Micro and Macro Indexes of Economic Activity: Multiple Indicators and Multiple Methods Using Bangladesh as a Test Case

Mariha Tahsin

A thesis submitted for the degree of Doctor of Philosophy (Ph.D.) in Economics

Department of Economics

University of Essex

2nd September 2022

Declaration of Authorship

I, Mariha Tahsin, confirm that the research included within this thesis is my own work.

I attest that I have exercised reasonable care to ensure that the work is original, and does not, to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the university and/or examiners have the right to use plagiarism detection software to check the electronic version of the thesis. I also confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Acknowledgements

The PhD journey is long, arduous, and often lonely. It feels surreal to have reached the end of the road. Writing this section is an opportunity to count my blessings, which often came in the form of people and relationships.

I would like to start by expressing my sincere gratitude to my supervisors Professor Marco Francesconi, Andreau Javier, and Professor Thomas Crossley. They each entered my PhD journey at different points, and guided me, while still allowing my research to take its own course.

I would also like to thank Dr. Fouzul Kabir Khan for providing me with the opportunities that eventually led me down the PhD route. His contributions to my academic, professional, and personal growth are countless.

On a personal note, I am grateful to my family – both immediate and extended, for providing me with ample support and going out of their way to make my life easy. My husband, Mashkur, is himself deserving of an award for picking up my share of the responsibilities; so, I had the time and energy to pursue my goals. Finally, I would like to dedicate this body of works to my parents and grandparents: Abbu, Ammu, Dada, Dadu, Nanabhai and Nanua.

Abstract

The first chapter explores, the use of night-time lights as a proxy for estimating annual GDP per capita and subsequently the GDP per capita growth rate. It is observed that even though, Bangladesh's, GDP per capita is under-estimated, the annual growth rate is over-estimated.

The second chapter explores the quality of the household surveys conducted in Bangladesh through the application of the Benford's Law and triangulation against administrative data. Sampling errors are detected in all rounds of the household surveys. The results showed that the micro dataset over-sampled wealthier households. This indicated that the income and expenditure levels of the three lowest quartiles, estimated from the household surveys, is likely over-estimated.

The results of the first two chapters are then combined to comment on the state of inequality in Bangladesh. It is observed that GDP per capita is higher than expected, while the income of the lowest three income quartiles is lower than estimated. Thus, true inequality is likely to be much higher than what is indicated by the published Gini-coefficients.

The fourth chapter assesses the accuracy of a proxy-means test, the Poverty Probability Index, in classifying household poverty, in the absence of sound data. Applicability of the PPI, over years and across population sub-groups, was tested. It was seen that the index overestimated poverty probability in both cases.

In the last chapter, machine learning algorithms are used to develop alternatives to the Poverty Probability Index, in the absence of extensive domain knowledge. These models out-performed the PPI by 3 percentage points in terms of accuracy and ROC-AUC

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Equations

Introduction

This thesis is separated into five chapters. The first chapter empirically investigates macro data quality by using a night-lights model to predict GDP (Gross Domestic Product) per capita. These results are then used to comment on the quality of macroeconomic data of Bangladesh. The second chapter explores the quality of the chief micro data source in Bangladesh i.e., the household income and expenditure surveys. Chapter 3 builds on the results of the first two chapters to comment on the status of income inequality of Bangladesh. Chapter 4 investigates the justification of using the proxy-means test- Poverty Probability Index (PPI) for household poverty classification. The performance of the PPI in classifying household poverty status across time and sub-populations is examined. The last chapter develops alternatives to the PPI, using machine learning algorithms. All five chapters use Bangladesh as a test case.

In the first chapter, the GDP per capita of 171 countries is estimated using a robust fixed-effects model taking night-time lights as a regressor, for the 1992 to 2020 period. Predicted log transformed GDP per capita values were within 18 percent of the true values for approximately 88 percent of the countries. However, 21 countries; all of whom were weak in terms statistical capacity and economic development; had mean residuals exceeding 18 percent. Bangladesh was one of these 21 countries. The mean log transformed per capita GDP estimated by the model, for Bangladesh, was between 9.16 and 9.23 per year. The government reported value was only 7.68. However, the annual per capita growth rate has been overestimated in recent years, thus, the true and predicted values appear to be converging over time. The differences are most likely a result of the large informal economy that is gradually being factored into GDP calculations.

The second chapter evaluates the chief sources of micro data for Bangladesh, i.e., the household income and expenditure surveys (HIES)[1]. The main objective of this chapter is to assess the quality of the micro data that is available for Bangladesh for research and policymaking. To do this, the micro data, for mean income, mean expenditure and mean consumption; estimated using the household surveys, were compared against the corresponding macro data i.e., Gross National Income (GNI), total expenditure and Private Final Consumption Expenditure (PFCE). In most periods, the macro values were larger than the micro values. These divergences were substantially lessened by using a consistent national accounting system across all periods. However, this did not eliminate the difference. The macro values remained larger than the corresponding micro values. Two methods were applied to identify the source of these incongruities. Firstly, whether the distribution of households' monthly expenditure conformed to Benford's Law was evaluated. It was observed that for every period the distributions fail to follow the expected pattern. Non-conformity with the Benford's Law provided indication of the presence of sampling error or data manipulation in the household surveys. Secondly, the relationship of income status and income tax payment, electrification, and vehicle ownership were determined using logistic regression models. The models used a binary for the households' income quartile as the dependent variable and income tax payment, electrification, and vehicle ownership as the independent variables. It was determined that all three of the explanatory variables were strongly and positively correlated with households' position in the top income quartile. By comparing administrative data of the explanatory variables against the household survey data, sampling errors were identified. This exercise revealed that the household surveys over-sampled the top income quartile of the population but under-sampled the income tax-paying segment of the population.

The third chapter used the results of the first three as building blocks, to determine the true state of inequality in Bangladesh.

[1] The Household Income and Expenditure Survey is the largest national survey conducted by the government of Bangladesh. Since 1973, 14 rounds have been conducted. Each round collects income, expenditure, and consumption data for a nationally representative sample of people. Important progress indicators, such as per capita calorie consumption, average household income and poverty reduction are estimated from this survey.

The first three chapters establish that both micro and macro level data available for Bangladesh suffer from quality issues. The micro and macro data used for the analyses were all collected by the Government of Bangladesh (GoB). The GoB invested considerable financial and human resources in the data collection and vetting processes. Despite the high investment, reliability of data could not be ensured.

This raises the question of the dependability of data collected through less rigorous data collection methods. For instance, income and expenditure data are often collected by development organizations for beneficiary selection. These are instances in which only limited resources can be allocated to data collection and data quality checks. At the same time, respondents often lack the required knowledge to provide accurate income and expenditure data and/or have incentive to provide misinformation. These situations demand the use of proxy-means tests (PMTs). A popular PMT used across governmental and non-governmental organizations in Bangladesh is the PPI. The PPI is a household-level poverty classification tool that uses 10 socioeconomic indicators to assign households a score between 0 and 100. The likelihood of falling below specified poverty lines, at each score level, is estimated. The PPI's ease of application has led to the index being adapted by most major microfinance institutions and other large-scale development projects.

The fourth chapter of this thesis investigates whether the PPI is an appropriate measure for profiling poverty. This is done by testing the reproducibility, replicability, and the generalizability of the PPI. The poverty probabilities predicted by the PPI are re-estimated by replicating the PPI's construction methodology on the original database. The PPI is observed to over-estimate poverty probability within a 5-year period. The index also loses predictive ability when applied to population sub-groups. Additionally, the PPI is tested on a panel dataset for the first time. This exercise estimates the PPI's ability to track falling into poverty at 35 percent. On the other hand, it is observed that the PPI can correctly track graduating out of poverty about 75 percent of the time. Given the widespread use of the PPI, this analysis has implications for development practitioners as well as researchers.

The construction of the PPI involved a complicated method, requiring in-depth domain knowledge and time-consuming stepwise regression methods to select the indicators and assign them appropriate weights. In the last chapter, machine learning algorithms are used to develop an alternative to the PPI in the absence of extensive domain knowledge. Effort is made to incorporate as little human judgment as possible. It is found that machine learning algorithms can build better classification tools with just as few indicators. These models outperformed the PPI by 3 percentage points on accuracy and ROC-AUC metrics. However, unlike humans, machines are unable to distinguish between easily verifiable data and hard to quantify data. Thus, the best models developed by the algorithms included features that would not make good indicators, e.g., calorie consumption from a specific food.

1 Assessing the Reliability of Macro Data Using Night-time Lights Models: Bangladesh

Abstract: the GDP per capita of 171 countries is estimated using a robust fixed-effects model taking night-time lights as a regressor, for the 1992 to 2020 period. Predicted log transformed GDP per capita values were within 18 percent of the true values for approximately 88 percent of the countries. However, 21 countries; all of whom were weak in terms statistical capacity and economic development; had mean residuals exceeding 18 percent. Bangladesh was one of these 21 countries. The mean log transformed per capita GDP estimated by the model, for Bangladesh, was between 9.16 and 9.23 per year. The government reported value was only 7.68. However, the annual per capita growth rate has been overestimated in recent years, thus, the true and predicted values appear to be converging over time. The differences are most likely a result of the large informal economy that is gradually being factored into GDP calculations.

1.1 Introduction

Development statistics are relevant for politicians, policymakers, technocrats as well as donors. The ability to collect accurate information is not only important in directing policymaking, but also in ensuring maximum collection of taxes. Despite, the well documented need for accurate and reliable statistics, the issue of inaccurate statistics is a universal one (Jerven, 2013). Concerningly, poor nations that could most benefit from properly kept accounts, often have the most ill-managed ones, due to the lack of funds available to maintain and create them. The case of Bangladesh is no different. Complaints surrounding the quality of national accounts are frequent and many. Despite decades of news reports and anecdotes of the large discrepancies in real and reported numbers, very little has been done to take stock of the issue (Osmani, 2017).

The lack of reliable statistics negatively impacts decision making in several ways. Firstly, it hinders the monitoring of social and economic indicators. Secondly, it makes it difficult to assess the political and socioeconomic needs of a country, leading to misallocation of resources. Thirdly, evidence-based decision-making in investment, program design or policy development are obstructed.

1.1.1. Statistical Capacity

According to data published by the World Bank, Bangladesh had a Statistical Capacity Indicator (SCI) score of 60 out of 100 in 2020, which is considerably lower than the South Asian average of 69.81. The SCI scores reflect a country's ability to collect, analyze and disseminate high-quality data about its population and economy. The scores are based on performance in four categories - methodology, source data, periodicity, and timeliness; and assessed based on 25 indicators (World Bank, 2020). The data reveals that the SCI score for Bangladesh has fallen by 20 points in the previous 6 years. Further disaggregation of the data shows that a large share of this fall is attributable to deterioration in methodology, a small portion is due to periodicity. Overall Bangladesh ranked 75th (tied with 5 others) amongst the 141 countries the SCI score was calculated for. Most countries that ranked below Bangladesh are low-income or war-torn nations.

This, shortcoming in statistical capacity is evident in the mismatch of available data for Bangladesh. Values of important statistics such as GDP growth rate reported by national and international organizations fail to match. Figure 1.1 illustrates the difference in real GDP growth rates estimated for Bangladesh, by the Asian Development Bank (ADB), World Bank (WB), International Monetary Fund (IMF), and the Government of Bangladesh (GoB) for 2019-20 and for 2020-21. It is observed that for every period, the real GDP growth rates estimated differ for the four organizations that data is presented for. The estimates provided by WB and IMF tend to be more conservative than those provided by the GoB and ADB.

It is important to note that the WB's data comes from statistical systems of member countries. Thus, the quality of the data is dependent upon the performance of national systems (World Bank, 2022). Since the WB does not collect its own data, the difference in the GDP growth rate estimates of the WB and the GoB must result from methodological discrepancies. The same is true for ADB (Asian Development Bank, 2022) and IMF (International Monetary Fund, 2022). All these organizations source data from the Bangladesh Bureau of Statistics.

Figure 1.1: Real GDP Growth Rate Estimates

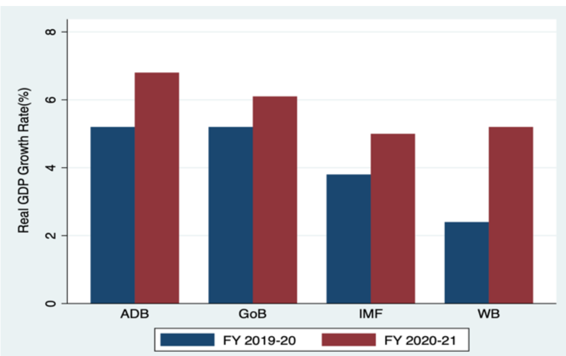


Fig 1.1. The figure compares the real GDP growth rate estimates published by the Government of Bangladesh (GoB), Asian Development Bank (ADB), International Monetary Fund (IMF) and World Bank (WB). It is evident that the estimates of the four organizations differ, and that the official government estimates tend to be on the higher side.

1.1.2. The Double Paradox of Bangladesh

In June 2021, GDP per capita for Bangladesh stood at USD 2,227 per annum. Meanwhile, the GDP per capita values of Pakistan and India, both countries Bangladesh had formerly been a part of, stood at USD 1,543 and USD 1,947 respectively (Sharma, 2021). While Bangladesh's GDP has been expanding at an accelerating pace over the past few decades, this is the first instance of the country overtaking India in per capita terms. This has fueled national and international debates regarding the validity and the quality of the national accounts of Bangladesh (Babones, 2020).

Inadequate governance, corruption and weak infrastructure have all been cited by economists and critics as proof that the government published national accounts are misleading at best and fabricated at worst. The sustained economic growth of the country, despite the poor state of governance and high-levels of corruption is described as Bangladesh's 'double-paradox' (Asadullah & Chakravorty, 2019).

Reported nominal per capita GNI and nominal GDP per capita of Bangladeshis have been growing steadily since the country's independence in 1971. In July 2015 the country joined the WB specified lower-middle-income country (LMIC) category, six years ahead of schedule. At the time of classification, countries with GNI per capita values between USD 1,026 and USD 4,035 were tagged as LMIC. Bangladesh's graduation to the status of LMIC, was heavily politicized. While the ruling party presented it as proof of their efficiency, critics claimed the numbers had been fabricated.

On the other hand, the country did not meet the requirements to graduate from Least Developed Country (LDC) under the United Nation's (UN's) classification system till November 2021. This is because, in addition to GNI per capita, the UN uses the Human Assets Index (HAI) and the Economic Vulnerability Index (EVI) to group countries (United Nations, 2021).

The World Bank uses its classifications to assess the credit worthiness of a country. An immediate result of becoming a middle-income country (MIC), is higher associated costs of external borrowing. The UN, uses its classification system to lend international support to reduce a country's structural deficits (World Bank, 2016).

Since the UN classification system uses two additional indexes to assess a country's position, it might be argued, that it gives a more holistic picture of the country's progress. It appears that despite unfavourable scores in the HAI and EVI, Bangladesh was trapped into paying higher interests on its external loans, due to its middle-income status. The issue becomes more pressing, when the changes to be faced due to LDC graduation are taken under consideration. Since Bangladesh successfully met the criteria set by the UN in 2021, the country is ear-marked for graduation from LDC in 2024. Once it is ear-marked, there is no option to opt out, and thus the facilities extended to Bangladesh were immediately revoked (Bhattacharya, 2018). In the area of development cooperation, the UN extends information, advocacy, Official Development Assistance (ODA) on special terms, capital and technological resources, tools to deal with the impact of climate change and Investment Support Programs (ISPs). The UN also provides diplomatic training, travel support to attend inter-governmental meetings and caps and discounts on the contribution to the United Nations system budgets (United Nations, n.d.).

Economists and experts in Bangladesh have argued that the increase in GNI and GDP experienced over the last decade have largely been a result of methodological changes and errors (Osmani, 2017). They are concerned that the errors have led to the inflation of GDP and GNI. The new calculation methodology allows for the quantification and inclusion of previously excluded sectors. Thus, the GDP and/or GNI figures calculated using the new methodology are no longer comparable to the old ones. This implies, that growth rates estimated using these inconsistent methodologies would be an over-estimation.

From an economic perspective, these statistics influence the actions of other agents in the economy-trade partners, multilateral organizations, and foreign investors. Thus, reliable estimates are imperative for a well-functioning economy. In the case of Bangladesh, these concerns are further amplified by the possibility that inflated GNI data was used to allow Bangladesh to graduate out of LDC status and into LMIC status. This could potentially create a scenario in which, Bangladesh loses the facilities available to other countries in its economic position, while paying higher interests on its external loans. Thus, placing the country in a disadvantaged position (Bhattacharya et al., 2019).

In this chapter, we explore whether the statistical concerns raised by critics hold merit, using night-time lights (NTL) models. Available literature is reviewed in the next section. That is followed by a description of the data used for analyses and the empirical strategy. The last two sections elaborate on the results and draw conclusions, respectively.

1.2. Literature Review

NTL have been used in recent economic literature to gage the quality of macroeconomic indicators and/or the level of economy activity. The relationship has been explored at the national and regional levels. The main findings of these papers are summarized below.

Firstly, change in visible lights from outer space is found to be a useful measure of GDP growth, when measurement error in light growth series is assumed to be uncorrelated to the measurement errors in traditional GDP measurement. Generally, the growth estimates of countries with poor statistical capacity had to be revised significantly when NTL methods were implemented. The technique was also capable of producing growth estimates at the city or regional levels when other data is unavailable. The optimal estimate of growth was found to be a composite with roughly equal weights on conventionally measured growth and growth predicted using NTL (J. V. Henderson et al., 2012). The revised estimates were up to 3 percentage points different from official data, annually. This method also enables growth measurement for sub and supranational regions.

The effects of population density and income per capita are easily observed through NTL. The visibility of light is dependent on both. For instance, despite having similar levels of income, the Northeast of the United States is much brighter than the West, due to being more densely populated. On the other hand, though India is more densely populated than Japan, it shows up as significantly dimmer, due to its lower level of economic activity. Thus, it can be posited that income per capita is one of the determinants of visible light, allowing the measurement of income through NTL. However, the NTL method is limited by the facts that light is not produced in a fixed ratio to output, and that true light is imperfectly measured by satellites. Humidity, reflectivity and excluded time periods are not consistent across the globe, differencing out some location-specific factors (J. V. Henderson et al., 2011). Several models are tested with time-fixed regression, on a set of 15-year panel data of 169 countries globally and regionally. In these, NTL is considered a consumer good to determine its relationship with GDP. It is found that light consumption propensity is affected by GDP per capita, latitude, spatial distribution of human activities and the gross saving rate. GDP per capita and light consumption per capita displayed an inverse-U relationship (Wu et al., 2013).

Using a parallelized spatial analytics platform to process twenty-one years of NTL data collected by the Defense Meteorological Satellite Program (DMSP), global long-term relationships between NTL and a series of socio-economic indicators are uncovered. It is found that NTL has the strongest correlation with electricity consumption, carbon emissions and GDP. Weaker correlations are observed with population, methane emissions, nitrogen dioxide emissions, poverty (inverse) and F-gas emissions. The variability in NTL can be explained to a great degree by electricity use in a basic logarithmic regression model (Proville et al., 2017).

As governments scale-back time and resource investment in data collection, developing alternative measures for urbanization, density, and economic growth have become imperative. One of these proxies has been NTL. Using a combination of correlation analysis and geographically weighted regressions, it was found that for Sweden, NTL was a good proxy for population and establishment density. A close relationship was deciphered between radiance light and economic activity. This relationship was weaker with saturated light. However, the relationship did slightly overestimate economic activity in urban areas, while underestimating activity in rural areas (Mellander et al., 2015).

Analysis of three time periods of NTL data for the US and India reveal that night lights can function as a proxy for local and sub-regional income at the 95 percent confidence interval. This is applicable for 25 to 33 percent of the lower 48 states of the US and 770 of the 3,100 plus counties of India. Though, the percentages of sub-regions for which the NTL method applies is small, this number as well as the spatial patterns remain nearly constant across the three time periods. This suggests a limited promise of substituting night lights with GDP. However, this number was much smaller for China, disabling the NTL method as a good proxy for GDP for regions in China. The failing was traced down to the light saturation level picked up by satellite sensors (Kulkarni et al., 2011).

Another study tested the relationship between GDP and NTL at the sub-national level for India using a multinomial non-linear regression technique. It found that GDP is significantly explained by NTL in the area. It was also observed that non-linearity is stronger for metropolitan cities, where GDP levels are higher than a linear model can explain. Conversely, in agriculture and forestry dependent areas, the use of NTL overestimates GDP (Bhandari & Roychowdhury, 2011).

Other papers that tested the relationship between regional data and NTL production reached weaker results. Using data from Brazil, and India; two large emerging economies; it was seen that the relationship between NTL growth and observed GDP growth varied, statistically and economically significantly, across regions. The same was observed for advanced economies like the United States and Western Europe. The only stable relationship that was observed was among urban counties in Brazil (Bickenbach et al., 2016).

Economic convergence among sub-national regions of Bangladesh were observed for the 1992-2013 period using an NTL model. The results showed an absolute divergence and conditional convergence in NTL intensity across the 544 upazilas (sub-districts) of the country. Relying on the NTL model, it appears that the less economically active regions are catching up, albeit at a very slow speed (Basher et al., 2021).

Elasticity of GDP figures to night-lights is systematically larger in non-democratic regions globally. The results indicated that the growth rates were inflated by 1.15 to 1.3 times in the most authoritarian regimes (Martinez, 2017). A similar model implemented to test the GDP data published by the government of Panama between 1996 and 2012, revealed that the GDP values reported were approximately 19 percent higher than predicted by the NTL model. This amounted to approximately USD 40 billion. The results suggest that governments may engage in political manipulation of government statistics to improve the appearance of government performance (Marx & Ziegler Rogers, 2017).

Overall, it appears that there is evidence to support that NTL can function as a reasonable proxy for GDP at the national level. The findings at the sub-national level are inconclusive at best. In fact, most of the research indicates that NTL is an unreliable model to estimate regional GDP. However, since it does show promise for use at the national level, it can be used to evaluate government reported national GDP and GDP growth. These findings can be extrapolated to comment on the data management/manipulation techniques of democratic and non-democratic governments.

Since the only paper that analyzed NTL data for Bangladesh, did not re-estimate national GDP or GDP growth using NTL, this paper will attempt to do so, and compare the values to those published by the GoB.

1.3. Data

The National Oceanic and Atmospheric Administration (NOAA) has been collecting data on average visible lights, average stable lights, and average cloud free coverage, using satellites since 1992. This data set is available for a 21-year period, from 1992 to 2013. The values are presented as cloud-free composites made using all the available smooth resolution data archived by the DMSP Operational Line-Scan System. For some years, data was collected using two satellites, thus, two composites were produced. The products are 30 arc second grids, spanning between -180 to 180 degrees longitude and -65 and 75-degrees latitude.

To collect the highest quality data, all data is collected from the center half of the 3,000 km wide OLS swaths. This was chosen because, lights in the center half have better geo-location, are smaller and have more consistent radiometry. Based on the angle of solar elevation, some sunlit data and glare are excluded. In addition, moonlit data are omitted based on the calculation of lunar luminance. Observations with clouds and lighting features from the aurora are also discounted.

Another night-time lights data set- Visible Infrared Imaging Radiometer Suite (VIIRS) is available from 2011 to early 2020. The VIIRS data set is known to be more detailed. It has much better low-light detection capabilities, with a dynamic range of seven orders of magnitude versus only two for DMSP (Gibson, 2020). Thus, the two datasets cannot be used together, without appropriate processing.

Li & Zhou (2017) developed a harmonized dataset that allows the NTL images from the two datasets to be used together. Using this dataset allows us to analyze a longer and more recent period.
The cleaned-up data contains the lights from cities, towns, and other sites with persistent lighting, including gas flares. Ephemeral events such as fires are discarded. Background noise is identified and replaced with values of zero. Illumination is measured in digital numbers (DN), which is a value assigned to pixels based on their brightness. The data ranges within values of 1 to 63, rising with higher degrees of illumination. Areas with zero cloud-free observations are represented as the value of 255 (DMSP OLS, 2014).

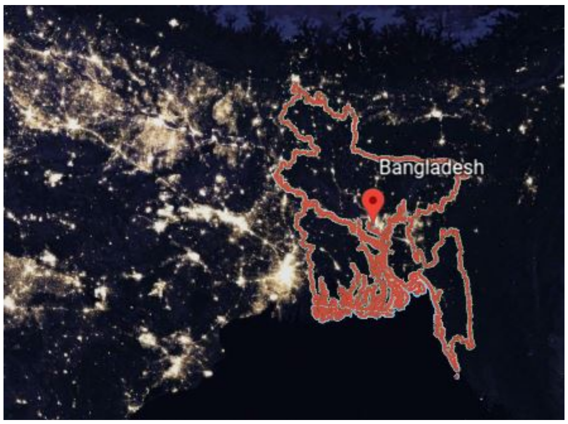
The data conversion process closely followed the steps recommended by Lowe (2014) . The lights data were disaggregated by country using the World Bank's international borders shape file (World Bank, 2021b). For each feature (country), 1,000 random points were generated, and pixel data were analyzed using the random point sampling tool. For years in which, night light data was collected by two satellites, the values were averaged.

The macroeconomic data required for these analyses: GDP per capita, PPP (constant international 2017 $), were collected from World Bank's data bank (World Bank, 2021a).

1.4. Empirical Strategy

Figure 1.2 below is an image of night-time lights of Bangladesh as of June 2021. Even though, pixel level data is not available for this period, it is evident that the areas in which economic activity is expected to occur, such as the capital of Dhaka, is much brighter than the rest of the country. It also appears that light clusters are relatively few and far apart. On the other hand, several larger and brighter night light clusters can be seen in the eastern states of India, despite the population density being lower in those regions. This indicates higher levels of economic activity. The Indian states on the east of Bangladesh which are less economically developed than the rest of India, are dark, as expected.

Figure 1.2: Night-time Lights (June 2021)



Source: The Earth at Night, Google Earth

To test the hypothesis that the degree of illumination of night-time lights is a close proxy for economic activity, panel data of real GDP per capita (World Bank, 2021a)[2] and night-time lights illumination (DMSP OLS, 2014) are used. The data is available for the period 1992 to 2020. There are 5,164,606 observations pertaining to 171 countries in the dataset. By grouping the observations by country and year, the number of observations is reduced to 4,785.

The target variable, i.e., GDP per capita[3], and the explanatory variable, i.e., night-time lights illumination, have right skewed distributions. Thus, by applying logarithmic transformations, their distributions are normalized. The global sample is split into two equal subsets of randomly selected observations. The models are analyzed for both subsets, this allows us to identify whether the models are overfit to their respective samples.

The process is repeated by applying a robustness check to the sample. The sample is restricted to South Asian countries. This group includes Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, and Sri Lanka. It is assumed that due to their proximity, these countries share several geographic, cultural, and economic factors.

1.4.1. Pooled Ordinary Least Squares Model

The first econometric model used to analyze the relationship between log-transformed real annual GDP per capita and log-transformed mean illumination, measured in DN, is a pooled-regression model. It should be noted that this model does not distinguish between the 171 countries in the global dataset, or the 8 countries in the South Asian dataset. By combining these countries, the heterogeneity and/or individuality that may exist among them goes undetected.

The predictive model estimated log-transformed[4] annual real GDP per capita using the form:

Eq 1.1

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

Where,

y: log transformed annual real GDP per capita
x_1: log transformed annual mean illumination measured in DN
β_1: rate of change of annual real GDP per capita in response to a single unit change in x1.
ε: error term

It is assumed that the measurement error in night lights is uncorrelated to the measurement error in real GDP per capita, i.e., $cov(\varepsilon_x, \varepsilon_y) = 0$.

1.4.2. Fixed Effects (FE) Regression Model

The pooled-OLS model fails to factor in characteristics that are constant over time and country. To correct for that a fixed-effects model is used. This

[2] GDP per capita values are presented in PPP (constant 2017 international $)
[3] The GDP per capita figures used in this chapter are PPP-adjusted.
[4] Log transformations were necessary to generate a normally distributed sample. It should be noted that negative values are dropped during transformations of this nature.

also helps avoid omitted variable bias. In our model, it is assumed that the 171 countries of the global dataset, and the 8 countries of the South Asian dataset for which data is available have different intercepts. The FE framework is given below:

Eq 1.2

$$y_{i,t} = \alpha_{i,t} + \beta_0 + \beta_1 x_{1i,t} + \varepsilon_{i,t}$$

Where,

i: country
t: time-period
$\alpha_{i,t}$: unobserved heterogeneity of country *i* at time

1.4.3. Random Effects (RE) Generalized Least Squares (GLS) Regression Model

For managing the multiple sources of random variability in the sample, random effects models are used.

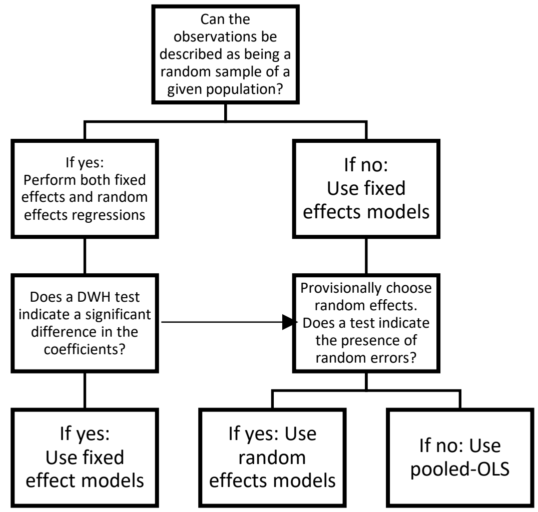
1.4.4. Model Selection

The regression results vary depending on the model used for analysis. To choose the best suited model, the first step is determining whether the sample is random. In this case, the illumination points[5] were chosen using a random point sampling tool.

Once the randomness of the sample is ascertained, fixed effects and random effects regressions are performed. The results of both are compared using the Durbin-Wu-Hausman (DWH) test with robust standard-errors. This tests whether the unique errors are correlated with the regressors. The null hypothesis is that they are not. If the test indicates significant differences in the coefficients, the FE model is better suited. If not, the Lagrange Multiplier (LM) test is used to select between the random effects model and the pooled-OLS model. If the LM test indicates the presence of random effects, the RE model will be chosen. Otherwise, the pooled-OLS model is considered the best option. Figure 1.3 charts the model selection framework.

[5] Illumination is measured in decimal numbers (DN)

Figure 1.3: Model Selection Framework



1.5. Results

Generating a scatter plot of the real GDP per capita against night-time lights illumination (measured in DN) shows that the two share a positive relationship. This is demonstrated in Figure 1.4.

Figure 1.4: Scatter Plot of Real GDP Per Capita and Night-time Lights Illumination

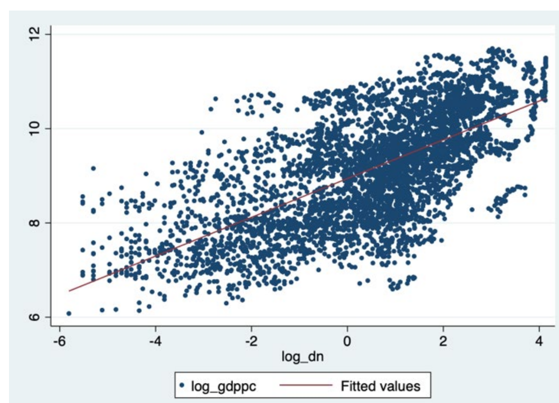


Fig 1.4 plots the log transformed DN data points against the log transformed real GDP per capita data. The figure shows that the best fit line is upward sloping and that there is a slightly higher concentration at the upper end.

1.5.1. Pooled Ordinary Least Squares Model

The pooled-OLS model included 2,357 observations in the training set and 2,352 observations in the test set for the global dataset. The p-value of the F-statistic is less than 0.01, indicating that the model has statistically significant explanatory power. The p-values of the t-statistics of the independent variable is also below 0.01. Thus, the night-time lights illumination is statistically significant even at the 99 percent confidence interval. The coefficient of log transformed DN is 0.41 (2 d.p.). Thus, a single unit shift in the mean of night-time light illumination would change the mean of the dependent variable by .41 times, when all other factors are kept constant[6]. These results are given in Table 1.1. Similar results are obtained for the validation set. These results are provided in Table 1.2.

When the model is run for the South Asian dataset, the number of observations was 96 in the training set and 123 in the test set. The coefficient of log transformed DN is 0.24 (2 d.p.). These results are given in Table 1.3. Similar results are obtained for the validation set. These results are given in Table 1.4.

1.5.2. Fixed Effects (FE) Regression Model

The fixed effects model included 2,357 observations in the test set and 2,352 observations in the validation set. These observations pertained to 171 countries of the global dataset. For this model too, the F-statistic and the t-statistic are statistically significant (p-value<0.01). Thus, it can be asserted with 99 percent certainty that the coefficients of the explanatory variables are not zero; implying that the model has explanatory power. Like the pooled regression model there is a positive relationship between log transformed DN and log transformed GDP per capita, though this coefficient is weaker. The regression results are provided in Table 1.1. In this case too, the results for the validation set are very close to that of the test set. These results are provided in Table 1.2.

When the model is re-run by restricting the sample to South Asian countries, it is observed that the coefficient is a little larger than in the POLS case. Here, it is 0.28 (2 d.p.). These results are given in Table 1.3. The results of the validation set are given in Table 1.4.

1.5.3. Random Effects (RE) Generalized Least Squares (GLS) Regression Model

The RE model also returns statistically significant results for both the global and the South Asian datasets. Thus, the model has explanatory power. The direction of relationship of the explanatory variable with the dependent variable is also the same as that of the FE model. The results are shown in Table 1.1 and Table 1.3, respectively. Like the previous two models, for the RE model too, the training and validation sets obtained similar results.

1.5.4. Model Selection

Firstly, the DWH test is applied to the FE and RE model estimates for the global dataset. The test is applied with robust standard errors to treat the heteroskedasticity in the sample. This test generated a significant result and thus the null hypothesis was rejected i.e., systematic difference in coefficients is detected. So, we accept that the fixed effects model is better.

It is necessary to test the presence of random effects by using the Breausch-Pagan Lagrange Multiplier. The results of this test are also statistically significant and indicate random effects. Thus, the pooled-OLS model is refused.

As the Hausman test refused the RE model and the LM test refused the pooled OLS model, we select with confidence the FE model. In the final step, the FE model is tested using the modified Wald test for groupwise heteroskedasticity. It is found that the p-value of the chi-squared is less than 0.01. Thus, the null hypothesis is rejected due to the presence of heteroskedasticity.

[6] Mean values of night-time lights illumination (DN) is estimated by averaging the DN values of each country by year.

To correct for the heteroskedasticity problem, a robust version of the FE model is run. The results from this are provided in Table 1.1. The table shows that there were 2,357 observations for 171 entities. The entities refer to the number of countries data is available for. The p-value for the F-statistic <0.01, so the coefficients in the model are different than zero. In this model, the errors are positively correlated to the regressors. The value of rho, also known as the intraclass correlation, is 0.95 (2 d.p.), indicating that 95 percent of variance is due to differences across panels. Since the p-values are lower than 0.01, it can be inferred that the explanatory variable has a significant influence on the dependent variable. From the coefficient of log of DN it can be seen that log of real GDP per capita has a significant and positive correlation with log of DN.

The process is repeated for the South Asia dataset. Again, by applying the DWH test to the FE and RE models, it is seen that there are systematic differences in the coefficients. Thus, the fixed effects model is deemed to be better.

The presence of random effects is then tested, by using the Breausch-Pagan Lagrange Multiplier. The results of this test are statistically significant and indicate random effects. Thus, the pooled-OLS model is refused.

As in the case of the global dataset, we select with confidence the FE model. The FE model is tested using the modified Wald test for groupwise heteroskedasticity. It is found that the p-value of the chi-squared is less than 0.01. Thus, the null hypothesis is rejected due to the presence of heteroskedasticity. Thus, a robust FE model is run. These results can be seen in Table 1.3 and Table 1.4. From the coefficient of log of DN it can be seen that log of real GDP per capita has a significant and positive correlation with log of DN. It is also observed that the coefficient estimated by the robust FE model for the global dataset is much larger in magnitude than the coefficient estimated by the robust FE model for the South Asia dataset.

Table 1.1: Pooled, FE, RE and Robust FE Results (Training Set)

	(POLS)	(FE)	(RE)	(Robust FE)
VARIABLES	log_gdppc	log_gdppc	log_gdppc	log_gdppc
log_dn	0.411***	0.150***	0.158***	0.150***
	-0.009	-0.006	-0.006	-0.02
Constant	8.921***	9.057***	9.049***	9.057***
	-0.019	-0.006	-0.065	-0.01
Observations	2,357	2,357	2,357	2,357
R-squared	0.444	0.236		0.236
Number of id		171	171	171

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 1.2: Pooled, FE, RE and Robust FE Results (Validation Set)

	(POLS)	(FE)	(RE)	(Robust FE)
VARIABLES	log_gdppc	log_gdppc	log_gdppc	log_gdppc
log_dn	0.411***	0.143***	0.149***	0.143***
	-0.01	-0.005	-0.005	-0.015
Constant	8.962***	9.100***	9.058***	9.100***
	-0.019	-0.005	-0.065	-0.008
Observations	2,352	2,352	2,352	2,352
R-squared	0.436	0.245		0.245
Number of id		171	171	171

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 1.3: Pooled, FE, RE and Robust FE Results FE (South Asia-Training Set)

	(POLS)	(FE)	(RE)	(Robust FE)
VARIABLES	log_gdppc	log_gdppc	log_gdppc	log_gdppc
log_dn	0.237***	0.270***	0.270***	0.270**
	(0.041)	(0.029)	(0.028)	(0.083)
Constant	8.118***	8.095***	8.166***	8.095***
	(0.067)	(0.031)	(0.232)	(0.058)
Observations	96	96	96	96
R-squared	0.261	0.503		0.503
Number of id		8	8	8

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

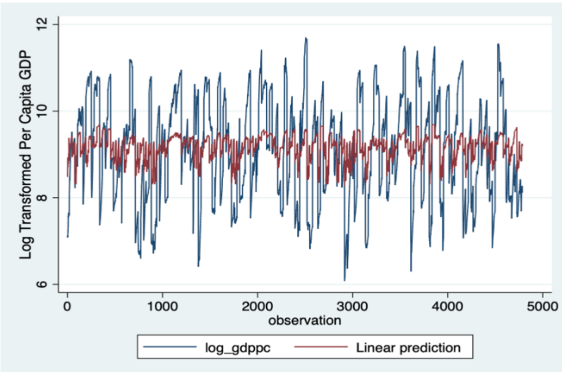
Table 1.4: Pooled, FE, RE and Robust FE Results (South Asia - Validation Set)

	(POLS)	(FE)	(RE)	(Robust FE)
VARIABLES	log_gdppc	log_gdppc	log_gdppc	log_gdppc
log_dn	0.276***	0.294***	0.294***	0.294***
	(0.033)	(0.023)	(0.023)	(0.048)
Constant	8.274***	8.264***	8.198***	8.264***
	(0.055)	(0.023)	(0.216)	(0.027)
Observations	123	123	123	123
R-squared	0.362	0.585		0.585
Number of id		8	8	8

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

1.5.5. Post-estimation Analysis and Regression Diagnostics

Figure 1.5:True and Predicted GDP Per Capita



Using the results of the robust FE model, the log transformed real GDP per capita values were re-estimated based on the log transformed DN coefficient of the global dataset. The predicted values are plotted against the true values in Figure 1.5. The plot indicates that the model can capture trends. However, its ability to predict exact GDP per capita values is limited.

Figure 1.6 plots the residuals of the robust fixed effects model using a histogram and a quartile-quartile plot. The mean of the residuals is calculated to be zero. The quartile graph plots the ordered values of the residuals against the quartiles of a normal distribution. It is observed that the distribution of the residuals is well-behaved, but not perfectly normal. In the quartile graph, it is seen, that the residuals are normally distributed for a large chunk of the sample. Deviation is mostly observed at the upper and lower ends. The residual distribution might be improved by removing the outliers. However, for the model to mimic real-world scenarios, the outliers were retained.

Figure 1.7 includes two panels. The left panel plots the residuals against the regressor (night-time lights illumination measured in DN). It reveals no discernible trend between the two. This indicates that there is a degree of randomness with the residuals, allowing us to not reject the model. The panel on the right plots the true and predicted values of the regressand (real GDP per capita). An upward sloping line appears to connect the two. This linear relationship indicates that the model possesses predictive capabilities.

Figure 1.6: Behaviour of Residuals

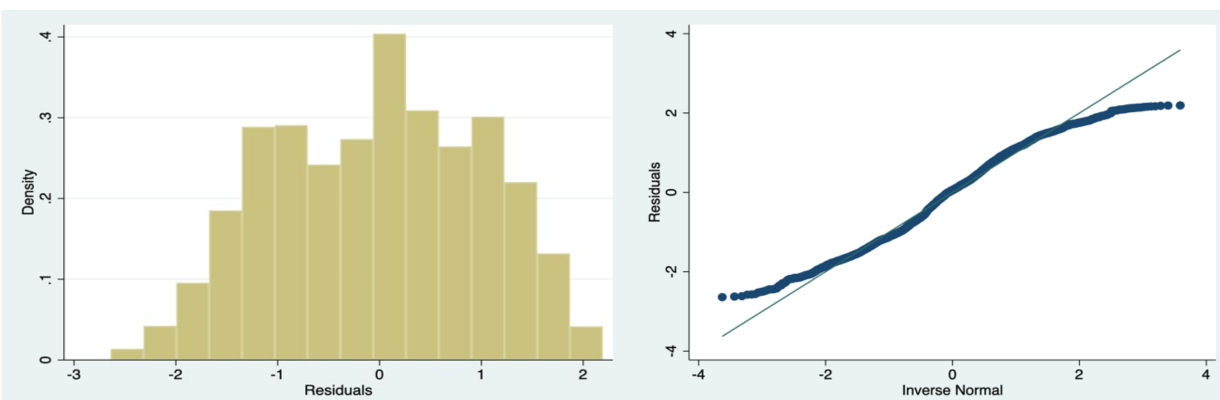


Fig 1.6 plots the residuals of the robust fixed effects model using a histogram and a quartile-quartile plot. The mean of the residuals is calculated to be zero. The quartile graph plots the ordered values of the residuals against the quartiles of a normal distribution. It is observed that the distribution of the residuals is well-behaved, but not perfectly normal. In the quartile graph, it is seen, that the residuals are normally distributed for a large chunk of the sample. Deviation is mostly observed at the upper and lower ends. The residual distribution might be improved by removing the outliers. However, for the model to mimic real-world scenarios, the outliers were retained. Note that these pertain to the global dataset.

Figure 1.7: Scatter Plots (a) Residual Against Night-time Lights Illumination (DN) (b)Predicted and True Values of the Log Transformed GDP Per Capita

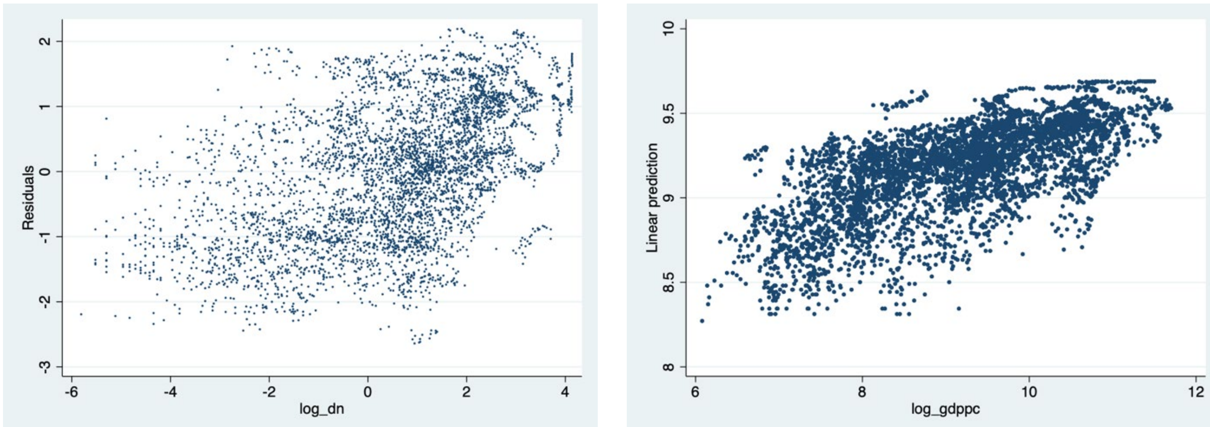


Fig 1.7 includes two panels. The left panel plots the residuals against the regressor (night-time lights illumination measured in DN). It reveals no discernible trend between the two. This indicates that there is a degree of randomness with the residuals, allowing us to not reject the model. The panel on the right plots the true and predicted values of the regressand (real GDP per capita). An upward sloping line appears to connect the two. This linear relationship indicates that the model possesses predictive capabilities. Note that the scatter plots pertain to the global dataset.

The predicted value of log real GDP per capita is calculated, along with the standard errors, using the lower bound and upper bound of the 95 percent confidence interval. This included estimation of the mean real GDP per capita. From this model, it is estimated that, at the 95 percent confidence level, the mean of the log of real GDP per capita of Bangladesh should be within 9.13 (2 d.p.) and 9.39 (2 d.p.). The true mean for this period is 7.84 (2 d.p.). This indicates that the log of real GDP per capita is underestimated for the 1992 to 2020 period for Bangladesh.

To make specific point forecasts, a larger confidence interval is generated, to factor in the greater uncertainty attached to making specific point predictions than there is in predicting means. This is illustrated in Figure 1.8. The panel on the left, plots the regression line, true values, and the mean prediction confidence interval, it is seen that most points fall outside of the interval. The panel on the right, plots the regression line, true values and forecast confidence interval. It is seen that most observations fall within the wider confidence interval. The standard error for forecasts is calculated using the equation below.

Figure 1.8: Confidence Intervals of Mean Prediction and Specific Point Forecast

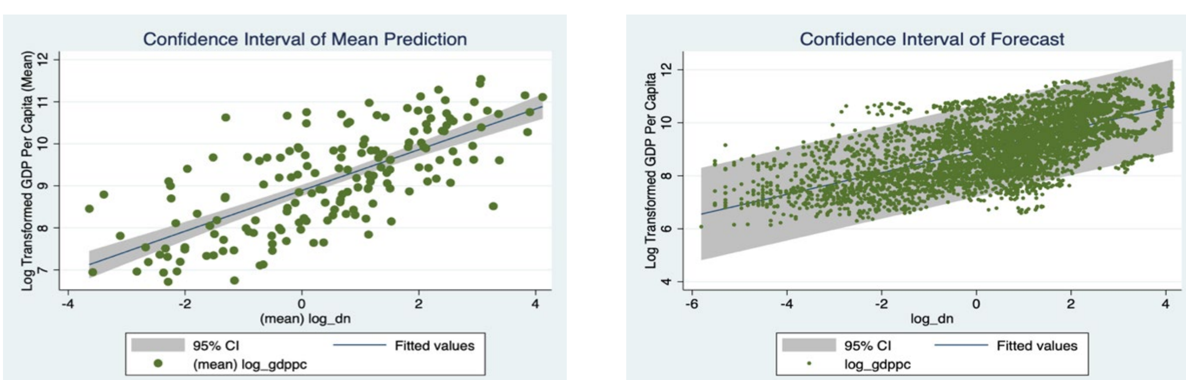


Fig 1.8 The panel on the left, plots the regression line, true values, and the mean prediction confidence interval, it is seen that most points fall outside of the interval. The panel on the right, plots the regression line, true values and forecast confidence interval. It is seen that most observations fall within this wider confidence interval. Note that the diagrams pertain to the global dataset.

Eq 1.3

$$square\ of\ standard\ error\ of\ forecast$$
$$= square\ of\ standard\ error\ of\ mean\ prediction$$
$$+ variance\ of\ residuals$$

In the analyzed data, 55 percent of the estimated values were within 10 percent of the true values. About 37 percent of the observations differed from the predicted values by 10 to 20 percent. And, only 8 percent of the predicted values differed from the true values by more than 20 percent. Of these observations, about 2 percent differed by 30-40 percent.

The true mean log transformed GDP per capita was within 1 percent of the predicted values for 12 of the 171 countries analyzed. It was within 5 percent for 54 of the 171 countries analyzed. It was within 10 percent for 95 of the 171 countries and within 20 percent for 158 of the 171 countries. The predicted value differed from the true values by more than 20 percent for only 14 (8 percent) countries. These results are provided in Appendix 1.1. The largest discrepancy seen for any country was 32. 4 percent.

The sample is restricted to the eight South Asian countries. It is observed that the discrepancy between the government published real GDP per capita and the real GDP per capita estimated by our model, is much larger. Where the largest discrepancy seen for any country in the global dataset was 32.4 percent, the highest discrepancy observed in the South Asia sample is 93 percent. It can be seen in Table 1.5 that amongst these countries, Bangladesh is the worst performing.

Table 1.5: Discrepancy in Per Capita GDP (South Asia)

Rank	Country	Discrepancy	Absolute Discrepancy
1	Bangladesh	-93.00	93.00
2	Bhutan	56.82	56.82
3	India	-56.05	56.05
4	Maldives	55.06	55.06
5	Afghanistan	-50.75	50.75
6	Pakistan	-30.77	30.77
7	Nepal	-29.80	29.80
8	Sri Lanka	29.08	29.08

1.5.6. Night-time Lights Model for Bangladesh

Bangladesh stood 152nd out of the 172 countries, in terms of discrepancy between the true and predicted values. The mean discrepancy for the 29-year period analysed was 18.3 percent. There are only 20 countries for which the model's predictions were worse. These countries were: Burundi, Mozambique, Central African Republic, Malawi, Rwanda, Ethiopia, Somalia, Niger, Sierra Leone, Togo, Uganda, Burkina Faso, Myanmar, Liberia, Kiribati, Tajikistan, Afghanistan, Chad, Madagascar, and Lesotho.

Interestingly, these are all countries that are on the lower rungs of economic development and/or war-torn. When the sample is restricted to South Asia, Bangladesh performs worst. However, given the small sample size, conclusions are based on the global sample.

The true and predicted value for Bangladesh displays a converging trend over time. At the beginning of the analysis period, the discrepancy stood at 24.3 percent. In 2020, it was 10.7 percent. This trend is observable in Appendix 1.2.

It should be noted though, that while the level of true GDP per capita appears to be underestimated, the government reported mean annual GDP per capita growth exceeded the predicted mean annual GDP per capita growth by over 400 percent. The true mean annual GDP per capita growth rate for the 1992 to 2020 period was 4.02 percent. In contrast the estimated mean annual GDP per capita growth rate was only 0.85 percent. This could indicate that the underestimation of GDP per capita is attributable to an economic sector that has been in existence since before 1992 but has not been fully integrated into the official GDP calculations yet. Thus, it is most likely attributable to a portion of the informal economy.

1.6. Extension

The results of the night-time lights model reveal an odd pattern. It shows that reported GDP per capita has been under-estimated for the entire period i.e., 1992 to

2020. It also shows that the discrepancy between the reported and the estimated GDP per capita values has been shrinking over time. At the same time, it demonstrates an over-estimation of the annual GDP per capita growth figures. In this section we explore how this phenomenon might have come about.

The national accounts of Bangladesh have been estimated since 1971-72. At the beginning the accounts were constructed using the UN System of National Accounts (SNA)-1968. Since then, it has been updated twice; once in 1991-92 when the SNA-1993 was adopted, and again in 2012-13 when the SNA-2008 was implemented. The original base year used to develop the national accounts for Bangladesh was 1972-73, this was shifted to 1984-85 from 1988-89. It was updated twice after that, once to 1995-96 and then to 2005-06. For the analyses in this this section, $CPI_{2005\text{-}06}$ adjusted, revised values were used.[7] The SNA-1968, which was used for constructing Bangladesh's national accounts till 1993 divided economic activity into 11 industries. It calculated GDP using the production approach and the expenditure approach. The production approach was more detailed, owing to the availability of data. Final GDP estimates for this period were derived by deducting intermediate consumption or production inputs. In addition, accounting and administrative data from public sector organizations were also incorporated into the national accounts statistics. Conversely, estimates by cost method were not possible due to data scarcity. The SNA-1993 was adopted in the 1993-1995 period. This comprised of improving the database and updating input-coefficients, by undertaking several surveys and studies. These surveys attempted to correct methodological flaws and bridge data gaps in various economic sectors. The change also entailed splitting the previously 11 industrial sectors into 15. Upward (mostly) adjustment was made to production data, using data from the most recent household surveys. Private final consumption expenditures were directly estimated using HES data. In addition, the calculation of Gross Capital Formation (GCF) or investment were slightly different for SNA-1993 and SNA-1968.

Eq 1.4

$$SNA - 1968: GCF$$
$$= Gross\ Fixed\ Capital\ Formation\ (GFCF)$$
$$+ change\ in\ inventories\ (\Delta S)$$

Eq 1.5

$$SNA - 1993: GCF$$
$$= GFCF + \Delta S + V\ (net\ aquisiton\ of\ valuables)$$

The SNA-2008, which was an update of the SNA-1993 was developed in 2006. To develop a national accounts framework based on SNA-2008, BBS collected data on selected segments. This included preparation of the updated Supply and Use Table (SUT) and external sector accounts. Starting 2014, recommendations from the SNA-2008 were incorporated into Bangladesh's national accounts, where possible. This included, categorizing military weapon systems, research and development expenditures, etc. as fixed assets. Owing to data limitations, the number of industrial sectors were kept at 15, instead of the SNA-2008 recommended 21. However, the sub-sectors were constructed to be compatible with those of the SNA-2008 (Bangladesh Bureau of Statistics, 2014). The estimation of GCF was also revised in the SNA-2008, the new estimation methodology is given in Eq 1.6. By convention the SNA-1993 treated the output of R&D as intermediate consumption, in SNA-2008 it is recommended that the value of R&D is valued at market price if purchased. If undertaken on own account than R&D is valued by summing the total production costs plus an appropriate mark-up including costs of production related fixed assets (Sim, 2011).

Eq 1.6

$$SNA - 2008: GCF +\ \Delta S + V + O$$

[7] The national accounts of Bangladesh for the 1972-73 to 1990-91 period, present macroeconomic indexes as two series: current prices and constant prices. Each series has two sets of data, the original dataset, and the revised data set. The revised set shows the difference in values when more recent surveys are used for estimation (Bangladesh Bureau of Statistics, 1993). The revision did not include the adoption of newer national accounting systems. Yet, it is observed that the incorporation of more recent surveys, censuses, and administrative data for estimation purposes pushes GDP values downwards. The proportion of GDP shrinkage decreases progressively with time. On average, the revised estimated are 5 percent smaller than the original estimates. Note that in this paper, only the revised estimates have been used for this period.

$$Where,$$
$$O = weapons\ systems, research\ and\ development\ and\ intellectual\ property$$

The changes implemented by SNA-2008, expanded the value of GDP in three ways. Firstly, through the capitalization of research and development. Secondly, through the valuation of outputs for own final use by households and corporations to include a return to capital. And lastly, by capitalization of expenditure on weapons systems. The refined method used for calculating Financial Intermediation Services Indirectly (FISIM), changes in recording of pension entitlement and the treatment of employee stock options also positively influenced the value of GDP (UNSIAP, 2014).

The above discussion, detailing SNA-1968, SNA-1993 and SNA-2008 demonstrates that each new national accounting system added new components to major macroeconomic indices such as the GDP. It is hypothesized that the adoption of a more recent national accounting system would cause a sudden spike in GDP growth rate in the year the new system is adopted. Drawing on the results of this paper, we know that GDP per capita values were under-estimated even after the adoption of the SNA-2008. This implies that a large section of the economy has still not been brought under official GDP calculations. It is also evident that this under-estimation was present even pre- SNA-2008. Though, the rate of under-estimation exhibits a downward trend as more and more of the SNA-2008 recommendations are incorporated.

This explains how, by gradually incorporating more of the informal economy into the formal GDP calculations, Bangladesh has been reporting higher GDP per capita growth rate than expected. This also explains why the discrepancy between reported and estimated GDP per capita has been shrinking over the years.

1.7. Limitations

The calculations of per capita real GDP growth are complex since they must account for population growth. In Bangladesh, population growth rate is estimated using several assumptions which are based on censuses that are carried out at 10-year intervals. These estimates are constructed by government bodies who have a strong incentive to under-report population growth as population control is a chief goal for a densely populated country like Bangladesh. On top of that, under-estimation of population automatically pushes GDP per capita values upward, which is another important objective for any government. Thus, it is difficult to have confidence in the traditional measures of GDP per capita.

Additionally, due to computing power limitations, a random sampling strategy was used to select 1,000 observation points from each country in each year. The average DN for each country in each period was estimated from these observations. Using pixel-level data to calculate average DN of a country in each period would have generated more accurate results.

1.8. Conclusion

The robust fixed-effects regression model was able to estimate log transformed real GDP per capita values that were within 20 percent of the trues values for 92 percent of the observations. When the observations are disaggregated by country, the model still performs well, and can predict values that are within 20 percent of the true values, for 92 percent of the countries the data is available for.

However, Bangladesh ranks lowly. Only 20 countries perform worse than Bangladesh in terms of discrepancy between true and predicted values. This difference does, however, decrease over time. This could have been the result of one of several things (i) methodological differences between how each country reported its GDP values (ii) data error and/or manipulation (iii) economic structure.

The results show that Bangladesh has been underestimating its GDP per capita values for the entire period analyzed. However, in the same period it has over-estimated its GDP per capita annual growth rate. This could be a result of the GDP calculations gradually incorporating more and more of the informal economy. This could explain the gradual convergence of the true and predicted values over time.

All 20 countries for which the model was worst performing are countries in lower stages of development with weak statistical capacity. Thus, it is likely that the GDP values reported by these governments is flawed. This coupled with the fact that these countries have larger proportions of economic activity in informal sectors, it is likely that GDP is under-estimated. Interestingly, all 20 of these countries are economically weak and/or politically unstable.

It is noted that the years in which the model predicts negative economic growth are 1994, 2001, 2003, 2006, 2009, 2012, 2016, 2019 and 2020. In 1994 opposition political parties began a parliamentary boycott, eventually leading to the dissolution of the government. In 2001, once again, opposition parties demonstrated forcing the prime minister to resign and hand over governance to an interim government. In 2003, the government launched *Operation Clean Heart* in response to bombings across the country. In 2006, there was another takeover by an interim government, followed by military rule. In 2009, the country suffered the repercussions of the forced removal of military leadership. In 2012, the opposition parties again formed a coalition to remove the ruling party, resulting in violence across the country. All these political crises were accompanied by strikes that halted economic activity in the country for long spans of time. It is evident that the model can pick up on this. Similarly, the model predicts negative growth in the year 2020. This was the year where the economy was impacted by the Covid-19 pandemic. The fact that the negative growth rates coincide with major political and economic events lends the econometric model additional validation.

Since the model used night-lights to predict growth in GDP, the night-light generation capacity of the economic sectors is important to the results. Some economic sectors generate more night-lights than others. For instance, agricultural work is done mostly during the day, and would not generate night-time lights. On the other hand, manufacturing industries, could operate round-the-clock and generate night-time lights. Hence, the distribution of economic activity could affect this model. In 1992, 68.9 percent of the working population of Bangladesh, were employed in agriculture; by 2020 this fell to 37.5 percent. This could also explain the growth in GDP over this period. However, this would have led to much less night-time light generation when the economy was more reliant on agriculture. Thus, the predicted values would be smaller. Hence, this theory is rejected.

1.9. Appendices

Appendix 1.1:Global Discrepancy Between Published Real GDP Per Capita and Night-time Lights Estimates

Rank	Country	Discrepancy	Absolute Discrepancy
1	Burundi	-32.40	32.40
2	Mozambique	-30.76	30.76
3	Central African Republic	-26.93	26.93
4	Malawi	-26.63	26.63
5	Rwanda	-26.53	26.53
6	Ethiopia	-26.26	26.26
7	Somalia	-24.95	24.95
8	Niger	-23.50	23.50
9	Sierra Leone	-22.31	22.31
10	Togo	-21.09	21.09
11	Uganda	-21.02	21.02
12	Burkina Faso	-20.96	20.96
13	Myanmar	-20.58	20.58
14	Liberia	-20.01	20.01
15	Kiribati	-19.81	19.81
16	Tajikistan	-19.66	19.66
17	Afghanistan	-19.65	19.65
18	Chad	-19.53	19.53
19	Madagascar	-18.96	18.96
20	Lesotho	-18.52	18.52
21	Bangladesh	-18.31	18.31
22	Gambia	-17.90	17.90
23	Tanzania	-17.77	17.77
24	Luxembourg	17.34	17.34
25	Guinea	-16.88	16.88
26	Guinea-Bissau	-16.76	16.76
27	Qatar	16.57	16.57
28	United Arab Emir	16.38	16.38
29	Australia	16.10	16.10
30	Cambodia	-15.70	15.70
31	Nepal	-15.61	15.61
32	Brunei Darussalam	15.61	15.61
33	Norway	15.58	15.58
34	Mali	-15.45	15.45
35	Iceland	15.39	15.39
36	Canada	15.03	15.03
37	India	-14.51	14.51
38	Switzerland	14.38	14.38
39	Comoros	-14.30	14.30
40	Saudi Arabia	14.08	14.08
41	Benin	-13.99	13.99

42	Timor-Leste	-13.94	13.94
43	Ireland	13.68	13.68
44	San Marino	13.51	13.51
45	Kuwait	13.39	13.39
46	Senegal	-13.37	13.37
47	Zambia	-13.20	13.20
48	New Zealand	13.10	13.10
49	Haiti	-13.08	13.08
50	Sweden	13.07	13.07
51	Austria	13.00	13.00
52	Denmark	12.97	12.97
53	Singapore	12.77	12.77
54	Pakistan	-12.64	12.64
55	Finland	12.47	12.47
56	West Bank and Gaza	-12.40	12.40
57	The Bahamas	12.38	12.38
58	Vanuatu	-12.14	12.14
59	Oman	12.11	12.11
60	Marshall Islands	-12.04	12.04
61	Germany	11.85	11.85
62	Tuvalu	-11.66	11.66
63	Ghana	-11.65	11.65
64	Netherlands	11.44	11.44
65	Vietnam	-11.34	11.34
66	France	11.26	11.26
67	Solomon Islands	-10.87	10.87
68	United Kingdom	10.79	10.79
69	Uzbekistan	-10.71	10.71
70	Nigeria	-10.65	10.65
71	Kyrgyz Republic	-10.65	10.65
72	Italy	10.63	10.63
73	Belgium	10.55	10.55
74	Cameroon	-10.25	10.25
75	Bahrain	10.20	10.20
76	Spain	10.17	10.17
77	Japan	10.14	10.14
78	Zimbabwe	-9.92	9.92
79	Cabo Verde	-9.62	9.62
80	Cyprus	9.47	9.47
81	Greece	8.95	8.95
82	Slovenia	8.94	8.94
83	Honduras	-8.81	8.81
84	Kenya	-8.67	8.67
85	Papua New Guinea	-8.41	8.41
86	Portugal	8.41	8.41
87	Estonia	8.36	8.36

88	Argentina	8.34	8.34
89	Czech Republic	8.22	8.22
90	Israel	8.20	8.20
91	Suriname	8.18	8.18
92	Tonga	-8.12	8.12
93	Nicaragua	-8.06	8.06
94	Russian Federation	7.98	7.98
95	Chile	7.66	7.66
96	Gabon	7.38	7.38
97	Lithuania	7.21	7.21
98	Djibouti	-6.96	6.96
99	Philippines	-6.84	6.84
100	Latvia	6.76	6.76
101	Kazakhstan	6.66	6.66
102	Panama	6.63	6.63
103	Hungary	6.63	6.63
104	China	-6.57	6.57
105	Morocco	-6.52	6.52
106	Libya	6.37	6.37
107	Croatia	6.33	6.33
108	Uruguay	6.32	6.32
109	Botswana	6.21	6.21
110	Sudan	-6.10	6.10
111	Malta	6.03	6.03
112	Malaysia	5.76	5.76
113	Samoa	-5.72	5.72
114	Turkey	5.68	5.68
115	Seychelles	5.66	5.66
116	El Salvador	-5.58	5.58
117	Mexico	5.46	5.46
118	Slovak Republic	5.24	5.24
119	Palau	4.99	4.99
120	Armenia	-4.89	4.89
121	Equatorial Guinea	4.85	4.85
122	Poland	4.85	4.85
123	Romania	4.67	4.67
124	Nauru	-4.48	4.48
125	Sri Lanka	-4.38	4.38
126	Trinidad and Tobago	4.27	4.27
127	Guatemala	-4.00	4.00
128	Azerbaijan	-3.99	3.99
129	Montenegro	3.91	3.91
130	Iraq	-3.85	3.85
131	Brazil	3.77	3.77
132	Guyana	3.60	3.60
133	Antigua and Barbados	3.58	3.58

134	Belize	-3.42	3.42
135	Kosovo	-3.35	3.35
136	Albania	-3.20	3.20
137	Moldova	-3.20	3.20
138	Bulgaria	3.11	3.11
139	Costa Rica	3.08	3.08
140	Bosnia and Herzegovina	-3.07	3.07
141	Turkmenistan	-3.02	3.02
142	Georgia	-2.95	2.95
143	Jamaica	-2.80	2.80
144	Indonesia	-2.63	2.63
145	Bhutan	-2.54	2.54
146	South Africa	2.41	2.41
147	Namibia	2.40	2.40
148	Bolivia	-2.01	2.01
149	Tunisia	-1.90	1.90
150	Fiji	1.88	1.88
151	Mongolia	1.82	1.82
152	Maldives	1.81	1.81
153	Colombia	1.81	1.81
154	Total	-1.71	1.71
155	Belarus	1.65	1.65
156	Paraguay	1.64	1.64
157	Thailand	1.50	1.50
158	Algeria	1.38	1.38
159	Mauritania	-1.36	1.36
160	Lebanon	1.05	1.05
161	Jordan	-0.96	0.96
162	Angola	-0.88	0.88
163	Mauritius	0.85	0.85
164	North Macedonia	0.72	0.72
165	Ecuador	-0.52	0.52
166	Serbia	0.52	0.52
167	Dominica	-0.34	0.34
168	Barbados	0.25	0.25
169	Peru	0.21	0.21
170	Grenada	0.16	0.16
171	Ukraine	-0.06	0.06
172	Dominican Republic	-0.04	0.04

Appendix 1.2: True and Predicted Log Transformed GDP Per Capita Values for Bangladesh

Country	Year	Log Transformed GDP Per Capita			Annual GDP Per Capita Growth Rate (%)		
		True	Predicted	Discrepancy (%)	True	Predicted	Discrepancy (%)
Bangladesh	1992	7.37	9.16	-24.35			
Bangladesh	1993	7.39	9.22	-24.74	2.44	5.91	-3.47
Bangladesh	1994	7.41	9.19	-24.07	1.68	-2.87	4.55
Bangladesh	1995	7.44	9.19	-23.61	2.86	0.14	2.72
Bangladesh	1996	7.46	9.20	-23.32	2.29	0.68	1.62
Bangladesh	1997	7.48	9.20	-23.01	2.28	0.47	1.81
Bangladesh	1998	7.51	9.21	-22.67	2.97	1.11	1.86
Bangladesh	1999	7.54	9.22	-22.37	2.54	0.90	1.64
Bangladesh	2000	7.57	9.26	-22.32	3.21	3.51	-0.30
Bangladesh	2001	7.60	9.22	-21.36	3.07	-3.55	6.62
Bangladesh	2002	7.62	9.25	-21.44	1.94	2.99	-1.05
Bangladesh	2003	7.65	9.22	-20.51	2.90	-3.59	6.49
Bangladesh	2004	7.68	9.23	-20.16	3.49	1.51	1.99
Bangladesh	2005	7.73	9.25	-19.59	4.85	1.42	3.43
Bangladesh	2006	7.78	9.17	-17.81	5.11	-7.75	12.86
Bangladesh	2007	7.84	9.21	-17.52	5.59	4.32	1.28
Bangladesh	2008	7.89	9.23	-17.01	4.69	1.49	3.20
Bangladesh	2009	7.92	9.22	-16.36	3.81	-0.70	4.50
Bangladesh	2010	7.97	9.23	-15.90	4.30	1.32	2.98
Bangladesh	2011	8.02	9.24	-15.28	5.12	1.01	4.11
Bangladesh	2012	8.07	9.24	-14.51	5.16	-0.24	5.41
Bangladesh	2013	8.12	9.32	-14.77	4.68	7.45	-2.77
Bangladesh	2014	8.16	9.37	-14.73	4.74	5.12	-0.37
Bangladesh	2015	8.22	9.37	-14.10	5.23	0.79	4.44
Bangladesh	2016	8.27	9.36	-13.13	5.78	-1.43	7.21
Bangladesh	2017	8.33	9.40	-12.78	5.96	3.81	2.14
Bangladesh	2018	8.40	9.41	-11.99	6.52	0.77	5.75
Bangladesh	2019	8.47	9.40	-11.05	6.81	-0.40	7.21
Bangladesh	2020	8.49	9.40	-10.68	2.45	-0.37	2.82

1.10. Bibliography

Asadullah, M. N., & Chakravorty, N. N. T. (2019). Growth, governance, and corruption in Bangladesh: A re-assessment. *Third World Quarterly*, *40*(5), 947–956. https://centaur.reading.ac.uk/84160/1/Bangladesh-growth without governance V-final ALK Edit FINAL 16 Mar 2019_FINAL23 March %281%29.pdf

Asian Development Bank. (2022). *Key Indicators Database|ADB*. https://kidb.adb.org/economies/bangladesh

Babones, S. (2020, October 20). Angst and Denial in India as It's Now Officially Poorer Than Bangladesh. Foreign Policy. https://foreignpolicy.com/2020/10/20/india-falls-behind-bangladesh-gdp/

Bangladesh Bureau of Statistics. (1993). *Twenty Years of National Accounting of Bangladesh (1972-73 to 1991-92)*. Statistics Division, Ministry of Planning. https://www.worldcat.org/title/twenty-years-of-national-accounting-of-bangladesh-1972-73-to-1991-92/oclc/30513466&referer=brief_results

Bangladesh Bureau of Statistics. (2014). *National Accounts Statistics of Bangladesh (provisional estimates of GDP 2013-14 and final estimates of GDP 2012-13)*. Statistics Division, Ministry of Planning. https://www.amazon.com/National-Accounts-Statistics-Provisional-Estimates/dp/9845190200

Basher, S. A., Behtarin, J., & Rashid, S. (2021). Convergence across Subnational Regions of Bangladesh – What the Night Lights Data Say? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3853803

Bhandari, L., & Roychowdhury, K. (2011). Night Lights and Economic Activity in India: A study using DMSP-OLS nighttime images. *Proceedings of the Asia-Pacific Advanced Network*, *32*(0), 218. https://doi.org/10.7125/APAN.32.24

Bhattacharya, D. (2018). *Bangladesh Becoming a Middle-Income Country, Ceasing to be a Least Developed Country: Clarifying Confusion*. https://www.semanticscholar.org/paper/Bangladesh-Becoming-a-Middle-Income-Country-%2C-to-be-Bhattacharya/d36f5ba2e711a29baedbd78d3e46e36432b19dee

Bhattacharya, D., Khan, S. S., Moazzem, K. G., Arno, A., Rahman, M., Bari, E., Khatun, F., Pervin, S., Rahman, Md. M., Khan, T. I., & Kamal, M. (2019). *Bangladesh's Graduation from the Least Developed Countries Group: Pitfalls and Promises*. Routledge & CRC Press. https://www.routledge.com/Bangladeshs-Graduation-from-the-Least-Developed-Countries-Group-Pitfalls/Bhattacharya/p/book/9780367665326

Bickenbach, F., Bode, E., Nunnenkamp, P., & Söder, M. (2016). Night lights and regional GDP. *Review of World Economics*, *152*(2), 425–447. https://doi.org/10.1007/s10290-016-0246-0

Data on Statistical Capacity | The World Bank. (n.d.). Retrieved September 2, 2022, from https://datatopics.worldbank.org/statisticalcapacity/

DMSP OLS: Nighttime Lights Time Series Version 4, Defense Meteorological Program Operational Linescan System | Earth Engine Data Catalog. (2014). Google Developers. https://developers.google.com/earth-engine/datasets/catalog/NOAA_DMSP-OLS_NIGHTTIME_LIGHTS

Gibson, J. (2020, June 8). *Shedding light on night lights data: DMSP vs. VIIRS*. https://blogs.worldbank.org/opendata/shedding-light-night-lights-data-dmsp-vs-viirs

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *American Economic Review*, *102*(2), 994–1028. https://doi.org/10.1257/aer.102.2.994

Henderson, V., Storeygard, A., & Weil, D. N. (2011). A Bright Idea for Measuring Economic Growth. *American Economic Review*, *101*(3), 194–199. https://doi.org/10.1257/aer.101.3.194

International Monetary Fund. (2022). *IMF Data: Access to Macroeconomic and Financial Data*. https://data.imf.org/?sk=388dfa60-1d26-4ade-b505-a05a558d9a42

Jerven, M. (2013). *Poor Numbers: How We Are Misled by African Development Statistics and What to Do About It*. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(14)60209-9/fulltext

Kulkarni, R., Haynes, K., Stough, R., & Riggle, J. (2011). Revisiting night lights as proxy for economic growth: A multi-year Light based growth indicator (LBGI) for China, India and the U.S. *George Mason University*, 21. https://www.researchgate.net/publication/228217352_Revisiting_Night_Lights_as_Proxy_for_Economic_Growth_A_Multi-Year_Light_Based_Growth_Indicator_LBGI_for_China_India_and_the_US

Li, X., & Zhou, Y. (2017). A Stepwise Calibration of Global DMSP/OLS Stable Nighttime Light Data (1992–2013). *Remote Sensing*, *9*(6), 637. https://doi.org/10.3390/rs9060637

Lowe, M. (2014). *Night Lights and ArcGIS: A Brief Guide*. 20. https://darrylmcleod.com/wp-content/uploads/2016/06/Night-Lights-and-ArcGIS-A-Brief-Guide.pdf

Martinez, L. R. (2017). How Much Should We Trust the Dictator's GDP Estimates? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3093296

Marx, A., & Ziegler Rogers, M. (2017). Analysis of Panamanian DMSP/OLS nightlights corroborates suspicions of inaccurate fiscal data: A natural experiment examining the accuracy of GDP data. *Remote Sensing Applications: Society and Environment*, *8*, 99–104. https://doi.org/10.1016/j.rsase.2017.08.005

Mellander, C., Lobo, J., Stolarick, K., & Matheson, Z. (2015). Night-Time Light Data: A Good Proxy Measure for Economic Activity? *PLOS ONE*, *10*(10), e0139779. https://doi.org/10.1371/journal.pone.0139779

Osmani, S. R. (2017). *Review of National Accounts Estimation in Bangladesh*. International Growth Centre (IGC). https://www.theigc.org/wp-content/uploads/2017/09/Osmani-2017-Final-report.pdf

Proville, J., Zavala-Araiza, D., & Wagner, G. (2017). Night-time lights: A global, long term look at links to socio-economic trends. *PLOS ONE*, *12*(3), e0174610. https://doi.org/10.1371/journal.pone.0174610

Sharma, M. S. (2021, June 1). Bangladesh rises to be South Asia's standout star as India, Pak fall behind. *Business Standard India*. https://www.business-standard.com/article/economy-policy/bangladesh-rises-to-be-south-asia-s-standout-star-as-india-pak-fall-behind-121060100237_1.html

Sim, B. (2011). *2008-SNA - Main Changes from 1993 SNA*. UN Statistics Division. https://www.unescap.org/sites/default/files/2008%20SNA-Main%20Changes%20from%201993%20SNA.pdf

The World Bank Group. (2022). *The World Bank| Data*. https://data.worldbank.org/about

United Nations. (n.d.). *Trade-related international support measures (ISMs) for LDCs | LDC Portal*. Retrieved December 8, 2021, from https://www.un.org/ldcportal/category/trade-international-support-measures/

United Nations. (2021). *LDC Identification Criteria & Indicators | Department of Economic and Social Affairs*. https://www.un.org/development/desa/dpad/least-developed-country-category/ldc-criteria.html

UNSIAP. (2014). *Lesson 2: 2008 SNA - Changes from 1993 SNA |Fourth Intermediate Level E-learning Course on System of National Accounts*. United Nations Statistical Institute for Asia and the Pacific. https://www.unsdglearn.org/courses/intermediate-e-learning-course-on-integrated-economic-statistics-to-support-2008-sna-implementation/

World Bank. (2016, July 1). *New country classifications by income level: 2016-2017*. https://blogs.worldbank.org/opendata/new-country-classifications-2016

World Bank. (2020). *Data on Statistical Capacity | The World Bank*. https://datatopics.worldbank.org/statisticalcapacity/Home.aspx

World Bank. (2021a). *World Bank Open Data*. https://data.worldbank.org/

World Bank. (2021b). *World Country Polygons , Very High-Definition*. https://maps.worldbank.org/

World Population Review. (n.d.). Bangladesh Population 2021 (Demographics, Maps, Graphs). Retrieved December 15, 2021, from https://worldpopulationreview.com/countries/bangladesh-population

Wu, J., Wang, Z., Li, W., & Peng, J. (2013). Exploring factors affecting the relationship between light consumption and GDP based on DMSP/OLS nighttime satellite imagery. Remote Sensing of Environment, 134, 111–119. https://doi.org/10.1016/j.rse.2013.03.001

Zheng, E., Tan, Y., Goes, P., Chellappa, R., Wu, D., Shaw, M., Sheng, O., & Gupta, A. (2017, August 20). When Econometrics Meets Machine Learning. https://www.sciendo.com/article/10.1515/dim-2017-0012

1.11. Acronyms

ADB	Asian Development Bank
DMSP	Defense Meteorological Satellite Program
DN	Digital Numbers
DWH	Durbin-Wu-Hausman
EVI	Economic Vulnerability Index
FE	Fixed effects
FISIM	Financial Intermediation Services Indirectly
GCF	Gross Capital Formation
GDP	Gross Domestic Product
GFCF	Gross Fixed Capital Formation
GLS	Generalized Least Squares
GNI	Gross National Income
GoB	Government of Bangladesh
HAI	Human Assets Index
IMF	International Monetary Fund
ISP	Investment Support Program
LDC	Least Developed Country
LM	Lagrange Multiplier
LMIC	lower-middle income country
MIC	Middle income country
NOAA	National Oceanic and Atmospheric Administration
NTL	Night-time lights
ODA	Official Development Assistance
RE	Random Effects
S	Inventories
SCI	Statistical Capacity Indicator
SNA	System of National Accounts
SUT	Supply and Use Table
UN	United Nation
USD	Unites States Dollar
VIIRS	Visible Infrared Imaging Radiometer Suite
WB	World Bank

2 Framework for Testing the Reliability of Micro-level Data: Using Data from Bangladesh

Abstract: This paper develops a multi-pronged framework to test the reliability of micro-level data. Firstly, micro-level data on income, expenditure and consumption are compared against the corresponding macro variables i.e., Gross National Income, total expenditure, and Private Final Consumption Expenditure. It is found that in most periods the macro values are larger than the micro values. These differences are lessened when an older system of national accounting is used to calculate the macro variable values. However, it does not mitigate the difference, and the macro values remain larger than the corresponding micro values. Secondly, conformity of the micro-level data to the Benford's distribution is checked. It is found that in all periods, the distribution deviated, and the deviations were statistically significant, flagging potential data errors. Lastly, through triangulation against administrative data it is found that the household surveys over-sampled the wealthier segment of the population but under-sampled the income-tax paying segment of the population.

2.1 Introduction

The results in Tahsin (2022a) indicate that the national accounts of Bangladesh are compromised. It is shown, that when the night-time lights model is used to estimate GDP per capita, the figures are grossly under-estimated. However, the annual GDP per capita growth rate is overestimated. Thus, making any data published by the Government of Bangladesh (GoB), suspect.

Macro data, such as the national accounts statistics are important for economic analysis and policymaking. But it should be kept in mind, that several important indexes pertaining to a country's economic performance, are generated through micro data, such as household surveys. In the case of Bangladesh, the national household surveys are used to calculate the Gini-coefficient. Additionally, several components of the national accounts are estimated through surveys e.g., estimates of agricultural production. Hence, it is evident that the reliability of micro data is not only intrinsically important but also vital for national accounts calculations. In this chapter we will construct a framework to evaluate the reliability of the household surveys of Bangladesh.

The rest of this chapter is organized as follows: Section 2.2 summarizes the available literature in this research area and identifies the contribution of this paper. In Section 2.3, the sources of data are described. Section 2.4 elaborates on the empirical strategy used. This is followed by Section 2.5, in which the results are summarized. Section 2.6 discusses the limitations of the study. The concluding remarks are presented in Section 2.7.

2.2 Literature Review

Existing literature in the area attempts to answer several pertinent questions. Researchers have studied the defining characteristics of micro data. They have then deliberated on its strengths and limitations, to understand what purpose it is better suited for. Other studies have compared micro data with their macro counterparts, to gage whether they validate each other, and if not, what that indicates. These triangulation exercises allow researchers to draw conclusions regarding the reliability of the data. The methods and findings of key studies in this area are discussed in this section.

Estimates of per capita consumption, per capita income and per capita expenditure are generated by both national account statistics and household surveys. In theory, each estimate should hold the ability to authenticate the other. Many nations; especially those with lower statistical capacities; make use of consumption and expenditure data from their respective national accounts to augment the data generated by household surveys and vice versa.

Using national accounts data and a household-level survey of the USA, Amel et al. (2008) pointed out how micro-level data can be used to isolate crucial information regarding heterogeneity across different types of households. This heterogeneity is often missed by macro data. The paper compared micro and macro data to assess whether they validate each other. Since micro-level data is generated at infrequent intervals with time lags; the data collected for national accounts and those for household surveys usually correspond to different time points. Hence, for the data to align, comparable series must be constructed. The paper concluded that macro and micro data sources are comparable when appropriate adjustments are made. Thus, researchers and policy makers may make macroeconomic inferences from micro data sources, such as household-level surveys.

Maki & Nishiyama (1993) drew the same conclusion by comparing the micro and macro data series of Japan. The ratio of values derived from household-level surveys to those derived from national accounts were estimated to be around 80 percent. It should be noted though, that the ratio for Japan was closer to 1, than it was for other nations.

While Amel et al. (2008) and Maki & Nishiyama (1993) found micro and macro data to be analogous; others found that micro and macro data series could only be reconciled after data processing measures were taken to minimize the sources of the divergence.

Ravallion (2003) found that for developing and transitional economies, the ratio of private consumption per capita derived from national accounts to average private consumption derived from household-level surveys did not maintain a steady ratio. The gap between average private consumption derived from household-level surveys and per capita consumption derived from national accounts grew steadily over time. However, the exceptions to this rule displayed strong regional and methodological effects. For instance, it was seen, that the difference in levels was attributable more to income surveys than expenditure surveys. Much of the divergence was also attributed to data problems in the (contracting) transition economies. However, despite the discrepancy between the two series at the individual country level, the group survey mean in the case of the expenditure surveys did not appear to be significantly different from the national accounts mean. Ravallion's argument is based on testing whether the ratio of survey to national accounts consumption mean was significantly different from one. The mean ratio for his sample of countries was 0.931, which was not significantly different from one (t= -1.21).

Robilliard & Robinson (2003) presented another approach to reconcile household surveys and national accounts data. This study applied the entropy measure of information to data from Madagascar. The survey household weights were treated as a prior. Using the additional information, new weights were estimated. It was concluded that this approach can efficiently use information from a variety of sources.

In contrast, Karshenas (2003) observed no consistent trend in the ratio of average consumption and income derived from household surveys and national accounts data using global panel data.

Even though a plethora of research is available in the subject area, very few studies have attempted to apply these methodologies to validate data from Bangladesh. The only relevant paper, Khan (2005), is over 15 years old. The paper compared the four household surveys conducted in Bangladesh between 1991/92 to 2005. The author re-estimated household income and consumption using "more appropriate" deflators, definitions, and sample weights. Through this exercise it was concluded that the household consumption estimates were unreliable.

The review of the available literature revealed three chief knowledge gaps. Firstly, the source of the discrepancy between micro and macro data in Bangladesh has not been identified. Secondly, whether micro data or macro data is more reliable has not been explored. Thirdly, the limited analyses that have been conducted only used pre-2006 data. Thus, the evolution of data quality in recent years remain unmapped. This paper will attempt to answer these questions.

2.3 Data

2.3.1 Micro Data

Household-level data was collected from the Household Expenditure Surveys (HES) and the Household Income and Expenditure Surveys (HIES) collected by the Bangladesh Bureau of Statistics (BBS) from 1973-74 to 2015-16. During this period BBS conducted 18 rounds of nationally representative household surveys. The first 14 of these surveys were expenditure surveys, whereas the most recent 4, were income and expenditure surveys.

The surveys collected current information on the expenditures of household consumption items to track changes in patterns of household consumption expenditure. Information regarding the attributes of the household members (earners, occupation, monthly and yearly income, etc.); expenditure on food and drinks; apparels, textile, and footwear; housing and household operations; personal cares and effects; medical expenses; educational expenses; laundry and cleaning; transport and traveling; recreation and reading; tax, interest and fines and radios and musical instruments were collected.

All survey rounds covered the whole geographic area of Bangladesh and households of every size, social status, and economic class.

The survey structure and data collection methods were improved gradually over the years. This included replacing the 'Recall' method with the 'Diary' and 'Schedule' methods. Training surveyors to ask structured questions verbatim. Using varying reference periods for different expenditure classes in line with their frequency of purchase. Introducing cross-verification of data improve accuracy; and digitizing the data collection process to minimize inconsistencies and errors (Bangladesh Bureau of Statistics, 2011).

Recent survey rounds also added survey modules for crises and coping measures, micro-credit, migration and remittance, social-safety nets, and disability (Bangladesh Bureau of Statistics, 2017b).

Key information of each survey round is summarized in Appendix 2.1.

The three main micro variables of interest to this study, are monthly household income per capita, monthly household expenditure per capita and monthly household consumption per capita Appendix 2.2. They have been gathered from the HES reports (1973-74, 1981-82, 1983-84, 1985-86, 1988-89, 1991-92 and 1995-96) and HIES reports (2000, 2005, 2010 and 2016-17). It should also be noted that income, expenditure, and consumption figures were published in 'per household' terms[8]. They were transformed to per capita figures using the average number of household members recorded in each round of the household survey.

2.3.2 Macro Data

The three main macro variables of interest: per capita GNI, per capita Private Final Consumption Expenditure (PFCE) and per capita Total Expenditure were gathered from the national accounts Appendix 2.3.

Bangladesh has been compiling national accounts statistics, since 1972. The framework developed by BBS followed the same methodology as the System of National Accounts (SNA) - 1968 of the United Nations. This ensured statistical consistency of the primary data generated through diverse sources. It also helped in harmonizing the concepts, definitions, classifications, and systems, with international standards. To aid comparability and relevance, the base year was changed from 1972-73 to 1984-85. (Bangladesh Bureau of Statistics, 1993).

After the adoption of the SNA-1993, the national accounts from 1989-90 onwards were revised. To generate better quality data, BBS conducted several surveys and studies, and used their results to revise the data and methodology. The number of production sectors was increased from eleven to fifteen. The more comprehensive coverage increased GDP by about 30 percent and per capita income by nearly 28 percent compared to the old method. The final consumption expenditures were based on the HES data. The base year was updated from 1984-85 to 1995-96.

From FY 2012-13, recommendations from System of National Accounts 2008 (2008 SNA) were incorporated, where possible. This included (i) classifying military weapon system, research, and development expenditures etc. as fixed assets; (ii) aligning the 21 sub-sectors of the 15 production

[8] Bangladesh Bureau of Statistics (1978) defines households as 'a person or group of persons who normally live and eat together'. All related persons, helpers, boarders, and lodgers who have no other usual place of residence are considered members of the household. Guests and visitors who consider their usual place of residence anywhere other than the household being surveyed are not considered members of the household.

sectors with the 21 sectors of ISIC revision 4; (iii) Allocating Financial Intermediation Services Indirectly Measured (FISIM) in the 2005-06 based GDP; and (iv) incorporating results from the Economic Census 2013 to improve data quality (Bangladesh Bureau of Statistics, 2014).

For the exercises conducted in this chapter, macro aggregates corresponding to the key micro variables: per capita monthly consumption expenditure and per capita monthly expenditure, had to be selected. None of the macro indices are exactly equivalent to the household level data they were compared to. For each micro variable several macro variables could have been chosen for comparison. Hence, the most relevant macro variable was identified through logical inference. This is discussed in Appendix 2.3.

Since the micro and macro data were collected and compiled for distinct purposes, they were not readily comparable and had to be transformed. The data transformation process required construction of a date-adjusted series for the macro-level variables.

National accounts are constructed to align with the fiscal year, July to June. On the other hand, household surveys are conducted for a period of 12 months, not necessarily aligned with the fiscal year. Hence, to compare national accounts data to household survey data a compatible series had to be constructed. Since, monthly data is unavailable for national accounts of Bangladesh, all figures were readjusted by assigning weights according to the proportion of the survey that was conducted in each fiscal year. For instance, the 1991-92 HES was conducted from September 1991-August 1992. For comparability, values of all macro variables for the period were estimated using Eq 2.1.

Eq 2.1

$$W_{1991-92} * V_{1991-92} + W_{1992-93} * V_{1992-93}$$

Where,

$W_{1991-92} = weight\ assigned\ to\ the\ macro\ variable\ for\ FY\ 1991-92 = 10/12$

$W_{1992-93} = weight\ assigned\ to\ the\ macro\ variable\ for\ FY\ 1992-93 = 2/12$

$V_{1991-92} = value\ of\ macro\ variable\ in\ FY\ 1991-92$

$V_{1992-93} = value\ of\ macro\ variable\ in\ FY\ 1992-93$

Secondly, owing to structural changes in the economy and other inflation inducing factors; comparing income, expenditure or consumption at current prices fails to show the real growth experienced by the economy. Thus, the values needed to be adjusted to ensure comparability across time. So, using data from Bangladesh Bureau of Statistics (2017a), all values were CPI adjusted[9]. It should be noted that CPI values using base year 2005-06, are only available for the 2006 to 2017 period. The values for the years preceding 2006 were estimated using conversion factors. These conversion factors were estimated by using CPI values of years, for which CPI values of at least two base years were available. The CPI values available were presented using 4 base years: 1973-74, 1985-86, 1995-96 and 2005-06. The conversion factor from base year 1995-96 to 2005-06, was calculated using Eq 2.2.

Eq 2.2

$$Conversion\ factor\ 1995-96\ to\ 2005-06: \frac{CPI_{2005-06}}{CPI_{1995-96}} = 100.00/164.21$$

Where,

$CPI_{2005-06} = CPI\ calculated\ using\ base\ year\ 2005-06 = 100.00$

$CPI_{1995-96} = CPI\ calculated\ using\ base\ year\ 1995-96 = 164.21$

Using similar methodology, the conversion factors for converting from base year 1973-74 to 1985-86 and for converting from 1985-86 to 1995-96 were also

[9] Important insights regarding real growth of income, expenditure, and consumption; and trends can only be obtained using CPI adjusted values. CPI is chosen instead of the GDP deflator for inflation adjustment for two reasons. Firstly, the CPI is reported more frequently, thus, it is possible to estimate CPI values for the months corresponding to the household surveys. Secondly, the CPI is deemed more relevant to the average consumer, as it dispenses with investment, net exports and government expenditure, all of which are components of the GDP deflator (Green, 2017).

estimated. These values were 0.25 and 0.51, respectively.

2.4 Empirical Strategy

Three analytical tools are implemented in this chapter, to reach conclusions regarding the quality of the primary micro data available in Bangladesh. Firstly, selected micro-level variables are compared to their macro equivalents. Comparable series are constructed to ensure that the time periods aligned. Secondly, Benford's Law is applied to the micro-level variables to check whether the leading digit distribution conform with Benford's distribution. Lastly, to identify whether the weaknesses in the micro-level data arose from sampling errors, selected household level indicators are compared against corresponding administrative data. The indicators are chosen on the strength of their correlation with income quartiles.

2.4.1 Micro vs. Macro Data

Theoretically, it should be possible to estimate macro values using their micro equivalents and vice versa. This is especially true when data from one is fed into the other. For instance, the national accounting procedure in Bangladesh relies in large part on data acquired through the HIES conducted by BBS. Though, this also limits the usefulness of using one data source to validate the other; it does still provide some legitimacy to the data.

Thus, in this chapter, the CPI adjusted micro variables are compared against the corresponding CPI adjusted macro variables. Note, that all the data used were first transformed to ensure consistent measurement units and time periods. The micro variables used were, per capita monthly household income, per capita monthly household expenditure and per capita monthly household consumption expenditure. These variables were compared with the following macro indexes: per capita monthly GNI, per capita monthly total expenditure and per capita monthly PCFE, respectively (refer to Appendix 2.3).

2.4.2 Benford's Distribution

It has been suggested that Bangladeshi data is manipulated to align with the goals of different stakeholders (Osmani, 2017). Thus, use of tools that can detect irregularities in data supply is appropriate. While use of such tools is common for auditors; economists utilize them infrequently. One such tool is the application of the Benford's Law, also known as the Newcomb-Benford Law. The tool is reliable enough, that it is considered legally admissible in several criminal courts across the world, including those of the United States (Singleton, 2011).

Benford's Law is a mathematical theory of leading digits. The theory posits that in data sets, the leading digit is distributed in a specific, non-uniform manner. It is best suited to be applied to exponentially growing data. However, it is also applicable to other data sets. The application of this law can recognize probabilities of highly likely or highly unlikely frequencies of numbers in a dataset. However, there are limitations to its application. For instance, it cannot be used in cases in which the numbers are preset to begin with a limited set of digits, or only cover one or two orders of magnitude. It is also unadvisable to apply the law to small sets of data (Singleton, 2011). It should be kept in mind that deviation from Benford's Law does not prove fraud or manipulation, however, it does flag potentially fraudulent activity.

Benford sets are insensitive to the unit of measurement i.e., if a data set complying to Benford's Law is multiplied by a non-zero constant, the new data series will also be Benford compliant (CaseWare IDEA, 2007). Benford's Law can be proven empirically. A mathematical formula (Eq 2.3) can be used to derive the probable frequency of occurrence of any leading digit or any numerical combination.

Eq 2.3

$$P(d) = log\ (1 + \frac{1}{d})$$

P(d) stands for the probability that a number starts with the digit 'd'.

For our analyses, Benford's Law is applied to the monthly household expenditures for ten rounds of the household surveys, for which household level data is available. This is done by plotting the distribution of leading digits of theses variables against the Benford distribution. Divergences from the expected distribution, i.e., the Benford's distribution are noted, as this could indicate sampling error or data manipulation. The expected distribution of leading digits is available in Table 2.1.To test whether the results vary significantly from theoretical expectations, the Kolmogorov-Smirnov (K-S) test is applied.

Table 2.1: Benford's Distribution

Digit	1st Position	2nd Position	3rd Position
0	N/A	0.120	0.102
1	0.301	0.114	0.101
2	0.176	0.109	0.101
3	0.124	0.104	0.100
4	0.097	0.100	0.100
5	0.079	0.097	0.100
6	0.067	0.093	0.099
7	0.058	0.090	0.099
8	0.051	0.088	0.099
9	0.046	0.085	0.098

Source: (Singleton, 2011)

2.4.3 Administrative Data

Since the micro-level data collected through the household surveys is used to estimate the macroeconomic indices of Bangladesh, any errors present in the micro data would recur in the national accounts. Thus, triangulating the micro-level data collected from the household surveys against a third source of data, i.e., administrative data, is useful. This exercise is particularly helpful in isolating sampling errors.

For these analyses to provide worthwhile insights, the household survey and administrative data indicators compared, need to be strongly correlated with households' economic status. The indicators chosen, are required to be strongly correlated with only the top income quartile, while simultaneously displaying much weaker or reverse relationships with the other quartiles. These correlations are estimated using logistic regression models. The methodology used to estimate the correlation coefficients and determine the correlated groups is discussed in the subsequent sections. Thus, using these comparisons, under-sampling or over-sampling of income quartiles is determined.

It should be noted that there are two underlying assumptions. Firstly, it is assumed that the errors in the indicator data are uncorrelated to the sampling errors. Thus, the magnitude of the correlation coefficients is relatively insensitive to sampling errors, i.e., they were distribution neutral. The second assumption is that the administrative data is less error-prone than the household survey data.

The micro-level variables and their administrative counterpart variables are provided in Table 2.2. The last column of the table corresponds to the income group with which each variable displayed strong correlation.

Table 2.2: Household Survey Data Compared Against Administrative Data

Household Survey Data	Administrative Data	Correlated Income Group
Number of income taxpaying households	Number of income taxpayers	Top 25%
Number of cars owned by households	Number of registered private cars	Top 25%
Number of motorbikes owned by households	Number of registered motorbikes	Top 25%
Number of electrified households	Number of electric connections in the country	Top 25%

Source: BBS

2.4.3.1 Income-tax Payers

Firstly, the households in each round of the household surveys are categorized into four income quartiles. A dummy variable is generated for each quartile. For each survey round, four models are run, where the dependent variable is being in either income quartile one, income quartile two, income quartile three or income quartile four (bottom quartile). Logistic regression models of the form in Eq 2.4, are used determined whether income-tax paying status is strongly correlated with being in a specific income quartile.

Eq 2.4

$$y_i = \beta_0 + \beta_1 x + \varepsilon$$

$Where,$

$y_i = income\ quartile\ 1, 2, 3\ or\ 4$

$\beta_0 = intercept$

$\beta_1 = coefficient\ of\ x$

$x = income - tax\ paying\ status$

After establishing that income-tax paying status is a decent indicator for detecting sampling-errors in the top 25 percent of the income earners. The number of income-tax payers in the household surveys are extrapolated. This is then compared to the number of income-tax payers recorded by the National Board of Revenue (2018).[10]

2.4.3.2 Income-tax Amount

Applying the same principles and methodology as before, a comparison is drawn between the total income-taxes collected by the government and the total income-taxes paid by the households sampled in the household surveys. This analysis also aided in commenting on the sampling distribution.

2.4.3.3 Electrification and Vehicle Ownership

The methodology used in determining the correlation of income-tax paying status with income group, is also applied in this case. The administrative data for these indicators were drawn from Bangladesh Bureau of Statistics (2003b), Bangladesh Bureau of Statistics (2019) and Bangladesh Road Transport Authority (2018). It should be noted that data was not available for the entire period of analysis i.e., 1981 to 2017, for any of these three indicators.

The number of electric connections provided to private dwellings was only recorded for the 1981 to 2001 period. Post-2001, the electrification data is recorded in terms of power consumption. On the other hand, vehicle ownership is only disaggregated by type of vehicle (private cars and motorbikes) from HIES 2000 onwards.

2.5 Results

This section includes the results generated using the empirical strategy discussed in Section 2.4. The organization of the results follows the same structure as the previous section.

2.5.1 Micro vs. Macro Data

The CPI-adjusted micro values were compared to the CPI-adjusted (revised) macro values. The selected micro variables were, per capita monthly household income, per capita monthly household expenditure and per capita monthly household consumption expenditure. The macro indexes these micro variables were compared to were, per capita monthly GNI, per capita monthly total expenditure and per capita monthly PFCE.

Appendix 2.4 depicts per capita monthly household income, per capita monthly household expenditure, and per capita monthly household consumption expenditure in the graph on the left. It is seen that, income exceeds expenditure and expenditure exceeds consumption expenditure, in almost all periods. The expenditure and consumption expenditure lines almost converge. But there is a noticeable gap between the income line and the other two lines. However, in the two most recent household survey rounds, the three variables began exhibiting a tendency towards convergence.

On the other hand, the graph on the right of Appendix 2.4 depicts the levels of the selected macro-indices, i.e., the per capita monthly GNI, the per capita monthly total expenditure and the per capita monthly PFCE. Though, the rankings were not static throughout the observed period, it is seen, that mostly, GNI is at the top, followed by total expenditure and PFCE, respectively.

[10] Income tax returns filed are used as a proxy for the number of income-tax payers in the country.

Hence, it is noted that generally, the micro and the corresponding macro variables selected for the analyses in this paper; demonstrate the same rank against each other. This provides a degree of validity to the equivalency rationale developed in Appendix 2.3. However, two important distinctions also emerged. Firstly, while the micro variables have been converging in recent years, the corresponding macro variables have been diverging. Secondly, in the case of the macro-indices, GNI and total expenditure displayed stronger co-movement than total expenditure and consumption expenditure.

Having validated the equivalency rationale, further analyses is carried out by comparing the selected variable pairs. The graphs in Appendix 2.5 plot the micro variables against their corresponding macro indices. The data presented extends from 1974 to 2017. It is evident that the trend lines for macro-sourced and micro-sourced data are divergent. The gap between the micro and macro variables grew progressively over the period. This is true for income, expenditure as well as, consumption.

The graph in Appendix 2.6 plots the information of Appendix 2.5 as ratios. The three lines correspond to the ratio of micro to macro income, the ratio of micro to macro expenditure and the ratio of micro to macro consumption expenditure. Overall, it appears that the ratios for all three variables, shrunk over the 1974 to 2017 period, i.e., data drawn from household surveys has become less and less representative of the national accounts. In other words, the macro indices, generated through the national accounts, grew at a faster rate than the micro variables generated through the household surveys.

In the initial years, there are instances of the micro-data, exceeding its macro counterparts (ratio > 1). These were most likely due to a combination of survey errors and gaps in national accounting methodology.[11] It is also observed that prior to 1993 the micro-macro gap was smallest for income data, followed by consumption expenditure data. The gap is largest for expenditure data. However, since 1993, the consumption expenditure ratios are above the income ratios. Expenditure ratios remained at the bottom. In the very last period, micro-macro income ratio is the smallest.

Till 1986 the micro-macro income ratio had not dropped below 0.87 or exceeded 1.09. This indicated that the micro variable varied by less than 13 percent from its macro counterpart. However, starting 1993, this ratio fell to 0.69 and has not improved since. In 2017, when the most recent round of the household survey was conducted, the micro variable was smaller than 40 percent of its macro equivalent.

Similar trends are observed for the expenditure data. Till 1989 the micro-macro ratio for expenditure had not dropped below 0.7. This indicated that the variable varied by less than 30 percent from its macro counterpart. However, starting 1993, this ratio fell to 0.6 and has only improved once, since. In 2017, the micro variable was smaller than 30 percent of its macro equivalent.

The data for the ratio of micro-macro consumption expenditure also exhibited similar behaviour to that of income and expenditure. Till 1996, the micro-macro ratio for consumption expenditure remained within 0.8 and 1.2. Thus, the micro data were within 20 percent of their macro counterparts. However, starting 2001, this ratio fell to 0.7 and only improved once since. In 2017, the micro variable was smaller than 60 percent of its macro equivalent.

These analyses revealed that of the three pairs of variables, the ratio of micro-macro consumption expenditure was the closest to 1, followed by the ratio of micro-macro income and the ratio of micro-macro expenditure. However, it should be noted, that the national accounting methodology for estimating PFCE; which is the macro-equivalent of household consumption expenditure; heavily incorporates consumption expenditure data gathered through the household surveys. Thus, it is expected that this pair of variables will have smaller gaps.

[11] It was recorded that the data from the first 3 rounds of the HES were compromised. The weaknesses were considered serious enough to discard the datasets. Only selected aggregate data from the survey were published separately (Bangladesh Bureau of Statistics, 1988a).

Across the 14 rounds of household surveys that data is available for, the micro-macro ratio for per capita monthly household consumption expenditure is 0.83. The micro-macro ratio for per capita monthly household income is 0.79 and the micro-macro ratio for per capita monthly household expenditure is 0.64. For all three pairs of variables, the largest divergence is observed in 2017. It is also observed that all three variable pairs experience a drop in performance from 1993 and again in 2017.

The results of this section indicate that in the initial years, income, consumption, and expenditure values produced through household surveys were good reflections of the national aggregate income, consumption, and expenditure, respectively. However, since 1993, the difference between the two has been quite large and getting progressively bigger, the most recent two rounds show the worst performance. This could be a result of the following (i) methodological errors or inconsistencies; (ii) under-sampling of the rich in household surveys; (iii) over-sampling of the poor in household surveys; (iv) increased economic inequality, in which the wealth of the super-rich can significantly disrupt the averages; (v) under-estimation of population growth; and (vi) intentional data manipulation by government agencies.

The subsequent sections will discuss the results of the other empirical strategies employed in this paper. These strategies will help determine whether the micro data for Bangladesh is reliable. It will also help identify, which of the above causes are contributing to the discrepancy between the micro and macro data of Bangladesh.

2.5.2 Benford's Distribution

Disaggregated household level data is available for 10 rounds of household surveys. These are HES 1981, HES 1983, HES 1985, HES 1989, HES 1992, HES 1996, HIES 2000, HIES 2005, HIES 2011 and HIES 2017. The data quality of these 10 rounds was assessed by applying Benford's Law to the monthly household expenditure values collected from each round. These results are given in Appendix 2.7.

The results show that for every round of the household surveys, the observed distribution was different from the expected distribution. To test whether the deviations, indicate issues with data reliability, the K-S non-parametric test was applied. The results of the K-S tests are given in Table 2.3.

The results indicate that there are fundamental issues present in the micro data collected through the household surveys. The fact that the null hypothesis can be rejected for every single round, further cements that theory. The next section explores the possible sources of the weaknesses in the household data.

For all 10 rounds, p-value < 0.005, implying that the null-hypothesis should be rejected. In other words, the distribution of the data differs significantly from theoretical expectations.

The results indicate that there are fundamental issues present in the micro data collected through the household surveys. The fact that the null hypothesis can be rejected for every single round, further cements that theory. The next section explores the possible sources of the weaknesses in the household data.

Table 2.3: Kolmogorov-Smirnov Test Results

Survey Round	D- statistic	p-value
HES 1981	D = 10.575	p-value < 0.01
HES 1983	D = 17.111	p-value < 0.01
HES 1985	D = 19.348	p-value < 0.01
HES 1989	D = 16.228	p-value < 0.01
HES 1992	D = 14.469	p-value < 0.01
HES 1996	D = 09.574	p-value < 0.01
HIES 2000	D = 08.362	p-value < 0.01
HIES 2005	D = 13.632	p-value < 0.01
HIES 2011	D = 14.432	p-value < 0.01
HIES 2017	D = 10.867	p-value < 0.01

2.5.3 Administrative Data

The analyses in the previous sections determined that the primary sources of data for Bangladesh cannot be triangulated by comparing micro data sets against their corresponding macro indices. It is also seen that the gap between micro and macro data has widened over time.

To identify whether the source of the discrepancy lies within the micro data, Benford's Law was applied to a selected variable from the household surveys. The results indicate that the surveys do not conform to Benford's distribution. Thus, it is evident, that there are weaknesses in the household survey data. This section explores whether the weaknesses are a result of sampling errors.

As discussed in Section 2.4.3, logistic regression models were used to determine the relationship of the four chosen indicators with each income quartile. The results revealed that income tax-paying status, car ownership, motorbike ownership and household electrification were good predictors of a household falling in the top income quartile for all 10 household survey rounds. The only deviation from this trend was observed in the data for HES 1983, in which there appears to be no consistent relationship between income quartiles and tax-paying status. However, even that year, the relationship with electricity maintains its general course. The logistic regression results showed that each of these indicators were strongly and positively correlated with the first income quartile. At the same time, the correlations were either negative or less than one-fourth in strength for the third quartile onwards. Their correlations weakened progressively as we moved down the income quartiles. All results were statistically significant ($p < 0.01$). These regression results are attached in Appendix 2.8.

2.5.3.1 Income Taxes

Though income tax returns are filed at the individual level in Bangladesh, the income tax data available from the household surveys were collected at the household level. Thus, it was not possible to directly compare the household survey data to the administrative data. However, it was possible to extrapolate the total income taxes paid by households, from the household surveys. This information was then compared to the total income tax amount realized by the GoB in that respective year.

Appendix 2.9 includes two graphs. The one on the left shows the difference between the true income tax revenue and the estimated income tax revenue. It shows that the total amount of income tax extrapolated using the household survey data is much lower than the actual amount collected by the government. The discrepancy between the two has gotten increasingly bigger over the analyzed period. The graph on the right plots the ratio of estimated income tax revenue to true income tax revenue. It shows that the ratio has been small throughout the analyzed period. The highest value of the ratio was 8 percent and the lowest was 0.4 percent.

This indicates that the income tax paying population was under-sampled in the survey. Since, it has been established that income taxpayers are most likely from the top income quartile, it suggests under-sampling of the richest 25 percent of the population in all periods, except 1983 (Appendix 2.8).

2.5.3.2 Electricity

For the 1981 to 2001 period, administrative data is available for the number of electric connections provided to private households. When this data is compared to the extrapolated data from the household surveys, it shows that for all periods, there was over-sampling of electrified households (Appendix 2.10). Since, it has been established that household electrification is strongly correlated with being in the top income quartile, this indicates over-sampling of households in the top income quartile.

2.5.3.3 Vehicle Ownership

Prior to 2001, the household surveys did not collect disaggregated data on vehicle ownership. Thus, comparison of survey data and administrative data of vehicle ownership could only be done for the 2001 to 2017 period. This analysis also revealed over-sampling of the top income quartile of the population from 2001 to 2011. However, for 2017, the numbers for private car and motorbike ownership exceeded the number estimated from the household survey. Thus, indicating under-sampling of households in the top income quartile. These comparisons are shown in Appendix 2.11.

The comparison of the survey data against the administrative data reveals that for all survey periods, except 2017, households in the top income-quartile were over-sampled. However, it also reveals that the estimated income tax revenue for all periods, was a lot lower than the true income tax revenue. This implies that either the income tax-paying segment was under-sampled or the highest income taxpayers were under-sampled. Since, less than 2 percent of the population pay income taxes (National Board of Revenue, 2018), both are probable scenarios.

On the other hand, in the last period, there appears to be over-sampling of households in the top income quartile. This contradicts the markedly larger discrepancy between micro and macro data in this period. However, results also show that even in this period there was under-sampling of taxpayers and/or under-sampling of the highest taxpayers. A possible explanation might be under-sampling of the 'highest earners' while over-sampling of the 'top quartile'.

2.6 Limitations

Government published population estimates have been used throughout this chapter. The last population census of Bangladesh was conducted in 2011. At the time, government statistics reported population to be 142 million, while the Central Intelligence Unit (CIA) estimated the population to be 158.6 million (World Population Review, n.d.). Under-estimation of population would have over-estimated the macroeconomic indices and thus effected the results published in this paper.

Secondly, administrative data is frequently found to be unreliable. One glaring example, is the regular discord between the MoA (Ministry of Agriculture) and the BBS (Bangladesh Bureau of Statistics), on the issue of calculating crop output. Invariably, the amount reported by the MoA is higher, presumably due to in-built upward bias, resulting from a need to reach set targets (Osmani, 2017).

Thirdly, it is possible that the administrative data available for private vehicle ownership in 2017 is inaccurate. It was observed that, in the last period the number of privately registered motorbikes and motor cars owned shot up. This was most likely caused by the introduction and growth of ride-sharing services. It is estimated that the annual sales of motorbikes more than doubled due to the introduction of these services. Vehicles for these services are often purchased as privately owned assets instead of business assets. This distinction might have been picked up by enumerators during surveys, resulting in the large gap. If this is the case, it would affect the inferences drawn from the results.

2.7 Conclusion

This study was able to incorporate a range of multi-disciplinary tools in an innovative manner. By doing so, it was able to not only conclude that there are inconsistencies in the data available in Bangladesh, but also identify some of the sources of weakness.

Firstly, it establishes that triangulating micro and macro data is unable to validate the data available for Bangladesh. Secondly, by applying Benford's Law, it is shown that the micro dataset does not conform to theoretical expectations of leading digit distributions. Lastly, using logistic regression models, it shows that the household surveys over-sampled the richest quartile of the population, while under-sampling income taxpayers. The only period for which this is not the case, is 2017. In this period, both the richest quartile and income taxpayers were under-sampled.

The growing difference between micro and macro data, despite the over-sampling of households in the top quartile, point to even more unfavourable conclusions regarding income inequality in Bangladesh. This is compounded by the fact, that for all periods income taxpayers were under-sampled. It may be postulated, that correcting for these sampling errors would show much wider income equality than is currently known.

2.8 Appendices

Appendix 2.1: Sources of Micro-level Data

Sl.	Round	Rural Households	Urban Households	Total Households	No. of Individuals	Household Level Data Availability
1	HES 1973-74[12]	9,536	2,237	11,773	68,636	+
2	HES 1974-75					+
3	HES 1975-76					+
4	HES 1976-77[13]					+
5	HES 1977-78[14]					+
6	HES 1978-79[15]					+
7	HES 1981-82[16]	5,949	3,614	9,563	55,173	✓
8	HES 1983-84[17]	2,112	1,728	3,840	22,173	✓
9	HES 1985-86[18]	2,112	1,728	3,840	22,844	✓
10	HES 1988-89[19]	3,804	1,871	5,630	31,284	✓
11	HES 1991-92[20]	3,840	1,920	5,696	30,499	✓
12	HES 1995-96[21]	5,040	2,380	7,420	39,051	✓
13	HIES 2000[22]	5,040	2,401	7,441	38,518	✓
14	HIES 2005[23]	6,040	4,040	10,080	48,969	✓
15	HIES 2010[24]	7,840	4,400	12,240	55,776	✓
16	HIES 2016-17[25]	32,096	13,980	46,076	1,86,067	✓

[12] Bangladesh Bureau of Statistics, 1978
[13] Bangladesh Bureau of Statistics, 1981
[14] Bangladesh Bureau of Statistics, 1983
[15] Starting from 1973-74, BBS conducted expenditure surveys every year, till 1978-79. However, the results of the surveys conducted in the first three years did not meet the test of acceptability and remained unpublished (Bangladesh Bureau of Statistics, 1986)
[16] The digital versions of HES 1976-77, HES 1977-78 and HES 1978-70 are unavailable. Unfortunately, detailed reports of these surveys were also never published due to delay in data processing. However, some selected tables of the surveys 1976-77, 1977-78, and 1978-79 were published in and. Thus, only aggregate data is available for these years. Crucial information, such as, sample size and survey beginning and end dates are, hence missing. It was assumed, that the dates corresponded to fiscal year beginning and end dates. (Bangladesh Bureau of Statistics, 1991)
[17] Bangladesh Bureau of Statistics, 1988a
[18] Bangladesh Bureau of Statistics, 1988b
[19] Bangladesh Bureau of Statistics, 1991
[20] Bangladesh Bureau of Statistics, 1988a
[21] Bangladesh Bureau of Statistics, 1998
[22] Bangladesh Bureau of Statistics, 2003a
[23] Bangladesh Bureau of Statistics, 2007
[24] Bangladesh Bureau of Statistics, 2011
[25] Bangladesh Bureau of Statistics, 2017b

Micro Variable	Definition	Equations[26]
Monthly Household Income Per Capita	Total income from agriculture, business and commerce, professional wages and salary, housing services, and gifts and remittances amongst others. Taxes and other withholdings are not deducted from this figure.	*(i) household income = agriculture + business + wages and salary + housing services + gifts and remittances*
Monthly Household Expenditure Per Capita[27]	Total expenditure on food and beverages, clothing and footwear, housing and house rent, fuel and lighting, household effects, medical expenses, education, taxation, insurance, hajj/pilgrimage, and marriage. Since investment expenditure is made to generate income for future consumption, factoring it in would amount to double counting and is thus excluded.	*(i) household expenditure = value of all expenditures made by household members from their own income and loans* *(ii) household expenditure= consumption expenditure + new housing + transfer payments*
Monthly Household Consumption Expenditure Per Capita[28]	Unlike 'expenditure' calculations, 'consumption' does not consider large, one-off expenditures, e.g., purchase of durable goods. In addition, payment of tax, insurance, expenses of pilgrimage/hajj, marriage, etc. are excluded.	*(i) household consumption expenditure = value of all consumption by household members from their own income, loans, home production and gifts.* *(ii) household consumption expenditure = food and beverage + clothing and footwear + housing and house rent + fuel and lighting + household effects + miscellaneous*

[26] Bangladesh Bureau of Statistics (2011)

[27] Gifts and remittances received which should only be added in consumption, were added in household expenditure calculations. Due to lack of information, this error could not be corrected.

[28] Household expenditure and consumption are both likely overestimated in the household surveys. This is because 'gifts and remittances' received and given out are clubbed together and added to both household expenditure and household consumption, leading to double counting; and depreciation of durable goods was not factored in.

Appendix 2.3: Macro Variables - Equivalency & Equations

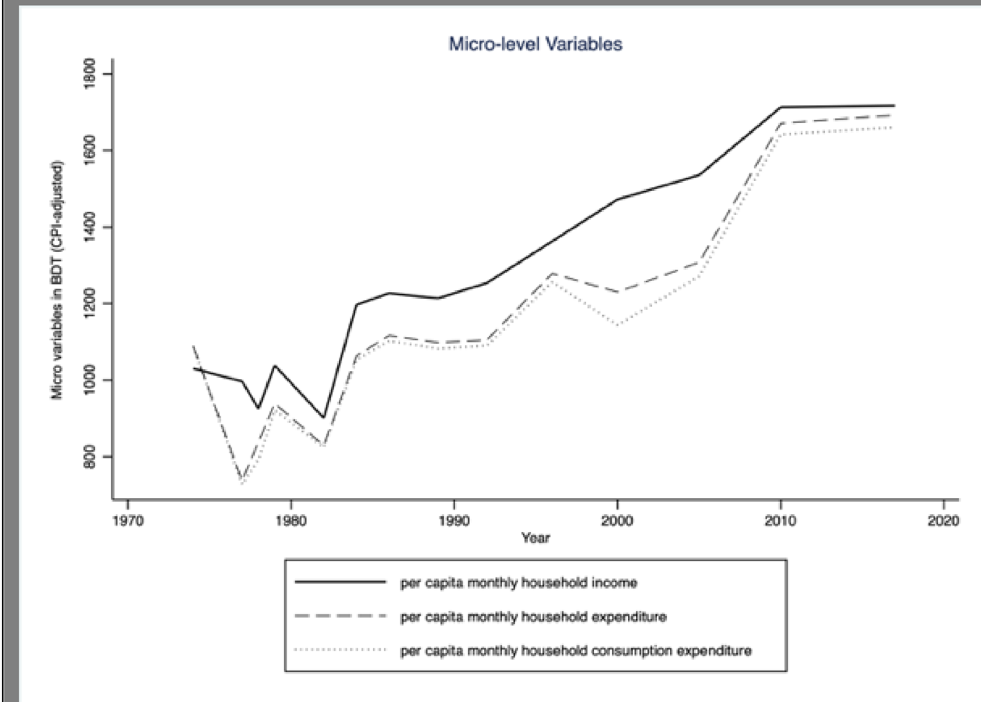
Macro Variable	Equivalency Rationale	Equations[29]
Per Capita Monthly Gross National Income	The household level income data were generated using a survey that only covered households within the borders of Bangladesh. Theoretically, the total production level of an economy should translate to total income earned by its residents. Since GDP is a metric that measures the total value of goods and services produced in a country, it may be assumed that GDP is a comparable aggregate for household income. On the other hand, GNI is the total value of all goods and services produced by a country and the income its residents receive from home and abroad. Additionally, GNI deducts money flowing out of the country, whereas GDP does not. The household survey, also, factored in foreign remittance and gifts received and transferred, in its calculation of household income. Thus, it was logically inferred that it is more reasonable to compare GNI to household income, than GDP.	*(i) GNI = GNDI - net international transfers* *(ii) GNI = GDP at current market price + net primary income from rest of the world*
Per Capita Total Expenditure	It is taken into cognizance that GDE[30] only considers expenditure made within the borders of the country. On the other hand, household expenditure calculated using the surveys factored in remittance to members living overseas and consumption of imported goods. Thus, it was inferred that total expenditure, which is the sum of GDE and imports, is the better macro counterpart.	*(i) total expenditure = GDE + imports* *(ii) GDE = C (consumption) +G (government expenditure) + K (gross fixed capital formation) + S (change in inventories) - X (exports) + M (imports)* *(iii) GDE = gross value of output - value of intermediate consumption*
Per Capita Private Final Consumption Expenditure (PFCE)[31]	There are two macro variables that might be comparable to the micro data generated for consumption. The first, PFCE, does not take government consumption expenditure into account and the second, final consumption expenditure, does. Households likely consume, the products and services produced through government spending e.g., public schools and hospitals. Social security, wages and salaries of government employees are components of government consumption expenditure. However, government expenditures also include other dimensions that are not factored into the calculation of household consumption expenditure e.g., public infrastructures, research, roads, public transport, military expenditures, etc. All these together likely make up a large fraction of government expenditure which household surveys cannot account for. Thus, it was decided that private final consumption expenditure would be a better macro aggregate for household consumption expenditure than final consumption expenditure.	*(i) PFCE = final consumption expenditure - government final consumption expenditure (GFCE)*

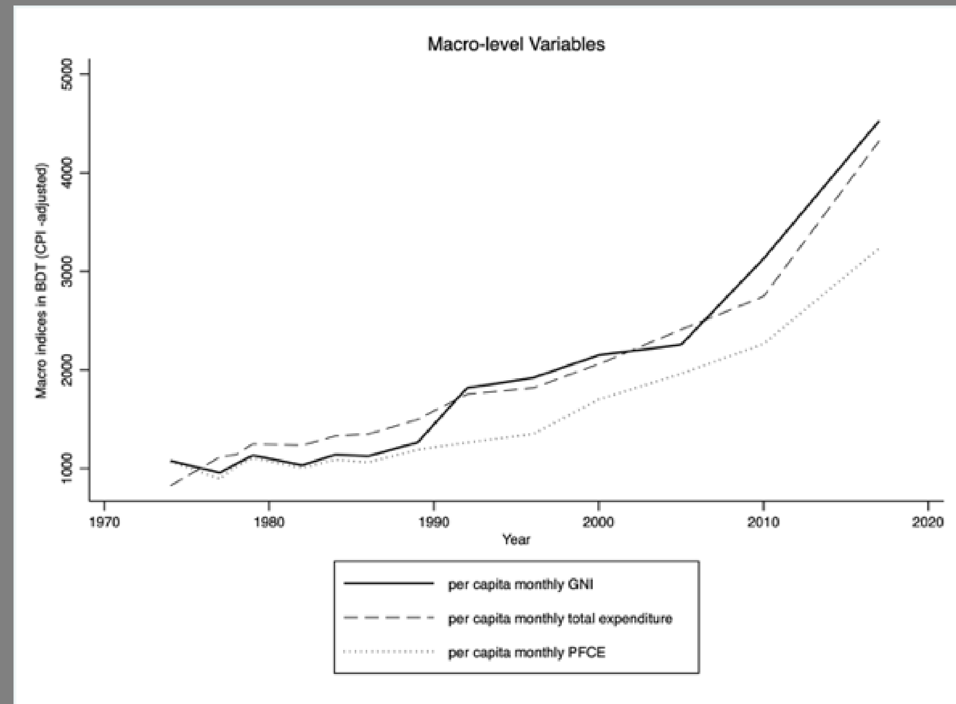
[29] Bangladesh Bureau of Statistics, 2014a

[30] There is a difference between GDP and GDE estimates. This difference occurs due to use of independent data sets and methodologies. The magnitude of the difference generally lies within 3-4 percent at current prices. This is termed as statistical discrepancy.

[31] These data were also published in terms of annual national figures and were transformed to monthly per capita figures. It should be noted that the report did not compile private consumption expenditure data directly due to lack of basic data on consumption. Even though, consumption aggregates were available from the HES, these surveys were not conducted every year and were inconsistent with production data. Thus, private consumption estimate for the period is residually determined and is equivalent to what remains after government expenditure and, public and private involvements are deducted from total national consumption expenditure.

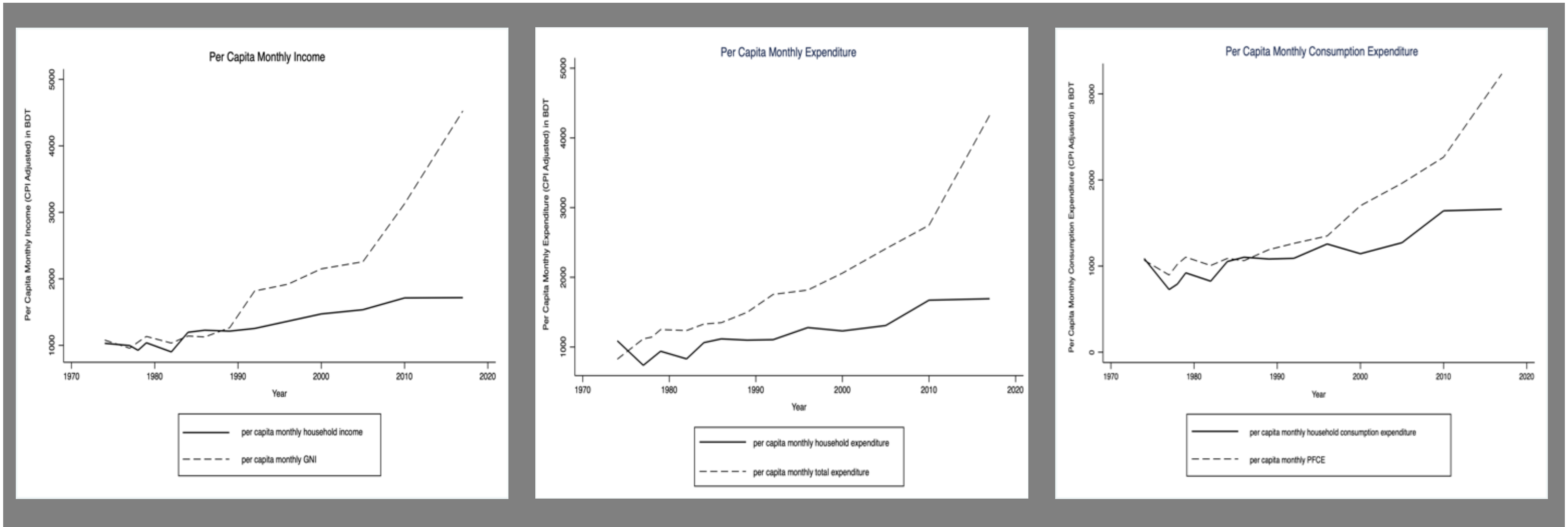
Appendix 2.4: Comparing Income, Expenditure & Consumption Levels





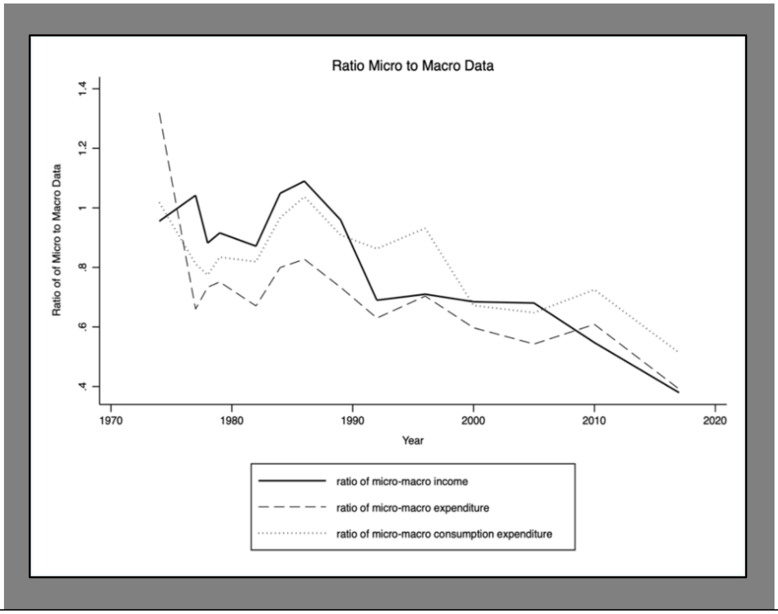
Appendix 2.4 depicts per capita monthly household income, per capita monthly household expenditure, and per capita monthly household consumption expenditure in the graph on the left. It is seen that, income exceeds expenditure and expenditure exceeds consumption expenditure, in almost all periods. The expenditure and consumption expenditure lines almost converge. But there is a noticeable gap between the income line and the other two lines. However, in the two most recent household survey rounds, the three variables began exhibiting a tendency towards convergence. The graph on the right depicts the levels of the selected macro-indices, i.e., the per capita monthly GNI, the per capita monthly total expenditure and the per capita monthly PFCE. Though, the rankings were not static throughout the observed period, it is seen, that mostly, GNI is at the top, followed by total expenditure and PFCE, respectively.

Appendix 2.5: Comparison of Micro & Macro Per Capita Monthly Income, Expenditure & Consumption



Appendix 2.5 consists of three graphs. These plot the micro variables against their corresponding macro indices. The data presented extends from 1974 to 2017. It is evident that the trend lines for macro-sourced and micro-sourced data are divergent. The gap between the micro and macro variables grew progressively over the period. This is true for income, expenditure as well as, consumption.

Appendix 2.6: Ratio of Micro to Macro Data



Appendix 2.6 plots the information of **Appendix 2.5** as ratios. The three lines correspond to the ratio of micro to macro income, the ratio of micro to macro expenditure and the ratio of micro to macro consumption expenditure. Overall, it appears that the ratios for all three variables, shrunk over the 1974 to 2017 period, i.e., data drawn from household surveys has become less and less representative of the national accounts. In other words, the macro indices, generated through the national accounts, grew at a faster rate than the micro variables generated through the household surveys.

Appendix 2.7: Deviation from Benford's Distribution

Leading Digit	Benford's Distribution	Observed Distribution									
		HES 1981	HES 1983	HES 1985	HES 1989	HES 1992	HES 1996	HIES 2000	HIES 2005	HIES 2011	HIES 2017
1	**30.1**	31.9	47.2	41.6	38.8	33.7	26.1	31.4	16.5	31.5	41.0
2	**17.6**	10.0	16.6	25.5	25.1	25.8	25.6	23.3	19.0	9.0	15.6
3	**12.5**	8.3	7.1	9.4	12.1	15.2	16.9	13.9	18.3	6.8	7.6
4	**9.7**	9.2	3.8	5.5	6.5	7.9	10.9	9.3	14.2	8.0	5.9
5	**7.9**	10.0	3.8	3.1	3.9	5.1	7.0	6.4	10.6	10.4	5.4
6	**6.7**	9.1	4.7	4.0	3.3	3.4	4.7	4.5	8.0	9.8	5.5
7	**5.8**	8.1	5.2	3.1	3.4	2.6	3.6	4.0	5.5	8.8	5.8
8	**5.1**	7.4	5.8	3.1	3.3	3.1	2.8	3.7	4.7	8.3	6.2
9	**4.6**	6.2	5.9	4.6	3.6	3.2	2.4	3.5	3.3	7.2	5.7

Appendix 2.7 assesses the quality of 10 rounds of HIES data by applying Benford's Law to the monthly household expenditure collected from each round. Disaggregated household level data is available for 10 rounds of household surveys. These are HES 1981, HES 1983, HES 1985, HES 1989, HES 1992, HES 1996, HIES 2000, HIES 2005, HIES 2011 and HIES 2017. The results show that for every round of the household surveys, the observed distribution was different from the expected distribution.

Appendix 2.8: Logistic Regression Results [32]

HES 1981

VARIABLES	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1
tax_payer	2.754***	-0.933***	-0.959***					
	(0.008)	(0.011)	(0.010)					
electricity					2.402***	-0.092***	-1.342***	-2.430***
					(0.002)	(0.002)	(0.003)	(0.005)
Constant	-1.612***	-1.136***	-0.944***	-0.797***	-1.852***	-1.134***	-0.883***	-0.715***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	14,784,668	14,784,668	14,784,668	14,702,135	14,784,668	14,784,668	14,784,668	14,784,668

HES 1983

VARIABLES	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1
tax_payer	-0.322***	-0.306***	0.205***	0.246***				
	(0.007)	(0.006)	(0.005)	(0.005)				
electricity					1.403***	0.357***	-0.975***	-1.358***
					(0.002)	(0.002)	(0.003)	(0.004)
Constant	-1.465***	-1.132***	-0.955***	-0.892***	-1.572***	-1.155***	-0.912***	-0.838***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	16,310,592	16,310,592	16,310,592	16,310,592	16,310,592	16,310,592	16,310,592	16,310,592

HES 1985

VARIABLES	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1
tax_payer	2.836***	-0.094***	-3.192***					
	(0.009)	(0.009)	(0.029)					
electricity					1.850***	0.089***	-0.613***	-1.561***
					(0.002)	(0.002)	(0.002)	(0.002)
Constant	-1.687***	-1.146***	-0.911***	-0.772***	-2.023***	-1.158***	-0.851***	-0.642***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	16,938,432	16,938,432	16,938,432	16,865,036	16,938,432	16,938,432	16,938,432	16,938,432

[32] inc_1 = lowest income quartile, inc_2 = second lowest (median) income quartile, inc_3 = second highest (median) income quartile and inc_4 = highest income quartile

(continued from previous page)

HES 1989

VARIABLES	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1
tax_payer	1.580***		-0.195***	-0.397***				
	(0.010)		(0.011)	(0.012)				
electricity					1.215***	0.117***	-0.598***	-1.222***
					(0.002)	(0.002)	(0.002)	(0.003)
Constant	-1.252***	-1.081***	-1.049***	-1.020***	-1.396***	-1.096***	-1.000***	-0.934***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	18,634,262	18,590,000	18,634,262	18,634,262	18,634,262	18,634,262	18,634,262	18,634,262

HES 1992

VARIABLES	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1
tax_payer	1.160***	0.178***	-0.802***	-0.915***				
	(0.011)	(0.012)	(0.016)	(0.016)				
electricity					1.675***	0.213***	-0.692***	-1.920***
					(0.001)	(0.001)	(0.002)	(0.002)
Constant	-1.411***	-1.116***	-1.008***	-0.896***	-1.728***	-1.146***	-0.929***	-0.731***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	20,110,112	20,110,112	20,110,112	20,110,112	20,110,112	20,110,112	20,110,112	20,110,112

HES 1996

VARIABLES	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1
tax_payer	2.318***	-1.434***	-0.366***					
	(0.009)	(0.016)	(0.010)					
electricity					2.185***	0.301***	-1.378***	-2.367***
					(0.001)	(0.001)	(0.002)	(0.002)
Constant	-1.343***	-1.134***	-1.003***	-0.938***	-1.925***	-1.195***	-0.817***	-0.690***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
Observations	22,121,949	22,121,949	22,121,949	22,063,776	22,121,949	22,121,949	22,121,949	22,121,949

(continued from previous page)

HIES 2000

VARIABLES	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1
tax_payer	3.474***	-1.486***	-3.079***	-2.971***								
	(0.007)	(0.009)	(0.018)	(0.016)								
electricity					1.934***	0.300***	-0.746***	-1.806***				
					(0.001)	(0.001)	(0.001)	(0.001)				
motorbike									2.037***	-0.165***	-1.665***	-2.047***
									(0.005)	(0.006)	(0.010)	(0.011)
Constant	-1.318***	-1.134***	-1.023***	-0.939***	-2.084***	-1.236***	-0.838***	-0.566***	-1.305***	-1.140***	-1.026***	-0.942***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	24,346,960	24,346,960	24,346,960	24,346,960	24,346,960	24,346,960	24,346,960	24,346,960	24,346,960	24,346,960	24,346,960	24,346,960

HIES 2005

VARIABLES	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1
tax_payer	2.298***	-1.267***	-0.639***									
	(0.007)	(0.011)	(0.008)									
car					2.525***	-0.885***	-1.798***	-2.627***				
					(0.007)	(0.009)	(0.012)	(0.018)				
motorbike									1.676***	-0.265***	-0.824***	-2.208***
									(0.003)	(0.004)	(0.005)	(0.009)
Constant	-1.123***	-1.095***	-1.097***	-1.080***	-1.125***	-1.095***	-1.095***	-1.080***	-1.139***	-1.095***	-1.091***	-1.070***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	28,644,100	28,644,100	28,644,100	28,528,700	28,644,100	28,644,100	28,644,100	28,644,100	28,644,100	28,644,100	28,644,100	28,644,100

(continued from previous page)

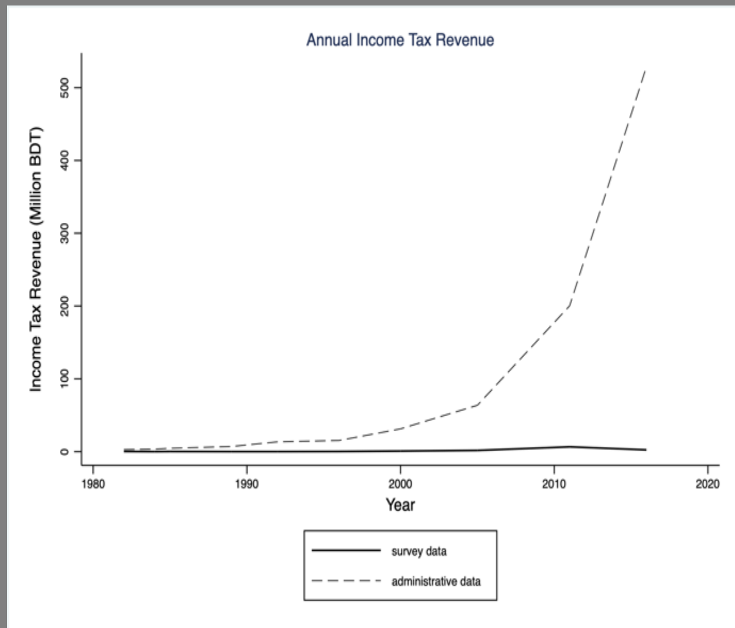
HIES 2010

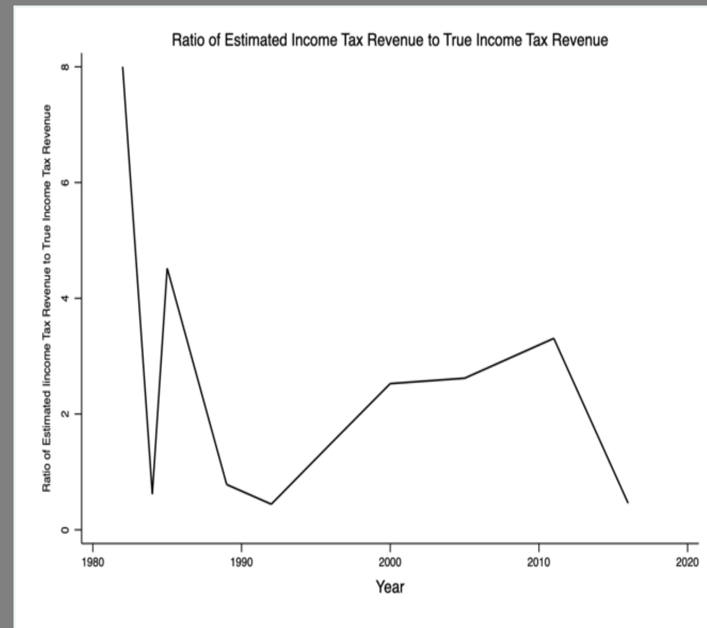
VARIABLES	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1
tax_payer	2.809***	-1.364***	-1.926***	-2.376***								
	(0.004)	(0.006)	(0.008)	(0.009)								
car					2.419***	-0.683***	-1.803***	-3.067***				
					(0.006)	(0.006)	(0.010)	(0.018)				
motorbike									1.996***	-0.324***	-1.639***	-2.048***
									(0.002)	(0.003)	(0.005)	(0.005)
Constant	-1.115***	-1.094***	-1.126***	-1.061***	-1.094***	-1.101***	-1.132***	-1.068***	-1.138***	-1.097***	-1.114***	-1.047***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	33,027,760	33,027,760	33,027,760	33,027,760	33,027,760	33,027,760	33,027,760	33,027,760	33,027,760	33,027,760	33,027,760	33,027,760

HIES 2017

VARIABLES	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1	(1) inc4_4	(2) inc4_3	(3) inc4_2	(4) inc4_1
tax_payer	0.900***	-0.116***	-0.529***	-0.822***								
	(0.004)	(0.005)	(0.006)	(0.006)								
car					0.849***	-0.276***	-0.445***	-0.564***				
					(0.005)	(0.006)	(0.006)	(0.007)				
motorbike									0.627***	0.136***	-0.311***	-0.885***
									(0.002)	(0.002)	(0.003)	(0.003)
Constant	-0.865***	-1.088***	-1.228***	-1.235***	-0.862***	-1.087***	-1.230***	-1.238***	-0.884***	-1.096***	-1.221***	-1.212***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	24,036,461	24,036,461	24,036,461	24,036,461	24,036,461	24,036,461	24,036,461	24,036,461	24,036,461	24,036,461	24,036,461	24,036,461

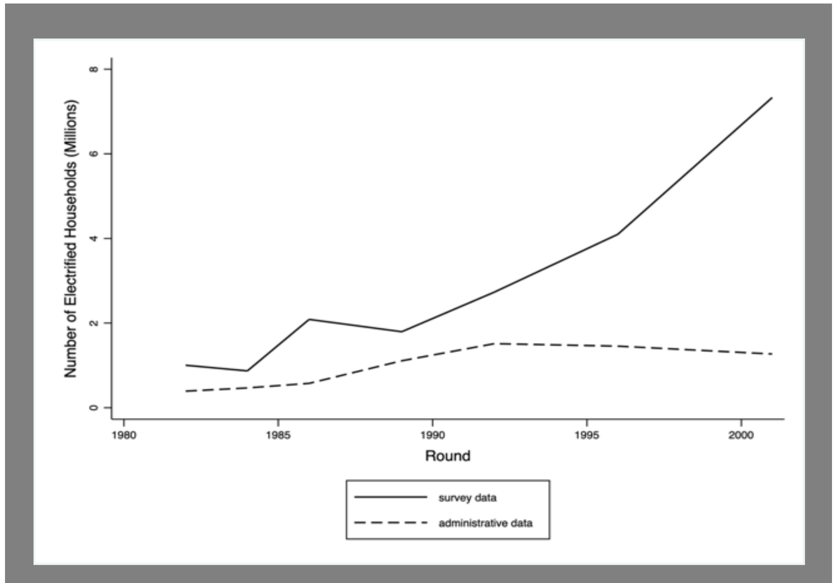
Appendix 2.9: True vs. Estimated Income Tax Revenue





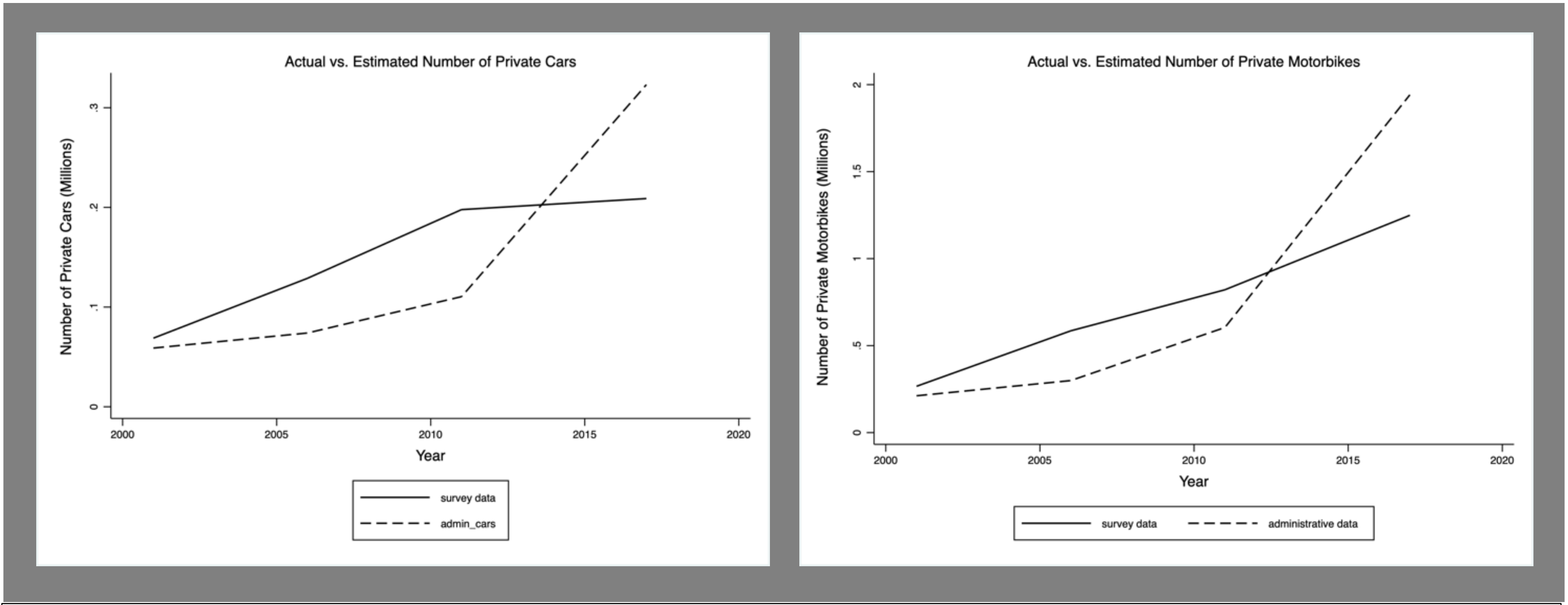
Appendix 2.9 includes two graphs. The one on the left shows the difference between the true income tax revenue and the estimated income tax revenue. It shows that the total amount of income tax extrapolated using the household survey data is much lower than the actual amount collected by the government. The discrepancy between the two has gotten increasingly bigger over the analyzed period. The graph on the right plots the ratio of estimated income tax revenue to true income tax revenue. It shows that the ratio has been small throughout the analyzed period. The highest value of the ratio was 8 percent and the lowest was 0.4 percent.

Appendix 2.10: True vs. Estimated Number of Electrified Households



Appendix 2.10 For the 1981 to 2001 period, administrative data is available for the number of electric connections provided to private households. When this data is compared to the extrapolated data from the household surveys, it shows that for all periods, there was over-sampling of electrified households. Since, it has been established that household electrification is strongly correlated with being in the top income quartile, this indicates over-sampling of households in the top income quartile.

Appendix 2.11: True vs. Estimated Number of Private Vehicles



Appendix 2.11 Prior to 2001, the household surveys did not collect disaggregated data on vehicle ownership. Thus, comparison of survey data and administrative data of vehicle ownership could only be done for the 2001 to 2017 period. The graph on the left depicts the actual number of private cars against the survey estimations. The graph on the right depicts the actual number of privately owned motorbikes against the survey estimates. In both cases it is see that till 2010 surveys exceeded actual figures. However, for 2017, the numbers for private car and motorbike ownership exceeded the number estimated from the household survey.

2.9 Bibliography

Amel, D. F., Kennickell, A. B., & Moore, K. B. (2008). *Banking Market Definition: Evidence from the Survey of Consumer Finances*. https://www.federalreserve.gov/econres/feds/banking-market-definition-evidence-from-the-survey-of-consumer-finances.htm

Bangladesh Bureau of Statistics. (1978). *A report on the household expenditure survey of Bangladesh 1973-74*. Statistics Division, Ministry of Planning. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2Fc60552ad-42b2-4f49-99ca-5a068d5429f2.pdf

Bangladesh Bureau of Statistics. (1981). *Statistical Yearbook of Bangladesh 1980*. Statistics Division, Ministry of Planning. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2F9782a070-1b6f-408e-863d-1f2abedd016a.pdf

Bangladesh Bureau of Statistics. (1983). *Statistical Yearbook of Bangladesh 1982-83*. Statistics Division, Ministry of Planning. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2F9b2545c7-786f-4d67-8c42-da039862a8f0.pdf

Bangladesh Bureau of Statistics. (1984). *Statistical Yearbook of Bangladesh 1983-84*. Statistics Division, Ministry of Planning. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2F44702778-6fc2-4c14-97cb-9fcdd621cb3f.pdf

Bangladesh Bureau of Statistics. (1986). A report on the household expenditure survey of Bangladesh 1981-82. Statistics Division, Ministry of Planning. http://203.112.218.73:8082/Account/Login

Bangladesh Bureau of Statistics. (1988a). *A report on the household expenditure survey of Bangladesh 1983-84*. Statistics Division, Ministry of Planning. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2F0039334f-6c8c-412b-a55d-f8b06bb88c7f.pdf

Bangladesh Bureau of Statistics. (1988b). *A report on the household expenditure survey of Bangladesh 1985-86*. Statistics Division, Ministry of Planning. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2F8aa8baa2-3b4d-4f89-919c-a854842ed686.pdf

Bangladesh Bureau of Statistics. (1991). *A report on the household expenditure survey of Bangladesh 1988-89*. Statistics Division, Ministry of Planning. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2F245a2c32-c421-4a11-bd77-39498be2fb28.pdf

Bangladesh Bureau of Statistics. (1993). *Twenty Years of National Accounting of Bangladesh (1972-73 to 1991-92)*. Statistics Division, Ministry of Planning. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2F50b17c27-c97c-4a0a-801a-ac43bac2f232.pdf

Bangladesh Bureau of Statistics. (1995). *Report on the household expenditure survey of Bangladesh 1991-92*. Statistics Division, Ministry of Planning. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2F8dbfd111-3d07-481d-9618-ce125f068d9b.pdf

Bangladesh Bureau of Statistics. (1998). *Report on the household expenditure survey of Bangladesh 1995-96*. Statistics Division, Ministry of Planning. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2Ff0371002-5f98-416f-9516-b5eece55db9b.pdf

Bangladesh Bureau of Statistics. (2003a). *Report on the household expenditure survey of Bangladesh 2000*. Statistics Division, Ministry of Planning. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2Fdf6b6926-2e55-4d7a-9d00-ae38644818f6.pdf

Bangladesh Bureau of Statistics. (2003b). *Statistical Pocketbook of Bangladesh (1978-2002)*. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2Fd45566c3-f8c7-4303-867a-88b6f6a82304.pdf

Bangladesh Bureau of Statistics. (2007). *Report on the household expenditure survey of Bangladesh 2005*. Statistics Division, Ministry of Planning. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2F4753a685-bbe3-4d08-8c49-601481545076.pdf

Bangladesh Bureau of Statistics. (2011). *Report on the household expenditure survey of Bangladesh 2010*. Statistics Division, Ministry of Planning. http://203.112.218.65:8008/WebTestApplication/userfiles/Image/LatestReports/HIES-10.pdf

Bangladesh Bureau of Statistics. (2014). *National Accounts Statistics of Bangladesh (provisional estimates of GDP 2013-14 and final estimates of GDP 2012-13)*. Statistics Division, Ministry of Planning. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2F5ee23343-0808-44b2-afb7-1e61584cc59e.pdf

Bangladesh Bureau of Statistics. (2014). Bangladesh National Accounts Statistics: Sources and Methods. Bangladesh Bureau of Statistics. http://203.112.218.65:8008/WebTestApplication/userfiles/Image/LatestReports/Na_Rebase_2005-06.pdf

Bangladesh Bureau of Statistics. (2017a). *Consumer Price Index (CPI), Inflation Rate, and Wage Rate Index (WRI) in Bangladesh*. RDP Section, FA and MIS Wing, Bangladesh Bureau of Statistics. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2F6f975a73-a7d4-4c77-8bf6-2355f351c814.pdf

Bangladesh Bureau of Statistics. (2017b). *Preliminary report on the household expenditure survey of Bangladesh 2016*. Statistics Division, Ministry of Planning. http://bbs.portal.gov.bd/sites/default/files/files/bbs.portal.gov.bd/page/b343a8b4_956b_45ca_872f_4cf9b2f1a6e0/HIES%20Preliminary%20Report%202016.pdf

Bangladesh Bureau of Statistics. (2019). *Statistical Pocketbook of Bangaldesh 2018*. http://203.112.218.73:8082/Home/DownloadPDF?filePath=%2FPDF%2F0f501abd-1868-4f34-8f07-c09d7454fa4c.pdf

Bangladesh Road Transport Authority. (2018). *Number of Registered Motor Vehicles in Bangladesh (Yearwise)*. https://brta.portal.gov.bd/sites/default/files/files/brta.portal.gov.bd/monthly_report/d4d56177_644f_44f8_99c4_3417b3d7b0f4/MV_statistics-bangladesh-march-18.pdf

CaseWare IDEA. (2007). *A Guide to Benford's Law*. CaseWare IDEA Research Department. https://silo.tips/download/a-guide-to-benford-s-law-a-caseware-idea-research-department-document

Fixler, D., Johnson, D., Craig, A., & Furlong, K. (2017). A Consistent Data Series to Evaluate Growth and Inequality in the National Accounts. *Review of Income and Wealth*, *63*, S437–S459. https://doi.org/10.1111/roiw.12324

Green, J. (2017, September 26). *The Differences in the GDP Deflator & CPI*. Bizfluent. https://bizfluent.com/info-8434694-differences-gdp-deflator-cpi.html

Karshenas, M. (2003). Global Poverty: National Accounts Based versus Survey Based Estimates. *Development and Change*, *34*(4), 683–712. https://doi.org/10.1111/1467-7660.00324

Khan, A. R. (2005). Measuring Inequality and Poverty in Bangladesh: An Assessment of the Survey Data. *The Bangladesh Development Studies*, *31*(3/4), 1–34. https://www.jstor.org/stable/40795713

Maki, A., & Nishiyama, S. (1993). Consistency Between Macro- and Micro-Data Sets in the Japanese Household Sector. *Review of Income and Wealth*, *39*(2), 195–207. https://doi.org/10.1111/j.1475-4991.1993.tb00447.x

National Board of Revenue. (2018). *NBR Annual Report Statistics*. https://nbr.gov.bd/uploads/publications/Income_Tax_Saroni_16-17.pdf

Osmani, S. R. (2017). *Review of National Accounts Estimation in Bangladesh*. International Growth Centre (IGC). https://www.theigc.org/wp-content/uploads/2017/09/Osmani-2017-Final-report.pdf

Ravallion, M. (2003). Measuring Aggregate Welfare in Developing Countries: How Well Do National Accounts and Surveys Agree? *The Review Of Economics And Statistics*, 9. https://www.jstor.org/stable/3211703

Robilliard, A.-S., & Robinson, S. (2003). Reconciling Household Surveys and National Accounts Data Using a Cross Entropy Estimation Method. *Review of Income and Wealth*, *49*(3), 395–406. https://doi.org/10.1111/1475-4991.00094

Singleton, T. W. (2011, May 1). *Understanding & Applying Benford's Law | ISACA Journal*. ISACA. https://www.isaca.org/resources/isaca-journal/past-issues/2011/understanding-and-applying-benfords-law

Tahsin, M. (2022a). *Assessing the Reliability of Macro Data Using Night-time Lights Models: Bangladesh*.

World Population Review. (n.d.). *Bangladesh Population 2021 (Demographics, Maps, Graphs)*. Retrieved December 15, 2021, from https://worldpopulationreview.com/countries/bangladesh-population

2.10 Acronyms

BBS	Bangladesh Bureau of Statistics
CIA	Central Intelligence Agency
CPI	Consumer Price Index
FISIM	Financial Intermediation Services Indirectly Measured
GDE	Gross Domestic Expenditure
GDP	Gross Domestic Product
GNDI	Gross National Disposable Income
GNI	Gross National Income
GoB	Government of Bangladesh
HES	Household Expenditure Survey
HIES	Household Income and Expenditure Survey
K-S	Kolmogorov- Smirnov
PFCE	Private Final Consumption Expenditure
SNA	System of National Accounts

3 How Unequal is Bangladesh?

Abstract: This paper comments on the state of income inequality in Bangladesh by drawing on results of the previous chapters. It first uses the night-time lights model from Tahsin (2022a) to conclude that the GDP values of Bangladesh are much larger than government published figures. It then uses the results of micro-macro variable comparisons made in Tahsin (2022b). Incorporating the results of the night-time lights model widens the discrepancy between macro and micro variables. The results of triangulating survey data with administrative data from Tahsin (2022b) are also used. These results indicate that the highest income-tax paying population was under-sampled, while the top income quartile was over-sampled. Since less than 2 percent of the population of Bangladesh pay income tax, it can be inferred that the difference results from the under-sampling of the small percentage of the population who are the highest taxpayers. Thus, the income equality situation in Bangladesh is much worse than suggested by the government published Gini-coefficient.

3.1 Introduction

The results of Tahsin (2022b) indicate that the household survey data available for Bangladesh are unreliable. Income inequality in Bangladesh is estimated using these surveys. Thus, it can be inferred that the published Gini-coefficients for Bangladesh are misleading.

In this paper, we combine the results of Tahsin (2022a) and Tahsin (2022b) to reach conclusions regarding the actual state of income inequality in Bangladesh. We propose a comprehensive method of understanding income inequality. Our method incorporates key insights from macro as well micro data allowing us to capitalize on the strengths of each, while circumventing their weaknesses.

3.2 Literature Review

Micro data, such as household surveys are often the only available data source for computing important economic indices. These include income inequality and poverty headcount. Findings of pivotal papers, that have combined micro and macro data sources for poverty rate estimation and income inequality estimation are summarized below.

Karshenas (2003) estimated the global poverty rate by estimating the number of people living under the WB's dollar-a-day poverty line. The author assumed, that since the national accounts consumption expenditure/income in developing countries are estimated independently from the household expenditure surveys, they should also be independent from measurement errors involved in the latter. Using the national accounts per capita private consumption

as the calibrating variable, and assuming a simple log-linear functional form, the author estimated a relationship between average private consumption derived from household-level surveys and per capita consumption derived from national accounts.

Sundaram & Tendulkar (2003) too developed an alternative methodology to estimate the poverty rate for India. The authors replaced average consumption from India's National Sample Survey (NSS) with private consumption per capita from the national accounts, while retaining the survey-based distributions. This accounted for a much greater rate of poverty reduction in the 1990's. To justify using national accounting data to estimate poverty rates, they assumed that most data errors were distribution neutral- i.e., the underestimation of consumption growth was largely for the non-poor. These numbers were generated against the backdrop of compelling data pointing to increasing consumption inequality between the rich and the poor, and urban and rural inhabitants. It was assumed that the income distribution derived from the survey was correct. Based on this postulation, they concluded, that on average people got richer, but the non-poor got richer faster than the poor. This, consequently increased inequality. It should be noted that the authors did not present enough evidence to justify their assumptions and confidently propose this methodology.

Another poverty-estimation method is discussed in Deaton (2005). In this paper, the ratio of the mean consumption generated from national accounts and the mean consumption generated from surveys was multiplied to the total consumption of each household. The number of people living in households below the poverty line is then calculated. When the ratio was less than one, no multiplication is done. This scaled-up household consumption values. It also, drove down the number of people estimated to be living below the poverty line, thus painting a rosier picture of the economy. Proponents of the methodology consider national accounts to be more accurate than national surveys, hence justifying the adjustment process. Skeptics of the method, question the basic assumption.

Similarly, Sala-i-Martin (2002) used aggregate GDP data and within-country income shares to assign a level of income to every individual in the world. Using a Gaussian Kernel Density Function distribution of income and world poverty rates were then calculated. Global income inequality was estimated using seven popular indexes: the Gini coefficient, the variance of log-income, two of Atkinson's indexes, the Mean Logarithmic Deviation, the Theil index, and the coefficient of variation. All indexes showed a reduction in global income inequality. Since the methodology used, was national survey reliant, the authors had assumed that growth measured in national accounts positively impact the economic status of the poor.

Other papers, such as Fixler et al. (2017), explored the impact of economic growth on income inequality. To do this, the authors, used micro and macro data of the United States for the period of 1979 to 2012. The primary variables used for this exercise were: National Income and Product Accounts (NIPA) consistent personal income distributions drawn from the census income; and economic growth rate calculated using the national accounts. The relationship between the two variables was observed to determine the effect of economic growth on income inequality. It was seen that these measures yield different levels and trends in the median and inequality than obtained using the usual money income measure. While these relationships did not hold for longer time periods, they were consistent with the absence of a consensus about the relationship between GDP per capita movements and income inequality.

Banerjee & Piketty (2005) adopted yet another technique to comment on the status of economic inequality in India. The paper used individual tax returns data of India for the 1922 to 2000 period. According to the tax data, the shares of the top 0.01 percent, 0.1 percent, and 1 percent, shrank substantially from the 1950's to the 1980's. However, this share rose again, till it plateaued at a point just below the 1930's level. Although the impact of inequality, was not large enough to fully explain the gap observed during the 1990's between average consumption growth, it was sufficient to explain 20-40 percent of it.

There are several studies on the causes of and trends in income inequality in Bangladesh. However, only one of these studies attempted to test the Gini-coefficient published by the government. Khan (2005) calculated Gini-coefficients for the distribution of several indicators that are known to have an impact on economic inequality. The magnitude of these coefficients signaled that the income inequality estimates derived from the household surveys were dependable. Nevertheless, it is important to note that this conclusion is only applicable assuming that the data errors were distribution neutral. The key weakness of this study was that it did not draw insights from the national accounts.

3.3 Empirical Strategy

The chief sources of information for the analyses in this chapter are the results of Tahsin (2022a) and Tahsin (2022b). We first use the results of the night-time lights model for GDP estimation of Bangladesh from Tahsin (2022a). We incorporate these results in the micro-macro variable comparisons conducted in Tahsin (2022b). We then account for the sampling errors in the household surveys to conclude whether the actual state of income inequality is better or worse than indicated by the government published Gini-coefficients.

3.4 Results

3.4.1 Night-time Lights Model Estimated GDP

Drawing from the results of Tahsin (2022a), we know that the estimated GDP per capita is higher than the GDP per capita values published by the government. For the analyzed period i.e., 1992 to 2020, the estimated GDP per capita ranges between 2.5 times and 6.2 times of the published values. These values are provided in Appendix 3.1.

In Tahsin (2022b), the corresponding macro equivalent selected for per capita household expenditure was total expenditure. Since, the night-time lights model was only used to re-estimate GDP per capita figures, actual calculation of the new micro-macro ratios could not be performed. However, it should be noted that, total expenditure values closely mirror GDP values. Total Expenditure and PFCE (Private Final Consumption Expenditure) are also expected to show close co-movement with GDP. Thus, it can be assumed that the micro-macro ratio is likely to be much smaller than estimated using the government published figures.

The micro values were much smaller than the corresponding government published macro values. Thus, using the estimated figures would widen the discrepancy between the two even more.

3.4.2 Administrative Data

The results discussed in Tahsin (2022b) indicate that for all rounds of the household surveys, the income tax-paying population or the highest income taxpaying population were under-sampled. It is estimated that approximately 1.2 percent of the population of Bangladesh filed income tax returns[33] in 2017 (National Board of Revenue, 2018). This was the highest proportion achieved till data. Of these taxpayers, about 13 percent bear 73 percent of the tax burden (Sarker, 2003). Thus, it may be assumed that the tax-paying population who were under-sampled in the surveys, represent a very small group. This group is likely to be wealthiest even amongst the top income-quartile.

On the other hand, the results also indicate over-sampling of the top income quartile. This conclusion is reached using household electrification, motorbike ownership and car ownership as indicators. From this, it can be extrapolated that the average per capita household income, consumption, and expenditure for the lowest three quartiles are likely to be lower than estimated. However, this might not be true for 2017, since there was over-sampling of the top quartile in that year's household survey.

3.5 Limitations

This paper is unable to quantify income inequality. Previous studies, such as Banerjee & Piketty (2005) were able to use income tax returns data to draw inferences regarding income inequality from the income distribution of the top 0.01, 0.1 and 1 percent of earners. However, tax data of this nature is unavailable for Bangladesh.

Other studies replaced the micro variables with the corresponding macro variables to re-estimate the Gini-coefficient. Since the key issue with the micro data for Bangladesh, was sampling error. It was not possible to use this method to quantify income inequality.

3.6 Conclusion

Incorporating the results of the night-time lights model widens the discrepancy between macro and micro variables. The comparison against the administrative data makes it apparent that this discrepancy is largely due to the under-sampling of the highest income-tax paying population. It is also clear that for most years, the top income quartile was over-sampled. Thus, the average income, expenditure, and consumption of the lowest three income quartiles is even lower than previously estimated.

The above analyses indicate that there is a large difference between the macro variables and their corresponding micro variables. This is despite the fact the top income quartile is over-sampled. Thus, it can be inferred that the difference results from the under-sampling of the small percentage of the population who are the highest taxpayers. Thus, the income

[33] It is important to note that the actual number of income taxpayers is likely to be less than the number of income-tax returns filed. This is because, a large proportion of these returns register zero income tax (National Board of Revenue, 2018)

equality situation in Bangladesh is much worse than suggested by the government figures.

These results suggest that many of the components that make up GDP, contribute very little to the economic well-being of the lower-income quartiles. In fact, it appears that most of it funnels back to a small group who are at the apex of the highest income earners. These findings contradict the findings of Asian Development Bank & Government of Japan (2004). The report had concluded that there is a strong and positive growth-poverty nexus in Bangladesh. However, the report relied on statistics collected by the BBS and did not endeavor to validate the data.

3.7 Appendix

Appendix 3.1: True and Night-time Lights Model Estimated GDP Per Capita

Year	True GDP per capita	Estimated GDP per capita	Ratio
1992	1587.6	9509.1	6.0
1993	1619.7	10097.1	6.2
1994	1652.4	9798.7	5.9
1995	1702.8	9798.7	5.8
1996	1737.1	9897.1	5.7
1997	1772.2	9897.1	5.6
1998	1826.2	9996.6	5.5
1999	1881.8	10097.1	5.4
2000	1939.1	10509.1	5.4
2001	1998.2	10097.1	5.1
2002	2038.6	10404.6	5.1
2003	2100.6	10097.1	4.8
2004	2164.6	10198.5	4.7
2005	2275.6	10404.6	4.6
2006	2392.3	9604.6	4.0
2007	2540.2	9996.6	3.9
2008	2670.4	10198.5	3.8
2009	2751.8	10097.1	3.7
2010	2892.9	10198.5	3.5
2011	3041.2	10301.0	3.4
2012	3197.1	10301.0	3.2
2013	3361.0	11159.0	3.3
2014	3498.2	11731.1	3.4
2015	3714.5	11731.1	3.2
2016	3904.9	11614.4	3.0
2017	4146.4	12088.4	2.9
2018	4447.1	12209.9	2.7
2019	4769.5	12088.4	2.5
2020	4865.9	12088.4	2.5

3.8 Bibliography

Asian Development Bank, & Government of Japan. (2004). *Economic Growth and Poverty Reduction in Bangladesh*.

Banerjee, A., & Piketty, T. (2005). Top Indian Incomes, 1922–2000. *The World Bank Economic Review*, *19*(1), 1–20. https://doi.org/10.1093/wber/lhi001

Deaton, A. (2005). Data and Dogma: The Great Indian Poverty Debate. *The World Bank Research Observer*, *20*(2), 177–199. https://doi.org/10.1093/wbro/lki009

Fixler, D., Johnson, D., Craig, A., & Furlong, K. (2017). A Consistent Data Series to Evaluate Growth and Inequality in the National Accounts. *Review of Income and Wealth*, *63*, S437–S459. https://doi.org/10.1111/roiw.12324

Karshenas, M. (2003). Global Poverty: National Accounts Based versus Survey Based Estimates. *Development and Change*, *34*(4), 683–712. https://doi.org/10.1111/1467-7660.00324

Khan, A. R. (2005). Measuring Inequality and Poverty in Bangladesh: An Assessment of the Survey Data. *The Bangladesh Development Studies*, *31*(3/4), 1–34. https://www.jstor.org/stable/40795713

National Board of Revenue. (2018). *NBR Annual Report Statistics*. https://nbr.gov.bd/publications/all-publication/eng

Sala-i-Martin, X. (2002). *The Disturbing "Rise" of Global Income Inequality* (No. w8904; p. w8904). National Bureau of Economic Research. https://doi.org/10.3386/w8904

Sarker, T. (2003). Improving Tax Compliance in Developing Countries via Self-Assessment Systems... 9(6), 35 https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.581.8719&rep=rep1&type=pdf

Sundaram, K., & Tendulkar, S. D. (2003). NAS-NSS Estimates of Private Consumption for Poverty Estimation: A Further Comparative Examination. *Economic and Political Weekly*, *38*(4), 376–384. https://policydialogue.org/files/publications/NAS_Estimates.pdf

Tahsin, M. (2022a). *Assessing the Reliability of Macro Data Using Night-time Lights Models: Bangladesh*.

Tahsin, M. (2022b). *Framework for Testing the Reliability of Micro-level Data: Using Data from Bangladesh*.

4 Is the Poverty Probability Index an Appropriate Measure for Profiling Poverty? Evidence from Bangladesh

Abstract: The Poverty Probability Index (PPI) is a household-level poverty classification tool that uses 10 socioeconomic indicators to assign households a score between 0 and 100. The likelihood of falling below a specified poverty line, at each score level, is estimated. Due to its ease of application, the PPI has been adapted by major microfinance institutions and other large-scale development projects. This study replicates the construction methodology of the PPI 2013 Bangladesh and finds that the indicator weights and the poverty likelihoods cannot be reproduced. The scorecard is applied to three micro datasets, to assess how the tool performs over time and on population sub-groups. It is observed that the poverty likelihoods change over time and for population sub-groups. Using these poverty likelihoods, the targeting performance of the tool, in terms of, precision, recall and F-measure, are also calculated. Lastly, the tool's performance in tracking households' falling-into-poverty or graduating out-of-poverty over time is evaluated. It is concluded that the tool performs reasonably well in tracking graduation out-of-poverty, but it does not perform well in tracking falling-into-poverty.

4.1 Introduction

4.1.1 Context

The discussions in the preceding chapters bring one key issue to the forefront: Reliable data is difficult to ensure. This is true even when considerable resources are invested in the data collection and quality check exercises. The Bangladesh Bureau of Statistics (BBS) has separate wings dedicated to calculating the national accounts and conducting the household income and expenditure surveys. Yet, from the results in Tahsin (2022a) and Tahsin (2022b), it is apparent that the data generated for the national accounts and the household surveys are not dependable. From this it can be deduced that lower-budget data collection activities probably produce untrustworthy information too.

Income is an important correlate for numerous phenomena (Micklewright & Sylke, 2010). Governments, non-government organizations, as well private entities often rely on household income data for policy making and resource allocation. This data is usually collected through rapid surveys and very little is done to ensure its quality and consistency. This leads to misallocation of resources, which are often limited.

Economic research has established that income data collected through single-questions is untrustworthy (Micklewright & Sylke, 2010). It is also confirmed that consumption is a better indicator of the economic status of the poor, than income (Meyer & Sullivan, 2003). Measurement error and under-reporting is higher for income data than for consumption data for poorer households. This is truer for households engaged in informal work. This is because, true income is the sum of earned income and imputed income. These calculations are complex; especially for households who are not on fixed salaries, as is the case of most poor households in developing countries. Thus, consumption measures are recommended for household poverty classification.

However, even consumption measures are difficult to compute. A standard household survey collects daily, weekly, monthly, and annual information of households' food items and non-food items consumption expenditures. This is done using the 'diary method' over a year. To smooth over seasonal peaks and drops in consumption, collection of the data over a long period of time is vital.

This, time, and resource intensive process of household consumption calculation is neither feasible nor practical in time-sensitive situations. For instance, if an aid program is responding to a crisis, or if the government is providing subsidised emergency or medical services to households under the poverty line. Even in the absence of a time constraint, most programs do not have the budget for a detailed, time-consuming survey. Thus, these programs often rely on single-questions, and collect income data using the 'recall method'. The result is usually unreliable data, misclassification of poor and non-poor households, and misallocation of resources.

4.1.2 The Poverty Probability Index: Users

It is to minimize the misallocation of resources due to household poverty misclassification, that the Poverty Probability Index (PPI) was developed. The PPI[34] is a scorecard used to predict the probability of a household falling below a specified poverty line. The index is currently used by over 600 development organizations from 61 countries, housing 90 percent of the world's poverty-stricken population (Innovations for Poverty Action, 2013a). Wide ranging projects in the sectors of micro-finance, employment generation, pro-poor market development and agriculture, use the PPI for beneficiary selection and to assess the effectiveness of their interventions. These projects, whose end objective, is usually, poverty alleviation, use the PPI to identify households who have high probability of being below specified poverty lines. They also use the index to track changes in poverty status of beneficiary households before and after treatment[35]. In comparison to the direct approach of poverty measurement via expenditure surveys, indirect poverty measurement[36] using the PPI scorecard is simple, easy to use and cost-effective. For instance, in 2010, the national household expenditure survey for Bangladesh ran 40 pages, whereas the PPI asks just 10 questions (Schreiner, 2015). Thus, the scorecard has gained rapid popularity, especially in the field of micro-finance. In 2009-10, over 14 percent of microfinance institutions who reported social-performance information to the microfinance's MIX Market used the PPI (Schreiner, 2015).

Countries where microfinance is an important driver of economic progress make use of the PPI frequently and in varying scenarios. Bangladesh is one such country. In 2014, 33 million of the 160 million people of Bangladesh were being served through microfinance institutions (Mia, 2017). These institutions used the index to ensure that deserving households, i.e., those falling below the poverty line, were selected to receive microcredit. Post-disbursement of the small loans, the index is used again to assess the impact of micro-credit on poverty status[37] of recipient households. Given the large outreach of such programs, it is imperative that the index used for beneficiary selection at least predicts households' probability of falling below the poverty line well, if not perfectly. Inaccurate beneficiary selection leads to misallocation of resources. On the other hand, tracking progress of inappropriately selected beneficiaries is futile, since, they might already have been above the poverty line. This leads to formulation of inappropriate policies and ill designed future interventions.

Presently, some of the most critical development schemes in Bangladesh are using the PPI. Users of the Bangladesh PPI 2013[38], have complained that the scores do not correctly predict the probability of a household falling below any of the specified poverty lines. Two of the largest users of the index in Bangladesh- pro-poor market development project Katalyst and BRAC microfinance, also shared these concerns. Because of its ease of use, PPI remains one of the most widely used proxy-means tools for evaluation of poverty status, despite concerns from a multitude of users.

4.1.3 The Poverty Probability Index: Construction

The PPI 2013 scorecard was created using data from the Household Income and Expenditure Survey (HIES) 2010 conducted by the BBS. The scorecard used 10 socioeconomic indicators and true income values, all of which were derived from the same

[34] "Poverty scorecards ("scorecards" for short) are also called "simple poverty scorecards", "Progress Out of Poverty Indexes®" or "PPIs®" (trademarks registered by Grameen Foundation). All the names refer to the same approach. Scorecards are available at no cost from progressoutofpoverty.org or microfinance.com. Copyright in a given scorecard is held by its sponsor and by Microfinance Risk Management L.L.C." (Schreiner, 2015)

[35] Treatment refers to the act of households being included in the projects.

[36] Indirect poverty measurement refers to any approach that attempts to determine level or likelihood of poverty using indicators other than income and expenditure. Indirect approaches of poverty measurement define people as poor if their physical, human, financial or social assets are below specified thresholds. However, expenditure-based definitions are most commonly used by governments and donors (Schreiner, 2015)

[37] Poverty status refers to the probability of a household falling below specified poverty lines.

[38] The PPI was most recently updated in November 2021. However, the analysis in this paper refers to the update prior to that (2013).

dataset. The highest scores attached to each of the indicators, add up to a total score of 100. The score to be assigned to each indicator is estimated using a logit model. The dependent variable in the regression model is a binary representing whether a household is above the USD 1.25, USD 1.75, USD 2.00, and USD 2.50 per person per day poverty lines. The explanatory variables are categorical variables representing the 10 socioeconomic indicators used to construct the scorecard. The coefficients are then converted to scores by replicating the methodology of Innovations for Poverty Action (2013b).

4.1.4 Objective

This study has five key goals. Firstly, using the same dataset as the original authors, the construction methodology is replicated to test the reproducibility of the PPI 2013 scorecard and the PPI 2013 poverty likelihood table. Secondly, since the tool was built using a nationally representative sample, the scorecard is applied to a rural sample from the same period to test the generalizability of the tool in population sub-groups. Thirdly, the tool is applied to a more recent nationally representative sample to evaluate whether the tool remains applicable over a 5-year period. Fourth, using a panel data set, the ability of the PPI 2013 to track households' movement in and out of poverty is evaluated. Lastly, the targeting accuracy and precision of the Bangladesh 2013 PPI for each of the 4 datasets is compared.

The next section elaborates on the existing literature on the PPI (Section 4.2). In Section 4.3, descriptions of the databases, data sets and the variables used for models are provided. Section 4.4 begins with a comparison of the original methodology used in Innovations for Poverty Action (2013b) to the assessment methodology used in this paper. Section 4.5 presents the results. Section 4.6 discusses the limitations of the methodology and how these might influence the results. This section also includes details of existing issues in income/expenditure calculation methodology. Section 4.7 summarizes the conclusions drawn from the main findings of the study and Section 4.8 recommends further avenues of research.

4.2 Literature Review

Studies testing the PPI have been conducted for several countries, including, El Salvador, Mexico, Zambia, Rwanda as well as Bangladesh. The key findings of these papers are summarized in this section. Additionally, the knowledge gaps in this area of research are identified.

The Committee on Sustainable Assessment (2015) compared 3 poverty measurement tools, including the PPI. They applied the 3 tools to a sample of the beneficiaries of a Catholic Relief Services Project in El Salvador. The study found PPI to be reasonably accurate when applied to nationally representative samples. However, the results were found to vary greatly when applied to a nationally non-representative sample. The committee also tested the index on farmers of a cocoa project in Nicaragua. By using a power analysis on the Nicaragua 2005 PPI it found that a minimum sample size of 339 individuals is required to estimate poverty rates with a .90 confidence level and a confidence interval of +/- 3%. It also required the assumption that these 339 farmers are representative of the national sample. While smaller samples were likely to reduce the statistical accuracy of the poverty rate for these groups, they could still be tracked over time utilizing the PPI. On the other hand, for the case of Mexico a correlation could be found for net income and the expected poverty rate calculated by PPI of each group, when using net income per capita and dividing the groups into quantiles. Similar results were derived for cocoa producers in Colombia.

Another study (Lawson-Mcdowall et al., 2017) conducted by the Catholic Relief Services applied the PPI to participants of a financial inclusion project in Kasama, Zambia. It compared the PPI's results with that of a wealth ranking index. The wealth ranking index divided households into 4 income categories- Well-off, Managing, Poor and Very Poor. It found that data followed the same trend as the four wealth groups: the Well-Off households had the lowest poverty likelihood with 76.2%, the Managing had a likelihood of 83.5% and the Poor and Very Poor households had the highest (90.3% and 90.6% respectively). However, it also found that the PPI cannot easily distinguish between groups of households that are almost all living below the poverty line. This suggests that the PPI should not be used for targeting when the intention is to reach the poorest-of-the-poor. The index does not do a good job of differentiating within the poor. It also suggests that the index should not be used to measure progress from one level of poverty to another, though it might still be used to track progress out of poverty.

Desiere et al. (2015) did a validity assessment for the PPI of Rwanda. The authors evaluated the PPI using the European Commission specified SMART criteria. They found the index to be S-specific, M-measurable, A-Available cost effectively and T-timely available. However, its R-relevance in distinguishing poor from nonpoor households and in capturing changes in poverty status over time remains questionable. It found that the index became unreliable within the 5-year period that is lapsed between each update. This resulted in gross overestimation of the number of poor households being targeted by a development program. The inaccuracy is especially stark in countries experiencing rapid economic growth and poverty reduction. Since most development projects are implemented in countries fitting that description, the usefulness of the PPI in evaluating the impact of development projects was found to be low.

While there is existing literature testing the accuracy with which the PPI predicts households' probability of falling below a specified poverty line, only 3 such studies have been conducted for the case of Bangladesh and none of these tested the scorecard since it was updated in 2013.

Schreiner (2015) reported that the PPI 2013 scorecard had out-of-sample/in-time bias for the USAID "very poor" poverty line of -0.3 percent and precision of 0.86 for estimates of poverty rates[39]. However, these estimates were quoted directly from the original design document of the PPI 2013 scorecard. No new calibration samples were used.

Katalyst (2012) tested the PPI scorecard for Bangladesh constructed in 2005. The results suggested that the scorecard would only be valid for the same year as the year of the household survey it was constructed using. It concluded that the relationship between indicators changes too fast for the index to remain relevant. As an example, the authors cite how radio-cassette players had become almost obsolete in Katalyst's beneficiary populations in the years following 2005, however, it had to remain an indicator in the scorecard till the next household survey was conducted in 2010.

Jalil & Azam (2014) was the second paper to dissect the PPI 2005 scorecard for Bangladesh. They concluded that it was a useful tool for comparing the level of pro-poorness of different projects. The writers also suggested that it could be used for more than simple poverty profiling. However, it should be noted that this study was based completely on case studies and experiences of one organization and that no statistical work was done to draw these conclusions.

From the literature that is available, it is clear, that each scorecard needs to be tested before application, since each scorecard's use could potentially be more restricted than is claimed by the developers of the index. The requirement of a minimum sized nationally representative sample implies that the population model is incorrectly specified. A correctly specified model would hold for all sub-groups. Since the index is generally used by development projects, it is usually applied to small and localized subgroups of people. Secondly, many projects, including BRAC microfinance use the PPI to not just target the poor, but the ultra-poor. As there is evidence from Zambia suggesting PPI might not be useful for such targeting; testing the PPI scorecard for Bangladesh is required. The Katalyst report suggests a very short time-period for which the PPI would remain applicable. This raises additional concerns in the case of Bangladesh since there are often lags in between household surveys and updating of the index. For instance, the dataset from the 2015-16 household survey was only made available in 2018. The updated PPI was not made available till the end of 2021. This means, that the Bangladesh 2013 PPI was used for 8 years and the data it used was 11 years old.

[39] Poverty rate refers to the average of participants' poverty likelihoods.

This is the first paper to calibrate the Bangladesh 2013 PPI scorecard against other data bases and 5 years into the future. In addition, it is also the first paper to check whether the Bangladesh 2013 PPI scorecard is a good tool to measure graduation out of poverty using panel data. The key contribution of this paper is that it will allow users to correctly interpret PPI scores when the tool is applied to sub-groups or to a more recent time-period. Given the large scope of use of the index in the country, the findings are important to both researchers and development practitioners.

4.3 Data

4.3.1 Databases

Four datasets from two databases were used for the purpose of this study. Descriptions of the data bases and the datasets are provided below.

HIES 2010-11, BBS: BBS, the statistics division of the government of Bangladesh conducts the HIES (usually at 5-year intervals). It is the core survey to provide income, expenditure, consumption, and poverty data. The first round of HIES was conducted in the year 1973-74 in the newly independent Bangladesh. Since then, including the latest survey in 2021, BBS has successfully completed 17 rounds of the household surveys. The HIES data series generated by BBS is the main data source for estimation of poverty and its correlates in Bangladesh. In 2010, the survey collected information for 12,240 households, comprising of 55,582 individuals. The sample surveyed was nationally representative. From the 2010 HIES, data pertaining to consumption of food and non-food items were collected by trained enumerators assigned to the respective areas. Data entry and digitization was also done at field level by the enumerators themselves to facilitate correction of errors and inconsistencies (Bangladesh Bureau of Statistics, 2011). The PPI 2010 scorecard was constructed using this dataset. Hence, it had all the information required for replication.

HIES 2015-16, BBS: The sixteenth round of the household income and expenditure survey carried out by the Government of Bangladesh was completed in 2016. Among other data, the survey collected income, expenditure, consumption, poverty, and calorie intake information for 46,080 households. The sample surveyed was nationally representative. Complete information required for calculation of PPI score was available for 45,767 of the households covered in the survey (Bangladesh Bureau of Statistics, 2017a).

Bangladesh Integrated Household Survey (BIHS) 2011-12, International Food Policy and Research Institute (IFPRI): The IFPRI BIHS was conducted by a food policy institute, it collects data on plot-level agricultural production and practices, dietary intake of individual household members, anthropometric measurements of all household members, and income, expenditure, consumption, and other important poverty correlates. For this study, we draw on the fourth component. The first round of BIHS conducted in 2011-12 covered 6,503 households. The sample is nationally representative for rural Bangladesh (International Food Policy Research Institute, 2016). Note that the dataset includes only 9 of the 10 indicators used to construct the PPI scorecard.

BIHS 2015-16, IFPRI: The second round of the BIHS survey interviewed 6,715 households including 5,659 households from the first round of BIHS in 2011-12 (International Food Policy Research Institute, 2016). Like the first round, details of income, expenditure and consumption were collected for each household in the sample, and this is the information that was used in this study.

4.3.2 Variables

The PPI scorecard uses 10 indicators to give households a score between 0 and 100 as shown in Appendix 5.1. This score is then used to predict poverty probability of the respective household. Information relating to the 10 socioeconomic indicators used to construct the PPI scorecard were extracted from each of the datasets described in Section 4.3.1.

The PPI score is calculated using:

Eq 4.1

$$\text{PPI score}^{40} = \sum w_i X_i$$

Where,
w_i: weight of each explanatory variable in the model
$X_i = \{x_1, \dots, x_n\}$: *set of binary variables*

4.4 Empirical Strategy

The index is tailored to each country. It uses statistical methods to draw correlations between socioeconomic indicators and household consumption expenditure. To construct the PPI scorecard and the look-up table for each country, the most recent national household income and expenditure survey is used. For the construction of the Bangladesh 2013 PPI scorecard and look-up table, data from the BBS HIES 2010-11 survey was used.[41] The dataset was randomly split into two sub-samples. The first sub-sample was used for construction and calibration of the selected indicators, and for associating scores with poverty likelihoods. The second sub-sample was used for validation of the index (Innovations for Poverty Action, 2013b).

4.4.1 Scorecard

As noted in Section 4.1.3, the PPI is a scorecard that combines 10 socioeconomic indicators to classify households as poor and non-poor. It does this by recording participants' responses to 10 simple questions[42]. Each response to the questions is attached to a pre-assigned weight. A household can score between 0 and 100.

Using the data set used to construct the PPI 2013 Bangladesh scorecard, we attempt to reproduce the scorecard generated in Innovations for Poverty Action (2013b). This is done by running a logistic regression model, in which, binaries of the PPI indicators are used as the explanatory variables, denoted as X_i in Eq 4.2 . The dependent variable y is a binary of the households' poverty status. This is specified as households' position above or below the poverty line. The logistic regression model is given in Eq 4.2.

Eq 4.2

$$\text{logit}(P_y) = \beta_0 + \beta_i X_i + u$$

We deviate from the original construction methodology in two respects (refer to Appendix 4.2 for scorecard construction details). Firstly, the original scorecard was constructed using one random sample. However, to reduce out-of-sample bias, we run 1,000 Monte Carlo simulations to select the calibration sample, i.e., 1,000 samples are drawn, with replacement from the dataset. Secondly, we opt to only include observations with complete information in our analysis. This reduces the data set from 12,240 households to 11,072 households. The original scorecard was built using the entire dataset. But the methodology used to impute the missing data is unknown (Innovations for Poverty Action, 2013b).

For all analyses, the household expenditure and poverty lines were calculated using the same methodology as Innovations for Poverty Action (2013b). For comparability, across countries and years, the poverty line is adjusted using the CPI and the PPP exchange rate. Thus, the USD 1.25 per capita per day poverty line, against which the scorecard is built, is estimated to be BDT 59.32 for HIES 2010-11.[43] The results of the logit model and the derivation of the corresponding scores are discussed in Section 4.5. To check that consistent parameters have been estimated, the model is then applied to the validation sample. The process is repeated for the other three datasets[44]. This allows us to assess the following (i)

---

[40] Note that each indicator is transformed into a binary variable for inclusion in the logistic model.

[41] At the time of construction of the PPI 2013, BBS HIES 2010 was the most recent household survey for Bangladesh.

[42] The methodology used to construct the scorecard is detailed in Appendix 4.2.

[43] $$\text{Poverty line for } 2011-12 = (2005 \text{ PPP Exchange Rate}) * \text{USD } 1.25 * \left(\frac{\text{CPI}_{2011-12}}{\text{CPI}_{2005}}\right) = \frac{\text{BDT } 25.29}{\text{USD } 1.00} * \text{USD } 1.25 * \left(\frac{295.86}{158.3}\right) = \text{BDT } 59.32$$

[44] It should be noted that the IFPRI BIHS 2011-12 and IFPRI BIHS 2015-16 data sets do not ask participants whether a household member had worked as a day labourer in the last 12 months. Instead, it asks whether any household member had worked as a day labourer in the previous week. Hence the variable 'daylabourer' is dropped from the calculations. Thus, before carrying out the analyses and comparisons, the scorecard and look-up table need to be

whether the scores can be replicated, (ii) whether indicator weights and relationships remain consistent over time, and (iii) whether indicator weights and relationships remain consistent for population sub-groups.

4.4.2 Look-up Table

The likelihood that a household falls below a specified poverty line, monotonically decreases as PPI scores rise, i.e., the higher the household's score, the less likely it is to fall below any of the poverty lines. The estimated poverty likelihood at each score level is summarized in a 'look-up' table.[45]

For our analyses, the original scorecard is applied to the BBS HIES 2010-11 dataset. The PPI scores and the total household expenditures are computed for each household. The likelihood of households falling below the USD 1.25, USD 1.75, USD 2.00, and USD 2.50 per capita per day poverty lines, at each score category is then estimated.

Like in the case of the scorecard, the household expenditures and poverty lines are estimated using the same methodology as Innovations for Poverty Action (2013b). The poverty line is adjusted using the CPI and the PPI exchange rate.[46]

The process is repeated for the other three datasets. Note that for assessing the performance of the PPI 2013 Bangladesh scorecard on the IFPRI BIHS datasets, the scores are scaled to 100 after removal of daylabourer indicator. Since the assigned score of the removed indicator is 8, all other scores are multiplied by 23/25 and rounded to the nearest digit.

This allows us to assess the following (i) reproducibility of the poverty likelihoods of the PPI 2013 Bangladesh, (ii) performance in terms of household poverty classification of the PPI 2013 Bangladesh scorecard over a 5-year period and (iii) performance in terms of household poverty classification of the PPI 2013 Bangladesh scorecard for rural populations.

4.4.3 Targeting

The primary use of the PPI is beneficiary selection. Thus, the accuracy with which the index can classify household poverty is its most important characteristic. To test targeting performance in terms of accuracy, precision and recall the scorecard is applied to all four datasets. Three cut-off points, 59, 69 and 79 are levied. The cut-off points are the scores that are to classify households by "poor" and "non-poor". This is done for the USD 1.25, USD 1.75, USD 2.00, and USD 2.50 poverty lines.

4.4.4 Graduation Out-of-Poverty

Panel data is available for 6,040 households in the IFPRI BIHS 2011-12 and IFPRI BIHS 2015-16 datasets. The samples are representative for rural Bangladesh. By calculating the PPI scores and corresponding per capita household expenditures for these households, in both years, we determine the usefulness of the PPI as a tool for tracking graduation out-of-poverty.

4.5 Results

4.5.1 Scorecards

The regression results that were transformed to generate these scorecards are attached in Appendix 4.4. The results reveal that several variables that were used in the construction of the PPI 2013 Bangladesh scorecard, did not have statistically significant relationships with the dependent variable. They also

appropriately transformed. In the first step of this transformation, the scorecard for BBS HIES 2010-11 is reconstructed using the 9 out of the 10 socioeconomic indicators of the PPI. The corresponding weights to be attached to each of these 9 indicators is re-estimated and a new scorecard is generated for BBS HIES 2010-11. Removal of the 'daylabourer' variable reduces the total score to 92. All the indicator scores are multiplied by 25/23 to scale the total score back to 100. Since we were unable to reproduce the results of the original scorecard, instead of re-running the model, we scale the scores to 100. Refer to Appendix 4.2 for details of scorecard construction.

[45] The methodology used to construct the look-up tables is detailed in Appendix 4.3

[46] Poverty line estimates: USD 1.25 = BDT 59.32, USD 1.75 = 83.05, USD 2.00 = 94.91 and USD 2.50 = 118.64. The estimates are in per capita per day terms.

show that the relationships between the dependent and the independent variable are reverse of what is presented in the PPI 2013 Bangladesh scorecard. For instance, vehicle ownership goes from having a positive relationship in BBS HIES 2010-11 to having a negative relationship in BBS HIES 2015-16, IFPRI BIHS 2011-12 and IFPRI BIHS 2015-16.

The re-estimated scorecards for all four datasets are shown in Table 4.1. When the logit model is used to reconstruct the PPI 2013 Bangladesh scorecard, it is not possible to reproduce the results.

Even with bootstrapped samples the estimated coefficients of each indicator differ for the training and validation sub-samples. This is indicative of the indicator relationships not being stable. Thus, the PPI 2013 Bangladesh scorecard appears to have been overfit to the sample.

When the scorecard is re-estimated for the BBS HIES 2015-16 dataset, the same phenomenon is observed. Additionally, even the direction of relationship of vehicle ownership and the explanatory variable is reversed. The positive relationship between vehicle ownership and the household being above the USD 1.25 per capita per day poverty line, changes to negative. This indicates that the relationship of the indicators and the explanatory variable changes dramatically over a 5-year period for a nationally representative sample.

When the scores are re-estimated for samples that are nationally representative at the rural level, it is observed that the indicator relationship change in the 4-year period between the collection of IFPRI BIHS 2011-12 and IFPRI BIHS 2015-16 datasets. In fact, even the direction of the relationship changes for two indicators. First, for IFPRI BIHS 2011-12, the relationship between number of rooms and households being above the USD 1.25 per capita per day poverty line is negative. This reverses to positive for IFPRI BIHS 2015-16. Second, while ownership of cultivable land is negatively related to being above the poverty line for IFPRI BIHS 2011-12, there is no relationship detected in IFPRI BIHS 2015-16. Thus, indicator relationships are not consistent over a 4-year period for rural level nationally representative samples.

Lastly, comparing estimated scores for BBS HIES 2015-16 and IFPRI BIHS 2015-16, show that that the magnitude of relationships between the dependent and independent variables change. It is observed that land ownership status is positively related to households' being above the USD 1.25 per capita per day poverty line for the nationally representative sample (BBS HIES 2015-16). However, the indicator has no bearing on the explanatory variable for the rural sample (IFPRI BIHS 2015-16) of the same year. This analysis indicates that the indicator relationships are not stable for population sub-groups.

4.5.2 Look-up Table

Table 4.2 lists the true poverty likelihoods calculated for four datasets (i) nationally representative dataset collected in 2010-11 (BBS HIES 2010-11), (ii) nationally representative dataset collected in 2015-16, (iii) rural-level nationally representative dataset collected in 2010-11 (IFPRI BIHS 2011-12), (iv) rural-level nationally representative dataset collected in 2015-16 (IFPRI BIHS 2015-16). These are compared to the poverty likelihoods predicted by the PPI 2013 Bangladesh look-up table. These comparisons allow us to predict how the PPI scores should be interpreted when targeting in a more recent sample or in population sub-groups.

The first set of poverty likelihoods are calculated using the BBS HIES 2010-11 dataset. This is the same dataset the PPI 2013 Bangladesh scorecard and look-up table was constructed using. It is observed that for all four poverty lines – USD 1.25, USD 1.75, USD 2.00, and USD 2.50 per capita per day, the true poverty likelihoods differed from the PPI 2013 Bangladesh look-up table. Thus, it is concluded that the look-up table could not be reproduced. For all score ranges the original look-up table generally over-estimates the poverty likelihoods of the households in the sample.

The process is repeated for the other three datasets. These results are summarized in Table 4.3. It is observed that when the PPI is applied to the BBS HIES 2015-16 dataset, the tool underestimates poverty likelihood till the PPI score of 49. This is in the case of the USD 1.25 poverty line. For PPI scores 50 and above, the tool overestimates the household poverty likelihood.

Similar under and over-estimations are observed at various score ranges and poverty lines, for all four datasets. Thus, it may be inferred that the application of the PPI should be tailored to the sample it is used for.

4.5.3 Targeting

Using the poverty likelihoods generated in Table 4.2 the targeting performance of the PPI 2013 Bangladesh scorecard is measured in terms of precision, recall and F-measure for all four datasets. These results are shown in Table 4.7.

The results indicate that when the cut-off score is set at 59 to target households below the USD 1.25 poverty line; precision, recall and the F-measure are highest for the BBS HIES 2015-16 data set.

When the cut-off score is raised to 69, precision, and the F-measure are highest for the IFPRI BIHS 2015-16 dataset. Recall is highest for the BBS HIES 2015-16 dataset. Similar results are observed when the cut-off score is raised to 79.

Similar trends are seen as we move up the other three poverty lines. The cut-off score and poverty line should be selected based on the specific requirements of the targeting exercise. For instance, when false inclusions are costlier than false inclusions; both the poverty line and the cut-off score should be lowered. On the other hand, when false exclusions are costlier than false inclusions, the poverty line and the cut-off score should be raised.

4.5.4 Graduation Out-of-Poverty

The IFPRI BIHS 2011-12 and IFPRI BIHS 2015-16 collected panel data for 6,040 households. By comparing these datasets, the PPI's performance as a tool to measure graduation out of poverty, is assessed. This is done by calculating the respective PPI scores of the households and generating a binary variable for whether a household has moved to the other side of a specified poverty line. From Table 4.4 we see that of the 6,040 households, 869 had fallen below the USD 1.25 2005 PPP poverty line from being above it, within this 5-year period. Of these only 34.5 percent had simultaneously experienced a decrease in their respective PPI scores. For the same poverty line 4,400 households experienced no change in poverty status. Yet, 88.7 percent of the households experienced a change in PPI scores. Similarly, 771 households rose above the poverty during the same period; and 75 percent of the households experienced an accompanying increase in their PPI scores.

Similar trends are seen for all four poverty lines. This indicates that while PPI performs relatively well as a tool to track graduation out-of-poverty; it does not perform well as a tool to track no change in poverty status or a deterioration in poverty status.

4.6 Extension

4.6.1 Vehicle Ownership

The generated scorecards show that vehicle ownership is positively related to households being above the USD 1.25 per capita per day poverty line. However, for the other three datasets the relationship is seen to be negative. This finding is not only contrary to the results of the PPI 2013 Bangladesh scorecard but also counter intuitive.

In the original model, cars, bicycles, and scooters were all clubbed together under the variable 'vehicle'. The logit regression is rerun by separating cars, bicycles, and motorcycles, for all three data sets. It should be noted, that the IFPRI data set does not collect data for car ownership. Thus, for the IFPRI data sets, the logit models unclub vehicle into motorbikes and bicycles. The results are provided in Table 4.5 below. From the regression results we observe that ownership of bicycles is negatively related to being above the poverty line, whereas owning a motorcycle or car is positively related. This is true for all four data sets. A possible explanation for this, maybe the popularity of rickshaws and motor driven three-wheelers in rural Bangladesh. These are relatively cheap modes of public transport that can be availed easily in all parts of Bangladesh. It is possible that households who are above the poverty line but still unable to afford motorcycles prefer to avail these services rather than buying a bicycle. Thereby bicycle becomes an inferior good. This could be a result of a combination of factors. Firstly, in many parts of the country women are socially discouraged from riding bicycles, thus if they can afford it, they prefer to use public transport. Secondly, many of these households may find it difficult to make the large upfront payment required to purchase a motorcycle even if they are above the poverty line. These households may prefer to use public transport at least till they have saved enough money to be able to afford a motorcycle. Even though rickshaws are more expensive than bicycles, relatively well-off households may opt to use them to avoid the physical strain of riding a bike. It is also possible that both bicycles and motorcycles are cheap enough that bicycles can be bought even when households are below the USD 1.25 poverty line and motorbikes are affordable enough to be bought as soon as a household has risen just above the USD 1.25 poverty line.

It is also seen that the negative correlation between bicycle ownership and poverty status got stronger over the 5-year period. At the same time, the positive relationship between ownership of motorcycles and cars weakened. Thus, it can be concluded that more households whose per capita daily income was less than USD 1.25 were now able to afford these vehicles.

4.6.2 Land Ownership

Ownership or use (through mortgage, sharecrop or rent) of 51 decimals or more land was positively correlated with the probability of a household being above the poverty line for the BBS 2010 data. This positive relationship is observed for BBS HIES 2015-16. However, the relationship is negative for IFPRI BIHS 2011-12. For IFPRI BIHS 2015-16 no relationship is detected, and the variable is not statistically significant.

These results are retested by decomposing the variable (use of over 51 decimals of land) into 3 separate variables- ownership, renting-in and renting-out over 33 decimals of land. The variable ’own_33’ represents household operating over 33 decimals of owned cultivable land. The variable ’rent_in_33’ includes renting in, leasing in, sharecropping in, or mortgaging in 33 decimals or more of cultivable land. And the variable ’rent_out_33’ represents renting out, mortgaging out, leasing out or sharecropping out over 33 decimals of cultivable land. The relationship between these variables against the probability of a household lying above the poverty line is observed

The threshold of 33 decimals was chosen for this analysis, because it is a commonly traded size of land in rural Bangladesh. It is large enough for small farmers to harvest by themselves and small enough that it remains within affordable range. The results show that ownership, renting-in and renting-out all have a positive relationship on being above the poverty line in 2010. These relationships are found to be statistically significant at the 5 percent significance level. In 2011-12, own operated land and renting-in show a negative correlation with poverty status. Renting-out cultivable land on the other hand, still displays positive correlation with falling above the poverty line. In a rural sample from 2015-16, the direction of correlations remains the same as it was in 2011-12. However, the negative correlation of having over 33 decimals of own operated cultivable land is a lot weaker than it was for the 2011-12 data set. This variable also stops being statistically significant in 2015-16. In a nationally representative sample of the same year, all three variables illustrate a positive relationship with poverty status.

Renting out shows positive correlation in all four data sets. It also retains statistical significance throughout. A possible explanation for this might be that renting land became more expensive, thus making, harvests less profitable for farmers farming leased or rented land. For 2015-16, the land ownership is not significant in predicting the probability of a household being above or below the poverty line. In this year, rural Bangladesh was inundated by floods, ruining crops, this might make have made land ownership an unimportant predictor of income status. More in-depth analysis and further testing is required to disprove this hypothesis. Another explanation might be a fall in agricultural profits, making it more profitable to rent out land than to use it for harvest. It should also be noted that the negative relationship of land ownership or use in any form is only found in the case of rural samples and not for nationally representative ones. This may indicate that rural households, who are more likely to be dependent on agriculture are only able to remain above the USD 1.25 per day poverty line, when they have enough excess land to be able to rent, lease or mortgage it out. These results are given in Table 4.6.

4.7 Limitations

While inferring the analyses in this paper, limitations resulting from the construction data, methodology and the application of the index should be taken into consideration.

4.7.1 Non-scorecard Sources of Inconsistency Predictive Model

It should be noted that the PPI is not designed to explain or estimate how the indicators affect the poverty status of households. The PPI scorecard is a predictive tool, thus the logit model used, only expresses a correlation between the dependent and independent variables. Careful consideration should be made to not infer causation from the model (Innovations for Poverty Action, 2013b).

Household Expenditure Calculation

In low-income countries, it is deemed easier to calculate consumption expenditure than income. In these economies, employment is predominantly informal making income difficult to calculate. On the other hand, savings are usually close to nil, thus consumption expenditure approximately equals income. Executing this theory, household expenditure is used instead of household income to assess poverty status. The PPI design document does not detail the methodology used for expenditure calculation. Thus, for all analysis in this paper, it is assumed that they used the same methodology as BBS.

For BBS HIES 2010 data set, BBS calculates household expenditure by summing up food and non-food consumption expenditures. Since investment expenditure is made to generate income for future consumption, factoring it in would amount to double counting. Hence investment expenditures were not included. This ensured that the results across years and databases were comparable. To this end certain calculation decisions were made that were not economically sound. Firstly, consumption expenditure included: inheritance, gifts, home production as well as purchases. To this list BBS also added the value of gifts and remittances transferred out of the household. Secondly, some families inherited assets upon the death of family members, this created an influx of assets in the respective year. The value of these assets is significant relative to the annual expenditures of poor households. Since depreciation calculations were not made (most likely because it would be too complicated), this might have led to an overestimation of household expenditure. The equations below show how the BBS methodology differs from standard practice.

Thus, the BBS calculation does not correctly value household expenditure or consumption. If the analysis is rerun by correcting the calculation of household expenditure, the correlation of PPI score and the probability of the household falling below the poverty line can be found and compared to the original findings. Since actual household expenditure should be lower, it is likely that the probabilities generated are underestimated. More accurate calculations would likely show that the probability of a household falling under the poverty line is higher at each score category. Unfortunately, the dataset does not separate the sources of the financing of each of the expenses. Thus, this analysis could not be carried out.

Management of Missing Data

Observations corresponding to all 10 variables required to calculate PPI score were collected for only 11,072 households. However, from the design documentation we know that PPI score was calculated for 12,240 households. Hence, approximately 9.4 percent of the sample had been mishandled or imputed. The management method of this missing information by the original authors is unknown. Thus, it is possible that the change in sampling distribution resulting from the elimination of the missing observations affected the derived results.

Data Quality

Some observations in the data set were obviously incorrect. For instance, some households were recorded to own 2500 mobiles and 2000 bicycles, these households were not eliminated from the sample.

This indicates that the quality of the underlying data might have been compromised, thus effecting the results derived from them.

Poverty Lines

The poverty lines used are adjusted using CPI. The CPI differs across regions, and changes over time. Hence, imperfections might result from cost-of-living adjustments. Since, the index is usually applied to sub-groups of individuals, it is expected that the CPI will differ from the national one.

Minimum Sample Size

The design document of the PPI details the methodology to be used to determine the minimum sample size required before applying the PPI. However, to apply the formula, users of the PPI would need to note the participants' population size, select a desired confidence level, assume about expected poverty rate (before measurement and know the degree of precision of the index). Most of these values can only be collected from the original construction report. Thus, when using those values, it must be assumed that the scorecard has remained relevant in the future and for nationally non-representative sub-groups.

Given the complicated process involved, users often don't bother determining minimum sample size. In practice, large-scale users of the PPI in Bangladesh, e.g., Katalyst does not apply this formula. Sample sizes are more often determined by budget restrictions than by statistical soundness. Frequently the sample sizes are as small as 20 households. However, this limitation does not apply to the analyses in this paper, since the minimum sample used here was of size n=5,659.

Cut-off Scores

While inferring the results it should be kept in mind that given the uses of the PPI, variation around cut-off points is of far more importance, than variance around all score ranges, or average variation.

4.7.2 Scorecard Sources of Inconsistency

Even though the scorecard is much easier to use than the direct survey approach, it should be kept in mind that this advantage comes at the cost of some level of inherent bias.

Correlation of Poverty Status with Socioeconomic Indicators

The scorecard must assume that the future relationship between indicators and poverty in all possible groups of households will be the same as the construction data. This is unlikely to hold since the correlation between poverty status and indicators is likely to change rapidly owing to technological advancement and prices.

Sampling Variation

Sampling variation will also lead to some differences between estimated likelihoods and true values. This is clearly seen when we applied the scorecard to the validation sample. For unbiased estimates the scorecard needs to be applied to households that are representative of the same population for which the scorecard was originally constructed. Unbiased means that in repeated samples from the same population, the average estimate matches the true value. The scorecard will be biased when applied after December 2010 or when applied with sub-groups that are not nationally representative.

Out-of-Sample Bias

To be able to apply the revised index to any given sample, it is assumed that the sample's distribution, mirrors the distribution of the construction sample of the index. This assumption cannot hold when the tool is being applied to participants of a local, pro-poor organization, since this sample must inherently be different from the national population.

Overfitting

Another source of differences between estimates and true values is overfitting. Though the scorecard is unbiased, it may still be overfit when applied to an independent sample. It might be the case that the scorecard fits the HIES 2010 data so closely, that it captures random patterns resulting from sampling variation which may not exist for the overall population of Bangladesh. It may also be overfit in the sense that it is not robust when relationships between indicators and poverty, change over time or when it is applied to nationally non-representative samples. Overfitting maybe mitigated by factoring in theory, judgement, and experience instead of basing it solely on data.

4.8 Conclusion

It was not possible to reproduce the results of the PPI 2013 scorecard or look-up table. Though there may be many reasons for this, it is most likely because of the elimination or imputation of incomplete observations. Our results indicate that the original scorecard was generally overestimating poverty likelihoods.

The comparisons in this chapter were made against the original scorecard (PPI 2013 Bangladesh). The assumption made is that the scorecard was unbiased in the beginning but might have lost predictive ability with time. When the scorecard is applied to a nationally representative sample, 5-years in the future (FY 2015-16), it is found that the scorecard, on average, overestimates poverty likelihood.

When the scorecard is applied to a rural-level nationally representative sample from FY 2015-16, we find that the scorecard overestimates poverty overall. It is also observed that the poverty likelihoods at every PPI score range are lower for the rural sample than the national sample.

However, in both cases, the PPI does under-estimate poverty likelihoods at the upper end of the PPI score scale. This is an important characteristic since the cut-off points are usually around the upper end.

The analysis of the targeting capabilities of the PPI reveals that, the cut-off scores and the poverty lines need to be carefully selected. These are closely linked with the degree of false inclusions and false exclusions experienced. Thus, users of the PPI need to be aware of the level of tolerance in terms of false negatives and false positives.

Lastly, change in the PPI scores and consumption expenditures of the households surveyed in the two surveys of the IFPRI BIHS database is tracked. Through use of this panel data set, whether the index can be used to track graduation out of poverty is checked. We see that in this period only 36 percent of the households who had fallen below the USD 1.25 2005 PPP poverty line had experienced a corresponding fall in PPI score. This percentage rises to 38 percent, 39 percent and 40 percent for the USD 1.75, USD 2.00 and the USD 2.50 poverty lines, respectively. On the other hand, 75 percent of households who had graduated above the USD 1.25 2005 PPP poverty line within this period, also experienced a corresponding increase in PPI score. This percentage fell to 71 percent, 71 percent and 69 percent as we move up the poverty lines. Thus, the index is not a perfect measure for tracking falling into or graduating out of poverty. However, it is more likely that households graduating out of poverty will experience increase in PPI score, than it is that households falling into poverty will experience a fall in poverty scores.

4.9 Figures & Tables

Table 4.1: Re-estimated Scorecards

Sl.	Indicator	Response	HIES 2010-11 (Schreiner) (1)	HIES 2010-11 (Training) (2)	HIES 2010-11 (Testing) (3)	HIES 2015-16 (4)	HIES 2010-11 (Schreiner) 9 Indicator (5)	HIES 2015-16 9 Indicator (6)	BIHS 2011-12 9 Indicator (7)	BIHS 2015-16 9 Indicator (8)
1	How many household members are 12 years old or younger?	Three or more	0	0	0	0	0	0	0	0
		Two	10	7	8	7	11	7	14	10
		One	16	12	16	15	17	15	24	17
		None	32	26	29	29	35	30	37	31
2	Do all household members ages 6-12 currently attend a school or educational institution?	No	0	0	0	0	0	0	0	0
		None 6-to-12	0	2	2	1	0	1	1	6
		Yes	6	1	2	3	6	3	2	13
3	In the past year, did any household member ever do work for which he/she was paid on a daily basis?	Yes	0	0	0	0				
		No	8	6	4	11				
4	How many rooms does your household occupy (excluding rooms used for business)?	One	0	0	0	0	0	0	11	0
		Two	3	5	6	0	3	0	1	8
		Three or more	5	9	12	4	5	4	0	9
5	What is the main construction material of the walls of the main room?	Hemp/hay/bamboo, or other	0	0	0	0	0	0	0	0
		Mud brick or C.I. sheet/wood	2	1	7	0	2	1	3	9
		Brick/cement	9	8	12	8	10	11	10	12
6	Does the household own any television?	No	0	0	0	0	0	0	0	0
		Yes	7	5	5	10	8	12	8	6
7	How many fans does the household own?	None	0	0	0	0	0	0	0	0
		One	4	8	7	8	4	8	10	8
		Two or more	7	9	9	12	8	14	11	16
8	How many mobile phones does the household own?	None	0	0	0	0	0	0	0	0
		One	8	8	7	6	9	5	5	5
		Two or more	15	16	17	15	16	15	11	10
9	Does the household own any bicycles, motorcycle/scooters, or motor cars, etc.?	No	0	0	0	5	0	6	5	3
		Yes	4	19	9	0	4	0	0	0
10		No	0	0	0	0	0	0	5	0

	Does the household own (or rent/sharecrop/mortgage in or out more decimals of cultivable agricultural land (excluding uncultivable land and dwelling-house/homestead land)?	Yes	7	0	1	3	8	5	0	0
	Total =		**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Table 4.1 summarizes the results of the re-estimated scorecards. Column (1) lists the scores from the PPI 2013 Bangladesh scorecard. Column (2) and (3) list the scores estimated by replicating the methodology of PPI 2013 Bangladesh on the same dataset (BBS HIES 2010-11). Since the PPI 2013 was constructed on a sub-sample of the dataset, the replication exercise was also conducted on a sub-sample. This is the training dataset in Column (2). These scores were then compared against the scores generated for the testing sample. These scores are shown in column (3). The scores against every indicator appear to have changed, showing that the PPI 2013 Bangladesh could not be reproduced. To add to that, it is seen that the scores generated for the test sample, differ from the scores generated for the training sample. This indicates that the indicator and poverty status relationships are not stable. Column (4) shows the scores estimated by applying the PPI 2013 Bangladesh construction methodology to the BBS HIES 2015-16 dataset. Again, it is seen that for every indicator the scores have changed. Interestingly, for one indicator – vehicle ownership, even the direction of the relationship is reversed. The positive relationship between vehicle ownership and the household being above the USD 1.25 per capita per day poverty line, changes to negative.

To test the applicability of the PPI 2013 Bangladesh on populations, it is applied to the IFPRI BIHS 2011-12 and IFPRI BIHS 2015-16 datasets. Unlike the BBS HIES datasets, the IFPRI BIHS datasets are only nationally representative at the rural level. However, these datasets do not collect information regarding households' daylabourer status. Thus, to compare the PPI scores, the daylabourer indicator is removed, and the PPI score is then scaled to 100. These scores are listed in column (5). This is compared to the scores estimated for BBS HIES 2015-16 in column (6), when only 9 indicators are included in the model. Again, it is seen that the scores change, and the direction of relationship with vehicle ownership reverses.

Column (7) and column (8) show the estimated scores when the PPI 2013 Bangladesh construction methodology is applied to the IFPRI BIHS 2011-12 and IFPRI BIHS 2015-16 datasets. Comparing column (7) and column (8) shows that for rural populations, the magnitude of the relationship between the selected indicators and households' poverty status changes over time. In fact, even the direction of the relationship changes for two indicators. First, for IFPRI BIHS 2011-12, the relationship between number of rooms and households being above the USD 1.25 per capita per day poverty line is negative. This reverses to positive for IFPRI BIHS 2015-16. Second, while ownership of cultivable land is negatively related to being above the poverty line for IFPRI BIHS 2011-12, there is no relationship detected in IFPRI BIHS 2015-16.

Lastly, when the scores in column (6) are compared to the scores in column (8), it is seen that the magnitude of relationships between the dependent and independent variables change. It is observed that land ownership status is positively related to households' being above the USD 1.25 per capita per day poverty line for the nationally representative sample (BBS HIES 2015-16). However, the indicator has no bearing on the explanatory variable for the rural sample (IFPRI BIHS 2015-16) of the same year.

Table 4.2: Poverty Likelihoods

	USD 1.25					USD 1.75					USD 2.00					USD 2.50				
Column	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
PPI Score	**PPI 2013 BD**	**BBS HIES 2010-11**	**BBS HIES 2015-16**	**IFPRI BIHS 2011-12**	**IFPRI BIHS 2015-16**	**PPI 2013 BD**	**BBS HIES 2010-11**	**BBS HIES 2015-16**	**IFPRI BIHS 2011-12**	**IFPRI BIHS 2015-16**	**PPI 2013 BD**	**BBS HIES 2010-11**	**BBS HIES 2015-16**	**IFPRI BIHS 2011-12**	**IFPRI BIHS 2015-16**	**PPI 2013 BD**	**BBS HIES 2010-11**	**BBS HIES 2015-16**	**IFPRI BIHS 2011-12**	**IFPRI BIHS 2015-16**
0-4	97.90%	83.6%	100.0%	71.8%	81.8%	98.8%	83.6%	100.0%	88.7%	90.9%	100.0%	85.9%	100.0%	90.1%	100.0%	100.0%	85.9%	100.0%	95.8%	100.0%
5-9	89.30%	79.1%	80.9%	60.6%	72.0%	98.2%	83.7%	91.5%	84.3%	96.0%	98.7%	83.7%	93.6%	89.8%	96.0%	99.7%	85.7%	95.7%	97.6%	100.0%
10-14	88.80%	77.9%	79.2%	57.5%	69.4%	98.2%	84.0%	95.8%	76.0%	86.1%	98.7%	84.3%	97.9%	84.9%	88.9%	99.7%	84.5%	100.0%	95.2%	91.7%
15-19	81.60%	73.0%	71.6%	50.0%	59.6%	96.9%	82.9%	92.8%	75.6%	84.8%	98.6%	83.6%	95.6%	83.2%	90.1%	99.7%	84.6%	97.2%	91.6%	97.8%
20-24	78.00%	67.6%	64.0%	34.5%	54.6%	96.3%	81.0%	90.0%	69.0%	78.9%	98.4%	82.8%	94.2%	77.2%	86.2%	99.7%	83.4%	97.7%	90.7%	93.4%
25-29	65.80%	61.6%	60.6%	35.0%	48.8%	91.6%	82.1%	89.1%	62.7%	74.2%	95.2%	84.1%	94.1%	72.4%	82.1%	98.7%	85.3%	98.0%	86.0%	91.5%
30-34	57.00%	52.6%	54.2%	28.4%	40.6%	87.9%	75.9%	83.5%	54.2%	65.6%	93.5%	80.4%	90.8%	65.3%	73.1%	98.2%	83.8%	96.9%	80.4%	84.2%
35-39	50.30%	45.1%	44.5%	29.9%	36.4%	83.6%	68.4%	79.2%	57.7%	62.9%	90.7%	73.6%	86.8%	67.8%	71.8%	96.9%	78.8%	94.6%	80.2%	85.7%
40-44	40.80%	40.1%	38.8%	19.9%	30.0%	79.6%	69.6%	74.4%	41.8%	56.6%	87.4%	76.6%	83.0%	51.8%	64.6%	94.9%	84.3%	92.9%	69.9%	78.2%
45-49	33.50%	28.6%	30.7%	23.0%	28.7%	68.8%	61.8%	65.4%	44.6%	52.6%	79.6%	70.2%	76.8%	54.8%	60.4%	91.5%	80.0%	89.5%	69.3%	77.1%
50-54	24.20%	20.4%	25.0%	12.9%	24.8%	60.3%	50.9%	58.8%	37.9%	44.5%	74.2%	61.2%	70.9%	48.7%	55.5%	87.9%	72.2%	85.8%	65.5%	70.7%
55-59	14.50%	15.4%	19.8%	14.4%	17.4%	50.4%	45.0%	51.0%	35.8%	39.6%	65.2%	56.4%	63.6%	49.1%	47.3%	84.3%	73.4%	81.1%	69.4%	66.2%
60-64	10.90%	12.2%	13.9%	12.2%	19.6%	40.4%	35.3%	40.2%	33.3%	36.7%	54.6%	47.2%	53.5%	45.2%	44.8%	73.2%	65.2%	71.8%	60.4%	63.0%
65-69	8.70%	6.4%	10.5%	11.0%	14.8%	32.2%	25.5%	33.0%	29.8%	32.3%	44.5%	36.1%	44.7%	38.7%	39.5%	63.3%	52.0%	63.7%	53.5%	53.8%
70-74	5.60%	8.1%	6.6%	14.3%	15.4%	31.5%	24.7%	26.0%	32.7%	33.8%	42.9%	33.5%	36.4%	42.9%	40.4%	60.4%	49.5%	58.5%	58.2%	58.1%
75-79	4.30%	3.8%	5.7%	6.0%	9.4%	25.8%	20.2%	20.6%	20.9%	27.8%	34.0%	30.2%	29.2%	29.7%	35.7%	50.7%	44.0%	47.4%	47.3%	50.0%
80-84	2.70%	2.9%	3.1%	10.3%	7.7%	19.7%	14.6%	16.1%	27.4%	21.4%	26.7%	22.3%	24.5%	40.2%	29.8%	40.9%	35.1%	41.0%	54.7%	45.8%
85-89	0.00%	1.0%	2.3%	14.3%	8.0%	10.7%	6.5%	11.6%	32.7%	21.4%	14.6%	18.8%	19.4%	38.8%	28.6%	33.3%	34.1%	33.5%	53.1%	40.2%
90-94	0.00%	0.0%	1.1%	7.7%	0.0%	5.1%	4.3%	6.8%	28.2%	12.5%	6.6%	9.9%	11.5%	41.0%	25.0%	12.3%	17.3%	22.5%	53.8%	37.5%
95-100	0.00%	0.0%	0.0%	12.5%	0.0%	0.0%	0.0%	0.0%	30.0%	12.1%	0.0%	0.0%	0.0%	32.5%	19.0%	0.0%	0.0%	0.0%	45.0%	32.8%

Table 4.2 depicts the true poverty probabilities at each PPI score range for four datasets – BBS HIES 2010-11, BBS HIES 2015-16, IFPRI BIHS 2011-12 and IFPRI BIHS 2015-16. Columns (1), (6), (11) and (16) show the poverty probabilities predicted by the PPI 2013 Bangladesh look-up table. These probabilities correspond to the USD 1.25, USD 1.75. USD 2.00 and USD 2.50 per capita per day poverty lines, respectively. Columns (2)-(5) depict the true poverty rates at each PPI score range, for the four datasets; these correspond to the USD 1.25 per capita per day poverty line. Columns (7)-(10), (12)-(15) and (17)-(20) also list the poverty likelihoods at each PPI score range, these correspond to the USD 1.75, USD 2.00, and USD 2.50 per capita per day poverty lines, respectively. Note that all the columns generally exhibit a monotonically decreasing trend. This implies that the poverty likelihoods decrease as the PPI score increases.

Table 4.3:Poverty Likelihood Summary

Dataset	Poverty Line	PPI Score	Performance
HIES 2010-11	USD 1.25	0-100	Overestimation
	USD 1.75	0-100	Overestimation
	USD 2.00	0-84	Overestimation
		85-100	Underestimation
	USD 2.50	0-100	Overestimation
HIES 2015-16	USD 1.25	0-49	Underestimation
		50-100	Overestimation
	USD 1.75	0-84	Overestimation
		85-100	Underestimation
	USD 2.00	0-84	Overestimation
		85-100	Underestimation
	USD 2.50	0-79	Overestimation
		80-100	Underestimation
BIHS 2011-12	USD 1.25	0-59	Overestimation
		60-100	Underestimation
	USD 1.75	0-79	Overestimation
		80-100	Underestimation
	USD 2.00	0-79	Overestimation
		80-100	Underestimation
	USD 2.50	0-79	Overestimation
		80-100	Underestimation
BIHS 2015-16	USD 1.25	0-49	Overestimation
		50-100	Underestimation
	USD 1.75	0-64	Overestimation
		65-100	Underestimation
	USD 2.00	0-74	Overestimation
		75-100	Underestimation
	USD 2.50	0-79	Overestimation
		80-100	Underestimation

Table 4.4: Poverty Graduation Tracker

Poverty Line	Poverty Status	Change in PPI Score		
		Decreased	Unchanged	Increased
USD 1.25 2005 PPP	Fell below poverty line	34.5%	11.4%	54.1%
	No change in poverty status	26.3%	11.3%	62.4%
	Rose above poverty line	16.0%	9.1%	75.0%
USD 1.75 2005 PPP	Fell below poverty line	37.6%	9.5%	53.0%
	No change in poverty status	25.4%	11.4%	63.3%
	Rose above poverty line	17.9%	11.3%	70.8%
USD 2.00 2005 PPP	Fell below poverty line	38.5%	10.8%	50.7%
	No change in poverty status	25.4%	11.1%	63.6%
	Rose above poverty line	18.1%	11.2%	70.7%
USD 2.50 2005 PPP	Fell below poverty line	39.9%	12.1%	47.9%
	No change in poverty status	25.3%	10.6%	64.1%
	Rose above poverty line	18.7%	12.1%	69.2%

Table 4.5: Logit Regression Results for Vehicle Ownership

VARIABLES	BBS HIES 2010	IFPRI BIHS 2011-12	IFPRI BIHS 2015-16	BBS HIES 2015-16
Bicycle	-0.262***	-0.287***	-0.201***	-0.431***
	(0.00778)	(0.0776)	(0.0729)	(0.0300)
Motorcycle	1.066***	0.674**	0.960***	0.833***
	(0.0312)	(0.286)	(0.237)	(0.107)
Car	1.759***			1.128***
	(0.0802)			(0.247)
Observations	611,000	6,503	5,659	45,767

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4.6: Logit Regression Results for Land Ownership

VARIABLES	BBS HIES 2010	IFPRI BIHS 2011-12	IFPRI BIHS 2015-16	BBS HIES 2015-16
own_33	0.0294	-0.277***	-0.0130	0.229***
	(0.0647)	(0.0796)	(0.0829)	(0.0348)
rent_in_33	0.0460	-0.241***	-0.195***	0.0877**
	(0.0596)	(0.0625)	(0.0645)	(0.0351)
rent_out_33	0.447***	0.283**	0.446***	0.411***
	(0.0963)	(0.114)	(0.102)	(0.0536)
Observations	11,072	6,503	5,659	45,767

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4.7: Targeting Performance

Cut-off Score	Poverty Line	USD 1.25					USD 1.75					USD 2.00					USD 2.50				
	Metric	PPI	HIES 2010-11	HIES 2015-16	BIHS 2011-12	BIHS 2015-16	PPI	HIES 2010-11	HIES 2015-16	BIHS 2011-12	BIHS 2015-16	PPI	HIES 2010-11	HIES 2015-16	BIHS 2011-12	BIHS 2015-16	PPI	HIES 2010-11	HIES 2015-16	BIHS 2011-12	BIHS 2015-16
59	True Positive	0.15	0.15	0.20	0.14	0.17	0.50	0.45	0.51	0.36	0.40	0.65	0.56	0.64	0.49	0.47	0.84	0.73	0.81	0.69	0.66
	False Positive	0.86	0.85	0.80	0.86	0.83	0.50	0.55	0.49	0.64	0.60	0.35	0.44	0.36	0.51	0.53	0.16	0.27	0.19	0.31	0.34
	False Negative	0.11	0.12	0.14	0.12	0.20	0.40	0.35	0.40	0.33	0.37	0.55	0.47	0.53	0.45	0.45	0.73	0.65	0.72	0.60	0.63
	Precision	0.15	0.15	0.20	0.14	0.17	0.50	0.45	0.51	0.36	0.40	0.65	0.56	0.64	0.49	0.47	0.84	0.73	0.81	0.69	0.66
	Recall	0.57	0.56	0.59	0.54	0.47	0.56	0.56	0.56	0.52	0.52	0.54	0.54	0.54	0.52	0.51	0.54	0.53	0.53	0.53	0.51
	F-measure	0.23	0.24	0.30	0.23	0.25	0.53	0.50	0.53	0.42	0.45	0.59	0.55	0.59	0.51	0.49	0.65	0.62	0.64	0.60	0.58
69	True Positive	0.09	0.06	0.10	0.11	0.15	0.32	0.26	0.33	0.30	0.32	0.45	0.36	0.45	0.39	0.39	0.63	0.52	0.64	0.54	0.54
	False Positive	0.91	0.94	0.90	0.89	0.85	0.68	0.74	0.67	0.70	0.68	0.56	0.64	0.55	0.61	0.61	0.37	0.48	0.36	0.46	0.46
	False Negative	0.06	0.08	0.07	0.14	0.15	0.32	0.25	0.26	0.33	0.34	0.43	0.33	0.36	0.43	0.40	0.60	0.49	0.59	0.58	0.58
	Precision	0.09	0.06	0.10	0.11	0.15	0.32	0.26	0.33	0.30	0.32	0.45	0.36	0.45	0.39	0.39	0.63	0.52	0.64	0.54	0.54
	Recall	0.61	0.44	0.61	0.43	0.49	0.51	0.51	0.56	0.48	0.49	0.51	0.52	0.55	0.47	0.49	0.51	0.51	0.52	0.48	0.48
	F-measure	0.15	0.11	0.18	0.18	0.23	0.39	0.34	0.41	0.37	0.39	0.47	0.43	0.49	0.43	0.44	0.57	0.52	0.57	0.51	0.51
79	True Positive	0.04	0.04	0.06	0.06	0.09	0.26	0.20	0.21	0.21	0.28	0.34	0.30	0.29	0.30	0.36	0.51	0.44	0.47	0.47	0.50
	False Positive	0.96	0.96	0.94	0.94	0.91	0.74	0.80	0.79	0.79	0.72	0.66	0.70	0.71	0.70	0.64	0.49	0.56	0.53	0.53	0.50
	False Negative	0.03	0.03	0.03	0.10	0.08	0.20	0.15	0.16	0.27	0.21	0.27	0.22	0.24	0.40	0.30	0.41	0.35	0.41	0.55	0.46
	Precision	0.04	0.04	0.06	0.06	0.09	0.26	0.20	0.21	0.21	0.28	0.34	0.30	0.29	0.30	0.36	0.51	0.44	0.47	0.47	0.50
	Recall	0.61	0.57	0.65	0.37	0.55	0.57	0.58	0.56	0.43	0.56	0.56	0.58	0.54	0.42	0.55	0.55	0.56	0.54	0.46	0.52
	F-measure	0.08	0.07	0.10	0.10	0.16	0.35	0.30	0.30	0.28	0.37	0.42	0.40	0.38	0.35	0.43	0.53	0.49	0.50	0.47	0.51

4.10 Appendices

Appendix 4.1: PPI 2013 Bangladesh Scorecard

ipa

ppi

POVERTY PROBABILITY INDEX

PPI® Scorecard for Bangladesh

To assist with collection, organizations can use the household roster located on the second page.

Entity	Name	ID		Date (DD/MM/YY)
Participant:			Date joined:	
Field agent:			Date scored:	
Service point:			# HH members:	

Indicator	Response	Points	Score
1. How many household members are 12-years-old or younger?	A. Three or more	0	
	B. Two	10	
	C. One	16	
	D. None	32	
2. Do all household members ages 6-to-12 currently attend a school/educational institution?	A. No	0	
	B. No one 6-to-12	0	
	C. Yes	6	
3. In the past year, did any household member ever do work for which he/she was paid on a daily basis?	A. Yes	0	
	B. No	8	
4. How many rooms does your household occupy (excluding rooms used for business)?	A. One	0	
	B. Two	3	
	C. Three or more	5	
5. What is the main construction material of the walls of the main room?	A. Hemp/hay/bamboo, or other	0	
	B. Mud brick, or C.I. sheet/wood	2	
	C. Brick/cement	9	
6. Does the household own any televisions?	A. No	0	
	B. Yes	7	
7. How many fans does the household own?	A. None	0	
	B. One	4	
	C. Two or more	7	
8. How many mobile phones does the household own?	A. None	0	
	B. One	8	
	C. Two or more	15	
9. Does the household own any bicycles, motorcycle/scooters, or motor cars etc.?	A. No	0	
	B. Yes	4	
10. Does the household own (or rent/sharecrop/mortgage in or out) 51 or more decimals of cultivable agricultural land (excluding uncultivable land and dwelling-house/homestead land)?	A. No	0	
	B. Yes	7	

By Mark Schreiner of Microfinance Risk Management, L.L.C. Score:

This PPI was created in March 2013, based on data from 2010. For more information about the PPI, please visit www.povertyindex.org

Appendix 4.2: Scorecard Construction Methodology

The first step in developing the index was the construction of the scorecard. To do this, 120 indicators from the areas of family composition, education, housing, ownership of durable assets, employment and agriculture were shortlisted. A logit model was run using just one indicator from the list of 120, on the construction sub-sample. The dependent variable of these models was a binary representing a household's position with respect to the USD 1.25 per capita per day PPP[47] poverty line. One of these indicators was then selected based on several statistical and non-statistical factors. These factors included the degree of correlation, improvement in accuracy, likelihood of acceptance by users (determined by simplicity, cost of collection, and "face validity" in terms of experience, theory and common sense), sensitivity to changes in poverty, variety among indicators, applicability across regions, relevance for distinguishing among households at the poorer end of the expenditure distribution, and verifiability (Innovations for Poverty Action, 2013b). It should be noted that the weights attached to non-statistical criteria used in indicator selection are not provided.[48]

The second step in the PPI construction process, was creating two-indicator regression models based on the one-indicator model selected in the first round, with a second indicator added. The steps used to determine the selection of the first indicator are then repeated. These steps are repeated until 10 indicators that work well together are selected. These 10 indicators form the PPI scorecard. The indicators' assigned weights are derived from the value of their coefficients. For ease of use, the logit coefficients are transformed into non-negative integers, such that total score ranges from 0 to 100 (Innovations for Poverty Action, 2013b).

Schreiner (2011) details the methodology of converting the coefficients β, of the x-variables of the logit regression, to scores. As noted above, the logit model relates households' expenditure-based poverty status y_i with the household's vector X_i of responses to survey indicators. y_i is a binary variable that has the value 1 when a household falls below a poverty line, and 0 when it lies above it. The x variables in the models were all categorical variables. Since β may have many decimal places or be negative, mathematical transformations were necessary to convert the coefficients into integer scores that added up to 100. The first step was calculating the "shifted coefficients"- γ. This was found by subtracting the minimum β coefficient from the responses of each indicator. The shifted coefficient is then multiplied by 100, this value is divided by the maximum possible value of γX and rounded to the nearest integer. If the highest possible score is not exactly 100, the rarest "least poor" response is modified (Innovations for Poverty Action, 2013b). This creates a scorecard in which all responses are non-negative integers. The lower the score, the higher the probability of a household being "poor". The transformation of the coefficients to scores are affine/ homothetic. It should be noted that the scorecard is constructed using just one sample, which is randomly drawn from the national household income and expenditure survey.

Appendix 4.3: Look-up Table Construction Methodology

The score range is first split into 20 equal categories, then the poverty likelihood of households at each score category is estimated. Each household's per capita consumption expenditure per day is then calculated by drawing information from the survey. These per capita consumption expenditures are then compared to the respective household's PPI score. At each score category, the PPI scores are calibrated with poverty likelihoods. The poverty likelihood is defined as the share of households in each score category, in the calibration sub-sample, who are below the specified poverty lines. For instance, if 1000 households from the sample were in the score range 35-39, and of these 445 were below the USD 1.25 per capita per day poverty line, it can be concluded that households at PPI score of 35-39 have a 44.5 percent likelihood of falling below the specified poverty line (Innovations for Poverty Action, 2013b).

Whether the index was under-estimating, or over-estimating likelihood was also tested, by applying the scorecard to the validation sample. The estimated poverty likelihoods were then compared against the true poverty likelihoods

[47] All references to poverty lines in this chapter should be assumed to be 2005 PPP-adjusted.
[48] This is likely to because these decisions were made based on subjective human judgment.

Appendix 4.4:Logit Regression Results for Scorecard Construction

VARIABLES	BBS HIES 2010-11 (Training) (1) y	BBS HIES 2010-11 (Testing) (2) y	BBS HIES 2015-16 (3) y	BBS HIES 2015-16 (16 variables) (4) y	IFPRI BIHS 2011-12 (5) y	IFPRI BIHS 2015-16 (6) y
x_1	2.252***	2.230***	1.782***	1.804***	2.166***	1.923***
	(0.133)	(0.131)	(0.0915)	(0.0906)	(0.135)	(0.137)
x_2	1.051***	1.197***	0.773***	0.774***	1.389***	1.049***
	(0.110)	(0.109)	(0.0727)	(0.0719)	(0.0969)	(0.105)
x_3	0.601***	0.600***	0.373***	0.362***	0.832***	0.636***
	(0.0934)	(0.0941)	(0.0704)	(0.0696)	(0.0899)	(0.102)
x_4	0.0862	0.171**	0.201**	0.226**	0.110	0.438**
	(0.0704)	(0.0698)	(0.0968)	(0.0956)	(0.0979)	(0.178)
x_5	0.179***	0.151***	0.0935	0.121	0.0481	-0.406***
	(0.0501)	(0.0485)	(0.106)	(0.104)	(0.0859)	(0.0886)
x_6	-0.537***	-0.307***	-0.692***			
	(0.0581)	(0.0584)	(0.0425)			
x_7	0.394***	0.463***	0.0117	0.0189	-0.595***	-0.480***
	(0.0659)	(0.0659)	(0.0504)	(0.0499)	(0.0833)	(0.0921)
x_8	0.769***	0.943***	0.292***	0.324***	-0.638***	0.0772
	(0.0785)	(0.0774)	(0.0578)	(0.0572)	(0.172)	(0.201)
x_9	0.0932	0.536***	0.108*	0.177***	0.188**	0.594***
	(0.0672)	(0.0672)	(0.0601)	(0.0594)	(0.0763)	(0.0987)
x_{10}	0.667***	0.879***	0.562***	0.718***	0.584***	0.774***
	(0.0976)	(0.0983)	(0.0767)	(0.0755)	(0.124)	(0.125)
x_{11}	0.458***	0.413***	0.610***	0.706***	0.458***	0.387***
	(0.0809)	(0.0827)	(0.0555)	(0.0548)	(0.0963)	(0.0894)
x_{12}	0.674***	0.514***	0.579***	0.585***	0.568***	0.493***
	(0.0816)	(0.0829)	(0.0535)	(0.0529)	(0.0945)	(0.0832)
x_{13}	0.737***	0.654***	0.823***	0.873***	0.622***	1.006***
	(0.0957)	(0.0963)	(0.0589)	(0.0583)	(0.111)	(0.0982)
x_{14}	0.728***	0.513***	0.341***	0.284***	0.305***	0.286***
	(0.0617)	(0.0612)	(0.0638)	(0.0629)	(0.0734)	(0.0933)
x_{15}	1.417***	1.304***	0.962***	0.934***	0.611***	0.646***
	(0.116)	(0.112)	(0.0703)	(0.0695)	(0.112)	(0.110)
x_{16}	1.610***	0.676**	-0.308***	-0.360***	-0.279***	-0.213***
	(0.403)	(0.275)	(0.0510)	(0.0505)	(0.0776)	(0.0730)
x_{17}	0.0423	0.111*	0.263***	0.398***	-0.315***	-0.0181
	(0.0609)	(0.0608)	(0.0495)	(0.0484)	(0.0662)	(0.0662)
β_0	-1.871***	-2.244***	-1.057***	-1.550***	-0.569***	-1.598***
	(0.141)	(0.143)	(0.127)	(0.122)	(0.113)	(0.213)
Observations	8,278	8,222	16,392	16,392	6,503	5,659

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 4.4 shows the results of the logistic regression models that were run to estimate the scores for each variable of the PPI scorecard discussed in Section 4.5.1. The model is specified as:

$$\text{logit}(P_y) = \beta_0 + \beta_i X_i + u$$

Where,

- y: household is above the USD 1.25 per capita per day poverty line
- x_1: no members below age 12 in household
- x_2: 1 member below age 12 in household
- x_3: 2 members below age 3 in household
- x_4: all members aged 6 to 12 attending school
- x_5: no members aged 6 to 12
- x_6: at least 1 household member worked as a day labourer in the last year
- x_7: household occupies 2 rooms
- x_8: household occupies 3 or more rooms
- x_9: mud brick/CI sheet/wood is the main construction material of the walls of the main room
- x_{10}: Brick/cement is the main construction material of the walls of the main room
- x_{11}: Household owns at least 1 television
- x_{12}: Household owns 1 fan
- x_{13}: Household owns 2 or more fans
- x_{14}: Household owns 1 mobile phone
- x_{15}: Household owns 2 or more mobile phones
- x_{16}: Household owns at least one bicycle/motorcycle/scooter/motor car, etc.
- x_{17}: Household owns/rents/sharecrops/mortgages in/mortgages out 51 decimals or more of cultivable agricultural land

Column (1) shows the results generated when the PPI 2013 Bangladesh methodology is applied to a sub-sample of the BBS HIES 2010-11 dataset. This is the training sample. It shows that x_9 and x_{17} are not statistically significant variables of the model. But the direction of relationship between the dependent and independent variables is as expected from the PPI 2013 Bangladesh scorecard.

Column (2) does the same thing for another sub-sample of the BBS HIES 2010-11. This is the validation sample. Unlike in the case of column (1), variable x_9 and x_{17} are statistically significant. Though, x_{17} is only significant at the 90% confidence level. Variable x_{16} also goes from being statistically significant at the 99% confidence level to the 95% confidence level. The drop in statistical significance of variables between sub-samples of the same dataset, indicates that the variable relationships might be unstable.

Column (3) shows the results of the logit model when applied to the BBS HIES 2015-16 dataset. This is a nationally representative dataset, collected 5 years after the BBS HIES 2010-11 dataset. Here, it is seen that variables x_5 and x_7 are no longer statistically significant. Variables x_4 and x_9 remain significant but at lower confidence levels. Interestingly, the relationship of the y variable with x_{17} reverses.

Column (4) shows the results of the logit model run using only 16 variables, on the BBS HIES 2015-16 dataset. This model is generated for comparison with the IFPRI BIHS 2011-12 and IFPRI BIHS 2015-16 datasets, which do not have information on variable x_6. Again, it is observed that variables x_5 and x_7 are no longer statistically significant. Variable x_4 remains significant at lower confidence level and the relationship of the y variable with x_{17} reverses. The only difference is that variable x_9 is statistically significant at the 99% confidence level, in contrast to 90% in column (3).

Column (5) shows the results of the logit model run using the available 16 variables on the IFPRI BIHS 2011-12 dataset. This is dataset that is nationally representative at the rural level. It is observed that variables x_4 and x_5 are not statistically significant. Variable x_9 is significant only at the 95% confidence level. The relationship of the y variable is reverse of what is seen in the PPI 2013 Bangladesh scorecard for two variables- x_{16} and x_{17}. Also note that unlike in the case of all the other models, the coefficient of x_8 is smaller than the coefficient of x_7; implying that households are more likely to be under the poverty line if they occupy 3 or more rooms.

Column (6) shows the results of the logit model run using the IFPRI BIHS 2015-16 dataset. This too is nationally representative at the rural level. For this model, variables x_8 and x_{17} are not statistically significant. Variable x4 is significant at the 95% confidence level. The relationship of the y variable is reverse of what is seen in the PPI 2013 Bangladesh scorecard for two variables- x_{16} and x_{17}. But the direction of relationships is the same as seen in column (5).

Overall, from column (1) and column (2) it appears that the results of the PPI 2013 scorecard cannot be reproduced. The relationship of the y variable with x_{16} is reversed for all datasets except BBS HIES 2010-11. For the IFPRI BIHS datasets the relationship of the y variable and x_{17} is also opposite of what is expected from PPI 2013 scorecard. It is also seen that variables x_4, x_5, x_7, x_8, x_9 and x_{17} (6 of 17 explanatory variables in the model) are not statistically significant for all the models. Variable x_{16} is statistically significant for all the models, but the confidence level varies.

4.11 Bibliography

Micklewright, J., & Sylke, S. (2010). How reliable are income data collected with a single question? *Journal of the Royal Statistical Society Series A*, *173*, 409–429. https://doi.org/10.1111/j.1467-985X.2009.00632.x

Meyer, B. D., & Sullivan, J. X. (2003). Measuring the Well-Being of the Poor Using Income and Consumption. *The Journal of Human Resources*, *38*, 1180–1220. https://doi.org/10.2307/3558985

Schreiner, M. (2015). A Comparison of Two Simple, Low-Cost Ways for Local, Pro-Poor Organizations to Measure the Poverty of Their Participants. *Social Indicators Research*, *124*(2), 537–569. https://doi.org/10.1007/s11205-014-0789-1

Innovations for Poverty Action. (2013a). *PPI*. PPI. https://www.povertyindex.org/

Mia, M. A. (2017). An Overview of the Microfinance Sector in Bangladesh. *Asian Journal of Business Environment*, *7*(2), 31–38. https://doi.org/10.13106/eajbm.2017.vol7.no2.31

Innovations for Poverty Action. (2013b). *Bangladesh 2010 Poverty Probability Index: Design Memo*. https://www.simplepovertyscorecard.com/BGD_2010_ENG.pdf

Committee on Sustainable Assessment. (2015). *Testing the "Progress Out of Poverty Index" Update to 2013-14 Report*. https://thecosa.org/wp-content/uploads/2015/04/Testing-Progress-Out-of-Poverty-Synthesis-Update-Report-..pdf

Lawson-Mcdowall, J., Beecher, S., & Stuart, G. (2017). *Expanding Financial Inclusion Project Catholic Relief Services Kasama, Zambia*. 36. https://www.crs.org/sites/default/files/tools-research/triangulation-report_6nov2017_final_0.pdf

Desiere, S., Vellema, W., & D'Haese, M. (2015). A validity assessment of the Progress out of Poverty Index (PPI)™. *Evaluation and Program Planning*, *49*, 10–18. https://doi.org/10.1016/j.evalprogplan.2014.11.002

Katalyst. (2012). *Poverty profiling using the progress out of poverty*. https://beamexchange.org/uploads/filer_public/bc/11/bc11c9d8-0530-4bda-8f54-6f4079485c35/ppi_index_study_compressed.pdf

Jalil, M. M., & Azam, M. N. (2014). Using PPI to assess pro-poorness of project activities: Experiences from the Katalyst Project in Bangladesh. *Enterprise Development & Microfinance*, *25*(4), 288–298. https://doi.org/10.3362/1755-1986.2014.027

Bangladesh Bureau of Statistics. (2011). *Report on the household expenditure survey of Bangladesh 2010*. Statistics Division, Ministry of Planning. http://203.112.218.65:8008/WebTestApplication/userfiles/Image/LatestReports/HIES-10.pdf

Bangladesh Bureau of Statistics. (2017a). *Preliminary report on the household expenditure survey of Bangladesh 2016*. Statistics Division, Ministry of Planning. https://www.academia.edu/37874731/Preliminary_Report_on_Household_Income_and_Expenditure_Survey_2016_BANGLADESH_BUREAU_OF_STATISTICS_BBS_STATISTICS_AND_INFORMATICS_DIVISION_SID

International Food Policy Research Institute. (2016). Description of the Bangladesh integrated household survey. USAID. https://2012-2017.usaid.gov/data/dataset/2296b12a-1b9d-42ae-a32e-a8c5cc121f99

Schreiner, M. (2011). Estimating Expenditure-Based Poverty from the Bangladesh Demographic and Health Survey. Bangladesh Development Studies, 31. https://bids.org.bd/uploads/publication/BDS/34/34-4/04Schreiner.pdf

4.12 Acronyms

BBS	Bangladesh Bureau of Statistics
BIHS	Bangladesh Integrated Household Survey
CPI	Consumer Price Index
HIES	Household Income and Expenditure Survey
IFPRI	International Food Policy Research Institute
PPI	Poverty Probability Index
PPP	Purchasing Power Parity

5 Using Machine Learning to Construct an Alternative to the Poverty Probability Index

Abstract - The Poverty Probability Index (PPI) is a household-level poverty classification scorecard that uses 10 socio-economic indicators to classify households as poor or non-poor. Due to its ease of use it has been adopted by governments, NGOs as well as financial institutions to target beneficiaries. Similar proxy-means test based tools could be used to solve other classification problems. However, the construction of the PPI involved a complicated method, requiring in-depth domain knowledge and time-consuming stepwise regression methods to select the indicators and assign them appropriate weights. This paper attempts to see whether it is possible to use machine learning algorithms, to construct similar tools using the same dataset as the PPI. Effort is made to exclude the incorporation of human judgment as far as possible. It is found that machine learning algorithms can build better classification tools with just as few indicators. These models out-performed the PPI by up to 3 percentage points in terms of ROC-AUC and 2 percentage points in terms of accuracy.

5.1 Introduction

The Poverty Probability Index (PPI) is a living standards measurement tool that uses a combination of 10 indicators as a proxy for a household's probability of falling below any of the World Bank specified poverty lines. The tool is constructed using logit regression models where both the explanatory as well as the independent variables are binary. It is tailored to each country, using the respective national household income and expenditure surveys.

The independent variables to construct the PPI are selected based on theoretical and operational criteria. Preference is given to indicators that are easily verifiable and time sensitive. Thus, the initial candidate indicators are chosen based solely on the researcher's judgment (Innovations for Poverty Action, 2013). These indicators are then ordered by the entropy-based "uncertainty coefficient". This coefficient measures the ability of the indicator to predict poverty on its own.

The construction of the PPI involves a tedious process of building separate scorecards for each candidate indicator. The indicator's ability to rank households by poverty status is measured as '*c*'. One of these one-indicator scorecards is then selected based on the accuracy, simplicity, poverty sensitivity and applicability across time and regions. This step is followed by the building of two-indicator scorecards. This is done by adding a second indicator to the indicator selected in the first round. The second indicator is then selected based on the aforementioned statistical and non-statistical criteria. The process is repeated till 10 indicators are selected.

The process of building this 10-indicator scorecard, using the above process, presents two complications. Firstly, the process is long and tedious; and secondly,

it is heavily reliant on human judgment. For instance, to build the PPI scorecard for Bangladesh (2013), the authors only used 120 independent variables (Innovations for Poverty Action, 2013). This is a very small subset of the data that was available from the BBS HIES 2010-11 which was used to construct the scorecard. It is possible that indicators that are better poverty predictors were dropped in the process.

This research paper is an exercise to test whether the use of machine learning algorithms can identify 10 indicators, which predict household poverty with more accuracy than the PPI, using the same dataset. The algorithms allow working with the entire available dataset. The use of human judgment, to shortlist indicators, can thus be avoided.

The three primary objectives of this paper are (i) using machine learning algorithms to develop a 10-indicator proxy-means test for household level poverty status classification; (ii) removing human bias from the indicator selection process; and (iii) identifying the best machine learning algorithm for the exercise.

5.1.1 What is Machine Learning?

Machine learning (ML) is a subset of Artificial Intelligence (AI). It allows a system to make inferences based on data. There are four categories of ML - supervised learning, unsupervised learning, reinforcement learning and deep learning (Singh et al., 2021).

Supervised learning algorithms are used when there is understanding of how the data is classified. If the data is continuous, then a regression algorithm is used. If it is discrete, classification algorithms are used. The regressions used in supervised learning help ascertain the degree of correlation between variables. Since the target feature of this study is a known binary, we will be using supervised learning classification methods to run our analyses.

5.1.2 Machine Learning vs. Econometrics

Statistical processes are the core of both econometrics and machine learning. This often makes it difficult to distinguish between the two, especially, when using supervised models.

The key difference between the two is that econometricians develop models based on economic theories, thus, those are theoretical models. On the other hand, machine learning statisticians only use data to develop models. Analysis of structures or background issues is not required. Thus, there are no assumptions made with respect to dependence or independence of the variables (Zheng et al., 2017).

5.2 Literature Review

In this section, literature pertaining to the use of machine learning in poverty classification is reviewed. The review is restricted to studies that relied on survey data. This is because, the exercise is to evaluate whether machine learning algorithms can perform better than the PPI to classify households by poverty status; and the PPI only uses survey data.

Some of the papers reviewed in this section tested the poverty classification performance of several machine learning algorithms against each other. Others tested the performance of machine learning algorithms against currently used tools for poverty classification i.e., proxy-means tests (PMT) such as the PPI and Poverty Assessment Tool (PAT). Since both the PMT tools were developed using traditional econometric methods, these comparisons help to assess whether machine learning algorithms are better than regular econometric models at the poverty classification task.

5.2.1 Comparing Machine Learning Algorithms

This sub-section summarizes the findings of papers that compared the performance of multiple ML algorithms in classifying household poverty.

Abu Bakar et al., 2020 using the random forest (RF) and decision tree models on data from Malaysia, individuals were classified into poor and hardcore poor groups. It was found that the RF model performed slightly better than the decision tree model-99 percent accuracy vs. 98 percent accuracy.

Another paper using household income and expenditure data of 99,546 Malaysian households, achieved promising results. The objective was to identify proxies that could reliably identify the bottom 40 percent of households using machine learning algorithms. Naive bayes, decision tree and k-nearest neighbour models were designed to classify the economically vulnerable population. It is observed that the decision tree model outperformed the other models (Sani et al., 2018).

Naive bayes, iterative dichotomiser 3 (ID3), decision tree, logistic regression, and k-nearest neighbours (KNN) classification models were used on household data from the Philippines. The models were evaluated on their poverty prediction abilities, based on the following performance metrics: accuracy, precision, recall, F1 score and AUC (Area Under the Curve). Naive bayes classifier was seen to be the most efficient model to distinguish poor and non-poor households. The model performed better than all the other classifiers in all metrics. The error rate was only 0.0014 (Talingdan, 2019).

In addition to the RF, decision tree, KNN and logistic regression models, authors have also attempted to classify household poverty using binarization, Synthetic Minority Over-sampling Technique

(SMOTE), wrapper feature selector, and feature explanation. When these models were applied to data from Costa Rica, it was found that different set of features are required to describe different levels of poverty. However, feature selection and classification are still concluded to be reasonable tools for poverty categorization (Mohamud & Gerek, 2019).

To synopsize, the most used machine learning algorithms in poverty classification using household survey data are, RF, decision tree, survey-weighted elastic net logistic regression, naive bayes, KNN and ID3.

5.2.2 Comparing Machine Learning and Econometrics

Usually, multiple imputations with variables selected by stepwise and lasso regression, are used for household poverty prediction. However, using 6 RF models on same-year data from Albania, Ethiopia, Malawi, Rwanda, Uganda, and Tanzania, it is observed that RF is often more accurate than traditional economic approaches in making out-of-sample poverty predictions (Sohnesen & Stender, 2017).

It has also been reported that the use of PMTs, relying on econometric approaches, for poverty estimation, has resulted in 25 percent exclusion and inclusion errors in Latin America and surrounding regions. This is mainly because traditional econometric methods are not optimized for out-of-sample prediction and are not able to easily leverage on non-linear relationships. Substantial improvement can be experienced by replacing these approaches with ML algorithms. In the case of Latin America, this could extend the coverage to nearly 2 million people in 2 countries (Carrillo et al., 2021).

This sub-section provides a brief of the results of papers that compared the performance of ML algorithms against PMTs developed using traditional econometric methods.

Kshirsagar et al., 2017 used nationally representative household survey data from Zambia, to rebuild the PPI using machine learning algorithms. Typically, simple stepwise logistic regression is used to construct PMTs. The major shortcoming of this method is the high variability in the model-selection step and consequent prediction errors. This is even more acute when the predictors are highly correlated, which is usually the case when these datasets are used. To circumvent this obstacle, a survey-weighted elastic net logistic regression, cross-validated over the overall penalty strength, is used to select active variables. The model is seen to be able to reasonably distinguish between poor and non-poor households

Another study with similar objectives is conducted using data for Bolivia, East Timor, and Malawi. Using quantile regression forests, household level poverty status identification is tested against the USAID PAT. It is found that the machine learning algorithm outperforms the currently used methods by 2 to 18 percent when tested for out-of-sample predictive ability. It is concluded that machine learning can improve the accuracy of poverty targeting tools, especially when stochastic ensemble methods are used (McBride & Nichols, 2015).

The same authors proved that the prioritization of minimization of out-of-sample error, identified through cross-validation and stochastic ensemble methods in PMT tool development, improves the out-of-sample performance of these targeting tools. Data from Bolivia, East Timor and Malawi are used to perform these tests. It should be noted that the loss of accuracy could be reduced by starting with a larger set of variables. However, this would necessitate spending more time to run stepwise regression for variable selection (McBride & Nichols, 2016).

Overall, it appears that machine learning algorithms performed better than both the PPI and the PAT at poverty classification. It should, however, be noted that all three studies were conducted using data from African countries. It is possible that performance will vary when the tools are applied to other regions. Since there has been no testing of data from Bangladesh, research is required to conclude whether machine learning will be better for household level poverty classification for Bangladesh.

5.3 Data

Household Income and Expenditure Survey (2010), Bangladesh Bureau of Statistics*:* The fifteenth round of the HIES, conducted in 2010, introduced many changes to improve data quality and gather more comprehensive socio-economic data. This included the addition of modules for crises and coping measures, micro-credit, migration and remittance, social-safety nets and disability (Bangladesh Bureau of Statistics, 2011b). It also covered more income sources, compared to earlier rounds. Like the previous three rounds, data-entry was instant and digital. This round covered 7,840 rural and 4,400 urban households. These households consisted of 55,776 members.

The performance of predictive machine learning algorithms is directly associated with the size of the data set. High dimensional data sets with many observations can make better predictions. The removal of human bias from feature selection also eliminates the option of relying on economic theory or human judgment for indicator shortlisting. Thus, all data from the dataset was used.[49]

5.4 Empirical Strategy

The objective of this paper was to construct a proxy-means test that is just as simple as the PPI but has greater predictive power in classifying households above and below the poverty line. Thus, the dataset used was the same one that was used for the construction of the PPI - BBS HIES 2010.

Since the PPI was constructed using traditional econometric methods, the authors relied on domain knowledge to shortlist indicators; and then deployed trial and error-based methods to select the 10 best indicators. To reduce the dependence on domain knowledge and the ensuing trial and error-based method, a machine learning model was designed using the available data. This process involved (i) feature engineering, (ii) feature selection, (iii) model training and testing and (iv) performance comparison with the original PPI.

5.4.1 Feature Engineering

Feature engineering is the process of using domain knowledge of the data to create features or variables to use in machine learning (Galli, 2017a). It is an umbrella term that includes multiple techniques to perform data transformation, including, filling missing values, encoding categorical variables, variable transformation and new variable creation from existing ones (Sole from Train in Data, 2020).

The original dataset included information on 12,240 households. Post feature engineering, the feature space had ballooned to 43,595 features including features associated with administrative locations: and 42,874 excluding administrative location features. Adjusting for the sample weights also expanded the sample size to 33,420 observations.

The following sections provide details of the specific feature engineering transformations that the dataset underwent for compatibility with the ML algorithms.

Unbalanced Dataset

The dataset used is not representative of the regional populations. Pre-assigned sample weights were provided with the dataset. These sample weights are to be factored in before any nationally representative estimates are made. Thus, the dataset was expanded using the sample weights prior to conducting the analyses.

Missing Data Imputation

The nature of the missing data was such that whether the data was missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR) could not be determined.

However, a significant number of 'not applicable' responses were generated against several questions in the survey. 'Not applicable' responses arise when a

[49] Due to the use of high-dimensional data, a significant portion of time and effort required for this study was allocated to data cleaning, processing, and transformation.

particular item is irrelevant to the subject. The relevance/irrelevance is usually random. Thus, it is assumed that the ‘not applicable’ responses should be counted as MCAR missing values. For instance, if a question is only to be answered by married persons, every unmarried person will generate a ‘not applicable’ response, irrespective of what his/her response would have been.

Any variable with a non-response rate above 20 percent was dropped. All other missing responses were coded to 0. Thus, missing, and non-applicable responses were both coded to zero. The value of ‘zero’ was also assigned to indicate the absence of things. For instance, if a family did not own a mobile phone, that would be denoted by a zero. Thus, non-responses, not-applicable responses and absences were also codified the same way.

It should be noted that more sophisticated forms of missing data imputations are often used. For instance, dummy variables could have been generated for ‘not applicable’ responses, distinguishing them from ‘none/zero’. This might have led to better inferences. However, this would have expanded the dimensions of the already large data frame. It would also add an additional aspect of complication to the process, which goes against the objective of this assignment - creating a simpler alternative to the PPI. The original authors of the PPI had adopted a similar missing data imputation methodology. Replacing that with a complex mechanism would go against the grain of this exercise.

Categorical Variable Encoding

The objective of this paper is to classify households by their position above or below the USD 1.25 per capita per day poverty line. Thus, it is a supervised classification problem where the target feature is known. Classification algorithms are used to solve such problems. These algorithms only accept numeric and binary variables. Thus, all features in the data set were transformed accordingly.

Since most machine learning models only accept numerical variables, dummy variables had to be generated against all categorical features. To avoid multi-collinearity, the n^{th} dummy variable is dropped, and n-1 dummy variables are retained. The categorical variable is also removed since they are redundant once the dummies have been created.

Distinction is made between single-select and multi-select variables, since the collinearity issue is absent in multi-select cases. Thus, all n dummies are kept. For instance, marital status is a single-select variable. An individual could be either married, unmarried, widowed, etc. However, the options are mutually exclusive. Thus, in this case n-1 dummies should be retained. On the other hand, a multi-select asset ownership variable would not preclude ownership of one asset from owning another. Thus, in these instances, all n dummies are retained.

Discretization

To ensure that the PMT developed is as simple as the PPI, only discrete variables are fed into the model. To do this, all continuous variables were categorized into 10 deciles and a separate category for 0. Eleven dummy variables were generated against these. Since these are mutually exclusive categories, the n^{th} dummy is dropped. Note that the decision to create 11 categories, was based purely on efficiency. It might have been better to create customized ranges for each variable, based on their frequency distributions. However, this process would require more time and computational power.

Outliers

The dataset used had already been cleaned prior to publishing. Thus, there were no outliers that needed to be considered (Bangladesh Bureau of Statistics, 2011b).

Feature Scaling

Features often vary in degrees of magnitude, range, and units. For machine learning algorithms to interpret the relationship of each feature with the target, these magnitudes need to be brought under the same scale. For this project, all features were transformed to binaries. Binaries take the values 0 or 1. Since the values are between a fixed range, there was no need for feature scaling.

Date and Time Engineering

Dates by themselves do not provide any useful information. They must be transformed using domain knowledge before the information they provide can be interpreted. The data set used here required several such transformations. For instance, the year households received certain services had been collected. However, dummy variables were generated for each of these years, to see whether beneficiaries listed in any specific year were more/less prone to living under the poverty line.

Feature Creation

In the interest of retaining simplicity, and to see whether machine learning algorithms could use the features already present to create a better tool, it was decided that no additional features would be created.

The original authors of the PPI did create features, using a combination of the features available from the questionnaire. These features were fed into the machine learning algorithms to see whether they appeared in the new tool even when feature selection is automated.

5.4.2 Feature Selection

The technique of extracting a subset of relevant features from a high-dimensional dataset is called feature selection. Feature selection enhances the interpretability of the model, speeds up the learning process and improves performance (Browniee, 2019). There are several methods that may be used in the feature selection process, each with its own set of advantages and disadvantages. Thus, the methods used for feature selection are incumbent upon the end goal.

Not all feature selection methods are model agnostic. Thus, different feature subsets render optimal performance for different machine learning algorithms. Repeatedly running different algorithms to test the best feature subset is impractical due to the computational expense involved. This makes it difficult to choose the best subset of features.

Feature selection algorithms are divided into three main categories (i) filter methods, (ii) wrapper methods and (iii) embedded methods.

Filter Methods

Filter methods are a category of feature selection methods that rely on the characteristics of the data. The selection of features is impartial to any machine learning algorithm. They are typically univariate, meaning that each feature is ranked independently of the feature space. Hence, these methods are computationally cheap. Due to the low computational power requirements, they are usually employed as the first step of the feature selection process.

Due to the high dimensions of the dataset used, filter feature selection methods had to be applied prior to conducting more sophisticated analyses. Both univariate and multivariate filtering methods were used to reduce the dimensions of the dataset. The filtering methods used are discussed in chronological order below.

Variance Threshold

This univariate filtering method uses the variance within a variable for feature selection. It was used to drop constant and/or quasi-constant features. The lack of variance implied that these features had little to no predictive power. Thus, these features could be safely eliminated.

This was the first method used to reduce the dimensions of the dataset. The tolerance level was set to 0.998, meaning that any feature that had variance less than 0.2 percent was dropped.

Duplicates

This multivariate filtering method scanned the dataset for duplicate features and only retained one of them. Since the information of the duplicates is redundant, retaining them does not add to the predictive ability of the model.

Correlation Co-efficient

Another multivariate filtering method that was used to reduce dimensionality, was the correlation co-efficient. This method was applied after dropping the duplicate features. It identified statistical relationships between two features in the feature space and dropped features that were highly correlated.

Since all the features in the dataset were dichotomous, the Pearson correlation coefficient was used. The coefficient was computed using the brute force approach. This meant that the algorithm scanned the features in the order they appeared. The correlation coefficient threshold was set at 0.8. Thus, if two features showed more than 80 percent correlation, the second feature was dropped.

Though this method is fast it should be noted that the process might have eliminated features with higher predictive power than those retained.[50]

Statistical Measures

After eliminating the correlated independent variables, further filtering was done using statistical tests. This method is also univariate in nature, however, this method factors in the relationship of each independent variable with the target feature.

Since both the independent features and the target feature are Boolean, the chi-squared statistical test was used. The test ordered features and then selected the 100 that ranked the highest. This reduced the dimensions of the dataset to a more manageable, 100 features, excluding the target feature.

Wrapper Methods

After using filter methods to eliminate constant variables, quasi-constant variables, duplicate variables, and correlated variables and ranking the best performing 100 features using the chi-squared test; wrapper methods were applied to the feature subset of 100 to identify the best performing 10 features.

Wrapper methods are unlike filter methods. They use predictive learning models to evaluate the performance of the feature subset. These methods train a new model on each feature subset evaluated; thus, they are computationally expensive. They provide the best performing feature subset for a given machine learning algorithm. However, since these methods are tied to the machine learning algorithms, they are tested on, they may not produce the best feature combination for other algorithms. For instance, a subset of features might provide the best prediction for a random forest model but not for a logistic regression model.

The recursive feature elimination wrapper method is used for the analysis in this paper. This wrapper method trains a machine learning model using all the features and ranks the importance of each feature by some metric. The least important feature is eliminated, and the process is repeated with the remaining features till all remaining features are above the preset performance drop threshold or the preset feature number is reached.

5.4.3 Model Training and Testing

For the analyses in this paper, recursive feature elimination is used in conjunction with 15 classification models, to select the 10 best features for identifying household poverty. To tackle the issues of overfitting and underfitting, the dataset is split into a training set and a testing set in a 7:3 ratio. The 10-fold cross validation procedure is then applied to the test set.

These models are evaluated based on mean ROC-AUC[51], standard deviation of ROC-AUC, mean accuracy, and standard deviation of accuracy.

The performance of these models is then compared to the performance of the PPI. Whether machine learning is a viable alternative to stepwise regression, for developing proxy means tests is then determined.

A second analysis is run using the 100 top ranked features. However, from this list, features that are difficult to compute are eliminated (e.g., calorie intake). Also, features associated with the same variable, that turned up multiple times in the top 100,

[50] Some features have higher predictive power in combination with other features. The brute force approach indiscriminately drops the second highly correlated feature, disregarding whether the feature has higher predictive power than its correlated pair, when feature interactions are factored in. Consequently, this could cause the final model to lose some of its predictive power.

[51] The Receiver Operating Characteristic (ROC)- Area Under the Curve (AUC) is a metric for assessing how well the ML algorithm is performing. The ROC plots the true positive rate against the false positive rate. The AUC measures the model's ability to distinguish between classes. The higher the AUC the better. An AUC score of 1 indicates that the model can distinguish the classes 100 percent of the time. Here, the ROC-AUC score represents the probability of the model correctly classifying households by poverty status.

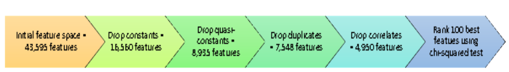
were aggregated. For instance, the consumption of lentils on each day of the week was aggregated to *'lentil consumption in the last 2 weeks'*. Lastly, we retrain the 15 classification models using the PPI feature subset.

5.5 Results

All household level data available from BBS HIES 2010 was used to set up the machine learning algorithm. After completing all the feature engineering steps, there were 43,595 features. Of these 721 features pertained to administrative geographic information, such as, divisions, districts, thanas, villages and wards.

Feature selection methods were used to reduce the dimensionality of the data frame. Several filter methods were deployed. Firstly, features with zero variance were dropped, since these features would not contribute to the model's predictive ability. Once this was done 16,560 features were left. The second step involved dropping the quasi-constant features. These were features with very little variance. For our analysis we set the tolerance level at 0.998. Hence, all features with variance <= 0.002 were eliminated, leaving 8,935 features. All duplicates were dropped next. This step reduced the dimension of the data frame to 7,548. Correlated features were then dropped from the feature set using Pearson correlation based on the brute force approach. In this step 2,598 correlated features were identified, leaving 4,950 features in the feature space. The last filter method used for feature selection was the chi-squared test. Using this test, the best 100 features for classifying households by poverty status were short-listed.

Figure 5.1: Dimensionality Reduction Using Filter Feature Selection Methods



Once the 100 best features were identified, the recursive feature elimination method was applied to the data frame to select a feature subset consisting of 10 features to predict households' poverty status.

5.5.1 Unrestricted Features

Table 5.1[52] includes the results of the 15 machine learning classes that were used in conjunction with the recursive feature elimination wrapper method, to select the 10 best features from the top ranked 100 features. These 10 features can best classify household poverty. The feature number was set to '10' to ensure that the resulting model is comparable to the PPI.

The results show that of the 15 models generated, Extra Trees Classifier, performed the best, in both evaluation-metrics, ROC-AUC and accuracy. It is also evident that the standard deviations for this model class are also the smallest for both. The model has a mean ROC-AUC score of 0.8299 and mean accuracy score of 0.7569.

The second-best model, in terms of our evaluation metrics is the Random Forest Classifier. Note that the metrics are generated for the test set, not the training set. Thus, metrics indicate how well the model will perform on data it was not trained on. [53]

Of the 10 features selected by the best performing model, 9 shared the top rank. The tenth feature ranked second. The features, in order of importance are provided in Table 5.2.

The features uncover two important trends in household poverty classification for Bangladesh. The first is that meat and spice consumption are strong indicators of household poverty status. Secondly, four of the ten indicators generated by the machine learning algorithm, overlap with the indicators in the PPI.

When comparing the performance of the PPI to the performance of the machine learning models, we observe that the PPI ranks eighth in terms of ROC-AUC and ninth in terms of accuracy. The best machine learning model, i.e., Extra Trees Classifier, is more than 3 percentage points better at classifying household poverty, when evaluated using the ROC-

[52] All tables attached at the end.

[53] Model performance is usually better for the training set. For instance, for the training set the 'Extra Trees Classifier' model has a ROC-AUC score of 0.8699.

AUC. In terms of accuracy, the Extra Trees Classifier performs 2 percentage points better. Table 5.1 shows these results.

It should be noted that the feature count for PPI does not mirror that of our model. The PPI clubs together, transformations of the same variable as one feature. For instance, the ownership of fans is broken down into three ranges- no fans in households, only one fan, or two or more fans. This counts as one feature in the PPI. However, in our model we counted these as separate features.

Nevertheless, it should be emphasised that the features used to develop our model, are less verifiable and more complicated to collect data for, than the features used in the PPI. Hence, it is possible that if the features are limited to those that are easily verifiable, the machine learning algorithms would lose their edge.

5.5.2 Restricted Features

The models are rerun using the same methodology, except that, the features are restricted in two ways.[54] Firstly, features that are difficult to compute are dropped e.g., calorie intake. Secondly, features pertaining to the same variable are aggregated in to one. For instance, consumption of a specific food item in week 1 and week 2, are clubbed together as 'consumption over the last 2 weeks'. These transformations reduced the feature subset to 43.

The results generated using the restricted feature subset are shown in Table 3. It is evident that even a restricted feature subset can generate models that have predictive powers like those of the unrestricted feature subset, and consequently better than the PPI. The best performing model is the Decision Tree Classifier. The ROC-AUC score for this model is 0.8298, which is just slightly lower than the best model using the unrestricted feature subset. In fact, when rounded to 3 d.p. the two models perform at the same level.

The 10 features selected by the best performing model were, in order of importance are given in Table 5.2.

[54] Note that these restrictions were applied after the feature set of the top ranked 100 features was generated using the chi-squared test.

The feature list generated from the restricted subset shares five features with the feature list generated from the unrestricted subset.

When comparing the performance of the PPI to the performance of the machine learning models, we observe that the PPI ranks eleventh in terms of ROC-AUC and tenth in terms of accuracy. The best machine learning model, i.e., Decision Tree Classifier, is more than 3 percentage points better at classifying household poverty, when evaluated using the ROC-AUC. In terms of accuracy, the Extra Trees Classifier performs 2 percentage points better. Table 5.3 shows these results.

5.5.3 Training ML Models Using PPI Feature Subset

The last set of analyses run for this study, retrains the 15 classification models using the PPI feature subset. Three key findings were generated from these analyses. Firstly, the Random Forest Classifier produces better results using this subset than Logistic Regression, even though the model was developed using the latter. Secondly, both, the restricted subset and the unrestricted subset were able to produce better models than the PPI feature subset. Lastly, some classification models produced better models using the PPI feature subset, than they did using the other feature subsets, e.g., Bernoulli NB. These results are provided in Table 5.4.

5.6 Limitations

There are seven major limitations of this study. The first limitation lies in the fact that the data used to conduct the analyses is compromised. It is known that the dataset likely suffers from sampling error (Tahsin, 2021).

In addition to that, of the 12,240 observations in the dataset, only 11,072 were complete. Missing information in incomplete observations were recoded as zero during the feature engineering process. This imputation method was chosen in keeping with the methodology that was used during the construction of the PPI. This meant that three types of data were clubbed together as zero: missing observations, not applicable observations, and observations for which the response was zero. This method of missing data handling might have affected the inferences drawn. It should be noted that limitations arising from errors in the dataset applicable to the PPI as well.

The brute force approach was used to drop highly correlated features. This approach scans the features space in the order it is arranged in and drops any feature that is correlated to a feature it has previously scanned. The issue with this method is that it can lead to the algorithm dropping features that have more predictive power, while retaining those that are worse predictors. The other option is to group correlated features, keeping only the ones that show the highest predictive ability. This is known as the model performance-based approach. It requires far more computational power and could not be run.

The number of observations in the dataset is also quite small for machine learning standards. Usually, machine learning algorithms outperform traditional econometric methods when big data is being used.

Due to time and computational constraints, this paper does not explore all the possible combinations of feature selection methods and machine learning classes. The only wrapper method used for feature selection was the recursive feature elimination method. It is likely that the use of other wrapper methods might have generated models with higher predictive power.

The 'Restricted Features' models (Section 5.5.2) endeavoured to remove features that would be difficult to compute. However, no effort was made to remove features that would be difficult to verify. Thus, the machine learning models would not be advantageous in situations in which respondents have incentive to misinform.

Lastly, food item-consumption, as predictors has the additional shortcoming of being overly reliant on memory recall.

5.7 Conclusion

The analyses conducted in this paper illustrated several key points. Firstly, it is seen that feature engineering is an integral part of any machine learning project. The accuracy of machine learning algorithms depends on the correct transformation of features.

Secondly, it is seen that even very high-dimensional datasets can be reduced to manageable dimensions provided appropriate feature selection methods are used. In this case, feature selection techniques were implemented to reduce dimensionality from 43,595 features to 10 best features.

Thirdly, we see that by using machine learning algorithms, even without adequate domain knowledge, it is possible to construct an alternative to the PPI that has higher predictive power. The best models scored 0.8299 and 0.7569 on ROC-AUC and accuracy respectively. In comparison, the PPI scored 0.7987 and 0.7358 in the same performance metrics. Thus, the best model performed three percentage points better than the PPI, in terms of ROC-AUC. However, these models had the drawback of including features that are not easily verifiable. This might prevent adoption in scenarios where respondents have incentive to misinform. Thus, a model that only included features as easily quantifiable as those of the PPI had to be selected.

When features were restricted to ensure easy interpretability, the machine learning algorithms still performed better than the PPI. The ROC-AUC and accuracy scores were 0.8298 and 0.7567 respectively. Most indicators in this model are relatively easy to collect. Hence, it can be concluded that machine learning algorithms can be used to develop proxy means tests, as easy to use, as the PPI.

These algorithms were developed without having to run several trial-and-error models. Neither was there any need to create new variables using domain knowledge. Thus, if desired, an alternative tool could potentially be built with limited human input.

5.8 Extensions

Machine learning and its uses in the development of proxy means tests is a vast area of study. Several avenues could be explored in this arena. Some research topics that could build directly on top of the work done in this paper are:

(i) Using more powerful computers, several permutations of the feature selection methods, feature subsets and training algorithms could be trained and tested to assess whether a better model can be constructed.

(ii) The same methodology could be applied to national level household survey data from Bangladesh for the 50-year period such surveys are available for. This study could help identify a feature or a feature subset that is able to classify poor and non-poor households well across time.

(iii) The methodology could also be applied to a panel dataset to identify a feature subset that may be used to track households' falling into poverty or graduating out of it.

(iv) A research study could be designed to check whether more powerful predictive models can be developed for sub-groups of the population, for instance, by dividing the sample into rural and urban groups.

5.9 Tables and Figures

Table 5.1: Evaluation of Machine Learning Models Using the Unrestricted Feature Subset

Sl.	Class	ROC-AUC		Accuracy	
		Mean	Standard deviation	Mean	Standard deviation
1	Extra Trees Classifier	0.8299	0.0504	0.7569	0.0425
2	Random Forest Classifier	0.8262	0.0518	0.7539	0.0468
3	Gradient Boosting Classifier	0.8164	0.059	0.7485	0.053
4	Ridge Classifier	0.8124	0.0552	0.7455	0.0452
5	Linear Discriminant Analysis	0.8124	0.0552	0.7447	0.0469
6	Ridge Classifier CV	0.8124	0.0552	0.7455	0.0452
7	Decision Tree Classifier	0.8051	0.0665	0.7381	0.0534
8	PPI	0.7987	0.0621	0.7358	0.0508
9	Adaboost Classifier	0.7971	0.0583	0.7378	0.0458
10	Linear SVC	0.789	0.0591	0.7213	0.0399
11	SGD Classifier	0.7885	0.0522	0.7235	0.0413
12	Logistic Regression CV	0.7859	0.0614	0.7179	0.0476
13	Logistic Regression	0.781	0.055	0.713	0.0425
14	Perceptron	0.745	0.0454	0.6821	0.045
15	Passive Aggressive Classifier	0.7398	0.0713	0.6939	0.0507
16	Bernoulli NB	0.7334	0.0866	0.677	0.0793

Table 5.2: Evaluation of Machine Learning Models Using the Unrestricted Feature Subset[55] [56]

Sl.	Indicators: Unrestricted Features	Indicators: Restricted Features
1	Household consumption of meat in the last two weeks	Household's possession of separate kitchen
2	Household consumption of ginger in the first week of the last two weeks	Household consumption of meat in the last two weeks
3	Household consumption of aromatic seeds in the first week of the last two weeks	Household consumption of biris[57] over the last two weeks
4	Household expenses pertaining to electricity over the last month	Household consumption of fuel and lighting over the last month
5	Household expenses pertaining to telephone/ mobile/ internet over the last month	Household consumption of electricity over the last month
6	Household mobile phone ownership	Household mobile phone ownership
7	Household room occupancy	Number of household members below age twelve
8	Number of household members below age twelve	Household consumption of pulses over the last two weeks
9	Household member working as day labourer over the last 12 months	Household consumption of ginger over the last two weeks
10	Household expenses pertaining to, fuel generated from agri by-products, over the last month	Household consumption of aromatic seeds over the last two weeks

Table 5.3:Evaluation of Machine Learning Models Using the Restricted Feature Subset

Sl.	Class	ROC-AUC		Accuracy	
		Mean	Standard deviation	Mean	Standard deviation
1	Decision Tree Classifier	0.8298	0.0519	0.7567	0.0413
2	Extra Trees Classifier	0.824	0.0581	0.7461	0.0465
3	Random Forest Classifier	0.8212	0.0602	0.7486	0.0496
4	Ridge Classifier CV	0.813	0.0642	0.7453	0.0514
5	Linear Discriminant Analysis	0.813	0.0643	0.7466	0.0518
6	Ridge Classifier	0.813	0.0643	0.7458	0.0509
7	Gradient Boosting Classifier	0.8122	0.0692	0.7431	0.0585
8	Adaboost Classifier	0.8042	0.0512	0.7398	0.0391
9	Linear SVC	0.8016	0.0632	0.7308	0.0465
10	Logistic Regression CV	0.7995	0.0611	0.7352	0.0455
11	PPI	0.7987	0.0621	0.7358	0.0508
12	SGD Classifier	0.796	0.0531	0.7322	0.0426
13	Logistic Regression	0.7901	0.0571	0.7221	0.0395
14	Bernoulli NB	0.7297	0.0856	0.673	0.0772
15	Perceptron	0.7238	0.0763	0.6479	0.1003
16	Passive Aggressive Classifier	0.7125	0.1083	0.6191	0.1045

[55] Biris are combustible tobacco products. They contain almost five times more nicotine than regular cigarettes.

[56] Table 2 compares the 10 strongest predictors for the unrestricted and restricted datasets. The greyed rows are indicators that are common for both sets. It is seen that half of the indicators are the same for the two sets. However, their rankings change.

Table 5.4: Evaluation of Machine Learning Models Using PPI Feature Set

Sl.	Class	ROC-AUC		Accuracy	
		Mean	Standard deviation	Mean	Standard deviation
1	Random Forest Classifier	0.7956	0.0587	0.7422	0.0433
2	Extra Trees Classifier	0.7927	0.0679	0.7385	0.0525
3	Gradient Boosting Classifier	0.7912	0.0674	0.7387	0.0526
4	Logistic Regression CV	0.7899	0.0624	0.739	0.0525
5	Logistic Regression	0.785	0.0683	0.7353	0.0565
6	Ridge Classifier	0.7841	0.0684	0.7273	0.0572
7	Linear Discriminant Analysis	0.7841	0.0684	0.7271	0.0576
8	Ridge Classifier CV	0.7841	0.0684	0.7273	0.0572
9	Decision Tree Classifier	0.7834	0.0645	0.7247	0.0491
10	Linear SVC	0.7832	0.0676	0.7301	0.0522
11	Bernoulli NB	0.7824	0.0525	0.6941	0.0516
12	SGD Classifier	0.7809	0.0576	0.7285	0.0519
13	Adaboost Classifier	0.778	0.0633	0.7187	0.0510
14	Perceptron	0.7632	0.0806	0.6979	0.0680
15	Passive Aggressive Classifier	0.6908	0.0695	0.6357	0.0940

5.10 Bibliography

Abu Bakar, A., Hamdan, R., & Sani, N. S. (2020). Ensemble Learning for Multidimensional Poverty Classification. *Sains Malaysiana*, *49*(2), 447–459. https://doi.org/10.17576/jsm-2020-4902-24

Bangladesh Bureau of Statistics. (2011). *Report on the household expenditure survey of Bangladesh 2010*. Statistics Division, Ministry of Planning. http://203.112.218.65:8008/WebTestApplication/userfiles/Image/LatestReports/HIES-10.pdf

Browniee, J. (2019). *Feature Selection*. Machine Learning Mastery: Making Developers Awesome at Machine Learning. https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/

Carrillo, A., Cantu, L. F., Tejerina, L., & Noriega, A. (2021, April 12). *[2104.04148] Individual Explanations in Machine Learning Models: A Case Study on Poverty Estimation*. https://arxiv.org/abs/2104.04148

Galli, S. (2017a). *Feature Engineering for Machine Learning*. Udemy. https://www.udemy.com/course/feature-engineering-for-machine-learning/learn/lecture/7631540?start=75#overview

Innovations for Poverty Action. (2013). *Bangladesh 2010 Poverty Probability Index: Design Memo*. https://www.simplepovertyscorecard.com/BGD_2010_ENG.pdf

Kshirsagar, V., Wieczorek, J., Ramanathan, S., & Wells, R. (2017). Household poverty classification in data-scarce environments: A machine learning approach. *ArXiv:1711.06813 [Stat]*. http://arxiv.org/abs/1711.06813

McBride, L., & Nichols, A. (2015). *Improved poverty targeting through machine learning: An application to the USAID Poverty Assessment Tools*. https://www.semanticscholar.org/paper/Improved-poverty-targeting-through-machine-learning-McBridea-Nicholsb/6ad8406eb8d2c809810e0485639bb41d237352bd

McBride, L., & Nichols, A. (2016). Retooling Poverty Targeting Using Out-of-Sample Validation and Machine Learning. *World Bank Group*, 43. https://elibrary.worldbank.org/doi/10.1093/wber/lhw056

Mohamud, J. H., & Gerek, O. N. (2019). Poverty Level Characterization via Feature Selection and Machine Learning. *2019 27th Signal Processing and Communications Applications Conference (SIU)*, 1–4. https://doi.org/10.1109/SIU.2019.8806548

Sani, N. S., Abdul Rahman, M., Abu Bakar, A., Sahran, S., & Mohd Sarim, H. (2018). Machine Learning Approach for Bottom 40 Percent Households (B40) Poverty Classification. *International Journal on Advanced Science, Engineering, and Information Technology*, *8*(4–2), 1698–1705. http://insightsociety.org/ojaseit/index.php/ijaseit/article/view/6829/pdf_851

Singh, P., Singh, N., & Deka, G. C. (2021). *Prospects of Machine Learning With Blockchain in Healthcare and Agriculture* [Chapter]. Multidisciplinary Functions of Blockchain Technology in AI and IoT Applications; IGI Global. https://doi.org/10.4018/978-1-7998-5876-8.ch009

Sohnesen, T. P., & Stender, N. (2017, March 9). *Is Random Forest a Superior Methodology for Predicting Poverty? An Empirical Assessment—Sohnesen—2017—Poverty & Public Policy—Wiley Online Library*. https://onlinelibrary.wiley.com/doi/10.1002/pop4.169

Sole from Train in Data. (2020). *Feature Engineering for Machine Learning: A Comprehensive Overview*. Train in Data. https://trainindata.medium.com/feature-engineering-for-machine-learning-a-comprehensive-overview-a7ad04c896f8

Talingdan, J. A. (2019). Performance Comparison of Different Classification Algorithms for Household Poverty Classification. *2019 4th International Conference on Information Systems Engineering (ICISE)*, 11–15. https://doi.org/10.1109/ICISE.2019.00010

Zheng, E., Tan, Y., Goes, P., Chellappa, R., Wu, D., Shaw, M., Sheng, O., & Gupta, A. (2017, August 20). *When Econometrics Meets Machine Learning*. https://www.sciendo.com/article/10.1515/dim-2017-0012

5.11 Acronyms

AI	Artificial Intelligence
AUC	Area Under the Curve
BBS	Bangladesh Bureau of Statistics
HIES	Household Income and Expenditure Survey
ID3	Iterative Dichotomiser 3
KNN	K-Nearest Neighbours
MAR	Missing at Random
MCAR	Missing Completely at Random
ML	Machine Learning
MNAR	Missing Not at Random
PAT	Poverty Assessment Tool
PMT	Proxy-Means Test
PPI	Poverty Probability Index
RF	Random Forest
ROC-AUC	Receiver Operating Characteristic - Area Under the Curve
SMOTE	Synthetic Minority Oversampling Technique

Conclusion

This is a comprehensive, but not exhaustive, study of data quality and scarcity issues experienced by researchers and policymakers in Bangladesh. Though the paper relies on evidence from Bangladesh, the framework developed to assess overall data quality can be applied to other countries with similar socioeconomic dynamics.

The research conducted in the first half of this thesis, signals that both the micro and macro data available for Bangladesh are compromised. However, contrary to preconceived notions held by many, GDP per capita is found to be underestimated as opposed to overestimated. At the same time, the micro data used to compute the Gini-coefficient is seen to have overestimated the incomes of the lowest three quartiles of households. This implies that the additional income reflected in the national accounts funnels back to households in the top income quartile. Thus, the state of income inequality is likely to be much worse than previously thought.

While the night-time lights model developed to estimate GDP per capita, indicates an underestimation, it should be noted that this underestimation has persisted for at least 20 years. In fact, the rate of underestimation appears to have fallen over time. This is concerning, given that GDP per capita growth rate has been consistently overestimated in recent years. It implies that economic activities that were already in existence have only been recently included in formal GDP calculations. Thus, growth might only have been on paper.

The findings directly challenge the narrative of miraculous economic growth that has been delivered by the Government of Bangladesh.

The exploration of data quality and data availability issues in the first half of the thesis, establishes how difficult it is to ensure reliability and accessibility of data. Thus, in the second half, we explore the reliability of the Poverty Probability Index (PPI). A proxy means-test developed to identify household level poverty, in the absence of income and expenditure data.

Our research reveals that the PPI, while not perfect, does a reasonable job of classifying households by poverty status. However, the poverty likelihood changes considerably as the construction data becomes dated, or when applied to nationally non-representative populations. In both cases, poverty likelihood is overestimated at most score categories. Thus, it is recommended that when using a dated PPI scorecard or applying it to a rural sub-group, the cut-off scores be lowered. This would reduce false exclusions and increase accuracy.

Lastly, the research in this thesis demonstrates that alternatives to the PPI can be developed using machine learning algorithms. Moreover, the alternative models generated by the machine learning algorithms performed up to 3 percentage points better, in terms of ROC-AUC and 2 percentage points better, in terms of accuracy. In addition to having higher predictive powers, these models were also easier to develop, since they required very little domain knowledge.

Though this paper explores machine learning only in the context of micro data. The same principles could also be applied to construct leading, lagging and coincident indicators for macroeconomic indices.