Daniel Mahler (The World Bank), David Newhouse (The World Bank), and Raul Andres Castaneda Aguilar (The World Bank)

Nowcasting Poverty Rates in the Developing World

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 With the world changing at an ever more rapid pace, this lag risks painting an outdated picture of poverty in a country.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 March 2019, for the average country with international poverty data, the most recent survey with comparable poverty estimates was from 2013-14. 19 economies with a population greater than 1 million had no 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.

Both the World Bank and the Brookings poverty clock currently nowcast poverty by assuming that the distribution of welfare grows in accordance with Household Final Consumption Expenditure or GDP/capita taken from national accounts. Although growth is an important driver of poverty reduction, it is not the only factor that is useful to predict poverty. In addition, research has documented large discrepancies between income measured from national accounts and welfare aggregates from household surveys. Relying on national growth rates also implicitly assumes that growth is distribution-neutral, that is, that inequality is unchanged between the year of data and the year of the nowcast. Therefore, this approach can deliver misleading estimates.

In this paper, we use large scale data sets and machine learning methods to nowcast poverty throughout the developing world. We focus on regularization techniques and random forests. Our primary data set is the World Economic Outlook (WEO), which contains many country-level economic 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 variables to predict changes in poverty, we transform the data into observed spells of annualized changes in log mean consumption and annual changes or growth rates in the WEO predictor variables. We evaluate the predictions by withholding the final spell for each country. Machine learning methods are applied to previous rounds to predict welfare growth into the final year, which is used to generate estimates of poverty assuming distribution-neutral growth. Poverty estimates for the most recently available round are compared to the survey-based estimates.

For example, poverty estimates for Angola exist for 2000 and 2008. We withhold the 2008 rates from the data, predict the annual growth in welfare between 2000 and 2008 using the pooled cross-section covering all countries, and shift the 2000 distribution by the predicted growth in consumption to estimate poverty in 2008. Predicted poverty is then compared with measured poverty from 2008 survey. These validation results, using all countries, are used to evaluate different models.

To evaluate our predictions, we use two loss functions; the mean absolute error in the predicted poverty rate and the share of poverty rates where the sign of the change is predicted incorrectly. The first metric informs how far off our predictions are from the true estimates, while the second metric informs whether we predict the trend in the poverty rates correctly.

Preliminary results suggest that using non-linear techniques have the highest predictive performance. Using random forests reduces average prediction error in the poverty rate from 3.9 percentage points to 3.1 percentage points and reduces the share of trends incorrectly predicted from 32% of the spells to 18% of the spells.

This method has the advantages that the predictions are anchored in previous poverty estimates and that it can be used for any given poverty line. A downside is that it cannot be applied to countries without poverty estimates. In addition, it maintains the assumption that inequality does not change between the year with data and the nowcasting year.

We will explore extensions that make the nowcasts distribution-sensitive and add to the set of predictor variables. One approach to tackle the first issue is to have a separate model that predicts changes in the Gini coefficient and impose a linear growth incidence function that converts the welfare distribution of the latest poverty estimate to match the nowcasted predicted mean and the nowcasted predicted Gini. Another approach is to use a distributional assumption, such as log-normality, to convert the predicted mean and Gini to a full distribution. Preliminary results suggest that the latter approach has limited potential. Even if the mean and Gini were predicted perfectly by the model, assuming log-normality gives no more accurate estimates than distributional-neutral estimates. Time permitting, we will also explore using the World Development Indicators to vastly increase the set of variables that can be used for predictions. This would allow incorporating indicators from other surveys, such as labor force surveys and demographic and health surveys, which can help improve predictions in some countries.