RedNIE Red Nacional de Investigadores en Economía

Global Multidimensional Poverty Prediction using World Development Indicators

Rodrigo García Arancibia (Universidad Nacional del Litoral/CONICET)

Ignacio Girela (Universidad Nacional de Córdoba/CONICET)

Daniela Agostina González (Universidad Nacional de Córdoba)

DOCUMENTO DE TRABAJO N° 350

Enero de 2025

Citar como:
García Arancibia, Rodrigo, Ignacio Girela, Daniela Agostina González (2025). Global Multidimensional Poverty Prediction using World Development Indicators. Documento de trabajo RedNIE N°350.

Global Multidimensional Poverty Prediction using World Development Indicators

Rodrigo Garcia Arancibia[1,3,4], Ignacio Girela[2,3],
Daniela Agostina Gonzalez[2]

[1]Instituto de Economia Aplicada Litoral, Facultad de Ciencias Económicas, Universidad Nacional del Litoral.
[2]Facultad de Ciencias Economicas, Universidad Nacional de Cordoba.
[3]Consejo Nacional de Investigaciones Científicas y Técnicas, (CONICET) .
[4]RedNIE .

Abstract

Effective implementation, monitoring, and evaluation of targeted poverty reduction programs require accurate measurements of poverty levels and their changes over time. The Multidimensional Poverty Index (MPI) offers a more comprehensive measure compared to traditional income-based assessments. However, for many countries, MPI data are either unavailable or limited to a few years due to the high cost of conducting relevant surveys. This paper presents alternative methodologies to predict the Global MPI across different countries and time periods using the World Bank's World Development Indicators as predictor variables. Given that MPI construction involves proportions bounded within the unit interval, we tailor statistical learning methods accordingly. In a high-dimensional context, where the number of predictors exceeds the number of training observations, we evaluate methodologies such as dimension reduction, regularized models, and ensemble learning. We conduct cross-validation experiments to assess model performance, incorporating both measured and non-measured countries in the testing dataset.

Keywords: MPI, Beta Regression, Statistical Learning, Data Imputation, Global Poverty Assessment, High-Dimensionality.

JEL Classification: : C52 , C53 , I32 , O10

1. Introduction

The Sustainable Development Agenda for 2030 commits to "Ending Poverty in all its forms". Having comparable poverty measures is essential for monitoring the progress of such a goal. This was the main motivation for the World Bank $1-a-day global poverty line (now adjusted to $2.15-a-day).

However, ending poverty in *all its forms* entails the acknowledgment of the multidimensional nature of poverty. In this context, the global Multidimensional Poverty Index (MPI) has been launched and yearly published by the Oxford Poverty And Human Development Initiative (OPHI) and the United Nations Development Program (UNPD) [1, 2]. This internationally comparable measure was created with the aim of complementing the World Bank monetary approach and tracking changes worldwide in other fundamental human development dimensions, such as health, education and minimum living standards.

The global MPI has been estimated in more than 100 countries of the developing world at least once since 2000 based on nationally representative samples containing information on each dimension. In spite of significant advances in socioeconomic data availability, poverty data is still severely limited by frequency and coverage in comparison to other economic phenomena, such as inflation or trade balance [3]. This is mainly because of the lack of public resources, or even fragility, conflicts, and natural disasters, that characterize the developing world. Even in countries with consistent and periodical multi-topic survey data collection, the time-span between conducting the survey and publishing results can take more than a year. As a consequence, poverty estimates, particularly the global MPI, can be lagged or discontinued in time, thus becoming inaccurate as the current representation of poverty in a country.

This inherent problem in poverty measurement has motivated researchers to explore alternative development indicators, such as the GDP, for estimating poverty [e.g. 4–8]. However, much of the research on poverty prediction using development indicators has remained focused on monetary poverty measures, neglecting multidimensional poverty. Given that MPI measures gain prominence in supplementing income-based measures within poverty reduction strategies [1, 9], estimating a country's global MPI for a specific time period becomes meaningful for tracking progress toward poverty eradication. This information can then inform policy decisions and prioritization efforts at the national level. In order to extend the OPHI-UNPD estimates platform, we explore various methods for predicting global multidimensional poverty using the World Bank's World Development Indicators (WDI).

The present paper is encompassed in a growing literature that applies Machine Learning (ML) methods for predicting aggregate socioeconomic indicators that require an intensive data collection process (namely, poverty, unemployment, inflation, etc.). For instance, Felix et al. [10] compare traditional linear models with eight different ML-based algorithms to predict the size of informal economies in 122 countries from 2004 to 2014. As a result, not only ML models outperform the linear ones but also the determinants of the shadow economies size found in the former were consistent with the traditional linear models. Chakraborty et al. [11] integrate machine learning methods with traditional time series models to predict unemployment rates in developed countries. In their study, they show that ML-based methods are able to capture better linear and nonlinear tendencies present in data than traditional models.

In the work by Mahler et al. [6], various machine learning methods are tested for predicting global poverty rates for both the current and previous year using data from the World Development Indicators (WDI), the World Economic Outlook, and Google Earth Engine. The

authors go beyond directly predicting poverty rates; they also find that real GDP per capita is an exceptionally strong predictor of changes in poverty, with no other variable demonstrating comparable importance for predictive accuracy.

On the other hand, Alkire et al. [12] modeled the multidimensional poverty projections for 75 in order to evaluate which countries are on track to meet the goal of halving poverty incidence in 2030 in which they assert that projections should consider the initial conditions and recent trends of a country. In a rapidly changing world, recent trends can be severely affected by shocks (e.g., COVID-19 pandemic or climate change). Hence, a framework for estimating recent changes in multidimensional poverty can also be highly relevant for monitoring new trends and adjusting mid-term poverty reduction goals.

Aligned with Mahler et al. [6] study, this paper aims to assess a diverse range of methods for predicting global multidimensional poverty within a high-dimensional context for a bounded target variable.

Specifically, we compare machine learning approaches to identify the most effective model for multidimensional poverty, with the dual objectives of i) imputing data across countries and ii) estimating recent trends in poverty changes within individual countries. Our modeling approach encompasses not only various machine learning techniques (such as dimensionality reduction and tree-based models) but also considers the nature of the dependent variable. We tailor predictive statistical learning models to a bounded target represented by a rate or proportion (e.g., poverty rates), where assuming a beta distribution for the conditional response is more appropriate. This is particularly important because models suited for unbounded continuous response variables may yield implausible predictions, such as negative poverty rates, in extreme cases.

To carry out this analysis, we use the World Development Indicators (WDI) database from the World Bank, exploring different data dimensionalities obtained via web scraping. Specifically, we examine the trade-off between maximizing the number of observations with fewer predictor variables and incorporating additional predictor variables at the expense of sample size for model training.

To the best of our knowledge, this is the first paper that examines machine learning methods for predicting global multidimensional poverty. The rest of the paper is structured as follows. Section 2 outlines our methodology, focusing on the unique challenges arising from the bounded nature of our target variables. We provide detailed descriptions of the statistical learning models used for prediction, data exploration and analysis, as well as the experimental and validation strategies employed to assess the performance of various predictive models. In Section 3, we present and analyze the results obtained from our experiments. Lastly, in Section 4, we draw key conclusions based on our findings. Data and R codes necessary to run experiments and reproduce the results presented in this paper are available at this repository.

2. Methods and Materials

2.1. The Target Variables and Modeling

Our main target response variable is the multidimensional poverty measure constructed using the Alkire–Foster method [13], the so-called Adjusted Headcount Ratio (M_0), also referred in the literature as, the Multidimensional Poverty Index (MPI). We will treat both terms interchangeably in the text. This poverty measure consists of building a score of the weighted sum

deprivations in d poverty indicators for each person. Under the Alkire-Foster framework, a person is identified as poor if this score is greater or equal to a multidimensional poverty cut-off k that censors the non-poor out from the analysis. Then, M_0 is calculated as the average of the censored score over the whole population. That can be written as the product of two partial indices: H and A [14]. The first refers to the incidence of poverty (the proportion of the population identified as poor in multidimensional terms). In contrast, the second represents the intensity of poverty (the average weighted deprivation suffered by the poor). Note that A should be always greater or equal to k. In this way,

$$M_0 = A \times H, \tag{1}$$

Predicting M_0 identifies trends of multidimensional poverty. Furthermore, predicting H and A can help to elucidate the nature of these trends. In other words, an increase in the MPI can be attributed to either a rise in the proportion of the population experiencing poverty (H) or an intensification of deprivations among the already poor (A).
We aim to predict a target variable (namely, M_0, H or A), denoted as Y, for a country i at a specific time t_i using a set of p predictor variables $\mathbf{X}$ sourced from the World Development Indicators (WDI). For this purpose, consider the regression model

$$Y_{it_i}|\mathbf{X}_{it_i} = g(\mathbf{X}_{it_i}) + \varepsilon_{it_i}, \tag{2}$$

where $Y_{it_i} \in \mathbb{R}$ is the response variable, $\mathbf{X}_{it_i}^T = (X_{1,it_i}, X_{2,it_i}, \ldots, X_{p,it_i})$ is a vector of p predictor variables measured for the country i in the time period t_i, $g : \mathbb{R}^p \longrightarrow \mathbb{R}$ is an unknown function and ε_{it_i} is a zero-mean error term. As can be noted, the time period subscript, t_i, depends on i, indicating that the period in which the response and predictor variables are measured varies depending on the specific country i.
Although our target variables are real-valued, they are defined as proportions. In this vein, analyzing the outcomes constrained to the interval $(0, 1)$ and employing methods specifically tailored to bounded data is recommended. Notably, a flexible approach that assumes a beta distribution for the conditional response variable of the regression model (2) is considered. The beta distribution captures a variety of response shapes through its parameters: mean (μ) and precision (ϕ), as these are modeled as functions of the covariates $\mathbf{X}$ [15, 16].
Several prior studies employing machine learning (ML) for predicting socioeconomic variables expressed as proportions or rates have assumed a normal distribution of the response variable [6, 10, 11]. For this reason, we compare the predictive performance of traditional ML models that assume an unbounded continuous response with alternative models assuming a beta distribution for $Y_{it_i}|\mathbf{X}_{it_i}$, a more suitable choice for bounded data.

2.2. Statistical Learning Methodologies

Various methodologies can address the challenges of high-dimensional data. We consider three distinct ML approaches: supervised dimension reduction using Partial Least Squares (PLS), regularization and variable selection models, and sequential ensemble learning through Boosting algorithms.

2.2.1. Supervised Dimension Reduction

Dimension reduction in regression consists of estimating a $p \times d$ orthogonal projection matrix $\widehat{\mathbf{W}}$ of the p-dimensional predictors space with $d \ll p$. Then, the projected predictor space, $\widehat{\mathbf{W}}^T\mathbf{X}$ replaces the original design matrix of the model such that d is sufficiently smaller than n. The most commonly used dimension reduction technique is Principal Component Analysis (PCA), which, when applied to regression, is known as Principal Component Regression (PCR). However, PCA/PCR is unsupervised, as it does not use information from the response variable during the reduction process. In contrast, supervised dimension reduction methods—particularly Partial Least Squares (PLS)—have gained popularity because they incorporate the response variable to inform the projection directions [17]. This approach significantly improves predictive performance over unsupervised methods like PCA, especially for income and poverty prediction [18, 19].

Although PLS regression (PLSR) is associated with a linear regression model between Y and the predictor vector $\mathbf{X}$ (i.e. linearly specifying $g(.)$ of (2)), some research has extended PLS to cases where $Y \in (0, 1)$ using beta regression and proposing alternative algorithms for PLS estimation [20, 21]. Furthermore, Cook and Forzani [22] extended PLS to cases where $E(Y|\mathbf{X})$ is not necessarily a linear function of $\mathbf{X}$. This approach employs a two-step algorithm: first, a dimension reduction step that assumes that exist d linear combinations, $\widehat{\mathbf{W}}^T\mathbf{X}$, sufficient for capturing $E(Y|\mathbf{X})$. Once the predictors are reduced, these linear combinations can be used to predict Y with a general (potentially nonlinear) rule when d is sufficiently small. Cook and Forzani [22] demonstrated that, under mild conditions, these linear combinations correspond to the first d-PLS projections. In this way, we can employ beta regression using the d linear combination $\widehat{\mathbf{W}}^T\mathbf{X}$ as predictor variables.

As we will see later, global MPI measures exhibit a noticeable heterogeneity pattern, forming a bimodal distribution with two distinct clusters: one concentrated around very low poverty levels and the other around medium to high poverty levels. Generally, observable predictor variables, such as development indicators and regional dummy variables, are related to these clusters, so their inclusion in the regression model can help control for this heterogeneity. However, in some cases, additional latent, unobserved factors may contribute to group heterogeneity that cannot be fully captured through the mean and precision of beta models [23].

To account for both observed and unobserved sources of heterogeneity, a plausible approach is model-based recursive partitioning [24], which builds on the Classification and Regression Tree (CART) methodology. This technique recursively partitions the sample based on selected *partitioning* variables to capture parameter differences that describe the response distribution. Model-based recursive partitioning for beta regression is referred to as *beta regression trees* by Grün et al. [23], where the detailed algorithm is provided.

In our analysis, we incorporate beta regression trees as a prediction rule within the dimension reduction and regularized model frameworks. We use a dummy variable as the partitioning variable, constructed around a cutoff that marks a potential distributional shift, selected by visualization. Specifically, this cutoff is set at 0.2 for predicting M_0 and H and at 0.5 for predicting A.

In summary, using PLS as dimension reduction we compare the following methods:

1. `Linear-PLS`: PLS using linear regresion (PLSR) of Y_{it_i} on $\mathbf{W}^T\mathbf{X}_{it_i}$.

2. `Beta-PLS`: a generalized linear PLSR model of Y_{it_i} on $\mathbf{W}^T\mathbf{X}_{it_i}$ using the beta distribution family.
3. `Beta-Tree-PLS`: a generalized linear PLSR model of Y_{it_i} on $\mathbf{W}^T\mathbf{X}_{it_i}$ using the beta distribution family with tree model (model-based recursive partitioning).

For linear PLS, we used the standard NIPALS algorithm, extending it to accommodate a nonlinear mean function as suggested by Cook and Forzani [22], specifically for both beta regression models in our study. Due to PLS predictive models' flexibility in overcoming the so-called curse of dimensionality, we initially fitted non-parametric regressions based on kernels for $g(.)$. However, this approach did not yield superior results compared to the three cited models. Additionally, this predictive rule requires an optimal bandwidth choice, which incurs a correspondingly higher computational cost. For these reasons, it was excluded as a comparable methodological strategy in this paper. Nonetheless, in many cases, it could represent a viable approach when d is sufficiently small.

The optimal number of dimensions, d, serves as the hyperparameter in our dimension reduction methods and is selected via 5-fold cross-validation on the training sample.

2.2.2. Regularized models

An alternative approach to dimension reduction for high-dimensional data is regularized modeling, also known as shrinkage or penalized methods for variable selection.[1] This approach relies on the assumption of model sparsity; that is, only a small subset of predictors play an important role in the response or target variable. Under a parametric specification of $g(.)$, a penalty is applied to the parameters during estimation, shrinking the coefficients and simultaneously performing model and variable selection to enhance prediction accuracy.

One of the most widely used regularized methods is the so-called Lasso estimator [25]. For a linear regression model $g(\mathbf{X}) = \mathbf{X}^T\boldsymbol{\beta}$, the Lasso finds a solution for the constrained least squares (LS) problem subject to $||\boldsymbol{\beta}||_1 = \sum_{j=1}^{p} |\beta_j| \leq t$. The LS's Lagrangian form includes a penalization term $\lambda||\boldsymbol{\beta}||_1$ where the tuning parameter λ controls the strength of the penalization. Using the ℓ_1-norm, many coefficient estimates are exactly zero, which excludes the corresponding predictors from the model. In consequence, Lasso is highly effective in high-dimensional contexts where $p >> n$ [26]. Alternatively, using the euclidean ℓ_2-norm for penalization, as in the *Ridge regression* method (which predates Lasso), shrinks coefficients toward zero but does not perform variable selection. Nevertheless, in usual situations where $n > p$, Ridge regression could outperform Lasso in predictive terms [25].

Despite its well-deserved popularity, the Lasso method has certain limitations, some of which can be addressed through extensions. One notable limitation is that when $p >> n$, Lasso tends to select at most n variables, limiting its effectiveness for variable selection in high-dimensional settings. Another limitation, particularly relevant in this context, arises when groups of predictor variables are highly correlated. In such cases, Lasso exhibits a tendency to under-perform [26, 27]. Regarding WDI as predictor variables, there are groups that measure similar aspects, leading to high pairwise correlations. For example, between gross domestic product and national income variables, life expectancy and mortality rates, or employment and labor market indicators. In these situations, Lasso typically selects

[1] In some cases, these approaches can be complementary, where dimension reduction is performed jointly with variable selection via regularization, resulting in more interpretable coefficients for the directions of the reduction; see, for example, Duarte et al. [18].

only one variable from each correlated group, without regard for which specific variable is selected. To overcome this, [27] proposed a regularization method that combines the Ridge and Lasso penalties in a convex combination, obtaining a variable selection method that has the ability to reveal the grouping information doing grouped selection. This approach is called Elastic Net, as well as Lasso or Ridge, is a LS problem but with the elastic net penalty $\lambda\left(\alpha||\boldsymbol{\beta}||_1 + (1-\alpha)||\boldsymbol{\beta}||_2\right)$. Considering the characteristics of our data, we adopt this approach.

As we mentioned, traditional regularization methods, including Elastic Net, are built on a linear regression framework. Since traditional Elastic Net predictions can fall outside the interval $(0, 1)$, they do not align with the bounded nature of our response variables, making these out-of-bound predictions nonsensical in this context. Extensions of Elastic Net for Generalized Linear Models (GLMs) can be found in the literature, including applications to Cox proportional hazards models [28, 29]. However, since the beta distribution does not strictly meet GLM assumptions due to its more complex mean-variance relationship—where the variance depends on both the mean and a precision parameter—it cannot be directly incorporated into these GLM extensions. To address this, we rely on the idea of a *Flexible* Elastic Net, as introduced by Meinshausen [30] for Lasso, which combines the lasso estimator with the OLS estimator on selected variables, demonstrating strong performance across various scenarios [31]. For our application, we adapt this approach by applying beta regression in place of OLS on the active set identified through Elastic Net estimation. Rather than combining Elastic Net and beta regression coefficients—due to differences in the link function for linear predictions—we use the coefficients from the second-step beta regression directly for predictions. Empirical research also supports this procedure, showing that predicting with selected variables using a non-regularized model - typically OLS, but beta regression in our case - performs well in out-of-sample prediction [e.g. 32]. Additionally, as in the dimension reduction approach, we incorporate a beta regression tree model to capture potential nonlinearity and account for heterogeneity in the predictor-response relationship.

Therefore, within the regularized models approach, we consider the following methods:

1. `Elastic Net`: a Elastic Net model assuming a gaussian model for linear $g(.)$.
2. `Beta (elastic)`: Flexible Elastic Net for beta regression model, using active set variables form a first-step Elastic Net estimation.
3. `Beta-Tree (elastic)`: Flexible Elastic Net for beta regression tree model, using active set variables form the Elastic Net.

In these cases, the penalization λ and α are the tuning parameters or hyperparameters of these models, that are selected via 5-fold cross-validation on the training sample.

2.2.3. Ensemble learning and Boosting

Ensemble learning refers to a combination of a large number of M base models or weak learners during the fitting process, denoted by $\mathcal{E} = \{g_1(.), \ldots, g_M(.)\}$, where each weak learner $g_m(.)$ may utilize any modeling approach, ranging from linear models to decision trees. Generally, the final prediction $\hat{g}(\mathbf{X})$ at a new point $\mathbf{X}$ is determined by averaging the predictions of all M base learners, $\frac{1}{M}\sum_{m=1}^{M} g_m(\mathbf{X}))$, or by using a weighted combination,

$\sum_{m=1}^{M} w_m\, g_m(\mathbf{X})$. This methodology is particularly effective for high-dimensional problems, as each weak learner relies on a relatively simple model (e.g., a linear regression with a single predictor).

Ensemble methods are classified into three primary categories: Bagging, Stacking, and Boosting. Each category encompasses a variety of algorithms, allowing for diverse approaches to ensemble learning.

Bagging, which stands for Bootstrap Aggregating, is the simplest ensemble approach. It consists of sampling M random subsets of the predictors space with replacement, $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(M)}$, and train the base models independently [33]. Since this method relies on bootstrap samples, it often results in correlations among weak learners. A widely used bagging technique is Random Forest [34], which mitigates this issue by randomly selecting a subset of predictors at each split, thereby constructing an ensemble model composed of a *forest* of largely uncorrelated decision trees. This reduction in correlation among trees enhances the stability and accuracy of predictions.

Stacking consists of training a set of different independent machine learning algorithms and combining the predictions [35]. Hence, Staking is based on learning how to weight the different models' predictions. Note that the weak learners in this case are more sophisticated. For instance, combining a PLSR, an Elastic Net and a Random Forest in the same model would be possible. Sometimes the stacking models are called meta-models, where the base models are referred to as level-0 models and their ensemble is known as level-1 models [36], where the latter can be of any predictive rule. Commonly, boosting-based models are used to combined the level-0 models [37]. This ensemble approach gives the possibility to combine highly different modeling approaches that capture different patterns in data.

Boosting is a powerful ensemble approach that builds the ensemble model sequentially rather than independently as in Bagging or Stacking [38]. In each iteration, m, a base learner g_m is fitted based on the errors of the previous weak learners $g_1, \ldots, g_{m-1}$. The algorithm identifies the best subset of covariates to improve its prediction, which subsequently enhances the overall model's prediction. Therefore, boosting-based models naturally carry out variable selection and improve predictive accuracy by giving higher weights to those observations with prediction mistakes of the earlier $m-1$ weak learners. However, this approach presents a trade-off, as it can be sensitive to noise in data, increasing the risk of over-fitting, and the learning process can be slower than other ensemble methods.

There are different approaches to Boosting. The two most well-known are AdaBoost [39] and Gradient Boosting [40]. A variant of the latter, XGBoost [41], has recently emerged as a leading algorithm in the field of applied machine learning, achieving notable success by winning multiple Kaggle competitions [42]. XGBoost is a tree-based ensemble model that improves gradient boosting by including a penalization term, a maximum tree depth, a learning rate, and a subsampling to prevent the model from overfitting during the learning process.

While no single ensemble method strictly outperforms others across all tasks and data types, we focus on boosting-based models in this paper, given their proven high performance in applied machine learning contexts [43]. Specifically, we employ XGBoost due to its strong methodological advantages, including the efficiency in handling large datasets and the reduced need for feature engineering.

However, in the context of predicting poverty indices, XGBoost and other ensemble methods may produce out-of-bound predictions—a challenge when working with bounded response variables. To address this issue and provide a meaningful comparison with XGBoost, we apply a gradient boosting-based beta regression model. Mayr et al. [44] adapted boosting algorithms for classical beta regression, enabling beta models to be used effectively in high-dimensional settings.

Based on these considerations, we employ the following two methodologies within the boosting framework:

1. `XGBoost`: a tree-based gradient boosting model.
2. `BetaBoost`: an additive model with gradient boosting using the beta distribution family.

The main hyperparameter of boosting models is the learning rate η and the number of boosting iterations [45]. Additionally, there is a set of hyperparameters associated with decision tree models, including aspects like maximum depth, minimum child weight, and the number of leaves. As with other approaches, we determine these parameters through 5-fold cross-validation, aiming to minimize the mean squared error (MSE) within the training datasets.

2.3. Data

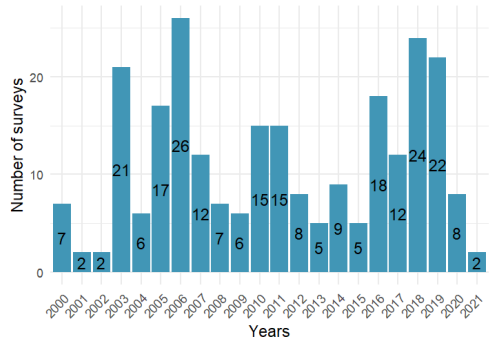
2.3.1. The global Multidimensional Poverty Index

Briefly, the global MPI is a measure of acute poverty in the developing world that accounts for deprivations in three key dimensions of human development: health, education and basic living standards. These deprivations are represented by ten indicators, each one is identical, or related, to a specific item of the Sustainable Development Goals (SDGs). The formal structure of the global MPI corresponds to equation (1) with a poverty cut-off of $k = 33\%$, which means that a person is multidimensionally poor if she is deprived in one out the three dimensions.

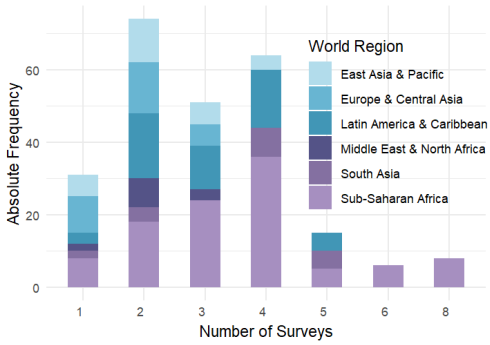
The global MPI is published every year by the Oxford Poverty and Human Development Initiative [46] for more than 100 developing countries. Our data takes the 106 countries and the years for which we have a global MPI measure (a period from 2000 to 2021). Its calculation relies on two main multi-topic household surveys: the Demographic and Health Survey (DHS) and the Multiple Indicators Cluster Survey (MICS). For some countries, national household surveys with similar content and questionnaires are standardized and used.

In spite of the great increase in data availability and quality, the frequency of multi-topic surveys for collecting data on poverty remains limited [3]. For this reason, the global MPI data has a remarkable feature: there is no country with a global MPI measure for each year in our time period (2000-2021). For most countries, we have two global MPI measures, at most 4 observations in some countries and between 5 and 8 in a few cases. Figure 1a (top-left panel) depicts the annual distribution of surveys, revealing a rise in multidimensional surveys during the mid-2000s, 2010s, and before the COVID-19 pandemic. The top-right panel of Figure 1 presents the distribution of countries by the number of survey rounds conducted. Notably, a majority of countries have undergone between 2 and 4 survey rounds, with a limited number exceeding this range across the entire study period (2000-2021). However, a significant geographic skew is evident in survey distribution. Sub-Saharan African countries contributed the

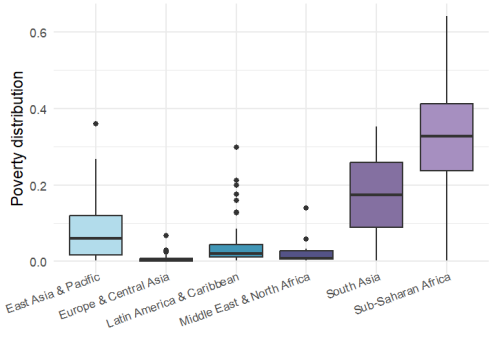
most surveys to the global MPI calculation (105 of 249), followed by Latin America & the Caribbean (54), Europe & Central Asia (30), East Asia & the Pacific (28), South Asia (19), and the Middle East & North Africa (13).



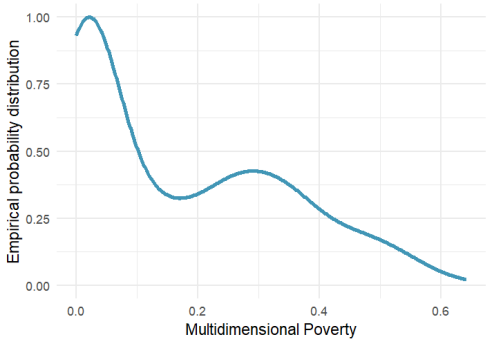
(a) Number of surveys by year



(b) Number of surveys by World Region



(c) Poverty distribution by World Regions



(d) Distribution of Multidimensional Poverty

Fig. 1: Multidimensional poverty: surveys frequencies and data analysis

Figure 1c shows the distribution of the global MPI by region, which further highlights the geographic disparities in multidimensional poverty. Sub-Saharan Africa and South Asia exhibit significantly higher levels of poverty on average, and also greater dispersion compared to other regions. While East Asia & the Pacific show some distinct patterns, the remaining regions (Latin America & the Caribbean with some exceptions) have lower and less dispersed poverty levels. These geographical disparities underscore the importance of including world regions as a key control variable in our prediction model, as discussed in Pasha [47]. Last but not least, 1d illustrates the kernel density distribution of the global MPI, showing a clear bimodality, with peaks near 0 and 0.3.

Given the world region disparities and the bimodal empirical distribution of the MPI, we postulate that more flexible methods such as the tree-based boosting models described in Section 2.2 will produce more accurate predictions. Furthermore, since many observations

are close to the lower boundary, beta regression models will yield meaningful predictions in these cases, avoiding negative poverty rate estimates.

2.3.2. Web scraping and World Development Indicators (WDI)

The predictor variables in this study come from the World Development Indicators database, which provides a comprehensive collection of 1485 national-level estimates across various topics including education, health, demographics, and other socioeconomic indicators, which is open-source and frequently updated. The construction of the predictors' dataset implements web scrapping as a technique that allows us to extract the data and information from the World Bank website. Then, each WDI is extracted matching the countries and years for which the global MPI is provided. However, the resulting dataset contains several missing values (denoted as "Not Available" or NA), particularly for developing countries. To address this issue and facilitate the data cleaning, we employ a multi-step process that results in 25 distinct datasets. Firstly, we begin with the complete dataset encompassing all available countries and years. Subsequently, we build another dataset removing countries with only one observation (one survey round), resulting in a dataset of 249 observations from 106 countries and 110 WDI. This filtering step ensures that each country has multiple target variable measures, allowing for a more robust evaluation of different methods.

To further explore the impact of missing data in our database, we proceed by iteratively removing countries with increasing proportions of NAs. For example, in the next dataset, we have an overall of 218 observations from 75 countries and 167 WDI. This process aims to evaluate the performance of the selected statistical methods in section 2.2 under a high-dimensionality scenario, characterized by a limited number of observations relative to the number of predictor variables, i.e., $n \ll p$. Table A1 in Appendix 1 describes the iterative process of removing countries with greater proportions of NAs in the WDI predictors[2].

Finally, in this highly unbalanced panel data, in order to account for systematic changes that occur across all countries over time, we model the temporal effects as a linear increasing trend in terms of the observed year to the minimum year observation in the dataset.

2.4. Experiments and Validation Metrics

In order to evaluate the alternative methods for MPI prediction, we select three different datasets: 1, 2 and 13 from Table A1. This selection allows to have three scenarios for assessing predictive performance. Within the first one, we train models with $n \gg p$; in the second one, $n \approx p$ in the training dataset; and the last one has $n \ll p$. This approach allows us to assess the precision of our models' predictions when transitioning to a high-dimensional context. We measure the predictive performance in the following ways:

- Experiment 1: For each dataset, we randomly divided the data into two subsets: 20% for testing and 80% for training and validation. The training and validation set are used to estimate the models and predict Y for the observations in the test dataset. This experiment is performed 50 times, measuring the prediction error using the Mean Square Error (MSE) in each repetition, and then analyzing the distribution of these errors. Specifically, for a

[2]This process can be examined in detail in https://github.com/agdaniela/GlobalMultidimensionalPovertyWDI

given predictive rule j, we compute

$$MSE_j^{(k)} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} (Y_i^{(k)} - \widehat{Y}_{i,j}^{(k)})^2, \qquad k = 1, \ldots, 50,$$

where Y_i^k is the actual value of the target variable for the i-th observation in the testing sample of size n_{test}, and $\widehat{Y}_{i,j}^{(k)}$ is the corresponding predicted value using the j-th methodology. To compare predictive performance we can analyze the distribution of MSE_j^k as well as its average and dispersion over the 50 replications.

- Experiment 2: Given the considerable heterogeneity among countries in terms of socioeconomic development and multidimensional poverty, it is crucial that predictions are as accurate as possible across the entire spectrum of low, medium, and high poverty levels. To evaluate prediction accuracy across the full range of MPI values, we analyze the empirical distributions of the predicted responses $\hat{Y}$ and the observed values Y for each dataset. We conduct a 10-fold cross-validation experiment, where we obtain $\widehat{Y}_{ij}$ for each j-th methodology and all $i = 1, \ldots, n$. Specifically, the data is partitioned into ten disjoint subsets; in each fold, one subset serves as the test set, while the remaining subsets are used to train the models and to predict the MPI values for the test set. We then compare the empirical distributions of the actual Y and the predicted $\widehat{Y}$. First, we estimate and compare the empirical probability density functions of Y and $\hat{Y}$, denoted by $\hat{f}_Y(y)$ and $\hat{f}_{\widehat{Y}}(y)$, respectively, using Kernel smoothing for density estimation. Second, we compute the Hellinger distance h [48] to quantify the similarity between the two density functions. For densities $f_Y(y)$ and $f_{\hat{Y}}(y)$, the Hellinger distance is given by

$$h(f_Y, f_{\widehat{Y}}) = \frac{1}{\sqrt{2}} \sqrt{\left(\int \left(\sqrt{f_Y(y)} - \sqrt{f_{\widehat{Y}}(y)} \right)^2 dy \right)}.$$

h is bounded to 0 and $\frac{\sqrt{2}}{2}$, where the closer to zero, the more similarity between f_Y and $f_{\widehat{Y}}$. The advantage of the Hellinger distance is its meaningful interpretation concerning probability distributions, making it particularly useful for non-Gaussian observations, as in our study [49].
- Experiment 3: For the selected countries, we predict MPI values for years with unavailable data to evaluate the behavior of the predicted series. Our methods estimate MPI across the entire period, enabling us to assess the plausibility of the results and evaluate precision through bootstrap prediction intervals. We focus on Senegal, Bangladesh, and Bolivia, which have varying amounts of MPI data over the period—specifically 6, 4, and 2 years of MPI measurements, respectively. To train the models, we use all observed MPI values from all countries and years, including those from the country of interest, and predict MPI for years without direct measurements by leveraging the country's WDI data available.

3. Results

3.1. Prediction for randomly selected countries and years

Table 1 presents the average MSE and the corresponding standard deviations from the 50 repetitions of experiment 1. The lowest predictive errors for each case are highlighted in boldface. Additionally, the corresponding box-plots are provided in the Appendix B.

For dataset 1, we have more observations than predictor variables to train the models. In this case, for the global MPI, M_0, the best predictive performance is achieved using gradient boosting with beta regression (`Betaboost`). However, `XGboost` as well as `Elastic Net` also show comparable average errors and dispersion. In fact, `Elastic Net` has the lowest error for the Intensity of multidimensional poverty, A. Regarding the Incidence of poverty, H (the proportion of poor people), the beta regression model again proves to be superior, particularly `Beta-Tree-PLS` and `Betaboost`, although the latter shows lower dispersion.

In the matter of dataset 2, we have a number of observations in the training dataset $n_{train} = 218 \times 0.8 \approx 175$. Considering the 167 WDI, the 6 dummy variables for regions and the time effect, we have a total of $p = 174$ covariates. Therefore, the models are trained using approximately the same number of observations as predictor variables. For M_0 prediction task, `Elastic Net`, `XGBoost` and `Betaboost` yield the lowest average MSE. For A, `Beta-Tree (elastic)` has the best performance. For H, `Betaboost` and `Beta-Tree (elastic)` yield the lowest average MSE albeit the former is more precise.

Finally, for dataset 13 where $n_{train} \ll p$, `Betaboost` once more achieves the best M_0 prediction. In fact, compared to the other datasets, these results show that including large information through more predictor variables enhances the prediction compared to the information provided by a bigger number of observed units (countries and years) in the training sample However, this pattern does not occur for all methodologies, as can be seen for `Beta-Tree (elastic)`, `Elastic Net` or `Beta-Tree-PLS`. Furthermore, when considering A's prediction, the best results are given by `Elastic Net`, with comparable results from `XGBoost` and `Beta-Tree (elastic)` models. Last but not least, `Betaboost` yields the best performance for H.

The box-plots of the MSE presented in Appendix B further support the description of these results, in which is possible to confirm that for M_0, the most precise methods are `Betaboost`, `XGBoost`, and `Elastic Net`. For H, the `Beta Tree (elastic)` yields the best prediction results for datasets 1 and 2, but not for dataset 3, where `Elastic Net` shows the best performance. For the A index, `Elastic Net` performs well across all datasets, but in dataset 2, `Beta-Tree (elastic)` outperforms it.

As hypothesized in Section 2.3.1, our results indicate that for the M_0 and H measures, which tend to be close to zero, beta distribution-based models significantly outperform models assuming continuous real-valued distributions. In contrast, for the A measure, commonly used methods like `Elastic Net` and `XGBoost` achieve better results since A is less likely to be close to any boundary. These findings underscore the importance of considering the bounded nature of our target variables when estimating multidimensional poverty predictive models.Broadly speaking, tree-based models also yield better results as suggested above, with some exceptions where `Elastic Net` has the lower average MSE.

Target Variable	Method	Dataset 1	Dataset 2	Dataset 13
M_0	`Linear-PLS`	0.0059 (0.0039)	0.0202 (0.0232)	0.0161 (0.0169)
	`Beta-PLS`	0.0125 (0.0076)	0.0065 (0.0038)	0.035 (0.024)
	`Beta-Tree-PLS`	0.0106 (0.0084)	0.0059 (0.0066)	0.0072 (0.0076)
	`Elastic Net`	0.0035 (0.001)	**0.0028** (6e-04)	**0.0015** (6e-04)
	`Beta (elastic)`	0.0125 (0.0409)	0.0041 (0.0016)	0.0211 (0.0258)
	`Beta-Tree (elastic)`	0.0096 (0.0324)	0.0033 (0.0014)	0.0264 (0.0302)
	`XGBoost`	0.0036 (0.001)	**0.0028** (8e-04)	0.0031 (0.0021)
	`Betaboost`	**0.0034** (0.0013)	0.0029 (0.001)	0.0025 (0.0014)
A	`Linear-PLS`	0.0031 (0.0015)	0.0062 (0.0118)	0.004 (0.0071)
	`Beta-PLS`	0.0058 (0.0022)	0.0049 (0.0032)	0.0061 (0.0035)
	`Beta-Tree-PLS`	0.0052 (0.0022)	0.0023 (0.002)	**0.0019** (8e-04)
	`Elastic Net`	**0.0026** (0.0011)	0.0019 (6e-04)	**0.0018** (0.0057)
	`Beta (elastic)`	0.0069 (0.0022)	0.0023 (7e-04)	0.0034 (0.0073)
	`Beta-Tree (elastic)`	0.0062 (0.0021)	**0.0015** (4e-04)	0.0029 (0.0053)
	`XGBoost`	0.0031 (0.001)	0.0021 (7e-04)	**0.0013** (6e-04)
	`Betaboost`	0.0057 (0.0017)	0.0035 (0.0015)	0.009 (0.0074)
H	`Linear-PLS`	0.0432 (0.0572)	0.0728 (0.1188)	0.1319 (0.411)
	`Beta-PLS`	0.0135 (0.0068)	0.0146 (0.0072)	0.0775 (0.0353)
	`Beta-Tree-PLS`	**0.0074** (0.0048)	0.014 (0.0088)	0.0316 (0.0308)
	`Elastic Net`	0.0101 (0.0039)	0.0077 (0.0018)	**0.0035** (0.0013)
	`Beta (elastic)`	0.0186 (0.0454)	0.0075 (0.002)	0.0371 (0.042)
	`Beta Tree (elastic)`	0.0104 (0.0319)	**0.0062** (0.0021)	0.0469 (0.0543)
	`XGBoost`	0.0081 (0.0018)	0.008 (0.0026)	0.007 (0.0033)
	`Betaboost`	**0.0075** (0.0015)	**0.0062** (0.0018)	0.0051 (0.0024)

Table 1: Average Prediction Errors (MSE) from 50 repetitions experiment. Standard errors of MSE in parenthesis

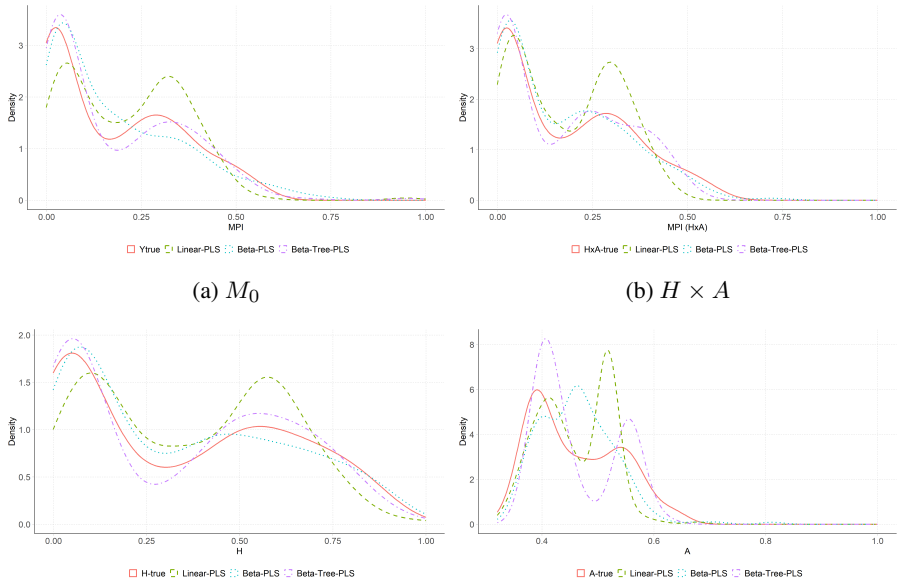
3.2. Distribution of predicted and actual MPI

In the second performance evaluation, we compared the probability distributions of the predicted MPI target variables (M_0, H, and A) from a 10-fold cross-validation experiment with the actual values in the testing samples. We also explore the possibility of estimating M_0 from $\widehat{H}$ and $\widehat{A}$, following equation (1).

A preliminary graphical examination of the empirical densities of the ground truth and the predicted values revealed that models assuming continuity were less accurate in predicting values close to the lower bound.

In Figure 2, we compare smoothed empirical density probability functions using dimension reduction and the three proposed predictive rules: linear regression (`Linear-PLS`), beta regression (`Beta-PLS`), and beta tree-based regression (`Beta-Tree-PLS`). The solid red line corresponds to the observed poverty rates. Figures 2a and 2b display the adjusted headcount ratios predicting directly (M_0), or indirectly via $\hat{H} \times \hat{A}$, respectively. From the visualization, it is clear that beta regressions outperform linear predictors. The empirical densities of M_0 and $\hat{M}_0$ are very close under beta models, whereas the linear regression predicted values diverge from the true MPI, especially in the more concentrated intervals; that is, for small MPI values (near zero) and in the middle mode (near 0.3). For the highest MPI values, the densities show that the linear predictor underestimates the true level of poverty, concentrating around the middle mode. Comparing `Beta-PLS` with `Beta-Tree`, we observe that

the empirical density of the `Beta-Tree` predictions aligns more closely with the density of the true M_0, again supporting our hypothesis that tree-based models fit better than linear models. However, the `Beta-PLS` performs better when the MPI is estimated via $H \times A$.



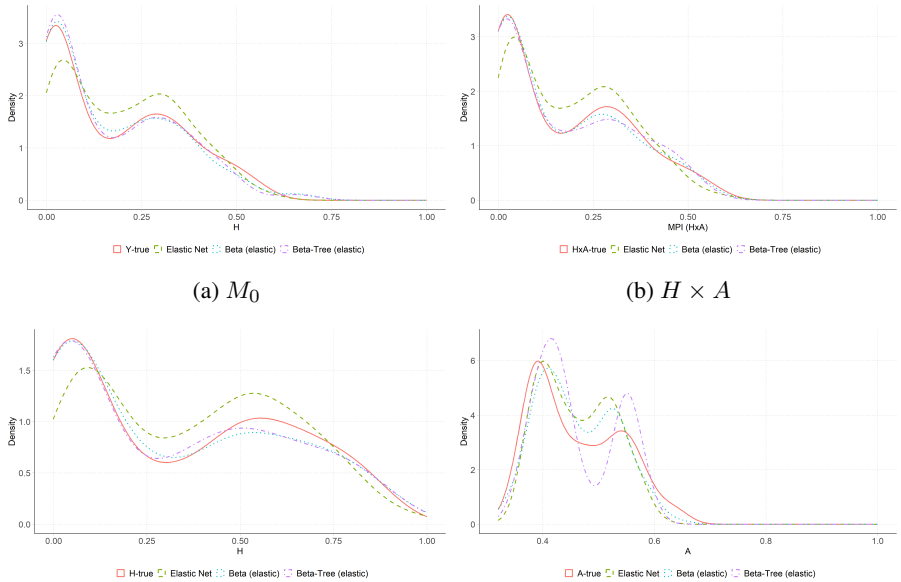
(a) M_0 (b) $H \times A$

(c) H (d) A

Fig. 2: Kernel Distributions for Actual and Predicted Multidimensional Poverty Rates using Dimension Reduction.

In Figures 2c and 2d empirical densities of H and A are plotted alongside the corresponding predictions. For H, the predicted and observed densities are similar, particularly for the regression models based on the beta distribution. However, for A more divergences are observed. In this case, `Beta-Tree-PLS` exhibits more pronounced peaks and valleys in the distribution, while `Beta-PLS` tends to concentrate the distribution in the middle values of A where, indeed, a valley is revealed from observed values of A. This poorer performance in predicting A can be attributed to the fact that its values have a lower bound at the cut-off $k = 0.3$, unlike H and M_0 which are lower bounded at zero.

The plots in Figure 3 show the empirical densities for predictions using regression models with elastic net regularization. These plots demonstrate a better performance compared to the PLS-based predictions. The linear specification (`elastic-net`) offers reasonably accurate predictions of MPI. However, similar to the linear PLS model, the prediction densities for low and mid-range MPI values deviate significantly from the true MPI. Notably, using beta models leads to substantial improvements. Both `Beta elastic` and `Beta-tree (elastic)` produce empirical distributions of MPI predictions that closely match the actual values of M_0 and $H \times A$. Greater divergences are observed in A predictions, with `Beta-tree (elastic)` exhibiting a more erratic pattern.



(a) M_0 (b) $H \times A$

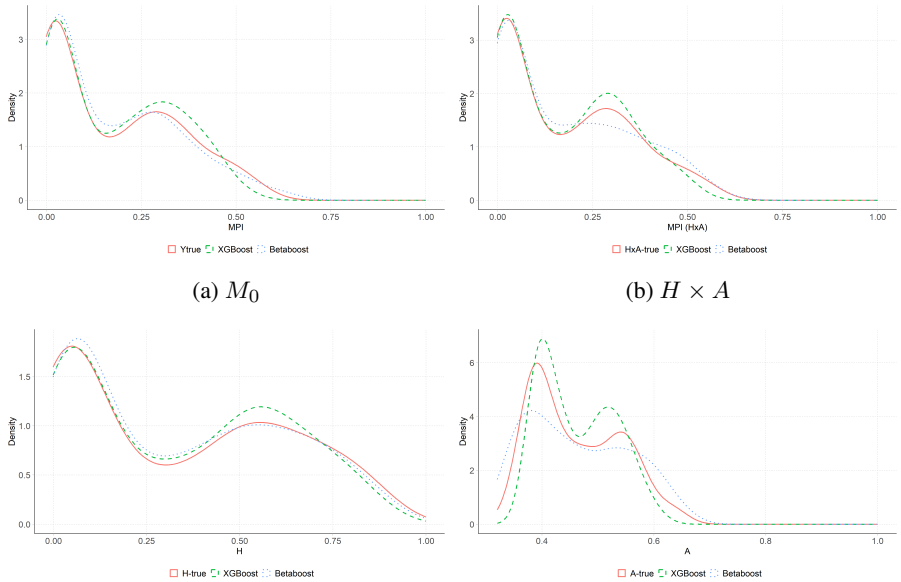
(c) H (d) A

Fig. 3: Kernel Distributions for Actual and Predicted Multidimensional Poverty Rates using Regularization for Variable Selection.

On the other hand, empirical densities from boosting models are shown in Figure 4. It is noticeable that `Betaboost` fits the true values of M_0 and A measures quite well, although `XGBoost` have comparable results. Again, the beta regression model fits poorly the true A values and, consequently, $H \times A$. In the same line as experiment 1 results, the tree-based boosting models yield better results.

It is possible to compare all methods across all target variables by observing Table 2, which displays the MSE and the Hellinger distance (h) for the 10-fold experiment in each of the selected datasets and the three poverty measures (M_0, H and A). Once more, the best results are highlighted in boldface.

Results related to MSE are consistent with the findings from the first experiment, indicating that beta regression models generally provide better predictions when the target variables are near their bounds, albeit with some exceptions. When evaluating performance using the Hellinger distance, which considers the overall distribution and, in particular, the tails where the target variables are bounded, we find that the predicted probability distribution of the beta regression models aligns more closely with the empirical distribution of M_0 and H. Furthermore, we find that every model under-perfoms for the variable A compared to the other two poverty measures. Finally, although it does not yield the best results in a few cases, on average, `Betaboost` model fits best M_0 and H.



(a) M_0 (b) $H \times A$

(c) H (d) A

Fig. 4: Kernel Distributions for Actual and Predicted Multidimensional Poverty Rates using Boosting for Model and Variable Selection.

3.3. Evolution of MPI for selected countries with predicted values

We have previously emphasized that, despite the increasing number of multi-topic surveys, available data for multidimensional poverty measurement remains scarce. As shown in Figure 1b, most countries have conducted fewer than four surveys over a twenty-year period.

For this reason, we predict the series of certain countries to account for the models' precision. As we mentioned, we selected three countries from different regions with middle to high levels of poverty and each one differs in the number of MPI measures in the training sample. The countries chosen are: Bangladesh, Bolivia and Senegal. Given that dimension reduction methods have shown to be less accurate, despite particular cases, we have limited our analysis to a few models that presented better performance. Accordingly, we show results for dataset 1 and the M_0 measure, given that in the previous experiment, all the models yielded poorer predictions for A. The rest of the results for all measures and data sets can be found in the cited repository[3].

Figures 5 to 8 showcase the out-of-sample bootstrap predictions of `Beta-Tree`, `Elastic Net`, `XGBoost` and `Betaboost`, respectively, with a number of B repetitions. The black-solid line refers to the average predicted value, while the blue-dotted lines are the 5% and 95% quantiles denoting the prediction interval over all the bootstrap repetitions. As it is possible to observe, we include the true multidimensional poverty measure marked with a cross.

[3]It is worth mentioning that this code is designed to be flexible, allowing for out-of-sample predictions in any developing country, not just Bangladesh, Bolivia, and Senegal.

Target Variable	Method	Dataset 1		Dataset 2		Dataset 13	
		MSE	h	MSE	h	MSE	h
MPI	`Linear-PLS`	0.0052	0.1079	0.0118	0.2226	0.0156	0.3614
	`Beta-PLS`	0.0130	0.2525	0.0068	0.2274	0.0251	1.5913
	`Beta-Tree-PLS`	0.0102	0.1613	0.0069	0.1515	0.0053	0.5066
	`Elastic Net`	0.0039	0.1523	**0.0028**	0.1116	**0.0015**	0.1329
	`Beta (elastic)`	0.0037	0.1098	0.0042	0.1346	0.0290	0.3844
	`Beta-Tree (elastic)`	**0.0031**	0.1496	0.0039	0.1373	0.0330	0.4539
	`XGBoost`	0.0032	0.1299	0.0029	0.1500	0.0024	0.1741
	`Betaboost`	0.0036	**0.1076**	0.0032	**0.1042**	0.0021	**0.0939**
A	`Linear-PLS`	**0.0026**	**0.1911**	0.0040	0.3927	0.0045	0.3984
	`Beta-PLS`	0.0045	0.3055	0.0043	0.3607	0.0109	1.6528
	`Beta-Tree-PLS`	0.0044	0.2932	0.0019	0.3274	0.0020	0.5867
	`Elastic Net`	**0.0026**	0.2734	0.0020	0.2909	0.0051	**0.1121**
	`Beta (elastic)`	0.0058	0.2956	0.0024	**0.1916**	0.0081	0.2414
	`Beta-Tree (elastic)`	0.0050	0.2377	**0.0015**	0.2991	0.0050	0.2173
	`XGBoost`	0.0034	0.2954	0.0019	0.2970	**0.0012**	0.2347
	`Betaboost`	0.0049	0.3130	0.0034	0.2456	0.0077	0.3302
H	`Linear-PLS`	0.0162	0.1345	0.0380	0.1917	0.0455	0.4261
	`Beta-PLS`	0.0171	0.1072	0.0144	0.0961	0.0655	1.0396
	`Beta-Tree-PLS`	0.0141	**0.0787**	0.0115	0.1035	0.0186	0.3456
	`Elastic Net`	0.0099	0.1579	0.0078	0.1253	0.0034	0.0876
	`Beta (elastic)`	0.0066	0.1179	0.0081	0.0639	0.0167	0.1156
	`Beta-Tree (elastic)`	**0.0048**	0.0912	**0.0058**	0.0571	0.0185	0.0902
	`XGBoost`	0.0083	0.1036	0.0069	0.0793	**0.0055**	0.12
	`Betaboost`	0.0079	0.1019	0.0062	**0.0519**	**0.006**	**0.0601**

Table 2: Errors from 10-Fold experiment for Multidimensional Poverty Indicators

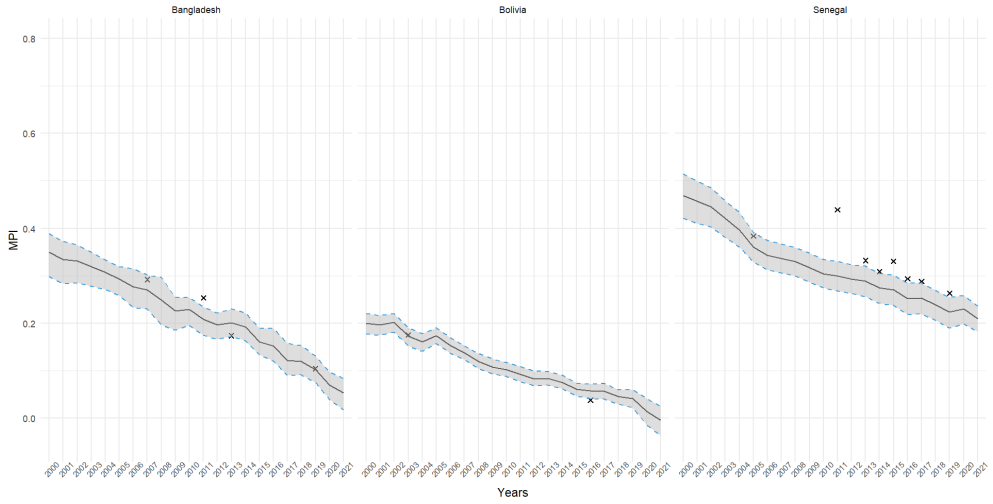


Fig. 5: MPI Series imputation using out-of-sample predictions with `Elastic Net`

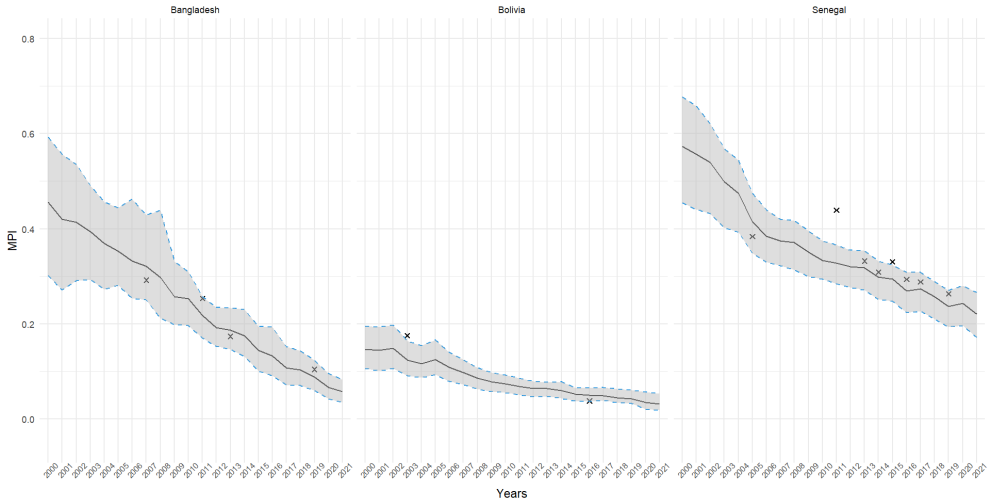


Fig. 6: MPI Series completation using out-of-sample predictions with `Beta-Tree (elastic)`

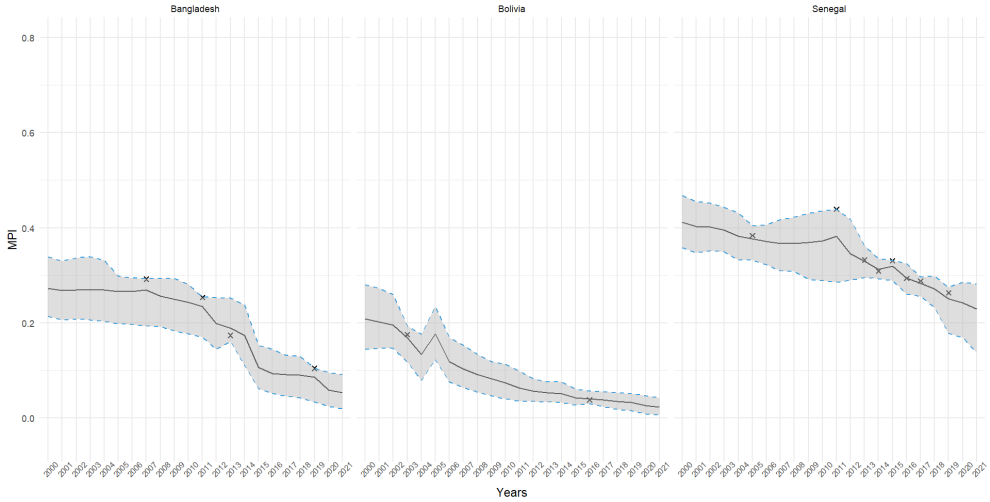


Fig. 7: MPI Series imputation using out-of-sample predictions with `XGBoost`

First of all, the average predictions in general terms are relatively close to the ground truth. We can order the methods on how well the average response predicts the true values as follows: `XGBoost`, `Betaboost`, `Beta-Tree` and `Elastic Net`. This supports our hypothesis and results of previous experiments that flexible models such as tree-based will fit data better than other parametric approaches.

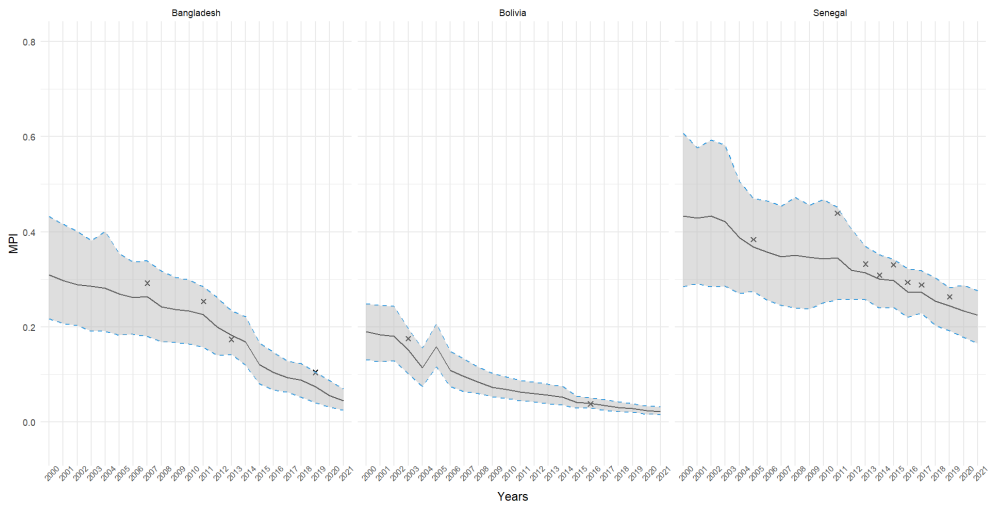


Fig. 8: MPI series imputation using out-of-sample predictions with `Betaboost`

Secondly, while penalized models (`Elastic Net` and `Beta-Tree`) effectively capture the overall decreasing trend in multidimensional poverty, they struggle to account for short-term fluctuations. As shown in Figures 5 and 6, these models consistently underestimated the significant rise in poverty levels in Senegal around 2010, whereas ensemble models were able to accurately capture this shift. Although penalized models provide narrower prediction intervals, indicating greater confidence in their estimates, this narrowness may obscure underlying uncertainty, especially during periods of rapid change.

On the other hand, `XGBoost` and `Betaboost` reveal interesting patterns in the behavior of multidimensional poverty levels in the other two countries, Bolivia and Bangladesh. In Bolivia, the models suggest a sudden increase in poverty prior to the structural changes of 2006, which was followed by a sharp decline, halving poverty levels in the subsequent years. Similar findings were reported by Villarroel and Hernani-Limarino [50]. Bangladesh also experienced significant reductions in multidimensional poverty between 2001 and 2021. According to the ensemble model predictions, this progress became particularly evident during the 2010s [51]. The observed increase in multidimensional poverty in Senegal around 2010 may reflect the lingering effects of the 2008 food crisis [52].

Finally, we highlight the importance of using beta regression in modeling. It is worth noting that beta regression models remain in the boundaries of the $(0, 1)$ interval, meanwhile other models that assume continuity predict negative poverty levels when the ground truth is close to zero. This is more evident in the predictions of Bolivia with `Elastic Net` where we observe a great proportion of negative predicted values. This phenomenon also happens with `XGBoost` but to a lower extent. Again, these findings support our modeling approach considering the bounded nature of the target variable.

4. Conclusions

This paper addresses the challenge of data scarcity in global multidimensional poverty measurement by applying advanced statistical and machine learning methodologies to open-source data, such as the WID, contributing to progress on various Sustainable Development Goals.

We explore several statistical learning approaches suited to high-dimensional contexts, comparing methodological strategies within several frameworks: supervised dimension reduction, regularization, and ensemble models. Given the bounded nature of our response variables, we also propose possible extensions to improve model accuracy. Our findings indicate that tree-based boosting models outperform dimension reduction and shrinkage methods, showing enhanced flexibility for predicting short-term poverty shocks—such as those observed in Bolivia and Senegal in 2005 and 2010, respectively. These models also effectively detect structural shifts in poverty trends, as evidenced by changing patterns in Bangladesh around 2010.

One limitation of our approach is that none of the models perform well in predicting poverty intensity. Consequently, when analyzing predicted multidimensional poverty values, it remains unclear whether reductions in poverty are driven by a decrease in the number of poor individuals (H) or by improvements in the well-being of the poor (A).

Lastly, in contrast to previous research, our results demonstrate that adapting machine learning methods to account for the bounded nature of poverty rates can significantly enhance prediction accuracy. Given that many variables in quantitative social science—particularly in fields like economics, sociology, public health, and political science—are defined as rates or proportions, we hope these findings encourage broader application of this approach across diverse domains.

References

[1] Alkire, S., Santos, M.E.: Measuring acute poverty in developing world: Robustness and scope of the multidimensional poverty index. World Development **59**, 251–274 (2014)

[2] Nations, U.: A/res/2029 (xx), un special fund with the un expanded programme of technical assistance. Technical report, United Nations Development Programme (1965)

[3] Alkire, S., Robson, M.: On international household survey data availability for assessing pre-pandemic monetary and multidimensional poverty in developing countries. Development Studies Research **9**(1), 277–295 (2022) https://doi.org/10.1080/21665095.2022.2141286

[4] Ravallion, M.: How long will it take to lift one billion people out of poverty? The World Bank Research Observer **28**(2), 139–158 (2013)

[5] Crespo Cuaresma, J., Fengler, W., Kharas, H., Bekhtiar, K., Brottrager, M., Hofer, M.: Will the sustainable development goals be fulfilled? assessing present and future global poverty. Palgrave Communications **4**(1) (2018)

[6] Mahler, D.G., Aguilar, R.A.C., Newhouse, D.: Nowcasting global poverty. World Bank Economic Review **36**(4), 835–856 (2022) https://doi.org/10.1093/wber/lhac017 https://elibrary.worldbank.org/doi/pdf/10.1093/wber/lhac017

[7] Lakner, C., Mahler, D.G., Negre, M., Prydz, E.B.: How much does reducing inequality matter for global poverty? The Journal of Economic Inequality **20**(3), 559–585 (2022)

[8] Balasubramanian, P., Burchi, F., Malerba, D.: Does economic growth reduce multidimensional poverty? evidence from low- and middle-income countries. World Development **161**, 106119 (2023) https://doi.org/10.1016/j.worlddev.2022.106119

[9] García Arancibia, R., Girela, I.: Graphical representation of multidimensional poverty: Insights for index construction and policy making. Social Indicators Research **172**(2), 595–634 (2024)

[10] Felix, J., Alexandre, M., Lima, G.T.: Applying machine learning algorithms to predict the size of the informal economy. Computational Economics, 1–21 (2024) https://doi.org/10.1007/s10614-024-10593-6

[11] Chakraborty, T., Chakraborty, A.K., Biswas, M., Banerjee, S., Bhattacharya, S.: Unemployment rate forecasting: A hybrid approach. Computational Economics **57**(1), 183–201 (2021) https://doi.org/10.1007/s10614-020-10040-2

[12] Alkire, S., Nogales, R., Quinn, N.N., Suppa, N.: On track or not? projecting the global multidimensional poverty index. Journal of Development Economics **165**, 103150 (2023) https://doi.org/10.1016/j.jdeveco.2023.103150

[13] Alkire, S., Foster, J.E.: Counting and multidimensional poverty measurement. Journal

of Public Economics **95**, 476–487 (2011)

[14] Alkire, S., Roche, J.M., Ballon, P., Foster, J., Santos, M.E., Seth, S.: Multidimensional Poverty Measurement and Analysis. Oxford University Press, USA, ??? (2015)

[15] Ferrari, S., Cribari-Neto, F.: Beta regression for modelling rates and proportions. Journal of applied statistics **31**(7), 799–815 (2004)

[16] Schmid, M., Wickler, F., Maloney, K.O., Mitchell, R., Fenske, N., Mayr, A.: Boosted beta regression. PloS one **8**(4), 61623 (2013)

[17] Cook, R.D., Forzani, L.: Big data and partial least-squares prediction. Canadian Journal of Statistics **46**(1), 62–78 (2018)

[18] Duarte, S., Forzani, L., García Arancibia, R., Llop, P., Tomassi, D.: Socioeconomic index for income and poverty prediction: A sufficient dimension reduction approach. Review of Income and Wealth **69**, 318–346 (2023)

[19] D'Iorio, S., Forzani, L., García Arancibia, R., Girela, I.: Predictive power of composite socioeconomic indices for targeted programs: principal components and partial least squares. Quality & Quantity **58**(4), 3497–3534 (2024)

[20] Huerta, M., Leiva, V., Lillo, C., Rodríguez, M.: A beta partial least squares regression model: Diagnostics and application to mining industry data. Applied Stochastic Models in Business and Industry **34**(3), 305–321 (2018) https://doi.org/10.1002/asmb.2278 https://onlinelibrary.wiley.com/doi/pdf/10.1002/asmb.2278

[21] Bertrand, F., Maumy-Bertrand, M.: Partial least squares regression for beta regression models. useR! (2021)

[22] Cook, R.D., Forzani, L.: PLS regression algorithms in the presence of nonlinearity. Chemometrics and Intelligent Laboratory Systems **213**, 104307 (2021)

[23] Grün, B., Kosmidis, I., Zeileis, A.: Extended beta regression in r: Shaken, stirred, mixed, and partitioned. Journal of Statistical Software **48**(11), 1–25 (2012)

[24] Zeileis, A., Hothorn, T., Hornik, K.: Model-based recursive partitioning. Journal of Computational and Graphical Statistics **17**(2), 492–514 (2008)

[25] Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology **58**(1), 267–288 (1996)

[26] Hastie, T., Tibshirani, R., Wainwright, M.: Statistical Learning with Sparsity: the Lasso and Generalizations. CRC press, ??? (2015)

[27] Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology **67**(2), 301–320 (2005)

[28] Friedman, J.H., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent. Journal of statistical software **33**(1), 1–22 (2010)

[29] Tay, J.K., Narasimhan, B., Hastie, T.: Elastic net regularization paths for all generalized linear models. Journal of Statistical Software **106**(1), 1–31 (2023) https://doi.org/10.18637/jss.v106.i01

[30] Meinshausen, N.: Relaxed lasso. Computational Statistics & Data Analysis **52**(1), 374–393 (2007) https://doi.org/10.1016/j.csda.2006.12.019

[31] Hastie, T., Tibshirani, R., Tibshirani, R.: Best subset, forward stepwise or lasso? analysis and recommendations based on extensive comparisons. Statistical Science **35**(4), 579–592 (2020)

[32] Zhang, Y., Ma, F., Wang, Y.: Forecasting crude oil prices with a large set of predictors: Can lasso select powerful predictors? Journal of Empirical Finance **54**, 97–117 (2019)

[33] Breiman, L.: Bagging predictors. Machine Learning **24**(2), 123–140 (1996) https://doi.org/10.1007/BF00058655

[34] Breiman, L.: Random forests. Machine Learning **45**(1), 5–32 (2001) https://doi.org/10.1023/A:1010933404324

[35] Wolpert, D.H.: Stacked generalization. Neural Networks **5**(2), 241–259 (1992) https://doi.org/10.1016/S0893-6080(05)80023-1

[36] Polikar, R.: Ensemble learning. In: Zhang, C., Ma, Y. (eds.) Ensemble Machine Learning: Methods and Applications. Machine Learning Series, pp. 134–142. Springer, Boston, MA, USA (2012). https://doi.org/10.1007/978-1-4419-9326-7_1

[37] Konstantinov, A.V., Utkin, L.V.: A generalized stacking for implementing ensembles of gradient boosting machines. Cyber-Physical Systems: Digital Technologies and Applications, 3–16 (2021)

[38] Schapire, R.E.: The strength of weak learnability. Machine Learning **5**(2), 197–227 (1990) https://doi.org/10.1007/BF00116037

[39] Freund, Y., Schapire, R.E.: A short introduction to boosting. In: Proceedings of the 16th International Joint Conference on Artificial Intelligence, pp. 1401–1406 (1999)

[40] Breiman, L.: Arcing classifiers. Annals of Statistics **26**(3), 801–824 (1998)

[41] Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016). https://doi.org/10.1145/2939672.2939785

[42] Bentéjac, C., Csörgő, A., Martínez-Muñoz, G.: A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review **54**, 1937–1967 (2021) https://doi.

org/10.1007/s10462-020-09896-5

[43] Mienye, I.D., Sun, Y.: A survey of ensemble learning: Concepts, algorithms, applications, and prospects. IEEE Transactions on Cybernetics **53**(10), 6355–6375 (2023)

[44] Mayr, A., Weinhold, L., Hofner, B., Titze, S., Gefeller, O., Schmid, M.: The betaboost package—a software tool for modelling bounded outcome variables in potentially high-dimensional epidemiological data. International Journal of Epidemiology **47**(5), 1383–1388 (2018) https://doi.org/10.1093/ije/dyy093

[45] Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Annals of Statistics, 1189–1232 (2001)

[46] OPHI: All Published Global Multidimensional Poverty Index (MPI), Results 2010-2022. Database, Oxford Poverty and Human Development Iniatiative, Univerity of Oxford (2022)

[47] Pasha, A.: Regional perspectives on the multidimensional poverty index. World Development **94**, 268–285 (2017) https://doi.org/10.1016/j.worlddev.2017.01.013

[48] Csiszár, I., Shields, P.C., *et al.*: Information theory and statistics: A tutorial. Foundations and Trends® in Communications and Information Theory **1**(4), 417–528 (2004)

[49] Zheng, Y., Yang, F., Duan, J., Kurths, J.: Quantifying model uncertainty for the observed non-gaussian data by the hellinger distance. Communications in Nonlinear Science and Numerical Simulation **96**, 105720 (2021) https://doi.org/10.1016/j.cnsns.2021.105720

[50] Villarroel, P., Hernani-Limarino, W.L.: La evolución de la pobreza en bolivia: un enfoque multidimensional. Revista Latinoamericana de Desarrollo Económico **20**, 7–74 (2013)

[51] World Bank: Bangladesh development update - new frontiers in poverty reduction. Technical report, World Bank, Dhaka (October 2023)

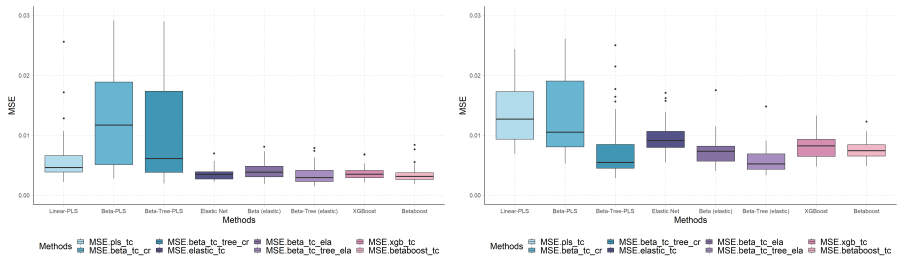
[52] Compton, J., Wiggins, S., Keats, S.: Impact of the global food crisis on the poor: what is the evidence. Technical report, Overseas Development Institute, London (2010)

Appendix A Possible datasets from World Bank web scrapping

Data sets	*Number of countries*	*Observations*	Number of WDI/predictors
1	**106**	**249**	**110**
2	**75**	**218**	**167**
3	63	165	226
4	58	155	258
5	**53**	**142**	**336**
6	48	130	345
7	46	125	351
8	**43**	**118**	**360**
9	39	108	425
10	**38**	**106**	**445**
11	36	102	465
12	33	95	474
13	**33**	**94**	**477**
14	27	75	526
15	**26**	**73**	**579**
16	26	72	581
17	26	71	582
18	18	54	607
19	16	50	610
20	16	49	612
21	14	44	618
22	13	40	635
23	12	36	644
24	11	33	651
25	11	32	652
26	7	17	751
27	4	8	858
28	1	2	1112
29	0	0	1478

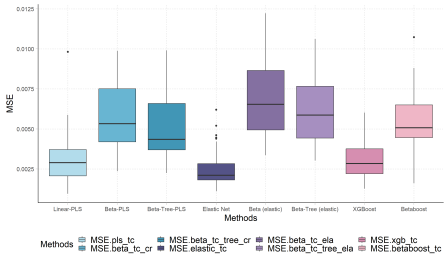
Table A1: Alternative Data sets to predict MPI using World Bank indicators

Appendix B Distribution of Prediction Errors from Experiment 1



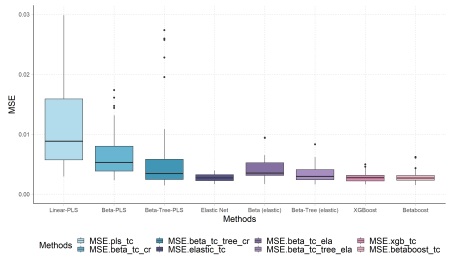
(a) M_0

(b) H

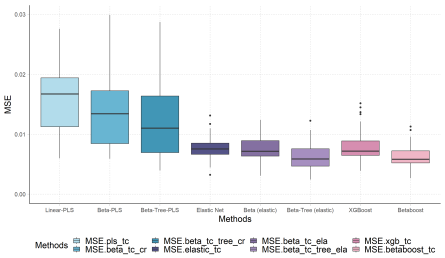


(c) A

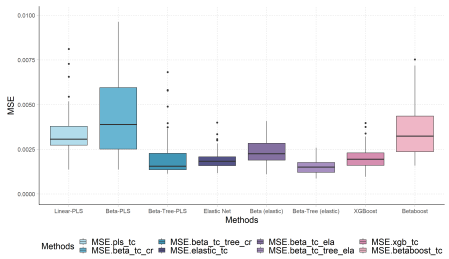
Fig. B1: Prediction Errors Distribution (MSE) for dataset 1 from 50 repetitions



(a) M_0

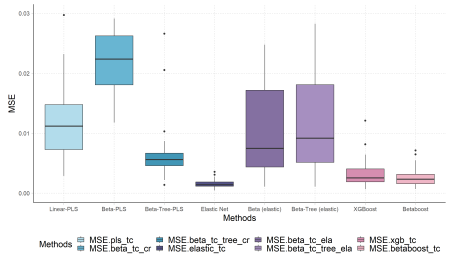


(b) H

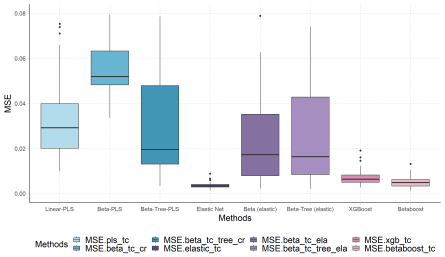


(c) A

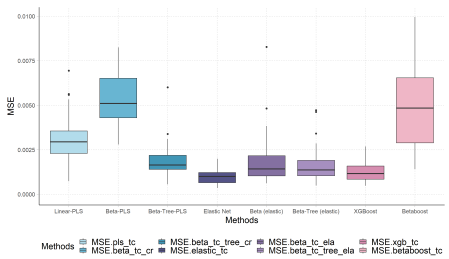
Fig. B2: Prediction Errors Distribution (MSE) for dataset 2 from 50 repetitions



(a) M_0



(b) H



(c) A

Fig. B3: Prediction Errors Distribution (MSE) for dataset 13 from 50 repetitions