; Pit latrine-without slab/open pit; Bucket toilet; Hanging toilet/latrine; Shared facilities; Open defecation (no facility/ bush/field); Other	Flush-to piped sewer system; Flush-to septic tank; flush-to pit latrine; Flush-don't know where; Pit latrine-ventilated improved pit (VIP); Pit latrine-with slab; Composting toilet
Floor materials	Earth; Sand; Clay; Mud; Dung	Tablets/wood plank; Palm, bamboo; Mat; Adobe; Parquet, polished wood; Vinyl, asphalt strips, floor mat, linoleum; Ceramic tiles, mosaic; Cement; Carpet; Stone; Bricks
Wall material	No wall; Cane/palm/trunks; Dirt; Mud and sticks; Tin/cardboard/paper/bags; Thatched/straw; Bamboo with mud; Stone with mud; Uncovered adobe; Plywood; Cardboard; Reused wood; Trunks with mud; Unburnt bricks; Unburnt bricks with plaster; Unburnt bricks with mud	Cement; Stone with lime/cement; Bricks; Cement blocks; Covered adobe; Wood planks/shingles; Burnt bricks with cement
Roof material	No roof; Grass/thatch/palm leaf; Sod; Straw; Rustic mat; Palm/bamboo; Wood planks; Cardboard; Tarpaulin, plastic	Metal; Wood; Calamine/cement fiber; Ceramic tiles; Cement; Roofing shingles; Asbestos/slate roofing sheets

Source: World Health Organization, Demographic and Health Survey program

Table 5.6: Feature set selection

Category	Variable
Population	population count, population density
Remoteness	travel time to nearest city
Light	nighttime light
Geography	elevation, slope
Vegetation	Normalized Difference Vegetation Index (NDVI)
Climate and Weather	rainfall, rainfall z-scores, temperature
Time	all year dummies

CHAPTER 6

RESULTS

6.1 Country-specific estimation

First, we execute the country-specific estimation and construct single-country structural poverty measures. Take Tanzania as an example. To construct asset-based national poverty measures in Tanzania, we use an unbalanced panel of 16,173 observations in 8,746 households across 5 waves (2008, 2010, 2012, 2014, 2019). Summary statistics of consumption expenditure and productive assets are given in Table 6.1.

6.1.1 Diagnostic statistics & confusion matrix

In Table 4.1 Step 1, we build ML models as $E_i = f(A_i)+\epsilon_i$, where E_i is HH-level consumption expenditure, A_i are assets of household i. To display the result of Step 1, we report out-of-sample diagnostic statistics and a confusion matrix of ML algorithms to evaluate the model performance. The full regression results are given in Appendix C.

This paper considers four out-of-sample metrics including R^2, RMSE, accuracy rate, and rank correlation coefficient as diagnostic statistics to evaluate the model performance and quantify how well a model fits the dataset in the testing sub-sample. Root mean squared error (RMSE) is the square root of the average of squared errors, telling us how far apart the predicted values are from the observed data on average (lower is better). Accuracy rate is the proportion of households which are correctly categorized into a poverty group (higher is better). Technically speaking, it is defined as the sum of diagonal elements of the confusion matrix divided by the sum of the elements in the whole confusion matrix. Another model performance criterion that matters in this care is Spearman's rank correlation coefficient between actual consumption and predicted consumption, which measures the degree of similarity between two rankings and can be used to assess the significance of the relation between them. This is a way of testing which model best preserves the welfare ordering of household observations. All diagnostic statistics including R-squared, RMSE, accuracy rate, and rank correlation coefficient are based on the testing sub-sample.

Table 6.2 displays evaluation metrics for OLS (1st and 2nd order approximation), household fixed effects model, LASSO and random forest model in Tanzania. Table 6.3 displays the same diagnostic statistics but using $\ln y$ as the dependent variables to address possible skewness in the expenditure distribution. Table 6.4 and Table 6.5 show the confusion matrix of each ML model based on the international poverty line and national poverty line respectively. To be more specific, in Table 6.4, we divide people into three groups based on \$1.90 and \$3.20 per capita per day, which are the standard international poverty lines, and Table 6.5 categorizes people based on the national food poverty line (e.g., in Tanzania TZS 33,748 per adult per month or \$1.48 in 2011 USD PPP-adjusted per capita per day) and national basic needs poverty line (TZS 49,320 per adult per month or \$2.16 in 2011 USD PPP-adjusted per capita per day) according to Bank (2019). Table 6.6 contrasts the individual country-level results in Tanzania, Uganda and Malawi.

There are three striking findings in the diagnostic statistics of country-specific estimation. The first is the consistency of general patterns in different countries. According to Table 6.6, the diagnostic statistics on predictive skill in each individual country are generally similar and comparable, demonstrating the generalizability of the approach across multiple settings.

The second finding is the superior performance of the random forest (RF) model. The RF model is strongly predictive of average household consumption expenditure across multiple African countries. Out-of-sample predictions based on models trained separately for each country explain 43% (Uganda) to 50% (Tanzania) of the variation in average household consumption across countries for which recent survey data are available. Compared with the benchmark ordinary least squares model and other machine learning methods, random forest always gives us highest R-squared, accuracy rate, rank correlation coefficient and lowest RMSE.

The third striking finding is the underlying relationship among different models. For example, the prediction performance of OLS 2nd order approximation decreases a lot from training set to testing sub-sample, even achieving negative R-squared in the test set. The negative R-squared in OLS 2nd order approximation is possibly attributed to the huge number of candidate predictors (as high as 277 variables in the right-hand side) and the resulting overfitting problem. LASSO model has very similar metrics with OLS 1st order approximation due to that we only add the penalty term that penalizes the sum of absolute values of the coefficients. Controlling for household fixed effects improves in-sample predictive performance significantly. The simple OLS regression controlling for household fixed effects can achieve up to 0.87 R^2, 80.8% prediction accuracy rate, 0.911 rank correlation coefficient and the lowest RMSE (1.574) in the training set, all of which are even much better than RF model. However, HH Fixed Effects can not serve as the dominant model in ML poverty prediction. In ML poverty prediction literature, few of the Earth Observation and remote sensing data products have panel structure. Additionally, the goal in ML poverty (or malnutrition) prediction is to be able to project into spaces that have not been sampled, but HH FE model cannot be used to predict out of sample. So under most circumstances (especially prediction tasks), it is impractical for practitioners and policymakers to control for household fixed effects as predictors in ML poverty prediction.

Additionally, we are also interested in which productive assets have the strongest predictive power to expenditure in the ML estimation. The variable importance of the Random Forest model is given in Figure 6.3. The identified best predictors are household size, land, and household head age, which intuitively indicates that among all selected productive assets, labor and land are the most informative predictors of flow-based poverty.

6.1.2 Kernel density & nonparametric regression

After training ML models, we generate household-level structural poverty estimates $\hat{E}_i$. Considering that the confusion matrices are coarse and could mask important patterns of the continuous poverty measure, we draw kernel density of predicted and true consumption expenditure ($\hat{E}_i$ and E_i). In Figure 6.1, we find that the kernel density distribution of $\hat{E}_{\mathrm{i,RF}}$ from the random forest model is much closer to that of true expenditure than other models. Consistent with both confusion matrix and our intuition, the random forest model often provides

Table 6.1: Summary Statistics in Tanzania

variable	mean	sd	min	max
PPP-adjusted exp. ($/day/person)	4.089	4.270	0.157	97.897
TLU	2.098	9.048	0.000	441.970
land (acres)	3.873	12.025	0.000	625.000
irrigated land (acres)	0.059	1.383	0.000	150.000
# of bicycles	0.509	0.782	0.000	16.000
# of motorcycles	0.060	0.259	0.000	6.000
# of boats	0.007	0.137	0.000	10.000
# of mobile phones	1.277	1.325	0.000	21.000
# of computers	0.075	1.402	0.000	124.000
# of plough	0.080	0.373	0.000	11.000
# of tractors	0.002	0.053	0.000	2.000
# of harvesters	0.001	0.036	0.000	2.000
# of sprayers	0.038	0.239	0.000	12.000
# of sewing machines	0.115	0.394	0.000	6.000
# of rooms	2.721	1.358	0.000	19.000
roof material	0.714	0.452	0.000	1.000
wall material	0.507	0.500	0.000	1.000
floor material	0.469	0.499	0.000	1.000
drinking water source	0.689	0.463	0.000	1.000
toilet facility	0.513	0.500	0.000	1.000
hh head gender	0.743	0.437	0.000	1.000
hh head age	45.510	15.710	16.000	108.000
hhsize	5.021	3.044	1.000	55.000

Notes: Categorical variable including roof/wall/floor material, drinking water source and toilet facility haven been recoded as standardized 0/1 dummy variable.

Table 6.2: Diagnostic statistics of out-of-sample estimation in Tanzania

Model	R-squared	RMSE	Accuracy Rate	Rank Correlation Coefficient
OLS (1st order approx.)	0.172	3.605	0.607	0.740
OLS (2nd order approx.)	-4.028	8.880	0.615	0.731
LASSO	0.190	3.565	0.605	0.737
Random Forest	0.503	2.793	0.619	0.751
HH Fixed Effects	0.870	1.574	0.808	0.911

Notes: 1) Diagnostic statistics are based on the testing sub-sample except HH Fixed Effects model. 2) R-squared is calculated as $1 - RSS_e/RSS_y$, where the numerator is calculated out of sample and therefore is not bounded above by RSS_y. 3) RMSE is most informative compared to the mean value of the dependent variable ($4.038/capita/day). 4) The accuracy rate is defined as the sum of diagonal elements of confusion matrix divided by the sum of elements in the whole matrix. 5) We do not include quadratic or interaction terms as candidate variables in LASSO model.

Table 6.3: Diagnostic statistics of out-of-sample estimation in Tanzania (use $\ln y$ as dep. variable)

Model	R-squared	RMSE	Accuracy Rate	Rank Correlation Coefficient
OLS (1st order approx.)	0.500	0.554	0.624	0.745
OLS (2nd order approx.)	-0.518	0.965	0.645	0.757
LASSO	0.500	0.554	0.622	0.742
Random Forest	0.572	0.513	0.628	0.752

Notes: 1) Diagnostic statistics are based on the testing sub-sample. 2) R-squared is calculated as $1 - \text{RSS}_e/\text{RSS}_y$, where the numerator is calculated out of sample and therefore is not bounded above by RSS_y. 3) RMSE is most informative compared to the mean value of the dependent variable (1.071). 4) The accuracy rate is defined as the sum of diagonal elements of confusion matrix divided by the sum of elements in the whole matrix. 5) We do not include quadratic or interaction terms as candidate variables in LASSO model.

researchers with the best prediction in our model set.

We also use nonparametric kernel regression to assess model performance graphically across the distribution of actual consumption per capita per day. The local linear regression results, with Epanechnikov kernel and bandwidth 0.6, are given in Figure 6.2, where the x-axis is the actual consumption expenditure (2011 USD PPP-adjusted per capita daily expenditure), the y-axis indicates the prediction error ($\hat{E}_i - E_i$), and the grey band is the 95% confidence interval. The general patterns in Figure 6.2 may be explained by transitory shocks and classical measurement error. Naschold and Barrett (2011) show that total economic mobility can be decomposed into two components, $y = y^* + \epsilon$, where y^* is structural economic mobility, ϵ is stochastic economic mobility, the sum of transitory shocks and measurement error. Structural economic mobility results from changes in household assets or from changes in the expected returns to assets (captured by asset-based predicted consumption expenditure $\hat{E}_i$ in Figure 6.2). In contrast, stochastic economic mobility, captured by prediction error in nonparametric regression, arises either from purely stochastic fluctuations in incomes (e.g., one-off government transfers) or from measurement error. The transitory shock and measurement error increase as the consumption expendi-

Table 6.4: Confusion matrix in Tanzania based on international poverty line

(a) OLS (1st order approx.)

Predicted \ Actual	<$1.90 a day	$1.90-$3.20 a day	>$3.20 a day
<$1.90 a day	**0.15**	0.04	0.01
$1.9-$3.2 a day	0.09	**0.07**	0.03
>$3.2 a day	0.08	0.14	**0.39**

(b) OLS (2nd order approx.)

Predicted \ Actual	<$1.90 a day	$1.90-$3.20 a day	>$3.20 a day
<$1.90 a day	**0.16**	0.05	0.02
$1.9-$3.2 a day	0.12	**0.09**	0.05
>$3.2 a day	0.04	0.10	**0.36**

(c) LASSO

Predicted \ Actual	<$1.90 a day	$1.90-$3.20 a day	>$3.20 a day
<$1.90 a day	**0.14**	0.04	0.01
$1.9-$3.2 a day	0.09	**0.07**	0.03
>$3.2 a day	0.08	0.14	**0.39**

(d) Random Forest

Predicted \ Actual	<$1.90 a day	$1.90-$3.20 a day	>$3.20 a day
<$1.90 a day	**0.15**	0.04	0.01
$1.9-$3.2 a day	0.14	**0.12**	0.07
>$3.2 a day	0.03	0.09	**0.35**

Notes: Confusion matrix is based on the testing sub-sample.

ture rises.

6.1.3 EA-level poverty estimates based on FGT framework

We next aggregate the predicted household-level consumption expenditures, $\hat{E}_i$, into enumeration area (EA) level poverty statistics based on the FGT framework. To be more specific, $P_{\alpha,s}$ = FGT($\hat{E}_i$| poverty line), $\forall i \in s$, where the

Table 6.5: Confusion matrix in Tanzania based on national poverty line

(a) OLS (1st order approx.)

Predicted \ Actual	<$1.48 a day	$1.48-$2.16 a day	>$2.16 a day
<$1.48 a day	**0.09**	0.04	0.02
$1.48-$2.16 a day	0.03	**0.02**	0.02
>$2.16 a day	0.08	0.12	**0.57**

(b) OLS (2nd order approx.)

Predicted \ Actual	<$1.48 a day	$1.48-$2.16 a day	>$2.16 a day
<$1.48 a day	**0.08**	0.04	0.03
$1.48-$2.16 a day	0.05	**0.04**	0.04
>$2.16 a day	0.07	0.10	**0.55**

(c) LASSO

Predicted \ Actual	<$1.48 a day	$1.48-$2.16 a day	>$2.16 a day
<$1.48 a day	**0.08**	0.04	0.02
$1.48-$2.16 a day	0.03	**0.03**	0.02
>$2.16 a day	0.08	0.11	**0.58**

(d) Random Forest

Predicted \ Actual	<$1.48 a day	$1.48-$2.16 a day	>$2.16 a day
<$1.48 a day	**0.04**	0.01	0
$1.48-$2.16 a day	0.11	**0.07**	0.06
>$2.16 a day	0.06	0.09	**0.55**

Notes: 1) Confusion matrix is based on the testing sub-sample. 2) The national basic needs poverty line for Tanzania in 2018 was TZS 49,320 per adult per month ($2.16 in 2011 USD PPP-adjusted per capita per day) and the food poverty line was TZS 33,748 per adult per month ($1.48 in 2011 USD PPP-adjusted per capita per day).

FGT() equation is given in Eq. 2.1, $\alpha \in \{0, 1, 2\}$, s is enumeration area. We estimate EA-level poverty prevalence $P_{0,s}$, poverty gap $P_{1,s}$ and squared poverty gap $P_{2,s}$, and draw maps to illustrate the spatial distribution of EA-level FGT measures of RF model prediction in Tanzania. In Figure 6.4-6.6, each point indicates an EA center and the color of the points indicates the poverty preva-

Table 6.6: Diagnostic statistics of country-specific estimation in eastern Africa

(a) Tanzania

Model	R-squared	RMSE	Accuracy Rate	Rank Correlation Coefficient
OLS (1st order approx.)	0.172	3.605	0.607	0.740
OLS (2nd order approx.)	-4.028	8.880	0.615	0.731
LASSO	0.190	3.565	0.605	0.737
Random Forest	0.503	2.793	0.619	0.751

(b) Uganda

Model	R-squared	RMSE	Accuracy Rate	Rank Correlation Coefficient
OLS (1st order approx.)	0.331	2.378	0.583	0.693
OLS (2nd order approx.)	-1.503	4.601	0.594	0.528
LASSO	0.330	2.379	0.592	0.696
Random Forest	0.431	2.193	0.653	0.717

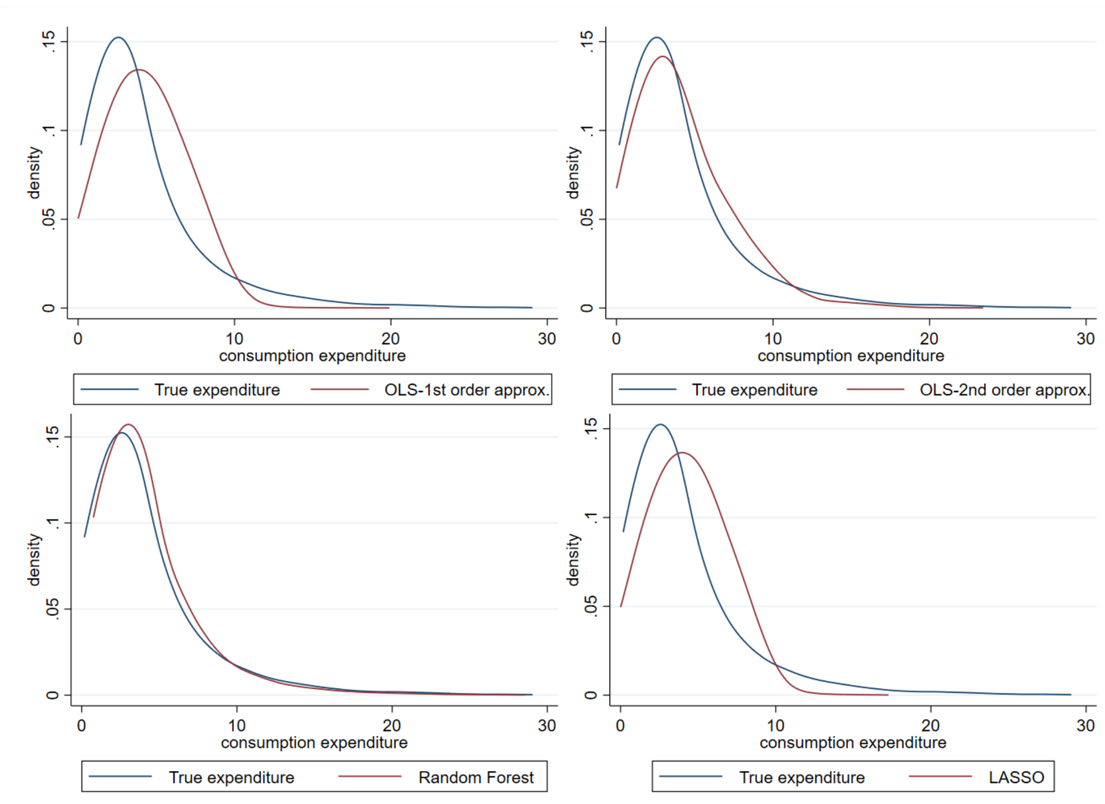
(c) Malawi

Model	R-squared	RMSE	Accuracy Rate	Rank Correlation Coefficient
OLS (1st order approx.)	0.048	14.500	0.558	0.614
OLS (2nd order approx.)	-0.210	16.349	0.556	0.445
LASSO	0.051	14.481	0.556	0.640
Random Forest	0.011	14.784	0.624	0.717

Notes: 1) Diagnostic statistics are based on the testing sub-sample. 2) R-squared is calculated as $1 - RSS_e/RSS_y$, where the numerator is calculated out of sample and therefore is not bounded above by RSS_y. 3) RMSE is most informative compared to the mean values of the dependent variable, which are 4.038, 2.280, 3.059 in Tanzania, Uganda, and Malawi, respectively. 4) The accuracy rate is defined as the sum of diagonal elements of confusion matrix divided by the sum of elements in the whole matrix. 5) We do not include quadratic or interaction terms as candidate variables in LASSO model.

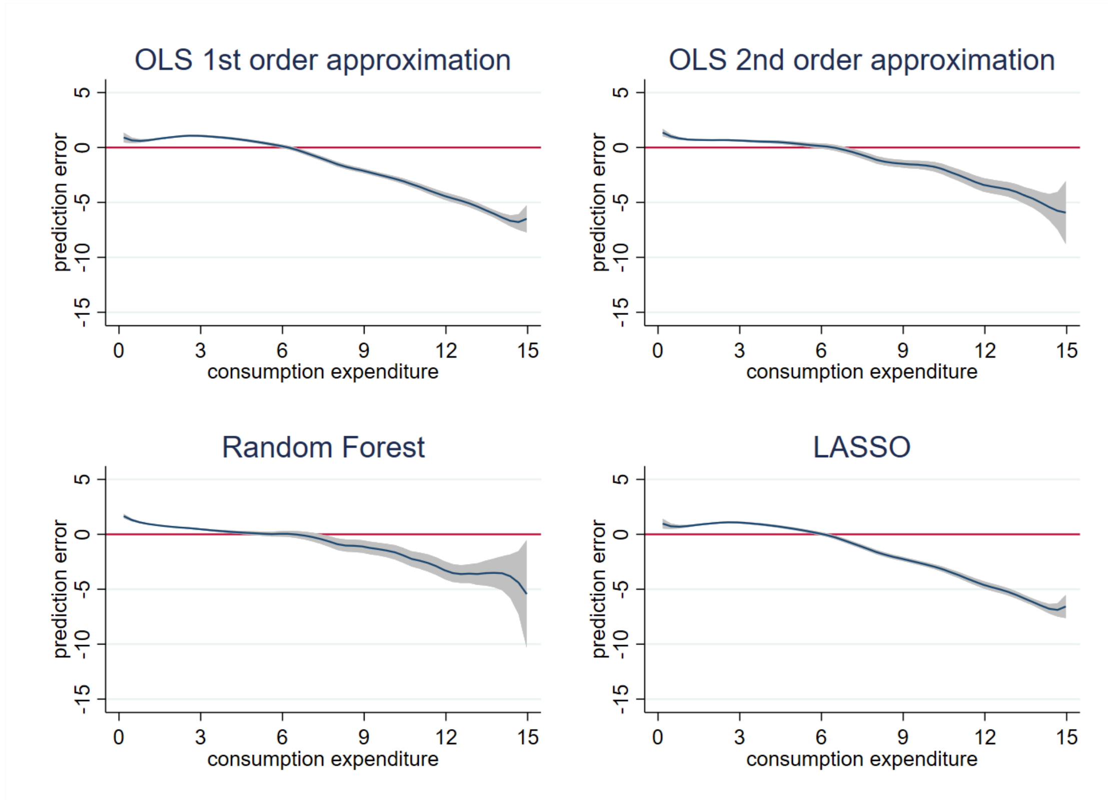
lence/poverty gap/squared poverty gap in this EA based on RF prediction. Additionally, we can find a clear spatial correlation among EA-level FGT measures. For example, in the northwest of Tanzania, the poverty prevalence is systematically higher than in the eastern region.

Figure 6.1: Kernel density estimation of true and predicted expenditure in Tanzania



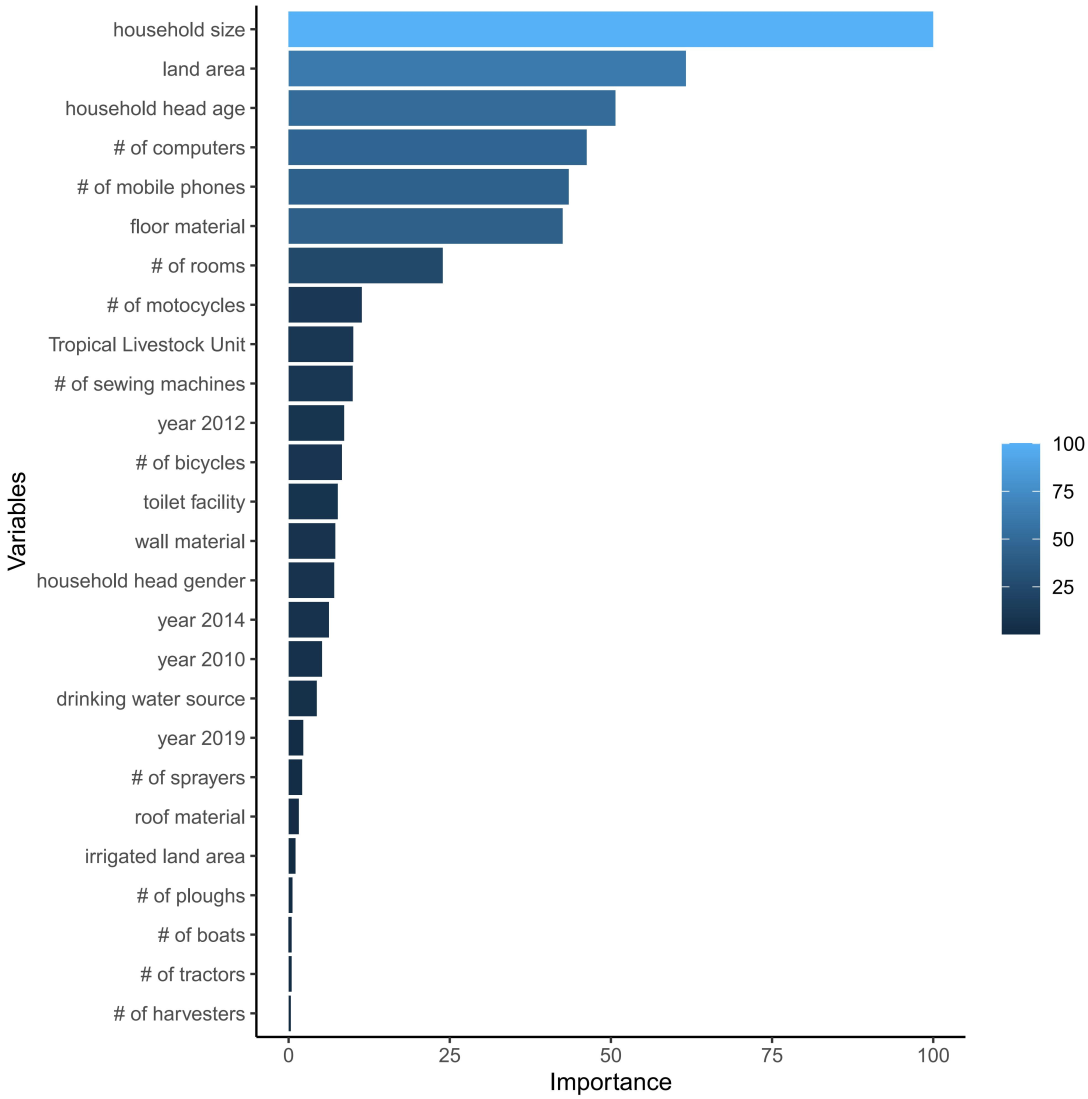
Notes: The x-axis is the actual consumption in Tanzania (2011 USD PPP-adjusted per capita daily expenditure), the y-axis is kernel density.

Figure 6.2: Nonparametric local linear regression in Tanzania



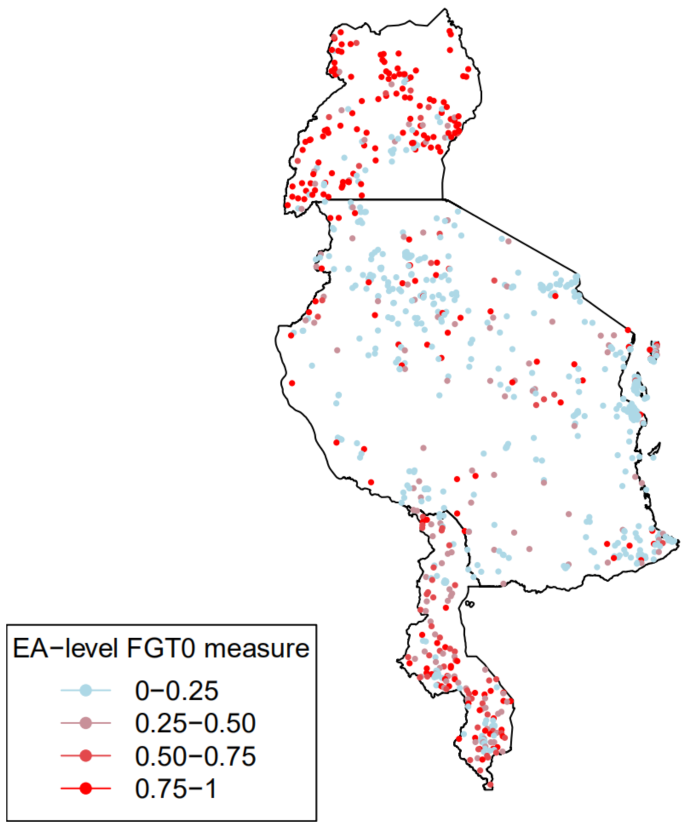
Notes: The x-axis is the actual consumption in Tanzania (2011 USD PPP-adjusted per capita daily expenditure), the y-axis is the prediction error in machine learning models $\hat{E}_i - E_i$, the grey band is the 95% confidence interval. We use Epanechnikov kernel with 0.6 bandwidth.

Figure 6.3: Variable importance of Random Forest in Tanzania



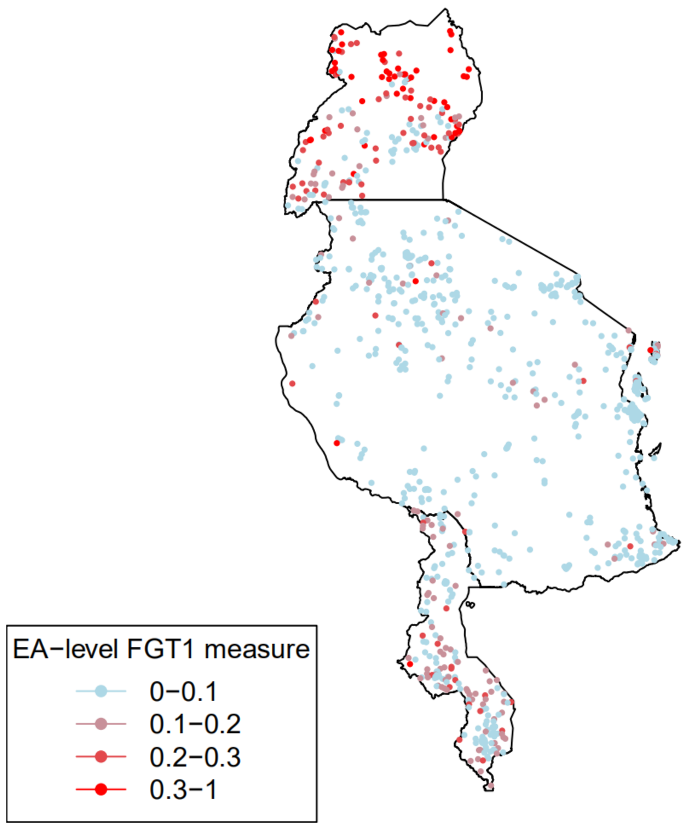
Notes: Variable importance is measured by total decrease in node impurities from splitting on the variable, averaged over all trees, and be scaled to [0,100] range.

Figure 6.4: EA-level FGT0 measure based on RF model in Tanzania, Uganda, and Malawi



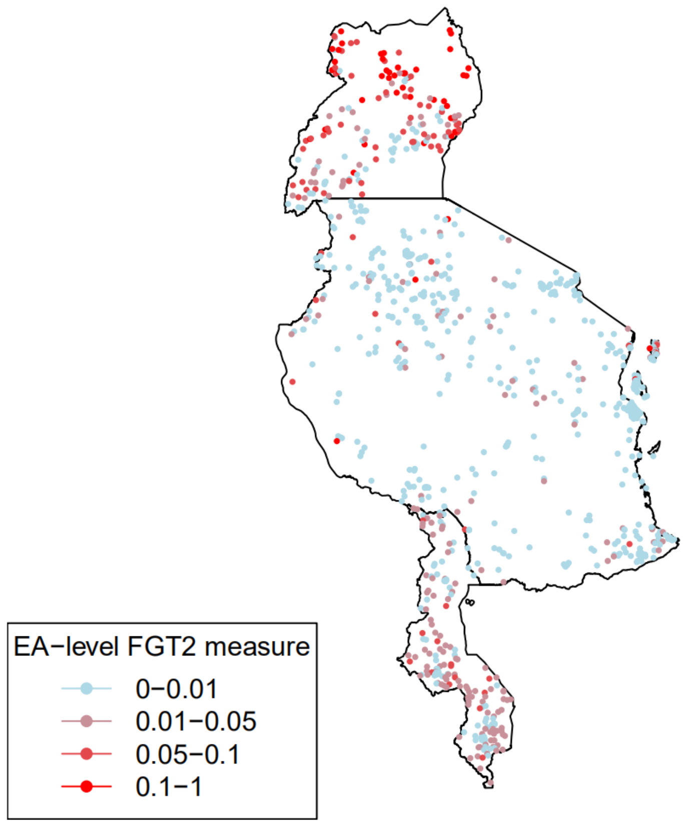
Notes: Each point indicates an enumeration area (EA) center. The EA-level FGT0 measure is calculated based on the predicted consumption expenditure in the testing sub-sample.

Figure 6.5: EA-level FGT1 measure based on RF model in Tanzania, Uganda, and Malawi



Notes: Each point indicates an enumeration area (EA) center. The EA-level FGT1 measure is calculated based on the predicted consumption expenditure in the testing sub-sample.

Figure 6.6: EA-level FGT2 measure based on RF model in Tanzania, Uganda, and Malawi



Notes: Each point indicates an enumeration area (EA) center. The EA-level FGT2 measure is calculated based on the predicted consumption expenditure in the testing sub-sample.

6.2 Pooled estimation

Models trained on pooled consumption or asset observations across all three countries (hereafter“pooled model”) perform similarly, with predictions preserving over 75% of the welfare ordering of household observations and 60% or greater accuracy rate (Table 6.7). The general patterns of diagnostic statistics on predictive skill, especially accuracy rate and rank correlation coefficient, are consistent between the nation-specific models and the pooled model. Compared with the benchmark OLS model, the random forest gives us the highest R^2, accuracy rate, rank correlation coefficient and lowest RMSE.

Comparing diagnostic statistics of nation-specific (Table 6.6) and pooled model (Table 6.7), we find that pooled model gets at the prospective trade-off between adding information in adjacent countries and loss of country-specificity. Intuitively, pooling data in different countries will compromise the model performance due to loss of country-specificity. For example, R-squared in country-specific model in Tanzania is 0.503, but it decreases to 0.033 in the pooled estimation. But we have a striking result that the pooled model gets better accuracy and higher rank correlation coefficients out of sample compared to country-specific model. That is not surprising because there exists significant spatial correlation of contiguous regions across country borders (Figure 6.4-6.6) and we make use of information from contiguous countries; therefore add prediction precision in the pooled estimation. Another possibility is that the inclusion of country dummy variables picks up nationwide differences (e.g., in social protection policies) and the larger sample size increases the precision of key coefficient estimates.

To reflect patterns of the continuous poverty measure, we also use nonparametric kernel regression to assess model performance across the distribution of actual consumption per capita per day. The local linear regression results, with Epanechnikov kernel and bandwidth 0.8, are given in Figures 6.7, where the x-axis is the actual consumption expenditure (2011 USD PPP-adjusted per capita daily expenditure), the y-axis indicates the prediction error ($\hat{y} - y$), and the grey band is the 95% confidence interval. The confidence interval should include zero line under unbiased prediction. Consistent with both the confusion matrix and our intuition, the pooled model is not as good as the nation-specific model. In our model set, the random forest provides researchers the best predictive model.

Table 6.7: Diagnostic statistics of pooled multi-country estimation

Model	R-squared	RMSE	Accuracy Rate	Rank Correlation Coefficient
OLS (1st order approx.)	0.019	18.828	0.583	0.746
OLS (2nd order approx.)	0.036	18.665	0.623	0.733
LASSO	0.018	18.831	0.574	0.734
Random Forest	0.033	18.686	0.640	0.778

Notes: 1) Diagnostic statistics are based on the testing sub-sample. 2) R-squared is calculated as $1 - \text{RSS}_e/\text{RSS}_y$. 3) RMSE is most informative compared to the mean value of the dependent variable (\$3.518/capita/day). 4) The accuracy rate is defined as the sum of diagonal elements of confusion matrix divided by the sum of elements in the whole matrix. 5) We do not include quadratic or interaction terms as candidate variables in LASSO model.

Table 6.8: Confusion matrix of pooled multi-country estimation

(a) OLS (1st order approx.)

Predicted \ Actual	<\$1.90 a day	\$1.90-\$3.20 a day	>\$3.20 a day
<\$1.90 a day	**0.22**	0.04	0.01
\$1.9-\$3.2 a day	0.14	**0.09**	0.03
>\$3.2 a day	0.07	0.13	**0.27**

(b) OLS (2nd order approx.)

Predicted \ Actual	<\$1.90 a day	\$1.90-\$3.20 a day	>\$3.20 a day
<\$1.90 a day	**0.27**	0.07	0.01
\$1.9-\$3.2 a day	0.12	**0.10**	0.05
>\$3.2 a day	0.04	0.09	**0.25**

(c) LASSO

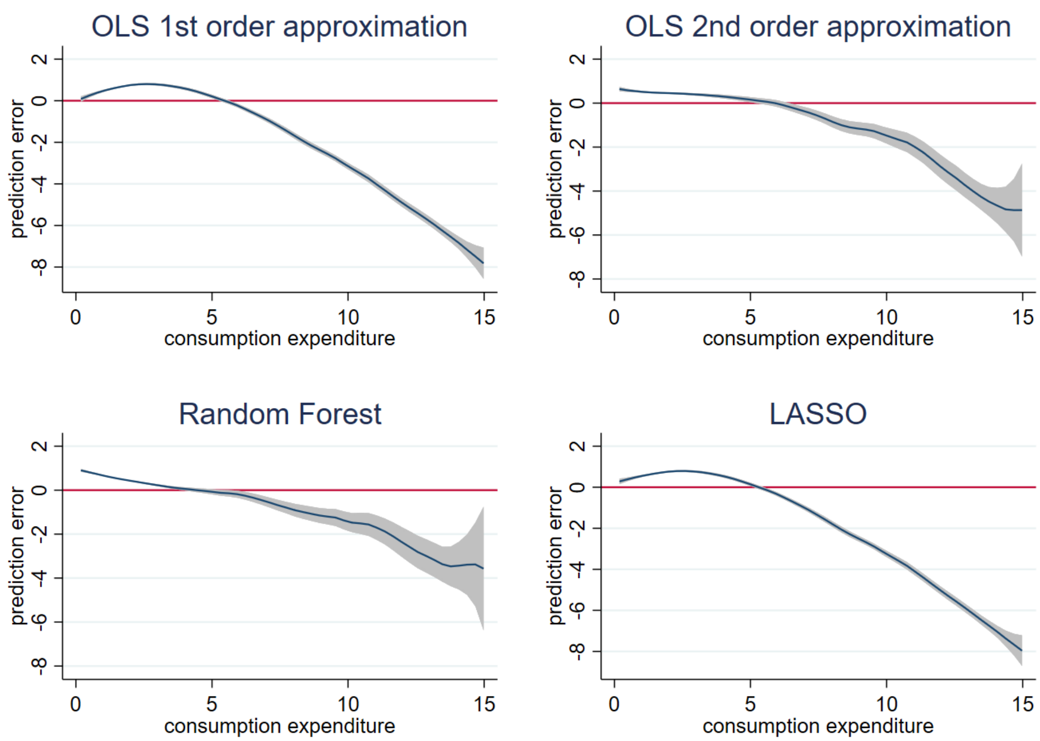
Predicted \ Actual	<\$1.90 a day	\$1.90-\$3.20 a day	>\$3.20 a day
<\$1.90 a day	**0.22**	0.04	0.01
\$1.9-\$3.2 a day	0.14	**0.09**	0.03
>\$3.2 a day	0.07	0.13	**0.27**

(d) Random Forest

Predicted \ Actual	<\$1.90 a day	\$1.90-\$3.20 a day	>\$3.20 a day
<\$1.90 a day	**0.27**	0.06	0.01
\$1.9-\$3.2 a day	0.13	**0.13**	0.06
>\$3.2 a day	0.02	0.07	**0.24**

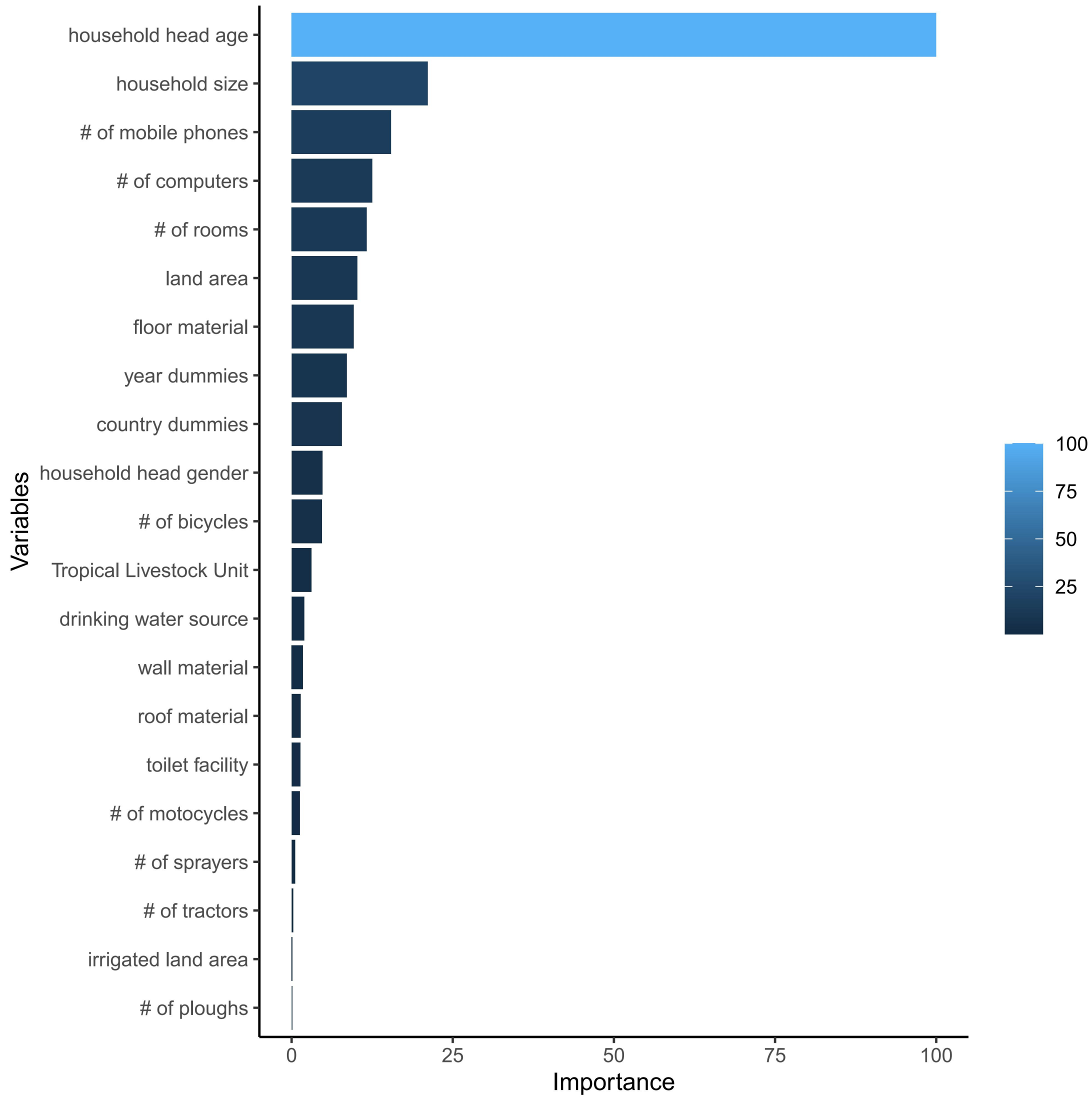
Notes: Confusion matrix is based on the testing sub-sample.

Figure 6.7: Nonparametric local linear regression in pooled model



Notes: The x-axis is the actual consumption in pooled country (2011 USD PPP-adjusted per capita daily expenditure), the y-axis is the prediction error in machine learning models $\hat{E}_i - E_i$, the grey band is the 95% confidence interval. We use Epanechnikov kernel with 0.8 bandwidth.

Figure 6.8: Variable importance of Random Forest in the pooled model



Notes: Variable importance is measured by total decrease in node impurities from splitting on the variable, averaged over all trees, and be scaled to [0,100] range.

Figure 6.9: Compare diagnostic statistics in Tanzania vs pooled model

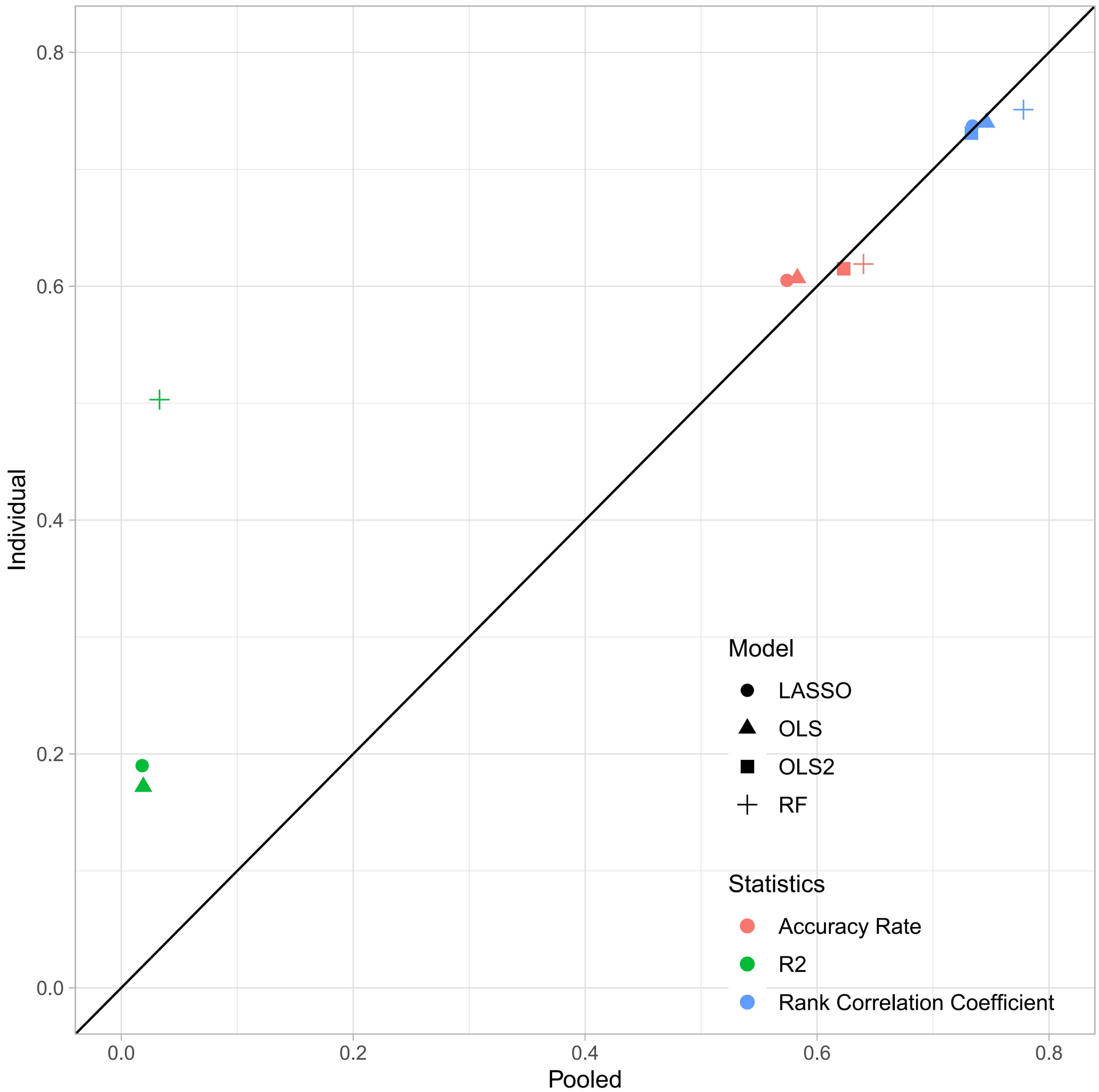


Figure 6.10: Compare diagnostic statistics in Uganda vs pooled model

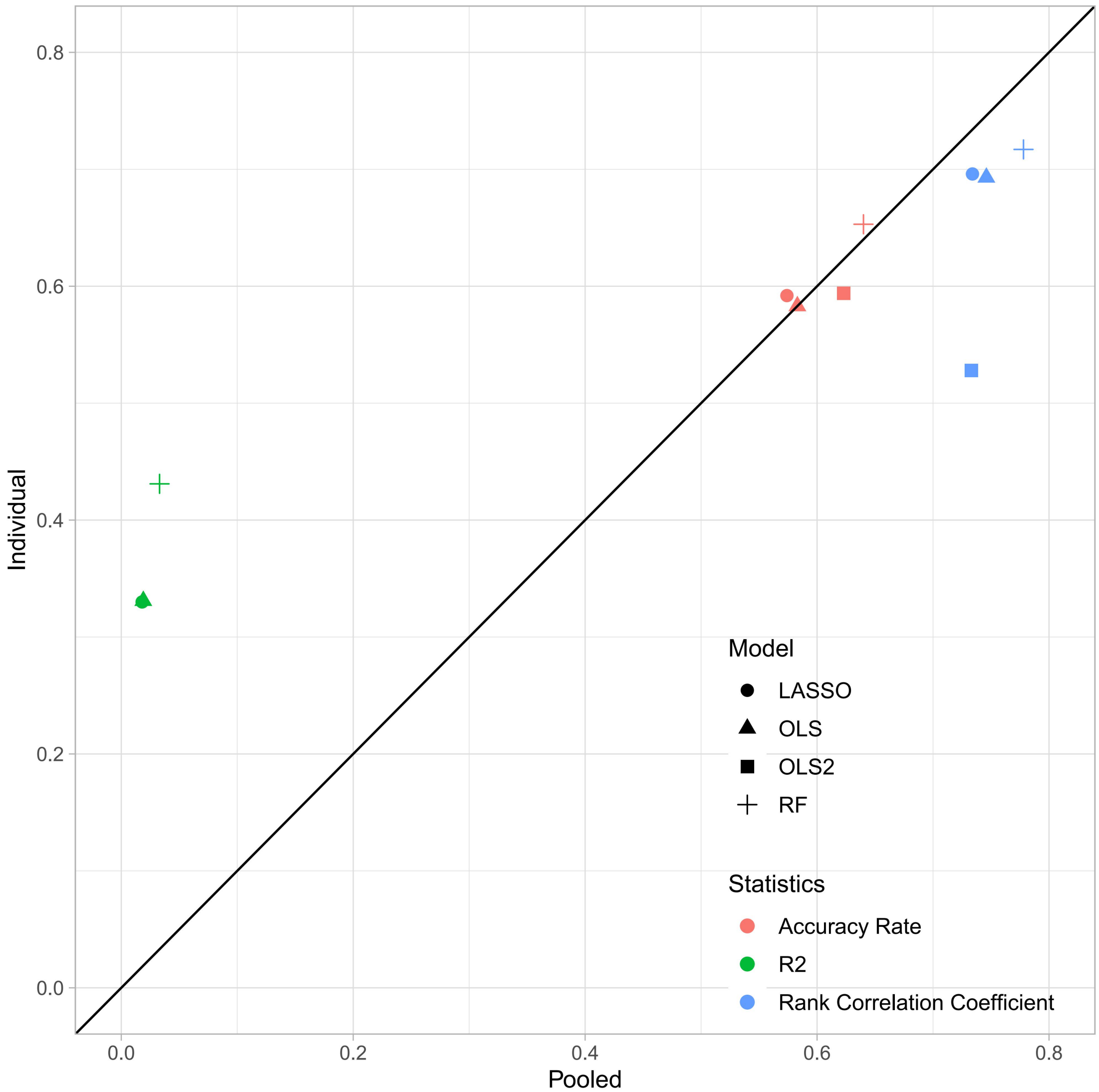
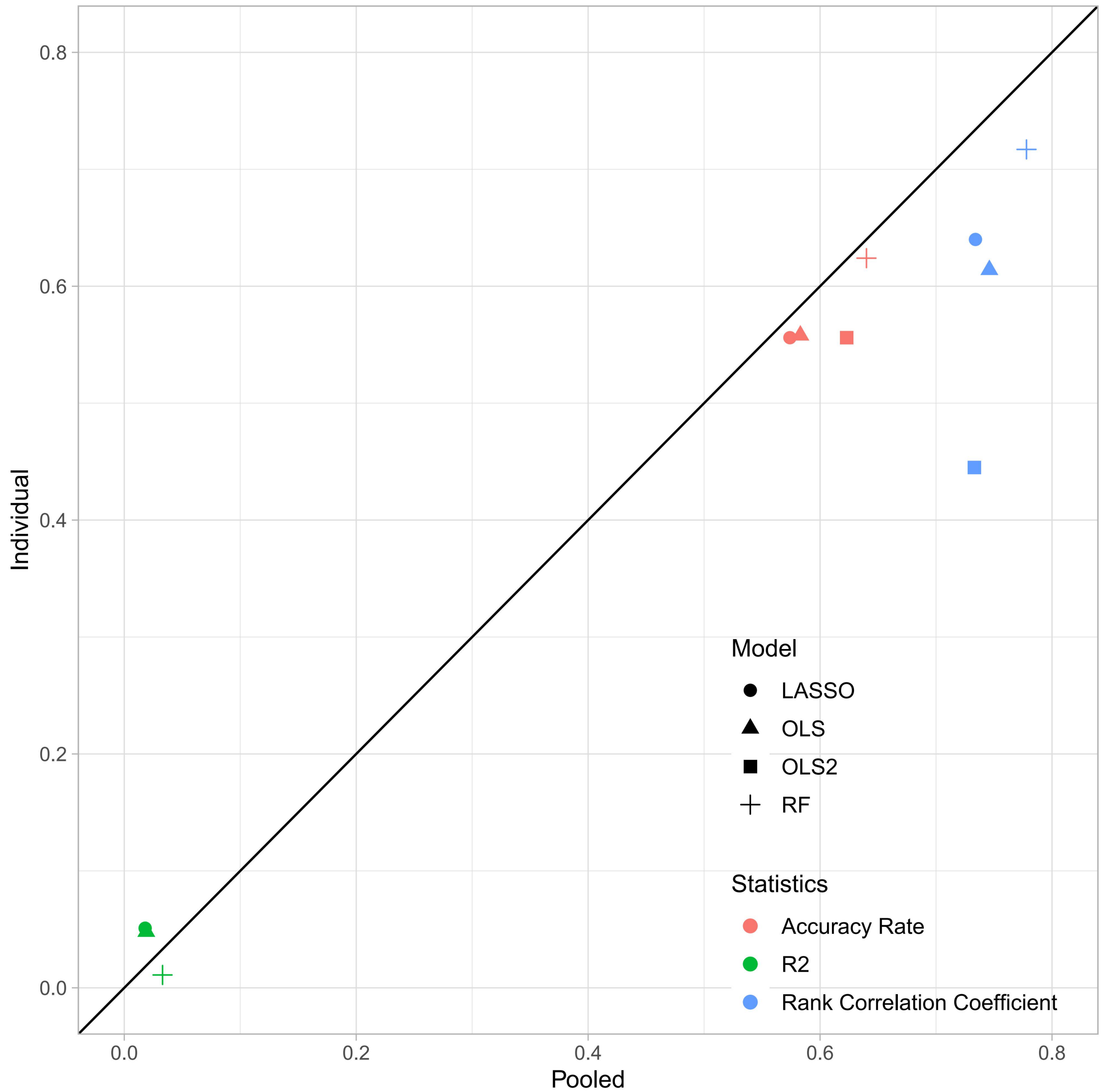


Figure 6.11: Compare diagnostic statistics in Malawi vs pooled model



6.3 Does it matter for ML poverty prediction?

Subsequently, this paper assesses whether the difference in poverty measures matters. We compare the predictive performance of ML methods in poverty prediction, i.e., does the diagnostic statistics (R-squared, RMSE, and rank correlation coefficient) vary from those in ML poverty prediction literature when using the same feature set, same ML algorithms, but a different LHS variable? Following Browne et al. (2021), this paper uses very similar feature sets and ML algorithms, but applies asset-based structural poverty estimates as the LHS variable. Intuitively, we hope the predictive performance is comparable to ML poverty prediction literature using alternative poverty estimates (e.g., DHS RWI). Additionally, we compare the FGT0 poverty measure with FGT1, 2 measures through rank correlation coefficient to see whether targeting based on poverty prevalence estimates would differ from those based on a more distributionally-sensitive poverty gap, or squared poverty gap measures.

In feature set selection, we focus on publicly available data in order to represent feasible, low-cost features for scholars and institutions that are unable to invest much in data collection or procurement. We finalize feature selection based on a range of prior studies (Browne et al., 2021; Elbers et al., 2003; Lang et al., 2013; Christiaensen et al., 2012) to have significant predictive power in explaining geographic patterns of poverty. The settled Earth Observation (EO) features we consider include population, remoteness, night light, geography, vegetation, climate and weather, and time (Table 5.6).

The diagnostic statistics of ML algorithms between asset-based poverty measures and feature set are given in Table 6.8. Fortunately, metrics (R-squared and RMSE) are comparable to ML poverty prediction literature. To be more specific, out-of-sample R-squared (0.49-0.56) and RMSE (0.02-0.23) are close to the existing poverty prediction literature, e.g., $R^2 = 0.31, \text{RMSE} = 0.12$ in Browne et al. (2021). The diagnostic statistics of other ML poverty prediction literature are given in Table 3.1. This promising finding indicates that our asset-based structural poverty estimates can serve as new, improved dependent variables for ML prediction.

Figure 6.12-6.14 illustrate feature importance in ML-based poverty prediction (FGT0, 1, and 2 measures as dependent variable, respectively). Among all the features, population and remoteness have strongest predictive power to poverty prevalence (FGT0) and poverty gap (FGT1), and time dummies are the weakest predictors. Another interesting result, shown in Figure 6.14, is that vegetation and weather variables serve as the highest predictive features in ML prediction of squared poverty gap (FGT2 measure). This may be because the poorest people heavily depend on primary agriculture and forestry, thus weather and vegetation reflect or influence primary productivity and thus earnings.

To assess whether the difference in poverty measures matters and how much predictive skill we sacrifice during the estimation procedure, we generate the rank correlation coefficient matrix across realized poverty, asset-based structural poverty, and feature-based structural poverty estimates (Table 6.10). Realized poverty is calculated based on HH-level true expenditure – which includes structural poverty, transitory shock, and measurement error – and aggregated to EA level. Asset-based structural poverty is generated from predicted HH-level expenditure – which only includes the structural component of poverty – and then aggregated to EA level. Feature-based structural poverty is predicted EA-level poverty measures using EO feature sets. Statistically, Spearman's rank correlation coefficient measures the degree of similarity between two rankings and can be used to assess the significance of the relation between them.

There exist two promising findings in Table 6.10. First, we consider the differences across FGT0, 1, 2 measures within the same step — do these additional metrics $P_{1,s}, P_{2,s}$ provide potentially interesting additional information for targeting poverty? The calculated rank correlation coefficients across $P_{0,s}$, $P_{1,s}$ and $P_{2,s}$ within the same step, highlighted gray in Table 6.10, are always 0.9 or greater, indicating that poverty gap and squared poverty gap largely preserve the welfare ordering of poverty prevalence very well and thus that these measures are good proxies for one another. Second, we assess whether the welfare ordering alters significantly during the ML estimation. The rank correlation coefficients across different stages, highlighted red in Table 6.10, lie between 0.7 to 0.9, indicating FGT poverty measures' ranking keep stable during the ML prediction using productive assets and feature sets. There is necessarily some loss of ordinal correspondence as one moves from poverty realizations to feature-based ML structural prediction. But that degradation seems reasonably modest.

Table 6.9: Diagnostic statistics of ML algorithms between EA-level poverty and feature set

(a) FGT0 poverty measure

Model	R-squared	RMSE	Rank Correlation Coefficient
OLS (1st order approx.)	0.398	0.246	0.678
OLS (2nd order approx.)	0.428	0.240	0.690
LASSO	0.284	0.269	0.572
Random Forest	0.486	0.227	0.720

(b) FGT1 poverty measure

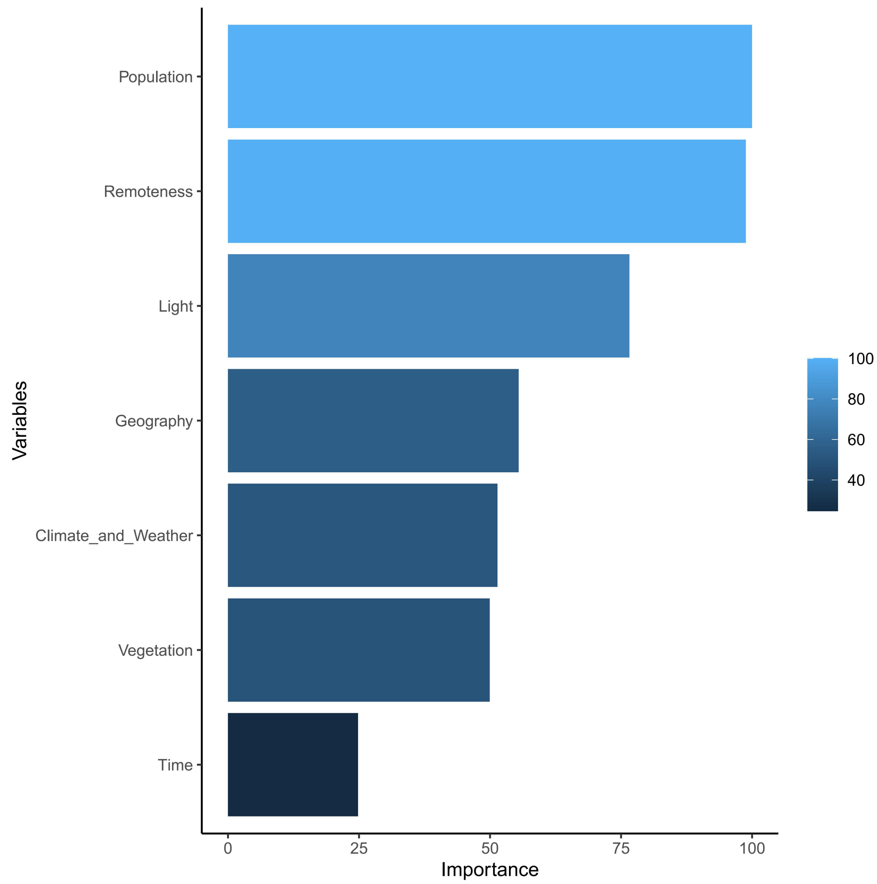
Model	R-squared	RMSE	Rank Correlation Coefficient
OLS (1st order approx.)	0.440	0.056	0.646
OLS (2nd order approx.)	0.466	0.055	0.650
LASSO	0.242	0.065	0.502
Random Forest	0.535	0.051	0.703

(c) FGT2 poverty measure

Model	R-squared	RMSE	Rank Correlation Coefficient
OLS (1st order approx.)	0.461	0.017	0.558
OLS (2nd order approx.)	0.516	0.016	0.571
LASSO	0.207	0.021	0.424
Random Forest	0.563	0.016	0.647

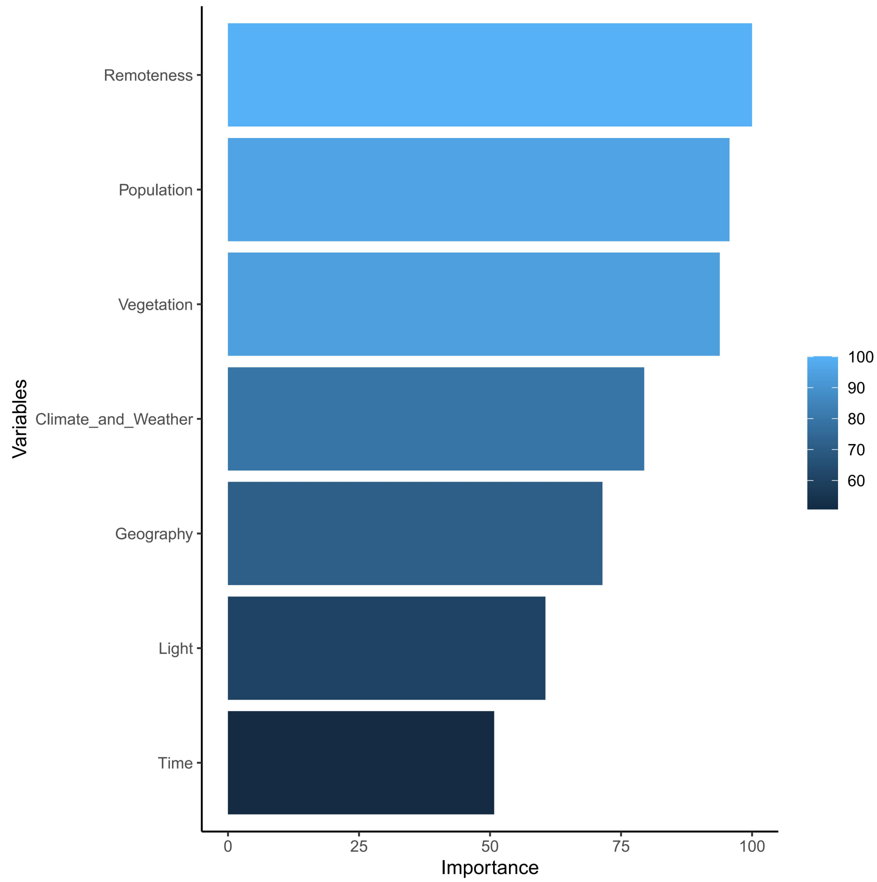
Notes: 1) Diagnostic statistics are based on the testing sub-sample. 2) RMSE is most informative compared to the mean value of the dependent variable. The mean of FGT0, 1, 2 measures are 0.252, 0.044, 0.011, respectively. 3) We do not include quadratic or interaction terms as candidate variables in LASSO model.

Figure 6.12: Feature importance in ML asset-based FGT0 poverty prediction



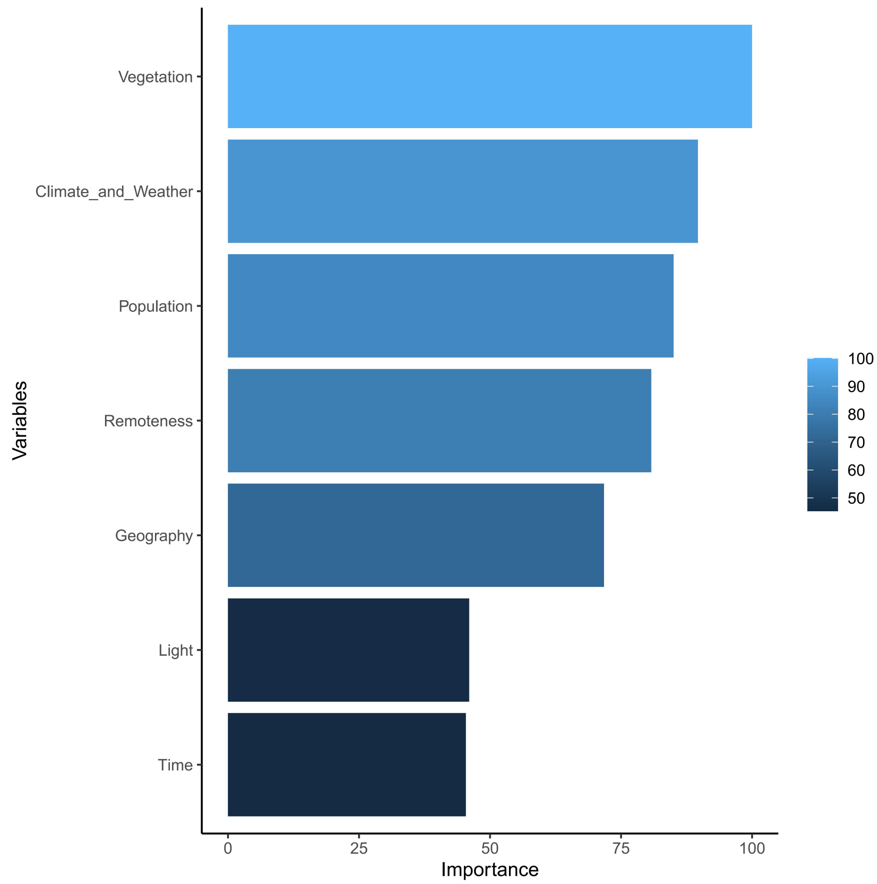
Notes: 1) Population refers to population density. Remoteness represents travel time to nearest city. Light indicates the nighttime lights received by satellites. Geography includes elevation and slope. Climate and weather include rainfall and temperature variables. Vegetation refers to NDVI. Time include all year dummies. 2) Variable importance is measured by total decrease in node impurities from splitting on the variable, averaged over all trees, and be scaled to [0,100] range.

Figure 6.13: Feature importance in ML asset-based FGT1 poverty prediction



Notes: 1) Population refers to population density. Remoteness represents travel time to nearest city. Light indicates the nighttime lights received by satellites. Geography includes elevation and slope. Climate and weather include rainfall and temperature variables. Vegetation refers to NDVI. Time include all year dummies. 2) Variable importance is measured by total decrease in node impurities from splitting on the variable, averaged over all trees, and be scaled to [0,100] range.

Figure 6.14: Feature importance in ML asset-based FGT2 poverty prediction



Notes: 1) Population refers to population density. Remoteness represents travel time to nearest city. Light indicates the nighttime lights received by satellites. Geography includes elevation and slope. Climate and weather include rainfall and temperature variables. Vegetation refers to NDVI. Time include all year dummies. 2) Variable importance is measured by total decrease in node impurities from splitting on the variable, averaged over all trees, and be scaled to [0,100] range.

Table 6.10: Difference between realized poverty, asset-based structural poverty, and feature-based structural poverty (measured by rank correlation coefficient)

		realized poverty			asset-based poverty			feature-based poverty		
		P_0	P_1	P_2	P_0	P_1	P_2	P_0	P_1	P_2
realized poverty	P_0	1	0.953	0.921	0.716	0.712	0.703	0.726	0.704	0.665
	P_1	-	1	0.992	0.747	0.755	0.749	0.745	0.726	0.685
	P_2	-	-	1	0.744	0.757	0.754	0.737	0.719	0.677
asset-based poverty	P_0	-	-	-	1	0.971	0.949	0.862	0.821	0.757
	P_1	-	-	-	-	1	0.995	0.840	0.837	0.788
	P_2	-	-	-	-	-	1	0.821	0.829	0.789
feature-based poverty	P_0	-	-	-	-	-	-	1	0.941	0.897
	P_1	-	-	-	-	-	-	-	1	0.949
	P_2	-	-	-	-	-	-	-	-	1

Notes: 1) Realized poverty is calculated based on HH-level true expenditure and aggregated to EA level. Asset-based structural poverty is generated from predicted HH-level expenditure and then aggregated to EA level. Feature-based structural poverty is predicted EA-level poverty measurements using feature sets. 2) P_0, P_1, P_2 indicate poverty prevalence, poverty gap, and squared poverty gap in the FGT framework respectively.

CHAPTER 7

CONCLUSION

The emergence of high-resolution, high frequency, low-cost Earth Observation data and powerful ML algorithms creates new opportunities to generate timely, spatially precise poverty measures to inform development and humanitarian programming in LMICs. But the usefulness of such measures depends fundamentally on the quality of the poverty measures used. This paper develops an approach to estimate internationally-and-intertemporally comparable, decomposable, structural asset poverty measures for a multi-country region in east Africa. These measures are founded in theory, link directly to official poverty lines, and are amenable to ML-based prediction using Earth Observation data.

The poverty measures we construct can serve as new, improved dependent variables and are well-suited to train ML models to predict where the world's poorest people live, and thereby enhance the usefulness of ML-based advances in poverty estimation for both science and society. Our random forest strategy shows the strong power that machine learning tools can be productively employed in the poverty prediction setting even when data on key outcomes of interest are scarce. Extending from country-specific models to a pooled multi-country model, we show that our model's predictive power declines only slightly or keeps stable. Despite differences in economic and socioeconomic culture across countries, productive assets appear to identify fundamental common structures that are sufficiently strong to explain intra-national as well as cross-national variation, suggesting that our approach could be used to fill in the large data gaps resulting from scarcity of representative national surveys in the poorest communities.

As with the majority of studies concerning comparable wealth/poverty measures, there exist several limitations for asset-based poverty estimates we construct. First, the productive assets are chosen manually and not all surveys include selected assets information. Selecting productive assets common to all nationally-representative household surveys indicates discarding much useful asset information collected in some but not all surveys. The information of common assets we consider are not necessarily available in other surveys or not perfectly applicable in other regions. Second, whether the asset-based approach to estimate poverty is fully generalizable is still unanswered. Intuitively, the asset-expenditure relationship can possibly vary across different regions and rural/urban division. Despite the promising initial results in east African countries, it is unsettled whether the approach can be generalized to other developing economies and the global estimates.

Our approach can have broad future applications in mapping, targeting, and monitoring poverty and malnutrition. Our ambition is to be able to project trained ML algorithms into spaces that have not been sampled. The future direction is to extend the region of interest from east Africa to other developing countries. Another possible extension is from agrarian regions to city context but we have to rethink the selection of productive assets carefully. The longer-term goal is to generate global estimates of socioeconomic outcomes of interest, for example, global estimates of poverty, malnutrition, value of ecosystem services or of exposure to natural hazards. Indeed, machine learning efforts towards poverty prediction are far from complete and could have promising future extensions across many scientific domains.

APPENDIX A

LSMS CONSUMPTION AGGREGATES

To check how consumption aggregates are constructed in each LSMS surveys, this appendix collects information about components of consumption, variable demarcation, expenditures not included in consumption aggregate, treatment on value of auto-provided services and comparability across different surveys.

A.1 Uganda (2011, 2013)

Consumption expenditure in LSMS Uganda 2011, 2013 are constructed similarly, both of which are separated into four parts: (1) Food, beverages, and tobacco; (2) Non-durable goods and frequently purchased services: rent of rented house/fuel/power, non-durable and personal goods, transport and communication, health and medical care, other services; (3) Semi-durable and durable goods and services: clothing and footwear, furniture, carpet and furnishing, household appliances and equipment, glass/table/utensils, education, others; (4) Non-consumption expenditure: income tax, property rates, user fees and charges, local service tax, pension and social security payments, remittances, gifts and other transfers, funerals and other social functions, interest on loans, others.

Variable demarcation: The key variable recorded in LSMS Uganda data set is consumption rather than monetary expenditures. Although the two are very close, they are not the same. Household consumption expenditures in cash, kind or through barter were recorded for the household only. For bartered items the value of the item paid for (not the value one got in exchange) was recorded. Food, beverages, or tobacco served to other members and guests in the household during the reference period were however included. The respondent for this section was the person (household member) who managed the household budget and was the best informed about the household's consumption expenditure.

Value of auto-provided services: In Part 2 (Non-durable goods and frequently purchased services), the survey collected information of rent of rented house, imputed rent of owned house, imputed rent of free house, maintenance, and repair expenses.

Comparability: The poverty estimates for Uganda surveys were derived following the methods applied to earlier surveys presented in Appleton (2001a, b). Thus, consumption and welfare measures are comparable across these surveys. Variable name: cpexp30 (Monthly HH Expenditure in 05/06 prices, Spatially/Temporally Adjusted in 11/12), equiv (Household Adult Equivalence Scale), welfare (cpexp30 adjusted by equiv-consumption aggregate) Note: Differentiate monthly versus annual consumption (need to be adjusted)

A.2 Tanzania (2008, 2010, 2012, 2014, 2019)

Creating the consumption aggregate in Tanzania is guided by theoretical and practical considerations. First, it must be as comprehensive as possible given the available information. Omitting some components assumes that they do not contribute to people's welfare or that they do not affect the ranking of the population. Second, market and non-market transactions are to be included, which means that purchases are not the sole component of the indicator. Third, expenditure is not consumption. For perishable goods, mostly food, it is usual to assume that all purchases are consumed. However, for other goods and services, such as housing or durable goods, corrections have to be made. Fourth, a common reference period should be chosen. Typically, each consumption module in a survey has a different reference period, for instance, education could refer to the last 12 months, food could refer to the last week, and health could refer to the last month. Following common practice in Tanzania, consumption will be reported per 28 days.

Consumption expenditure in LSMS Tanzania is separated into two parts: (1) Food expenditure: cereals, sugar, pulses, nuts and seeds, vegetables, fruits, meat and fish, milk, oil and fats, spices, and beverages; (2) Non-food expenditure: utilities (water, kerosene, lighting), furnishing and household expenses, health, transportation, communications, recreation, education etc.

Expenditure not included in consumption aggregate: The survey assumes that the consumption of non-food goods and services coming from own production, from gifts or from other sources is negligible and can be ignored. The information about some non-food goods and services needs to be excluded from the consumption aggregate because those items are not consumption. Payments of mortgages or debts are financial transactions and not consumption. Losses to theft are neither expenditure nor consumption. Remittances to other households are expenditures but not consumption. Expenditures on marriages, dowries, births and funerals are consumption but given their sporadic nature and the fact that the reported amounts are typically rather large, this consumption is left out to avoid overestimating the true level of welfare of the household. In terms of durable goods, the NPS only provides data on the number of durable goods owned by the household. Calculating this consumption component would have involved making assumptions about their age, their current value and their lifespan, which might have resulted in an extremely imprecise estimation. As a result, durable goods are excluded from the consumption aggregate.

Value of auto-provided services: In most developing countries, limited or non-existent housing rental markets pose a difficult challenge for the estimation and inclusion of this component in the consumption aggregate. As in the case of durable goods, the objective is to measure the flow of services received by the household from occupying its dwelling. When a household lives in a rented dwelling, and provided that rental markets function well, that value would be the actual rent paid. If enough families rent dwellings, imputations can be made for those families that own their dwelling. It is common to include a question for homeowners asking them to provide the hypothetical rent they would pay for renting their dwelling. These self-reported rents can in principle be used to value the consumption the household gets from occupying its dwelling, but these amounts are not always credible or usable, particularly in rural areas where very few households rent. If imputed rents cannot be estimated, actual rents must be excluded from the consumption aggregate for the sake of consistency. The NPS does not collect information on imputed rents and given that the number of households living in rented dwellings is fairly small, housing rental component was excluded from the consumption aggregate.

Comparability: The methodology used for the NPS2014/2015 is identical to the methodology used in the previous rounds (2008/2009, 2010/2011, 2012/2013) so that the aggregates in different waves are comparable over time.

Variable name: expm (total consumption, annual, nomimal), expmR (total consumption, annual, real), hhsize, aduleq (adult-equivalents in the household).Note that nominal consumption in each round of the NPS was adjusted for temporal and spatial price differences, thus real consumption is expressed in Tanzanian prices.

A.3 Malawi (2010, 2016)

Creating the consumption aggregate in Malawi is guided by theoretical and practical considerations. First, it must be as comprehensive as possible given the available information. Omitting some components assumes that they do not contribute to people's welfare or that they do not affect the rankings of individuals. Second, market and non-market transactions are to be included, which means that purchases are not the sole component of the indicator. Third, for perishable goods, mostly food, it is usual to assume that all purchases are consumed. But for other goods and services, such as housing or durable goods, corrections have to be made. Fourth, a common reference period should be chosen. Each consumption module in the survey has a different reference period, for instance, for education it is the last 12 months, for food it is the last week and for clothing it is the last three months. All components were converted into annual figures, thus consumption is reported per year.

Consumption expenditures in LSMS Malawi 2010, 2016 fall into four categories: (1) Food; (2) Non-food, non-consumer durables; (3) Consumer durable goods; (4) Imputed rental cost of housing.

Expenditure not included in consumption aggregate: Payments of mortgages or debts are financial transactions and not consumption. Losses to theft are neither expenditure nor consumption. Remittances to other households are expenditures but not consumption. Expenditures on marriages, dowries, births and funerals are consumption but given their sporadic nature and the fact that the reported amounts are typically rather large, this consumption is left out to avoid overestimating the true level of welfare of the household. Repairs to the dwelling and construction materials are excluded because the housing component of the consumption aggregate already takes into account any improvement to the dwelling.

Value of auto-provided services: In Malawi, the imputed housing rent is collected and included in consumption aggregate.

Variable name: rexpagg (total annual HH expenditure) (2004), pcrexpagg (total real annual consumption per capita - 2013 Prices, spatially & temporally adjusted) Source: IHS2 expagg note.pdf (2004), IHS3_report Appendix B (Page 226) (2010)

A.4 Summary

The aim of this appendix is to ensure the regressions we are about to run use aggregates that are created using consistent protocols. The key things it establishes is that auto-provided food is included, housing is treated in an internally consistent manner (i.e., either impute rents for those who own the residence they occupy and include it in consumption expenditures, as in Uganda, or exclude cash rental payments, as in Tanzania), and business expenses (e.g., spending on fertilizer or business or farm equipment) and debt repayment are excluded even though they involve monetary outlays. When we start pooling countries to estimate the consumption expenditures – asset relationship, please recognize that differences in how housing expenditures are handled will create a difference between countries automatically, hence the need for country dummy variables.

APPENDIX B

DATA CLEANING PROCESS

This appendix explains the cleaning protocol in detail. A reader should be able to fully reproduce results in this paper from raw LSMS data following the steps below.

Step 1: Select suitable data sets and rounds

- Choose LSMS data rounds including consumption expenditure and productive assets

Step 2: Process dependent variables

- Check how consumption aggregates are constructed in each survey

- Generate PPP-adjusted consumption expenditure

Step 3: Process independent variables

- Select region-specific productive assets

- Recode categorical variables to dummies

- Generate Tropical Livestock Units (TLU)

Step 4: Merge panel datasets

- Note that the panel households in Tanzania are identified and linked through upd4_hh_a.dta file in National Panel Survey 2008-2015, Uniform Panel Dataset. The upd4_hh_a.dta serves as the master file providing HHIDs of the same household in different waves.

Step 5: Fit ML models

- Preprocess: Data partitioning & Features demeaning

- Fit OLS/machine learning models

- Generate evaluation metrics (R-squared, RMSE, accuracy rate, Rank Correlation Coefficient) and confusion matrix

- Nonparametric regression of $\hat{E}_i - E_i$ versus E_i to test for prediction error

APPENDIX C

REGRESSION RESULTS IN THE TRAINING SUB-SAMPLE

Table C.1: OLS regression results in Tanzania

	Coef.	Std. Error	t value	Pr(>|t|)
Intercept	4.15	0.09	48.34	0.00
hhsize	-0.52	0.01	-39.66	0.00
TLU	0.03	0.00	7.69	0.00
land	0.01	0.00	2.66	0.01
land_irrigated	0.00	0.02	0.21	0.83
bicycle	-0.19	0.05	-4.16	0.00
motocycle	0.70	0.13	5.39	0.00
boat	0.32	0.24	1.34	0.18
mobile_phone	0.84	0.03	26.19	0.00
computer	0.74	0.06	13.22	0.00
plough	0.59	0.10	5.63	0.00
tractor	0.46	0.74	0.63	0.53
harvester	3.92	1.35	2.89	0.00
sprayer	0.01	0.14	0.06	0.95
sewing_machine	-0.06	0.09	-0.71	0.48
rooms	-0.02	0.03	-0.92	0.36
roof_material	0.21	0.09	2.25	0.02
wall_material	0.66	0.08	8.10	0.00
floor_material	1.21	0.10	12.67	0.00
drinking_water_source	0.17	0.08	2.08	0.04
toilet_facility	1.05	0.09	11.59	0.00
hh_head_gender	-0.00	0.08	-0.06	0.95
hh_head_age	-0.02	0.00	-8.56	0.00
year2010	0.14	0.11	1.21	0.23
year2012	0.27	0.11	2.43	0.01
year2014	-0.35	0.12	-2.96	0.00
year2019	-1.22	0.15	-8.14	0.00

Notes: Results come from the OLS regression (1st order approximation) in the training set for the country-specific model in Tanzania.

Table C.2: OLS regression results in Malawi

	Coef.	Std. Error	t value	Pr(>|t|)
Intercept	2.79	0.34	8.19	0.00
TLU	-0.06	0.08	-0.71	0.48
land	-0.12	0.11	-1.07	0.28
land_irrigated	0.28	1.31	0.22	0.83
sewing_machine	-1.21	0.68	-1.77	0.08
bicycle	0.25	0.24	1.05	0.29
motorcycle	0.19	1.12	0.17	0.86
car	15.47	0.71	21.72	0.00
computer	3.11	0.64	4.86	0.00
mobile_phone	-0.16	0.18	-0.89	0.37
sprayer	0.07	0.88	0.08	0.94
plough	0.17	1.42	0.12	0.90
tractor	-22.93	11.61	-1.97	0.05
rooms	0.14	0.16	0.93	0.35
roof_material	0.52	0.43	1.22	0.22
wall_material	0.11	0.37	0.30	0.77
floor_material	1.24	0.45	2.76	0.01
drinking_water_source	0.06	0.44	0.14	0.89
toilet_facility	0.28	0.35	0.80	0.42
hh_head_gender	0.21	0.36	0.58	0.56
hh_head_age	-0.02	0.01	-1.89	0.06
hhsize	-0.50	0.08	-6.25	0.00
year2016	0.16	0.38	0.43	0.67

Notes: Results come from the OLS regression (1st order approximation) in the training set for the country-specific model in Malawi.

Table C.3: OLS regression results in Uganda

	Coef.	Std. Error	t value	Pr(>|t|)
Intercept	2.30	0.09	25.68	0.00
TLU	0.02	0.02	0.87	0.38
land	0.01	0.02	0.46	0.64
land_irrigated	-0.04	0.20	-0.18	0.86
bicycle	-0.02	0.04	-0.46	0.64
motorcycle	-0.09	0.29	-0.33	0.74
boat	-0.09	0.88	-0.10	0.92
mobile_phone	0.29	0.10	3.10	0.00
computer	0.83	0.51	1.62	0.10
internet_access	8.28	1.37	6.02	0.00
ploughs	-0.00	0.22	-0.02	0.98
tractors	0.31	4.92	0.06	0.95
sprayers	-0.00	0.28	-0.01	0.99
rooms	0.37	0.08	4.43	0.00
roof_material	0.26	0.23	1.11	0.27
wall_material	0.19	0.24	0.80	0.42
floor_material	1.17	0.28	4.24	0.00
drinking_water_source	0.05	0.21	0.23	0.82
toilet_facility	0.20	0.22	0.92	0.36
hhsize	-0.33	0.03	-9.84	0.00
hh_head_gender	0.14	0.20	0.71	0.48
hh_head_age	-0.02	0.01	-3.09	0.00
year	0.23	0.11	2.05	0.04

Notes: Results come from the OLS regression (1st order approximation) in the training set for the country-specific model in Uganda.

Table C.4: OLS regression results in the pooled model

	Coef.	Std. Error	t value	Pr(>|t|)
Intercept	3.30	0.24	13.85	0.00
TLU	0.03	0.01	4.02	0.00
land	0.00	0.01	0.55	0.58
land_irrigated	0.01	0.04	0.24	0.81
bicycle	-0.12	0.04	-2.99	0.00
motorcycle	0.33	0.18	1.77	0.08
mobile_phone	0.97	0.04	23.98	0.00
computer	0.24	0.04	6.66	0.00
plough	0.38	0.15	2.56	0.01
tractor	3.46	1.15	3.02	0.00
sprayer	0.07	0.18	0.40	0.69
rooms	0.05	0.04	1.30	0.19
roof_material	0.25	0.11	2.26	0.02
wall_material	0.37	0.10	3.79	0.00
floor_material	1.49	0.11	13.01	0.00
drinking_water_source	0.17	0.10	1.70	0.09
toilet_facility	0.50	0.10	5.10	0.00
hhsize	-0.52	0.02	-30.59	0.00
hh_head_gender	0.12	0.09	1.25	0.21
hh_head_age	-0.01	0.00	-4.71	0.00
year2010	-0.23	0.19	-1.21	0.23
year2011	0.23	0.32	0.73	0.47
year2012	-0.04	0.18	-0.24	0.81
year2013	-0.18	0.29	-0.63	0.53
year2014	-0.67	0.20	-3.39	0.00
year2016	-0.36	0.24	-1.50	0.13
year2019	-1.42	0.26	-5.39	0.00
countryTanzania	0.76	0.19	4.06	0.00

Notes: Results come from the OLS regression (1st order approximation) in the training set for the pooled model.

BIBLIOGRAPHY

Ayush, K., Uzkent, B., Burke, M., Lobell, D., and Ermon, S. (2020). Generating interpretable poverty maps using object detection in satellite images. *arXiv preprint arXiv:2002.01612*.

Bank, W. (2019). Tanzania mainland poverty assessment.

Barrett, C. B. and Carter, M. R. (2013). The economics of poverty traps and persistent poverty: empirical and policy implications. *Journal of Development Studies*, 49(7):976–990.

Barrett, C. B., Garg, T., and McBride, L. (2016). Well-being dynamics and poverty traps. *Annual Review of Resource Economics*, 8:303–327.

Barrett, C. B., Marenya, P. P., McPeak, J., Minten, B., Murithi, F., Oluoch-Kosura, W., Place, F., Randrianarisoa, J. C., Rasambainarivo, J., and Wangila, J. (2006). Welfare dynamics in rural kenya and madagascar. *Journal of Development Studies*, 42(2):248–277.

Barrett, C. B., Reardon, T., and Webb, P. (2001). Nonfarm income diversification and household livelihood strategies in rural africa: concepts, dynamics, and policy implications. *Food Policy*, 26(4):315–331.

Baulch, B. and Hoddinott, J. (2000). Economic mobility and poverty dynamics in developing countries. *Journal of Development Studies*, 36(6):1–24.

Browne, C., Matteson, D. S., McBride, L., Hu, L., Liu, Y., Sun, Y., Wen, J., and Barrett, C. B. (2021). Multivariate random forest prediction of poverty and malnutrition prevalence. *PloS one*, 16(9):e0255519.

Carter, M. R. and Barrett, C. B. (2006). The economics of poverty traps and persistent poverty: An asset-based approach. *Journal of Development Studies*, 42(2):178–199.

Carter, M. R. and May, J. (1999). Poverty, livelihood and class in rural south africa. *World Development*, 27(1):1–20.

Carter, M. R. and May, J. (2001). One kind of freedom: Poverty dynamics in post-apartheid south africa. *World Development*, 29(12):1987–2006.

Chi, G., Fang, H., Chatterjee, S., and Blumenstock, J. E. (2022). Microestimates of wealth for all low-and middle-income countries. *Proceedings of the National Academy of Sciences*, 119(3).

Christiaensen, L., Lanjouw, P., Luoto, J., and Stifel, D. (2012). Small area estimation-based prediction methods to track poverty: validation and applications. *Journal of Economic Inequality*, 10(2):267–297.

Croft, T. N., Marshall, A. M., Allen, C. K., Arnold, F., Assaf, S., Balian, S., et al. (2018). Guide to DHS statistics. *Rockville: ICF*, 645.

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

Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, pages 761–766.

Foster, J., Greer, J., and Thorbecke, E. (2010). The foster–greer–thorbecke (fgt) poverty measures: 25 years later. *Journal of Economic Inequality*, 8(4):491–524.

Friedman, M. (1957). The permanent income hypothesis. In *A theory of the consumption function*, pages 20–37. Princeton University Press.

Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., Romero, B. E., Husak, G. J., Michaelsen, J. C., Verdin, A. P., et al. (2014). A quasi-global precipitation time series for drought monitoring. *US Geological Survey data series*, 832(4):1–12.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Hruschka, D. J., Gerkey, D., and Hadley, C. (2015). Estimating the absolute wealth of households. *Bulletin of the World Health Organization*, 93:483–490.

Jahnke, H. E. and Jahnke, H. E. (1982). *Livestock production systems and livestock development in tropical Africa*, volume 35. Kieler Wissenschaftsverlag Vauk Kiel.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., and Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794.

Lang, C., Barrett, C. B., and Naschold, F. (2013). Targeting maps: An asset-based approach to geographic targeting. *World Development*, 41:232–244.

Lee, K. and Braithwaite, J. (2020). High-resolution poverty maps in sub-saharan africa. *arXiv preprint arXiv:2009.00544*.

Li, X., Zhou, Y., Zhao, M., and Zhao, X. (2020). A harmonized global nighttime light dataset 1992–2018. *Scientific data*, 7(1):1–9.

Lipton, M. and Ravallion, M. (1995). Poverty and policy. in (j. behrman and tn srinivasan, eds.) Handbook of Development Economics.

McBride, L., Barrett, C. B., Browne, C., Hu, L., Liu, Y., Matteson, D. S., Sun, Y., and Wen, J. (2022). Predicting poverty and malnutrition for targeting, mapping, monitoring, and early warning. *Applied Economic Perspectives and Policy*.

Modeling, G. and (GMAO), A. O. (2015). Merra-2 tavg1_2d_slv_nx: 2d, 1-hourly, time-averaged, single-level, assimilation, single-level diagnostics v5. 12.4.

Nasa, J. (2013). Nasa shuttle radar topography mission global 1 arc second. *NASA EOSDIS Land Processes DAAC*, 10.

Naschold, F. and Barrett, C. B. (2011). Do short-term observed income changes overstate structural economic mobility? *Oxford Bulletin of Economics and Statistics*, 73(5):705–717.

Pokhriyal, N. and Jacques, D. C. (2017). Combining disparate data sources for improved poverty prediction and mapping. *Proceedings of the National Academy of Sciences*, 114(46):E9783–E9792.

Ravallion, M. (1992). *Poverty comparisons: A guide to concepts and methods*. The World Bank.

Rothman-Ostrow, P., Gilbert, W., and Rushton, J. (2020). Tropical livestock units: re-evaluating a methodology. *Frontiers in Veterinary Science*, page 973.

Rutstein, S. O. and Johnson, K. (2004). The dhs wealth index. dhs comparative reports no. 6. *Calverton: ORC Macro*.

Rutstein, S. O. and Staveteig, S. (2014). *Making the demographic and health surveys wealth index comparable*, volume 9. ICF International Rockville, MD.

Sen, A. (1976). Poverty: an ordinal approach to measurement. *Econometrica*, pages 219–231.

Sen, A. (1995). *Inequality reexamined*. Harvard University Press.

Smits, J. and Steendijk, R. (2015). The international wealth index (iwi). *Social indicators research*, 122(1):65–85.

Staveteig, S. and Mallick, L. (2014). *Intertemporal comparisons of poverty and wealth with DHS data: A harmonized asset index approach*. ICF International.

Sullivan, J. X., Turner, L., and Danziger, S. (2008). The relationship between income and material hardship. *Journal of Policy Analysis and Management*, 27(1):63–81.

Tang, B., Liu, Y., and Matteson, D. S. (2022). Predicting poverty with vegetation index. *Applied Economic Perspectives and Policy*.

Weiss, D. J., Nelson, A., Gibson, H., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., et al. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553(7688):333–336.

WorldPop and for International Earth Science Information Network (CIESIN), C. (2018). Global high resolution population denominators project.

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., and Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nature Communications*, 11(1):1–11.