2019

IARIW-World Bank

Special IARIW-World Bank Conference "New Approaches to Defining and Measuring Poverty in a Growing World" Washington, DC, November 7-8, 2019

Nowcasting Global Poverty

R. Andres Castaneda Aguilar

Daniel Gerszon Mahler

David Newhouse

Nowcasting Global Poverty[*]

R. Andres Castaneda Aguilar[†]
Daniel Gerszon Mahler[‡]
David Newhouse[§]

October 15, 2019

Abstract

Timely and comparable poverty estimates are vital to assess countries' development progress and track the first Sustainable Development Goal, to end extreme poverty by 2030. Yet timely and comparable estimates of poverty are lacking. For these reasons, initiatives that reliably and cost-effectively predict what the poverty rate is today (i.e., nowcasting) are crucial for informed high-level decision-making. In this paper, we discuss how to leverage large-scale datasets, such as the World Development Indicators, and statistical learning techniques to improve the accuracy of the World Bank's current poverty nowcasts. We apply these techniques and dataset to predict growth in mean welfare, and back out poverty rates by applying the predicted growth rates equally to all households in the last observed distribution. This is in line with how the World Bank's current nowcasts work. We find only minor gains in prediction accuracy but that progress in reducing global poverty is slower than current estimates indicate. Predicting headcount rates directly, rather than through growth in mean welfare, considerably reduces prediction accuracy. Prediction accuracy would be greatly improved if it were possible to accurately predict both growth in mean welfare and growth in the Gini coefficient.

Keywords: Poverty, Nowcasting, Machine Learning, Measurement.
JEL codes: C53, D31, I32, O10.

---

1. Introduction

Timely and comparable poverty estimates are vital to assess countries' development progress. International poverty estimates serve as a public good for researchers and inform the development community on efforts to meet the first Sustainable Development Goal, to end extreme poverty by 2030. Within international development organizations, they also inform the allocation of resources and the development of strategic priorities.

Yet timely and comparable estimates of poverty are lacking for many reasons. In some countries, fragility, conflict and violence make it difficult to conduct household expenditure surveys altogether, while in other countries, lack of monetary resources is the main obstacle. Even when surveys are frequently conducted, the time it takes to field a survey, collect, process and analyze the data, imply a two-year lag before the data are released. With the world changing at an ever more rapid pace, this lag risks painting an outdated picture of poverty in a country. As of October 2019, on average across the developing world, the most recent survey with international poverty data was from 2013. In addition, 19 economies with a population greater than 1 million had no international poverty estimates at all and 12 had only one estimate. For these reasons, initiatives that reliably and cost-effectively predict what the poverty rate is today ("nowcast") are crucial for informed and effective high-level decision-making.

The World Bank currently nowcasts poverty by assuming that the distribution of welfare as observed in the most recent household survey grows in accordance with Household Final Consumption Expenditure (HFCE) per capita or GDP per capita taken from national accounts (Prydz et al, 2019). Yet research has documented large discrepancies between income measured from national accounts and welfare aggregates from household surveys (Ravallion, 2003; Deaton, 2005; Pinkovskiy & Sala-i-Martin, 2016). In addition, though growth is an important driver of poverty reduction, it may not be the only factor that is useful to predict poverty.

In this paper, we use large-scale datasets and machine learning methods to nowcast poverty throughout the world. We focus on regularization techniques, random forests, and gradient boosting. Our primary data set is the World Economic Outlook (WEO) and the World Development Indicators (WDI), which contain nearly 2000 country-level variables that may be relevant for predicting poverty. We combine this data with the PovcalNet database, which contains more than 1500 international poverty estimates covering 164 countries.

To test the ability of the WEO and WDI variables to predict poverty, we transform the data into observed spells of annualized changes in log mean consumption and annual changes or growth rates in the WEO and WDI predictor variables. Machine learning methods are applied to historical spells to predict welfare growth into the final spell for each country. In line with the World Bank's current nowcasting method, this welfare growth is used to generate estimates of

poverty assuming distribution-neutral growth, that is, assuming that inequality is unchanged between the year of data and the year of the nowcast. The most recent spell for each country is used to evaluate the performance of these predictions.

We find that utilizing machine learning techniques and large-scale datasets only decreases the expected error in poverty rates by 6.7% -- from 2.82 to 2.66 percentage points. In other words, despite of the shortcomings of the current extrapolation method, it works surprisingly well. We discuss why using WEO and WDI only yields small improvements. We show that predicting the headcount rate directly (rather than growth in the mean combined with distribution-neutrality) is not the answer – with this approach prediction errors nearly double. Rather, we find that predicting growth in the mean joint with changes in inequality can reduce errors substantially. We discuss the best way in which one can implement changes in inequality under a variety of different distributional assumptions. We show that if one were to predict changes in the Gini coefficient as well as growth in the mean and use linear growth incidence curves to implement the changes in the Gini, then there is potential to reduce the prediction error by 71% (from 2.82 to 0.83).

Though a range of papers try to *forecast* poverty to 2030 (see for example Hillebrand (2009), Edward & Sumner (2014), Lakner et al. (2019)), their objective is not to *nowcast* global poverty. Nowcasting and forecasting are different exercises. With nowcasting, the features of the model are actual data at the target year. Forecasting, in turn, rely on features that themselves have been forecasted beforehand. Thus, the accuracy of poverty forecasting highly depends on the accuracy of the forecasts of the features, which in most cases are restricted to growth rates in GDP per capita. Among the papers that have pursued nowcasts at a less-than-global scale, includes Caruso et al. (2017), which test different methods of nowcasting poverty in Latin American and the Caribbean, Jean et al. (2016), which uses satellite imagery to nowcast poverty in five African countries, and Leventi et al. (2013), which predicts risk of poverty among EU countries using microsimulation tools. Attempts at nowcasting poverty in country settings include Blumenstock et al. (2015) and Newhouse & Vyas (2018).

We believe our contribution comes from being the first paper with a clearly worldwide focus: to nowcast poverty globally. In addition, we believe we add value by comparing a range of different methods to predict poverty as well as exploring, to this date, the largest set of possible features yet.

The rest of the paper is organized as follows. Section 2 outlines how the World Bank currently nowcasts poverty. Section 3 details the data we will use. Section 4 goes through the various models and methodological choices we make. Section 5 presents the main results. Section 6 offers a discussion of possible avenues to improve upon the main results. Section 7 concludes.

2. Current methods for nowcasting

In this section, we outline the World Bank's current method to nowcast poverty in countries, which is behind the official reporting on the first Sustainable Development Goal and will be referred to as the status quo in this analysis. This method will be used for benchmarking how other methods work in contrast to the status quo.

The status quo method to nowcast poverty in a country is based on the premise that there is a tight relationship between income or expenditure as measured in national accounts and the income or consumption observed in household surveys. The method works by taking the last observed distribution of welfare, collected at time t_{survey}, and scale the welfare of each household, h, by the growth observed in household final consumption expenditure (HFCE) from national accounts between t_{survey} and $t_{nowcast}$. The adjusted vector is used to estimate poverty at time $t_{nowcast}$. Using the international poverty line of $1.90 per capita per day in 2011 PPP to measure poverty (as we will do throughout this paper), the nowcasted poverty rates are given by:

$$headcount_{t_{nowcast}} = F\left[welfare_{h,t_{survey}}\left(1 + growth_{t_{nowcast},t_{survey}}\right) < 1.9\right]$$

Growth rates in GDP per capita are used for countries without data on household final consumption expenditure and countries in Sub-Saharan Africa (Prydz et al, 2019). Figure 1 illustrates how this scaling works in the case of Nigeria. The latest survey from 2009-10 gives a headcount rate of 53.3%. Between 2009-10 and 2015—the last year at which global poverty numbers currently are expressed—GDP per capita in Nigeria grew about 13%. When applying this growth to the entire welfare vector, the nowcasted poverty rate for 2015 becomes 47.0%.

There at three important assumptions behind this method:
1. It assumes that the growth observed in household final consumption expenditure is fully 'passed through' to the welfare observed in household surveys. The full passthrough is in contrast to empirical evidence showing that on average, only a fraction of growth in national accounts trickles down to household surveys (Ravallion, 2003; Deaton, 2005; Pinkovskiy & Sala-i-Martin, 2016).
2. It assumes that the only factor informative for changes in poverty is growth in national accounts. It is plausible that other variables might be more informative, such as changes in the employment rate, or, more likely, that a combination of variables matter.
3. It assumes that growth accrues to everyone equally, that is, without changing the distribution of welfare. This is problematic if growth was pro-poor or pro-rich in the intervening period.

Figure 1: Example of the World Bank's current nowcasts

Note: The figure shows how the World Bank currently does nowcasts (referred to as the status quo in this paper), which is by shifting the entire distribution of welfare at the last observed point by the growth observed in national accounts between the time of the survey and the nowcasting year.

The approaches we take try to tackle these three assumptions. Dealing with the first assumption can be rather straightforward; one would simply regress the annualized growth in mean consumption observed in surveys on the annualized growth in income from national accounts, as shown below.

$$g_{survey\ mean} = \beta * g_{HFCE/capita} + \varepsilon$$

The β-coefficient reflects the expected share of growth from national accounts that is passed through to growth observed in household surveys. Rather than shifting the entire welfare distribution by $1 + g_{HFCE/capita}$ it could be shifted by $1 + \beta * g_{HFCE/capita}$. We will use this method as well for benchmarking against other results.

3. Data

All poverty estimates used in this paper comes from PovcalNet, which contains the World Bank's official country-level, regional and global estimates of poverty. Most of the data in PovcalNet comes from the Global Monitoring Database (GMD), which is the World Bank's repository of multitopic income and expenditure household surveys used to monitor global poverty. PovcalNet contains more than 1500 surveys from 164 countries covering 97% of the world's population. The data available in PovcalNet are standardized as far as possible but differences exist with regards to the method of data collection, and whether the welfare aggregate is based on income or consumption. By relying on the PovcalNet database, we ensure consistency with the official numbers used by the World Bank and United Nations for monitoring poverty, inequality and related goals.

PovcalNet also contains some auxiliary variables that may be relevant for nowcasting poverty, such as the region to which the country belongs, whether income or consumption is used to measure poverty, GDP per capita, HFCE per capita, and the population size. To compliment these variables with other variables that might be relevant for nowcasting poverty, we utilize the World Economic Outlook (WEO) database and the World Development Indicators (WDI). The former is hosted by IMF and contains about 50 variables related to macroeconomic outcomes, such as inflation, government debt, unemployment and current account balance. The latter is hosted by the World Bank contains more than 1000 indicators obtained from a variety of sources related to a broad range of themes, such as health, agriculture, education, climate change, infrastructure and more.

For both WEO and WDI we consider all variables available that satisfy two criteria: (1) The variable has at least 50% non-missing in the nowcasting year (2018), and (2) the variable can be compared across countries. The former criterion assures that the variable is useful for the actual nowcasts. The latter removes non-sensical variables, such as variables reported in local currency, where exploiting cross-country variation is not meaningful.

Not all poverty trends within countries are comparable over time due to changes in the survey methodology or the poverty estimation. This matters for our predictions of poverty in a country, since there is no reason to believe that the variables from WDI and WEO follow the same trend as the growth in the survey mean when this growth is based on two incomparable surveys. Put in other words, even if we knew the exact causes of poverty in a country, if poverty estimates are not comparable, then we would not be able to predict the changes in poverty between two surveys. Our main results will cover all data points, even the ones that are not comparable over time, but as a robustness check we look at what happens if we restrict ourselves to comparable spells.

4. Methods

Several methods will be employed to nowcast poverty around the world in this paper. In this section we outline the various methods used, their advantages and disadvantages, and other important methodological choices.

The target variable

When discussing how to nowcast poverty, it seems intuitive that the target variable to predict is the poverty rate in each country. Yet, the status quo method only predicts the poverty rate indirectly. It first predicts the growth in the mean welfare and then shifts the entire distribution to back out a poverty rate. We will take this approach as our starting point, and thus likewise try to predict growth in the mean. To check the robustness of this choice, we show results when predicting the poverty rate directly.

Predicting the growth in the mean has some advantages: 1) The model can be applied to any poverty line, 2) the model takes full advantage of the previous data in sense that the entire distribution is leveraged, and 3) by anchoring the target variable to the past estimate, many counter-intuitive outcomes are avoided. The latter argument means that if economic and social conditions have improved since the last poverty estimate, then the mean will (in all likelihood) be expected to grow, and poverty will decline. If the poverty rate is predicted directly without anchoring to the past estimate, and if a country generally has higher poverty levels than its other economic variables would suggest, then it is possible that the nowcasted estimate will be higher than the latest estimate. This can be challenging from a policy dialogue perspective, which is often centered around changes from the last observed point.

Predicting growth in the mean also comes with some disadvantages. For one, predicting the poverty rate directly is arguably the most intuitive option; the poverty rates are what we care about at the end of the day. Predicting growth in the mean also has the disadvantage that, in contrast to predicting the poverty rate directly, it does not work for countries without any data at all. The methods which are based on the last observed survey need some assumptions about poverty levels in countries without data to arrive at global poverty rates. Finally, predicting growth in the mean requires some distributional assumption.

Algorithms

In order to predict growth in the mean (or the poverty rate directly), we rely on a number of frequently used machine-learning algorithms. In particular, we will use the lasso (Tibshirani, 1996), the post-lasso (Belloni and Chernozhukov, 2013), CART random forests (Breiman, 2001), conditional inference random forests (Hothorn et al., 2006) and gradient boosting (Friedman, 2001). These methods all have in common that they predict the outcome variable of interest while being agnostic about which variables are relevant for the predictions and how these interact.

Missing data

Since the variables we use suffer from a lot of missing data, it is necessary to find a strategy to deal with the missingness. Simply deleting rows with missing values is not feasible as this would leave no or very few observations left. For the tree-based methods (random forests and gradient boosting), we rely on imputation methods implicit in the algorithms to deal with missing data. Tree-based methods, generally, work by sequentially splitting the sample into two based on a variable deemed most predictive of the target variable. For example, it might judge that countries with a decline in the share of male workers in agriculture is predictive of growth in the mean. For country-years without data on the share of male workers in agriculture, the algorithm will search for the most similar variable in terms of how it relates to the target variable, which in this case could be the share of *all* workers in agriculture, and split the observations with missing values in the first variable based on the latter. Such methods of dealing with missing values are not possible for the lasso and post-lasso, where we will instead restrict ourselves to variables with no or very few missing values, and in case missing values are present, use the status quo method adjusted by a passthrough rate for the predictions.

An alternative approach is to multiple impute the entire dataset of features and thereby avoid missing values altogether (Rubin 1976, 1987; Schafer 1997). This has the advantage of working with all methods, even the ones that do not have their own implicit imputations and that it can deliver standard errors which reflect part of the uncertainty of the predictions. On the flip side, they add quite a bit of computing time. In this paper, we will focus on leveraging the imputation methods implicit in the algorithms but make one comparison with multiple imputation.

Organization of data

In order to compare the predictive accuracy of the various methods, they need to be evaluated on the same terms and on the same data. This matters particularly for the machine learning methods as they should not be trained and evaluated on the same data. For these purposes we construct a training set, on which all models will be tuned through cross-validation, and a test set which will use the tuned models for purposes of out-of-sample comparisons. The training set contains all estimates, except for the last spell for each country, which will be in the test set. The test set will thus contain one entry per country, while the training set could contain no rows for a particular country (if it only has two observations), or many, depending on the number of poverty estimates.[5] To assure that we do not fit the training data to the countries which happen to have many poverty estimates, we weigh each observation in the training set such that the sum for each country equals 1. This ensures that the (weighted) training set mimics the test set. Table 1 below

---

[5] Some countries in PovcalNet have estimates based on both consumption and income. If the latest survey with both income and consumption estimates is after 2000, we treat these two series as different countries, and thus put the latest spell of each method in the test sample. If the latest survey with either income or consumption estimates is before 2000, we consider it to be the case that the country no longer uses this method for poverty estimation and put all spells with this method in the training sample.

illustrates how this works for two exemplary countries. We also add a nowcasting sample, which contains the spell between the latest datapoint for each country and the nowcasting year. The nowcasting sample will be used to derive the final nowcast for each country.

Table 1: Example of organization of data

Original data			Spell data			
Country	*Year*		*Country*	*Spell*	*Sample*	*Weight*
Mali	1994		Mali	1994-2001	Train	0.5
Mali	2001		Mali	2001-2006	Train	0.5
Mali	2006		Mali	2006-2010	Test	1
Mali	2010	→	Mali	2010-2018	Nowcast	1
		Turning				
Bangladesh	1984	*dataset into*	Bangladesh	1984-1986	Train	0.143
Bangladesh	1986	*spells*	Bangladesh	1986-1989	Train	0.143
Bangladesh	1989		Bangladesh	1989-1992	Train	0.143
Bangladesh	1992	→	Bangladesh	1992-1996	Train	0.143
Bangladesh	1996		Bangladesh	1996-2000	Train	0.143
Bangladesh	2000		Bangladesh	2000-2005	Train	0.143
Bangladesh	2005		Bangladesh	2005-2010	Train	0.143
Bangladesh	2010		Bangladesh	2010-2016	Test	1
Bangladesh	2016		Bangladesh	2016-2018	Nowcast	1

Loss function

To evaluate the performance of the various methods on the test set, our primary loss function will be the mean absolute deviation in the predicted headcount rates:

$$MAD_{test} = \frac{\left|headcount_{predicted,test} - headcount_{true,test}\right|}{N_{test}},$$

where $N_{test,c}$ indicates the number of spells in the test set. Looking at the mean absolute deviation rather than the mean squared error tends to give less weight to how well outliers are predicted. We believe this is convenient since data incomparabilities can create some strong outliers at times, and we do not want to judge the methods by how well they predict these outliers. We focus on percentage point deviations rather than percentage deviations, as the latter sometimes can be huge for countries with low poverty rates. If a country has a poverty rate of 1% but our model predicts a poverty rate of 2% the error in percentage terms would be 100% which would give this observation a large impact. Our focus on percentage point deviations thus implicitly gives a larger focus on countries with high poverty rates.

5. Results

Before illustrating the results of the various models, we show how the status quo method performs in terms of predicting poverty rates in Figure 2a. Evidently, the current method has quite large prediction errors. On average in the test sample, the predicted poverty rate is 2.82 off the true rate. Though this may sound small for countries with poverty rates around 50 percent, this is the average including countries with little or no extreme poverty and including countries with short time between the nowcasting year and the latest datapoint. The expected prediction error is around half a percentage point when the extrapolation time – the time between the latest survey and the nowcast – is less than two years, but it increases to 8 percentage points when the extrapolation error is 10 years. The median error is a fifth of the average error (0.56 percentage points), suggesting that the mean error is driven by extreme values. Indeed, the figure shows that six countries have errors greater than 15 percentage points.
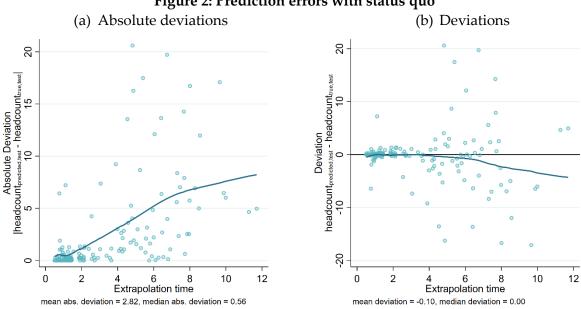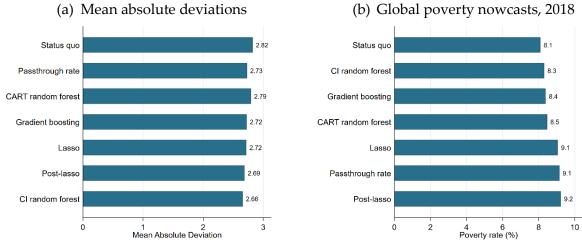
Figure 2: Prediction errors with status quo

(a) Absolute deviations   (b) Deviations

mean abs. deviation = 2.82, median abs. deviation = 0.56   mean deviation = -0.10, median deviation = 0.00

Note: The figure shows how well the status quo method predicts poverty rates, measured by the absolute deviation and the deviation between the survey estimate and the nowcasted estimate pretending the particular survey did not exist. Three observations with extrapolation time greater than 12 years are not shown in the figure. A small jitter has been added to the extrapolation time variable to prevent surveys with the same extrapolation time from overlapping.

The current method also appears to give biased predictions. As shown in Figure 2b, for extrapolation distances greater than four years, on expectation, the predicted headcount rate tends to be systematically lower than the true headcount rate. This is likely due to the assumption that the entire growth in national accounts is passed through to the welfare vector.

We next want to explore the extent to which the mean absolute deviation of 2.82 can be lowered by applying a passthrough rate and by leveraging large-scale dataset jointly with machine learning methods. Figure 3a shows the test errors from the various methods.

Figure 3: Comparison of performance and nowcasts of different models

| (a) Mean absolute deviations | (b) Global poverty nowcasts, 2018 |

Note: The left figure shows how well our various methods nowcast poverty, measured as the mean absolute deviation between nowcasted and true poverty rates. The right figure compares the global poverty nowcasts for 2018 with the different methods. *Status quo* is the method currently employed by the World Bank. *Passthrough rate* is the status quo method, but where only a fraction of growth is passed through to welfare observed in surveys. The rest are based on various statistical learning methods.
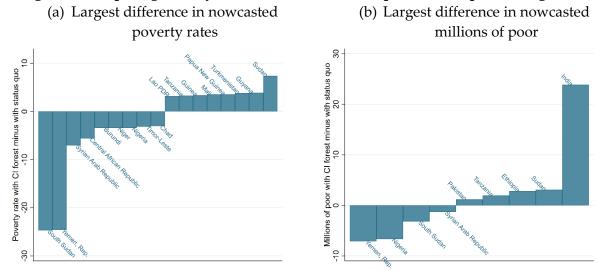
None of the methods we apply have a marked impact on the test error. Applying a global passthrough rate, the simplest fix, reduces it by 3.2% (to 2.73) while using conditional inference random forests – which gives the best results – reduces it by 5.7% (to 2.66). All other models perform in between the conditional inference random forest and the status quo method.

Though these reductions are small, they do matter for our global and country-level poverty nowcasts (Figure 3b). The status quo methods predicts a global poverty rate of 8.1% in 2019. This is lower than all other methods, whose estimates range from 8.3% to 9.2%. In terms of millions of poor, this means that there could be between 16 and 86 more millions of poor in the world. 86 million is more than the population of the Democratic Republic of Congo, Germany or Iran – hardly an insignificant number.

If we look into country specific nowcasts, some changes are even more stark (Figure 4a). The best performing method predicts poverty rates in South Sudan and Yemen that are more than 20 percentage points lower than the status quo method would suggest, and poverty rates 7 percentage points lower for Syria. These are all countries that have entered conflict since the last datapoint available. As part of this conflict, their GDP per capita has plunged. What these results suggest is that the GDP per capita falls faster than consumption per capita as countries enter conflicts.

When looking at the difference in the millions of poor between the status quo and the conditional inference random forest, India sticks out as the largest difference. The status quo method would predict 25 million less poor in India. This, like the example above, is because the status quo tends to make too large adjustments based on growth rates in GDP per capita. Consumption tends to change slower and less erratically than changes in GDP per capita.

Figure 4: Comparing country-level nowcasts in status quo and best performing model

(a) Largest difference in nowcasted poverty rates

(b) Largest difference in nowcasted millions of poor

Note: The left figure shows countries where the difference in the nowcasted poverty rate from the status quo method and using conditional inference random forests is greater than 3 percentage points. The right figure shows countries where the difference in the nowcasted millions of poor between the status quo method and using conditional inference random forests is greater than 1 million.

The smaller relevance of GDP per capita for conditional inference random forests begs the question of what other variables might be relevant for nowcasting poverty in this model. Figure 5 shows the ten variables most important for the predictions with this method ordered by their importance. The most important variable for the predictions is annualized growth in employment. This is followed by the annualized growth in GDP per capita and the annualized growth in the composite national accounts indicator (GDP for Sub-Saharan Africa and countries that lack HFCE, HFCE otherwise), which the World Bank is currently using to nowcast poverty. No other variable is of great importance, but changes in unemployment, employment share in industry, and government debt all have some relevance. In general, the variable importance plot points to why the status quo method, and particularly the passthrough rate method are doing fairly well: no other variable except for employment growth (which is highly correlated to growth in GDP per capita) matters much for the nowcasts.
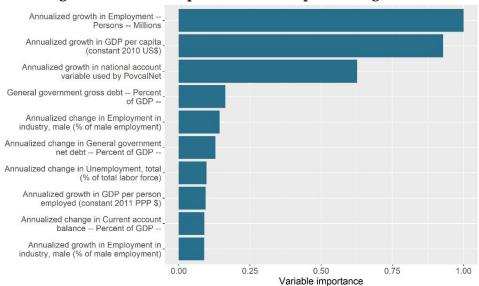
Figure 5: Variable importance for best performing model

Note: The figure shows the variable importance for the conditional inference random forest model. The importance measure has been standardized such that the variable with the highest importance equals 1.

6. Discussion

The results so far have shown relatively small reductions in the test error. In this section we discuss possible reasons why that is.

Incomparable data

One possible reason why the results do not give substantial decreases in the test error is that many of the survey spells are not comparable. If the spells are not comparable, there is no reason to believe that any variable can predict the growth in welfare. We test this by repeating our exercise but only including the spells in the test sample that are based on comparable data. Results are shown in Figure A.1 in the Annex. Only relying on comparable spells does not change the main conclusion: no method markedly reduces the test error. The best-performing method is now the post-lasso, which reduces the error by 6.7% (from 2.24 to 2.09).

Predicting the headcount rate directly

Another reason could be that the method of predicting growth in the mean rather than the headcount rate directly is not the best avenue. To check whether this is the case, we try to use our best-performing model, conditional inference random forest, but have the poverty rate be the target variable. With this approach, or test error nearly doubles from 2.82 with the status quo to 5.34 (Figure A.2 in the annex), suggesting that the indirect approach is not the reason why it is difficult to improve upon the status quo.

Using multiple imputation to account for missing values

Next, we try to account for missing values by using multiple imputed dataset. This is particularly relevant for the nowcasts relying on the lasso, where we so far have removed the variables with many missing values and used the passthrough rate method for rows which still had presence of missing values. With multiple imputed dataset, this is not of concern. Still, when comparing the mean absolute deviation in the test sample, using multiple imputed data yields a slightly higher error rate. This could be because most of the variables with missing data are not relevant for predicting growth in the survey mean.

Predicting growth in the mean better

Yet another reason could be that we are just very poor at predicting the mean. We can check whether this is the case by evaluating the performance of the status quo method under the hypothetical scenario that it predicts the growth in the mean perfectly. Supposing we have surveys for a country in 2009 and 2015, this means that we would scale the entire 2009 distribution until it reaches the same mean as the 2015 distribution, and see how different the headcount rate with this scaled-up 2009 distribution is to the headcount rate of the 2015 distribution. With this method, the average prediction error in the test sample decrease by 36% – from 2.82 to 1.80 – a marked improvement over the status quo and all of our machine learning methods (Figure A.2 in the annex). This could imply that we currently do not have the right features in terms of predicting growth in consumption. Given the richness of our current set of covariates and current set of methods, however, it is dubious whether this level can be obtained in practice.

Navigating the distribution-neutral assumption

A final reason why the predictions are not much of an improvement over the status quo could be that the distribution-neutral assumption does not hold. We have already seen that predicting the poverty rate directly probably is not a better avenue, but it could be that predicting growth in the mean and changes in the distribution – for example proxied by changes in the Gini coefficient – could give better outcomes.

A challenge with this approach is that the same numerical change in the Gini (or any other measure of inequality) can materialize in infinitely many ways. Hence, some further distributional assumptions are needed. We explore two ways of implementing changes in the Gini. One method is to use the predicted mean and predicted Gini together with a known two-parameter distributional shape that welfare could follow – such as the log-normal distribution – to back out, under this distributional shape, what the headcount rate would be.

We can again work under the hypothetical scenario that we perfectly predict the mean and the Gini and see how far off the poverty predictions would be from the true poverty rates. In other words, we compare the poverty rate from a given survey with the poverty rate one would recover by using the mean and Gini from that survey and approximating the full distribution with a

particular distributional shape. We try using the log-normal distribution, the gamma distribution, the Weibull distribution, and the Fisk distribution (also known as the log-logistic distribution).

Using log-normality reduces the test error by 51% relative to the status quo (from 2.82 to 1.39) (Figure A.2 in the Annex). The distribution that works best, though, is the Fisk distribution. If one were to predict the median and the Gini perfectly with the Fisk distribution, then the test error would decrease by 59% relative to the status quo (from 2.82 to 1.15).

The other method with which one could implement changes in the Gini is by using specific growth incidence curves. Growth incidence curves plot the growth in welfare as a function of the percentile of the initial welfare distribution. Downwards-sloping growth incidence curves reduce inequality and vice versa. Evidence shows that growth incidence curves often take an approximately linear form (Lakner et al., 2019). Given a growth in the mean and a change in the Gini, there is only one linear growth incidence curve. Hence, by restricting ourselves to this functional form we can implement predicted changes in the Gini.

We can again work under the hypothetical scenario that we perfectly predict the mean and the Gini and see how far off the poverty predictions would be from the true poverty rates with this method. In practice, this means that we take the old distribution and subject it to a linear growth incidence curve that is constrained such that the mean after applying this growth incidence curve will equal the new survey mean. We gradually change the slope of this growth incidence curve until we match the Gini of the new survey.

If we were to perfectly predict growth in the mean and changes in the Gini and use a linear growth incidence curve to implement these changes, the error in the predicted poverty rate decreases by 71% (from 2.82 to 0.83, as shown in Figure A.2 in the Annex). These results are encouraging because changes in poverty can largely be described by changes in the mean and Gini. These errors remain hypothetical in the sense that they are based on perfect predictions of the mean and Gini, yet they suggest that improvements are likely to be found by pursuing this avenue.

7. Conclusion

In this paper we leveraged statistical learning techniques combined with large-scale dataset – the World Economic Outlook database (WEO) and the World Development Indicators (WDI) – to nowcast poverty around the world. We benchmarked our nowcasts against the World Bank's current implicit method to nowcast poverty around the world, which only relies on growth data from national accounts.

Results showed little improvement from incorporating these large-scale data and utilizing statistical learning methods: The error in the predicted poverty rate decreased by 5.7% (from 2.82 to 2.66 percentage points). A number of reasons could be behind this. It could be that changes in poverty are not captured well by WDI and WEO indicators. Another potential reason could be that the predictions suffer from measurement error – either in the poverty data or in the WDI and WEO indicators. Yet another reason could be that the current method works rather well because it is theoretically guided. Finally, it could be because most of our methods impose no changes in the distribution between the year of the latest datapoint and the survey year.

We find that if we were to predict the Gini as well, and were able to predict both the survey mean and the Gini perfectly, then changes in poverty rates can be approximated quite accurately by applying linear growth incidence curves to the old distribution. This remains a hypothetical scenario in the sense that it is based on perfect predictions of the mean and Gini, yet it suggests that improvements are likely to be found by pursuing this avenue. In future work, we will try to predict the Gini directly to see if there is any merit to the approach highlighted above.

A number of other potential avenues will be considered going forward to improve the predictions. We will incorporate satellite data, starting with nighttime lights and rainfall data, to see if this helps the predictions. Next, we will fully utilize multiple imputed data to see if these methods can reduce predictions errors in certain cases. Finally, we will use other methods, such as dynamic panel data models, to see if these methods are better capable of capturing the information in the data.

References

Belloni, Alexandre, and Victor Chernozhukov. 2013. "Least squares after model selection in high-dimensional sparse models." *Bernoulli* 19 (2): 521-547.

Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. "Predicting poverty and wealth from mobile phone metadata." *Science* 350 (6254): 1073-1076.

Breiman, Leo. 2001. "Random forests." *Machine learning* 45 (1): 5-32.

Caruso, German, Leonardo Lucchetti, Eduardo Malasquez, Thiago Scot, R. Andres Castaneda. 2017. "But … What is the Poverty Rate Today? : Testing Poverty Nowcasting Methods in Latin America and the Caribbean." *Policy Research Working Paper* 8104, World Bank, Washington DC.

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

Edward, Peter, and Andy Sumner. 2014. "Estimating the scale and geography of global poverty now and in the future: How much difference do method and assumptions make?" *World Development* 58: 67-82.

Friedman, Jerome H. 2001. "Greedy function approximation: a gradient boosting machine." *Annals of Statistics*. 1189-1232.

Hillebrand, Evan. 2009. "Poverty, growth and inequality over the next 50 years." In *Expert Meeting on How to feed the World in 2050*.

Hothorn, Torsten, Kurt Hornik, and Achim Zeileis. 2006. "Unbiased Recursive Partitioning: A Conditional Inference Framework." *Journal of Computational and Graphical Statistics* 15 (3): 651–674.

Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. 2016. "Combining satellite imagery and machine learning to predict poverty." *Science* 353 (6301): 790-794.

Lakner, Christoph, Daniel Gerszon Mahler, Mario Negre, Espen Beer Prydz. 2019. "How Much Does Reducing Inequality Matter for Global Poverty?" *Policy Research Working Paper* 8869, World Bank, Washington DC.

Leventi, Chrysa, Jekaterina Navicke, Olga Rastrigina, Holly Sutherland, Erhan Ozdemir, and Terry Ward. 2013. "Nowcasting: estimating developments in the risk of poverty and income distribution in 2012 and 2013." Research Note 1/2013, Social Situation Monitor, European Commission.

Newhouse, David Locke, and Pallavi Vyas. 2018. "Nowcasting poverty in India for 2014-15: A survey to survey imputation approach." *Global Poverty Monitoring Technical Note* 6, World Bank, Washington DC.

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

Prydz, Espen Beer; Dean Mitchell Jolliffe, Christoph Lakner, Daniel Gerszon Mahler, and Prem Sangraula. 2019. "National Accounts Data Used in Global Poverty Measurement." *Global Poverty Monitoring Technical Note* 8, World Bank, Washington DC.

Ravallion, Martin. 2003. "Measuring Aggregate Welfare in Developing Countries: How Well Do National Accounts and Surveys Agree?" *The Review of Economics and Statistics* 85(3): 645-652.

Rubin, D.B. 1976. "Inference and missing data." *Biometrika* 63: 581–592.

_____. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley

Schafer, J. L. 1997. *Analysis of Incomplete Multivariate Data*. Boca Raton, FL: Chapman & Hall/CRC

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the lasso." *Journal of the Royal Statistical Society* 58 (1): 267–88.
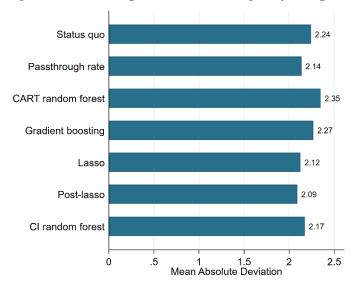
Annex

Figure A.1: Model performance using only comparative spells

Note: The figure shows the test error of different models when the test set is restricted to spells that contain two comparable surveys in terms of poverty estimation.
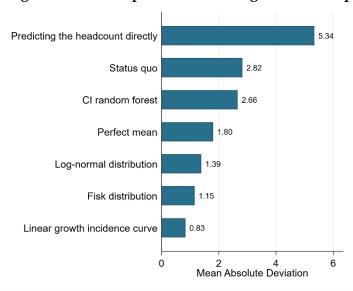
Figure A.2: Model performance using alternative approaches

Note: The figure compares the status quo and best performing test error (conditional inference random forest) with test errors when the poverty rate is predicted directly, and hypothetical test errors if we were perfectly able to predict the mean, and in the last three cases also the Gini while using various methods to implement the changes in the Gini.