RECENT EXPERIENCES WITH POVERTY NOWCASTING IN LATIN AMERICA

Xavier Mancero and Alvaro Fuentes, ECLAC, Statistics Division

Second technical workshop on nowcasting in international organizations

May 25-26, 2022

INTRODUCTION

This presentation is part of a study carried out with the objective of review and evaluate different alternatives for predicting poverty for Latin American countries.

- The base model was the one usually used by ECLAC to yearly predict poverty, which is based on micro simulations from the projected annual variation of income and the Gini index.

- Different poverty forecasting alternatives were evaluated considering:
  - the micro-macro nature of the data,
  - the temporal dynamics of the series,
  - the heterogeneity of the panel and
  - the use of machine learning techniques.

- The final evaluation was performed on a panel for 12 countries in the region between 2000 and 2019, comparing the projections of aggregate poverty and extreme poverty rates.

THE SET OF VARIABLES AND DATA USED

- Period: 2000-2019.
- An initial panel of 18 countries was constructed for 20 years.
- However, only for 12 of the 18 countries there was sufficient information to make the assessment.
- The relatively cyclical and smooth nature of the mean income, poverty and Gini values allowed the use of cubic splines for missing data interpolation.
- Although there are still missing values, mainly at the beginning and end of the period, the level of coverage increases from 87% to 92%. In any case, this slight imbalance will be considered in the estimation of several of the models proposed in this work.

Variables surveyed

Variable	Source
Average p/c household income, in local currency	SEDLAC (CEDLAS and World Bank)
GDP p/c at constant prices in dollars	CEPALSTAT
Gini Coefficient	CEPALSTAT & SEDLAC
Urban poverty line, in national currency	CEPALSTAT
Extreme poverty line, in national currency	CEPALSTAT
Poverty rate	CEPALSTAT
Extreme poverty rate	CEPALSTAT

THE MICRO-MACRO APPROACH TO ECLAC POVERTY FORECASTS.

The base model for prediction comes from the following formula,

$$\widehat{\boldsymbol{y_i^{t+1}}} = (\mathbf{1}+\boldsymbol{\beta})[(\mathbf{1}-\boldsymbol{\alpha})\boldsymbol{y_i^t} + \boldsymbol{\alpha}\boldsymbol{\mu^t}]$$

where:

- y_i^t is the income per person of each household i in year t ;
- μ^t is the average income per person of all households in year t ;
- α is an exogenous parameter that expresses the projected percentage variation in the Gini index
  - It takes a positive value in periods of economic contraction and zero in periods of economic growth;
- β is an exogenous parameter that accounts for the growth rate of income per person,
  - It is assumed to be equal to the ECLAC projected growth rate of GDP per capita (in constant dollars) for *t+1*.

The income projected for *t+1* from the household survey data at *t* allows us to identify poor people at *t+1* and calculate the respective poverty rate.

EVALUATION OF THE PREDICTION OF POVERTY AND EXTREME POVERTY FOR THE YEAR 2019 WITH THE BASE MODEL.

- The forecast performance measures used were mean error (ME), root mean square error (RMSE), mean absolute error (MAE), mean percentage error (MPE), and mean absolute percentage error (MAPE).
- These measures involve an aggregate assessment of the projections made for this set of twelve countries. Since the mean error is defined as the difference between the observed data and the poverty projection, negative errors mean that there is a tendency to over-predict both the poverty rate and the extreme poverty rate.
- The measures expressed in percentage terms (MPE and MAPE) show a worse performance of the extreme poverty rate forecasts with respect to the poverty forecasts. On average, there is a good fit of the projections for 2019 for the aggregate of the countries in the region.

Base model: forecast performance measures for 2019.

	Poverty	Extreme poverty
ME	-0,59	-0,24
RMSE	1,52	1,06
MAE	1,09	0,74
MPE	-1,49	-5,97
MAPE	4,23	11,03

DIFFERENT EXTENSIONS OF THE BASE MODEL WERE EVALUATED AS ALTERNATIVES TO THE ECLAC´S PREDICTIONS.

Base model	**Model 0:** ECLAC projections for year *t +1* using GDP per capita projections made in *t*.
Models that predict the per capita GDP growth rate and the Gini coefficient from the panel data instead of using ECLAC´S prediction of GDP	**Model 1 (PVAR):** PVAR(1) between pc GDP growth rate and Gini, by GMM, considering unbalanced panel and clustered SE by country.
	Model 2 (PVAR+ quadratic): Model 1 including exogenously a quadratic term of the first lag of the GDP growth rate.
	Model 3 (LASSO): PVAR(n) whose regressors and fixed effects are selected through LASSO.
Models that directly predict the poverty rate from the panel, without resorting to microdata.	**Model 4 (LASSO aggr):** PVAR(1) using LASSO for lag selection and country fixed effects.
Models introducing additional information as exogenous regressors	**Model 5 (LASSO + exog):** Model 3 including the growth rate of prices of agricultural commodities, energy and nominal exchange rate.
Pooling models, which combine two of the previous models	**Model 6 (Pooling 0 + 3):** a combination (pooling) of forecasts between model 0 and 3 was evaluated, taking simple averages.
	Model 7 (Pooling 0 + 5): a pooling of forecasts between model 0 and 5 was evaluated, taking simple averages.

METHODOLOGY FOR THE EVALUATION OF THE PREDICTIVE PERFORMANCE OF THE DIFFERENT MODELS.

To validate the results obtained from the forecasting models, a cross-validation exercise was carried out.

- We work with the original panel of data and iteratively extract one year at a time that is used as a validation set for the models.

- The exercise was not replicated for the first years of the panel because in those years there is a higher proportion of missing data in the sample. Our final panel goes from 2003 to 2019.

- Four performance measures were evaluated: the mean forecast error (ME), the root mean square error (RMSE), the mean absolute error (MAE) and the median absolute error (MdAE).

- Additionally, the statistical significance of the difference between the ECLAC prediction and the best prediction obtained for each model was analyzed. In the case that the ECLAC prediction was the best, the difference with the second best was studied.

RESULTS OF THE EVALUATION OF THE PREDICTIVE PERFORMANCE OF THE DIFFERENT MODELS.

As a result of the model evaluation process, the following conclusions were reached:

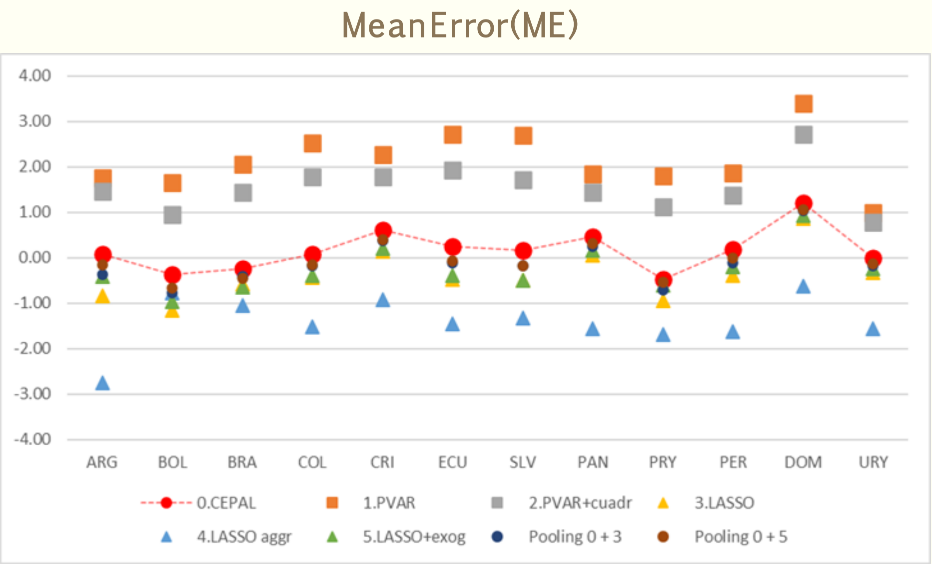
- For the poverty rate forecasts, the ECLAC projections achieve lower biases for Bolivia, Brazil and Uruguay. For the rest of the countries, improvements can be obtained with the model (3) estimated by LASSO or a combination of the ECLAC and LASSO projections. The results are more mixed when analyzing the biases of the extreme poverty rate forecasts.

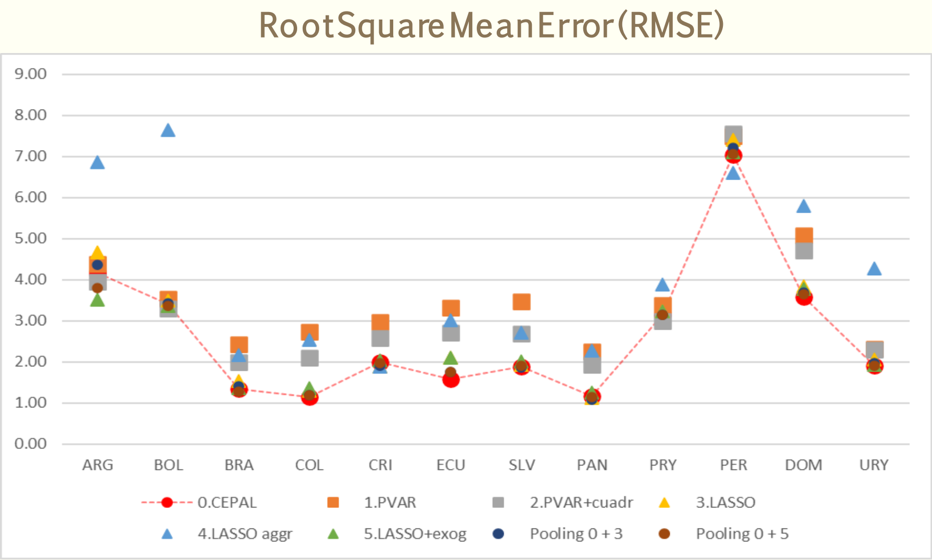
- No statistically significant differences were founded for RMSE and MAE between the ECLAC projections and the rest of the forecasting models, except for Argentina. For this country, an improvement in the performance of poverty rate forecasts is achieved when LASSO is used for the selection of predictors in models where exogenous macroeconomic variables are included.

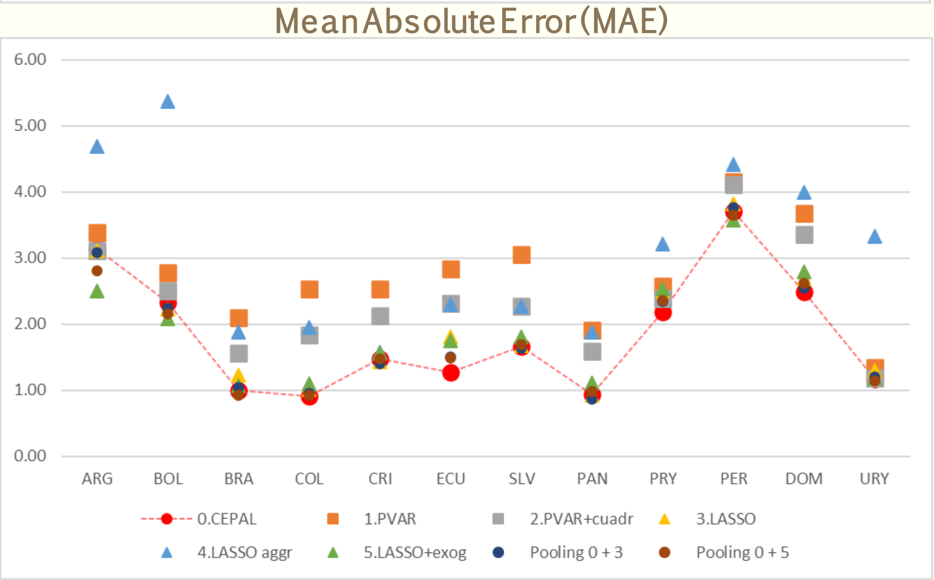
- By working with an aggregate absolute loss function based on the median, we obtain a measure that is more robust to the presence of outliers than when the mean is used. Although no statistically significant differences emerge except for Argentina, there is evidence of greater parity in relative performance between the ECLAC projections and the LASSO projections (or some combination of the two).

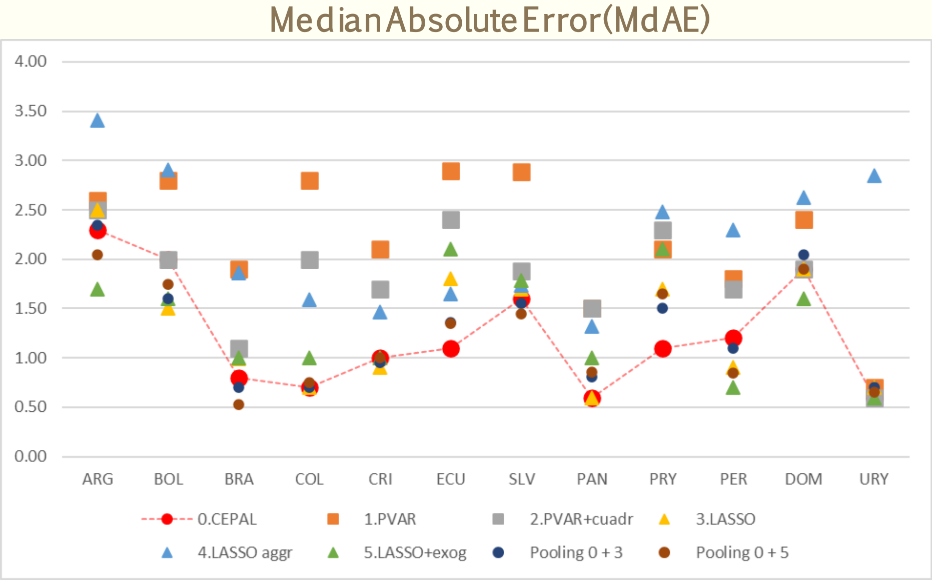
- In any case, we can see that working with the ECLAC projections in combination with the LASSO models would yield some predictive gain in 9 of the 12 countries, in the case of the poverty rate, and in 6 of the 12 countries, in the case of the extreme poverty rate.

RESULTS OF THE EVALUATION OF POVERTY PREDICTION MODELS.

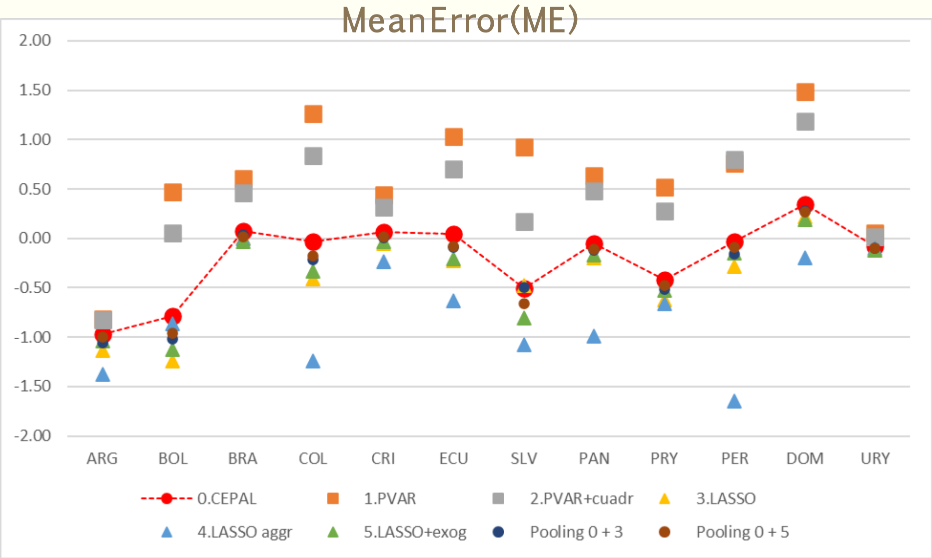


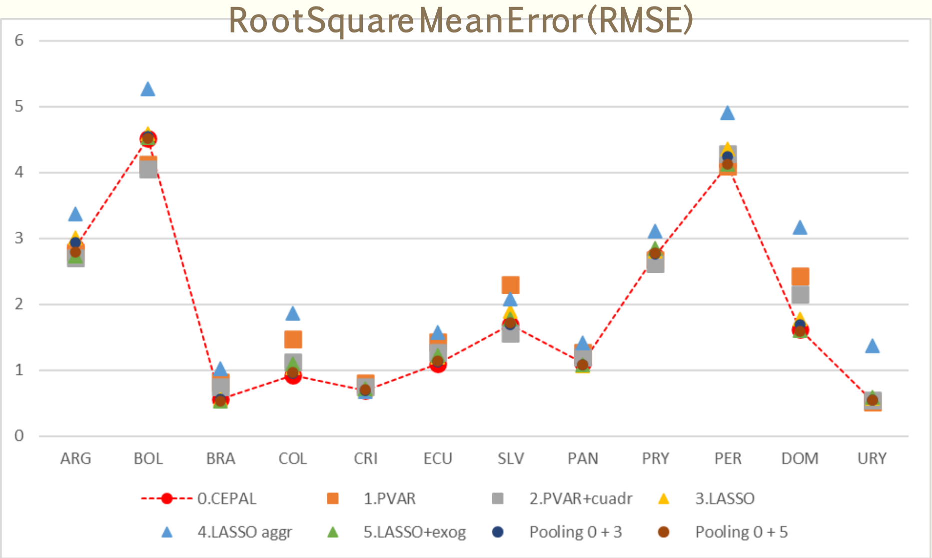


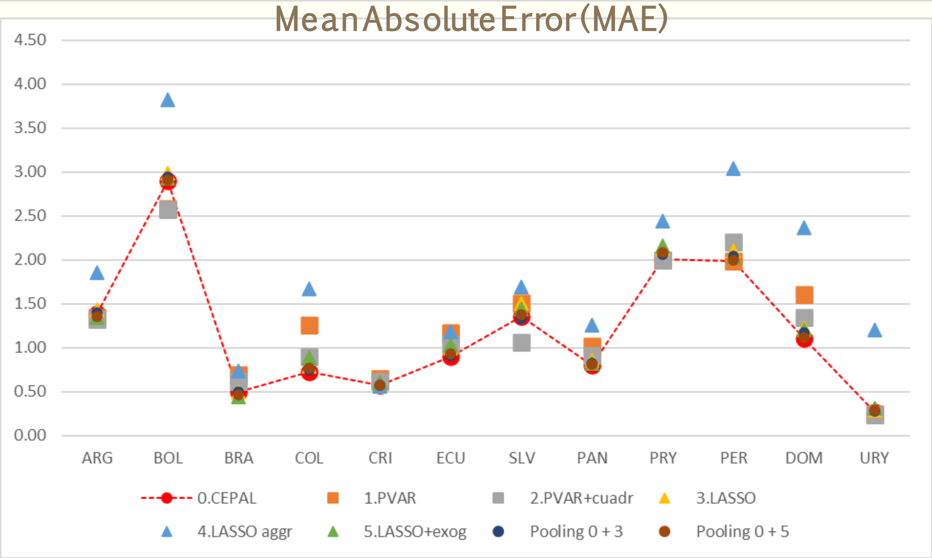


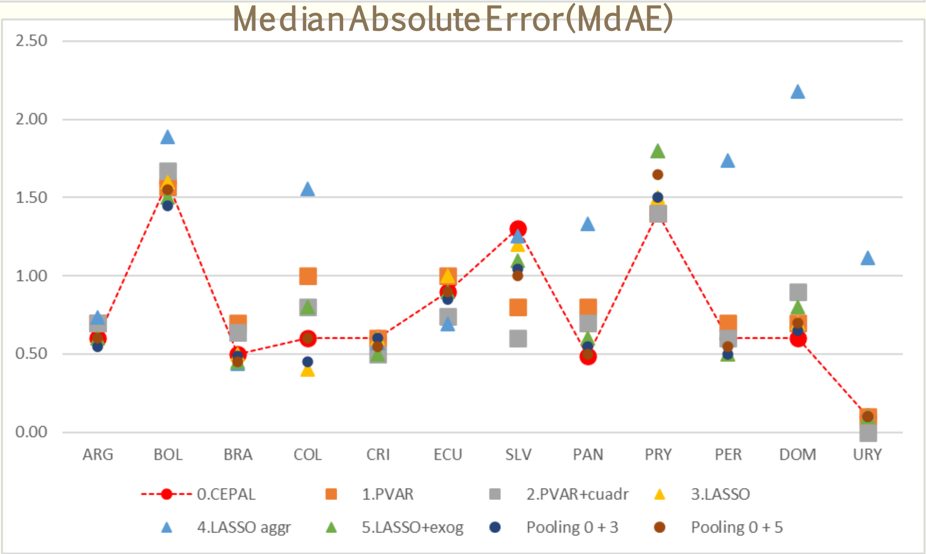


RESULTS OF THE EVALUATION OF EXTREME POVERTY PREDICTION MODELS.









A SUMMARY WITH THE BEST MODEL FOR EACH COUNTRY AND INDICATOR, INCLUDING STATISTICAL SIGNIFICANCE.

	Poverty[a]			
	ME	RMSE	MAE	MdAE
ARG	0.ECLAC	5.LASSO+exog[c]	5.LASSO+exog[c]	5.LASSO+exog
BOL	0.ECLAC[d]	2.PVAR+quadr	5.LASSO+exog	3.LASSO
BRA	0.ECLAC[b]	Pooling 0 + 5	Pooling 0 + 5	Pooling 0 + 5
COL	0.ECLAC	0.ECLAC	0.ECLAC	0.ECLAC, 3. LASSO, Pooling 0 + 3
CRI	3.LASSO[d]	4.LASSO aggr	Pooling 0 + 3	3.LASSO
ECU	Pooling 0 + 5[c]	0.ECLAC	0.ECLAC	0.ECLAC
SLV	0.ECLAC	Pooling 0 + 3	Pooling 0 + 3	Pooling 0 + 5
PAN	3.LASSO	Pooling 0 + 3	Pooling 0 + 3	0.ECLAC, 3. LASSO
PRY	0.ECLAC	2.PVAR+quadr	0.ECLAC	0.ECLAC
PER	Pooling 0 + 5[c]	4.LASSO aggr	5.LASSO+exog	5.LASSO+exog
DOM	4.LASSO aggr	0.ECLAC	0.ECLAC	5.LASSO+exog
URY	0.ECLAC[d]	Pooling 0 + 5	Pooling 0 + 5	0.ECLAC, 2. PVAR+quadr, 5. LASSO+exog

[a] The best model according to its performance for each country and indicator is included.
[b] $p<0.10$.
[c] $p<0.05$.
[d] $p<0.01$.

	Extreme Poverty[a]			
	ME	RMSE	MAE	MdAE
ARG	1.PVAR	2.PVAR+quadr	2.PVAR+quadr	Pooling 0 + 3
BOL	2.PVAR+quadr[d]	2.PVAR+quadr	2.PVAR+quadr	Pooling 0 + 3
BRA	3.LASSO[d]	Pooling 0 + 5	5.LASSO+exog	4.LASSO aggr
COL	0.ECLAC[c]	0.ECLAC	0.ECLAC	3.LASSO
CRI	Pooling 0 + 3[d]	4.LASSO aggr	4.LASSO aggr	2.PVAR+quadr
ECU	0.ECLAC	0.ECLAC	0.ECLAC	4.LASSO aggr
SLV	2.PVAR+quadr[d]	2.PVAR+quadr	2.PVAR+quadr	2.PVAR+quadr
PAN	0.ECLAC[b]	3.LASSO	0.ECLAC	0.ECLAC
PRY	2.PVAR+quadr[d]	2.PVAR+quadr	2.PVAR+quadr	0.ECLAC, 1.PVAR, 2.PVAR+quadr
PER	0.ECLAC[c]	1.PVAR	0.ECLAC	3.LASSO, 5.LASSO+exog, Pooling 0 + 3
DOM	5.LASSO+exog[b]	Pooling 0 + 5	0.ECLAC	0.ECLAC
URY	2.PVAR+quadr	1.PVAR	2.PVAR+quadr	2.PVAR+quadr

[a] The best model according to its performance for each country and indicator is included.
[b] $p<0.10$.
[c] $p<0.05$.
[d] $p<0.01$.

FINAL REMARKS.

- Although the results implemented by the ECLAC methodology to date could be considered qualitatively inferior to LASSO and PVAR, we conclude that, due to its simplicity and easy to implement and communicate, it remains a reasonable competitor.

- The purely "macro" approach generates predictions very similar to those of the "micro macro" approach, from which it follows that the "micro" stage is essentially a sort of "computational rule" for the calculation of poverty prediction. This suggests that the inclusion of more macro information in the dataset could help to improve predictive capacity.

- Attempts made to add purely macro information, however, resulted in only slight predictive improvements. This does not rule out the value of this type of information, if that which is integrated allows us to better capture the cyclical and episodic nature of poverty.

- The same methodology was applied to evaluate the prediction of poverty by subgroups of people according to their sex, age and activity status. For this, groups of people were built for each group made up of the combination of said variables, working with a total of 90 groups of individuals for each country and available year. Based on this, predictive improvements were obtained, especially in the disaggregation of poverty by age and by activity condition compared to ECLAC, using PVAR(1).

FINAL REMARKS.

- In the construction of the panel database, six countries were eliminated for which there were not enough temporary data on GDP per capita, Gini coefficient and poverty rates. For these countries, the variation rates of GDP per capita and the Gini were estimated using the models, obtaining predictive improvements for GDP with PVAR(n) + LASSO, which could improve the current poverty predictions made based on to the ECLAC methodology.

- Regarding the possibility of using more complex prediction models, it should be kept in mind that the use of more flexible methods, typical of the "machine learning" approach, is limited by the amount of information available, given the frequency and number of countries for which the forecast is made. In summary, for the set of information available and for the frequency and number of countries to forecast, there is little room to explore the route of complexity of predictive models.

- Other alternatives could include the use of the panel structure implicit in some household surveys, or at the pseudo-individual level. Additionally, "non-standard" information could be used by constructing other indicators that facilitate poverty monitoring, with a higher frequency or with a lower lag than the one naturally available to public statistical offices.

- In a parallel line of work, we are working to obtain poverty estimates disaggregated geographically and by population groups, based on the Small Area Estimation methodology, combining information from censuses, surveys and satellite images from Google Earth Engine, among other sources.

UNITED NATIONS

ECLAC