Predicting Poverty

Paolo Verme

August 2024

Premises

- The poverty rate remains one of the most important indicators for developing nations and accurate measurement is essential for achieving the sustainable development goals, international aid allocation and national social protection policies.
- Poverty measurement relies on survey data and monetary metrics such as income, consumption or expenditure that suffer from missing observations with the potential to bias this measurement significantly.
- The question of estimating statistics with sample surveys suffering from unit or item non-response has a long history in statistics and poverty specialists have adopted some of the solutions adopted in statistics and also proposed new solutions.

Missing observations

- Statisticians distinguish missing observations as: 1) Missing Completely At Random (MCAR); 2) Missing At Random (MAR) and, 3) Missing Not At Random (MNAR) - (Rubin 1986, 1996; Imbens and Lancaster, 1994)
- Money metrics such as income, consumption or expenditure typically suffer from MNAR observations
- Scientists have developed several methodologies to address missing data issues such as censoring, trimming, replacing, reweighting, single and multiple imputations, matching, and various machine learning methods
- Note that I refer to poverty *measurement* if measurement is based in censuses, poverty *estimations* if poverty is measured with sample surveys, and poverty *predictions* if poverty is measured with sample surveys containing missing observations

Poverty is Always Predicted: Some Examples

- Poverty profiles (Ex: World Bank Poverty Profiles)
- Targeting (Coady et Al., 2004)
- Poverty mapping (Elbers et Al., 2003, 2007; Tarozzi and Deaton, 2009))
- Cross-survey imputations (Dang et Al., 2019)
- Top and bottom income studies (Atkinson et al., 2011; Hlasny et al., 2021)

In all these cases poverty is measured with sample surveys containing non-random unit or item non-response - MNAR

Objective

- Based on the most recent poverty prediction literature, the objective of this paper is to conduct a laboratory experiment to compare the poverty prediction accuracy of classic econometric and machine learning methods in the presence of different types of missing data

Two Traditions - Similarities

- Social science tradition: **Regression Analysis (RA)**
- Comupter science tradition: **Machine Learning Analysis (MLA)**
- RA and MLA rely on the same statistical foundations and both traditions may use Frequentist or Bayesian statistics.
- Both traditions have been adapted to continuous and dichotomous dependent variable models
- These two traditions are converging. RA is used in most ML methods. Social scientists have, more recently, started to use ML methods

Two Traditions - Differences

- RA largely developed to address the question of causality. Great value is given to the understanding of the factors that explain good predictions. The end purpose is to devise policies that affect the factors that determine outcomes to improve outcomes. The focus is on predictors. *Ex: We want to know which teachers' training program is more effective in determining pupils learning.*
- MLA largely focused on improving prediction accuracy irrespective of whether the factors used for predictions cause outcomes. The end purpose is to come as close as possible to the true outcome. The focus is on outcomes. *Ex: We seek the best possible predictions of rice prices next week for budgeting purposes irrespective of what may determine rice prices.*

Baseline models

- **Dichotomous Dependent Variable models** where the dependent variable is poverty status (poor/non-poor). In this case, researchers a) Split the population in poor/non-poor groups using a poverty line; b) Predict the probability of being poor and c) Determine a probability thrsehold to assign predictions to poor/non-poor status.
- **Continuous Dependent Variable models** where the dependent variable is a monetary value of income, consumption or expendituure. In this case, researchers a) predict the monetary indicator of welfare and b) Adjust predictions to account for errors on the tails; c) Use a poverty line to split predicted observations into poor/non-poor observations.

Classic econometrics and machine learning models can be run in both settings, which provides a nice setting for comparisons

Predicting Poverty

Step 1 - Modeling

$$W_i = \alpha + \beta_1 X_i + \eta_i + \epsilon_i \quad (1)$$

$$P_i = \delta + \gamma_1 X_i + \nu_i + \psi_i \quad (2)$$

where i is the unit of observation (usually a household or an individual, household for short), $Wi =$ income, $P_i =$poor where P_i $=$1 if the unit is under the poverty line and $P_i = 0$ otherwise, X is a vector of household or individual characteristics, $\eta_i and \nu_i$ are random errors and ϵ_i and ψ_i are model fitting errors.

Step 2 - Prediction

$$\widehat{W}_i = \widehat{\beta}_1 X_i + \tilde{\eta}_i + \tilde{\epsilon}_i \tag{3}$$

$$\widehat{P}_i = \widehat{\gamma}_1 X_i + \tilde{\nu}_i + \tilde{\psi}_i \tag{4}$$

where $\widehat{W}_i$, $\widehat{P}_i$ are predicted welfare or poverty and $\tilde{\eta}_i, \tilde{\epsilon}_i, \tilde{\nu}_i, \tilde{\psi}_i$ are the estimated random and model fitting errors.

Step 3 - Classification

$$\begin{aligned} if \ \widehat{W_i} < z : i &= poor \\ else : \ i &= nonpoor \end{aligned} \tag{5}$$

$$\begin{aligned} if \ \widehat{P_i} > prob* : i &= poor \\ else : \ i &= nonpoor \end{aligned} \tag{6}$$

where z is the poverty line with $W_{min} \leq z \leq W_{max}$ and $prob*$ is an arbitrary probability cutpoint with $0 \leq p \leq 1$.

Confusion Matrix

All prediction methods result in a confusion matrix:

		Predicted Poverty	
		Non-Poor = 0	**Poor = 1**
True Poverty	**Non-poor = 0**	True Negative (TN) [1,1]	False Positive (FP) [1,2]
	Poor = 1	False Negative (FN) [2,1]	True Positive (TP) [2,2]

Note: [x,y] indicates row and column.

All prediction models can be estimated with continuous (welfare model) or dichotomous (poverty model) dependent variables.

Objective Functions

- The primary objective of any classification exercise is to maximize TP and TN and minimize FP and FN.
- Incorrect classifications result in errors Type I and Type II.
- There is a variety of objective functions:
  - FPR (Type I error)
  - FNR (Type II error)
  - True Positive Rate, sensitivity or recall (TPR=TP/(FN+TP)),
  - True Negative Rate or specificity (TNR=TN/(TN+FP)),
  - Precision (TP/(TP+FP))
  - False Discovery Rate (FP/(TP+FP)).
- All objective functions are based on the confusion matrix. The only difference is the weight they attribute to each cell of the matrix. This is a normative choice.

Objective Functions for Poverty Measurement

- Type I error refers to non-poor persons who are erroneously predicted as being poor. This error is also known as False Positive Rate (FPR), inclusion error or leakage rate and is defined as FP/(FP+TN).
- Type II error refers to persons who are poor but are erroneously predicted to be non-poor. This error is also known as False Negative Rate (FNR), exclusion error or undercoverage rate and is defined as FN/(FN+TP).
- In the case considered by this paper, the true poverty rate is known by design and models can be compared by testing the difference between the true and predicted poverty rate

Experiment

- We take a dataset of a middle income country with an exceptionally low non-response rate and reweight observations to clear the sample from any non-response issue. We consider this data set as a dummy data set clear of missing observations.
- We then generate from this data a series of new data sets featuring different types and size of missing data including MCAR, MAR, and MNAR patterns.
- We then compare the capacity of different poverty prediction models to predict poverty in the presence of these different types of missing observations
- This experiment allows to compare poverty predictions across models and type of missing data with the "true" poverty rate (the true counterfactual).

Data

- Morocco Consumption Survey, 2007
- Non-response rate of 2% corrected with Korinek et al (2007) correction method
- The outcome variable is household income per capita with only positive values and no missing observations
- The final data set contains 7,062 observations and 8 variables (gender, age, marital status, skills, employment status, employment sector, urban, and household size).

Objective Functions

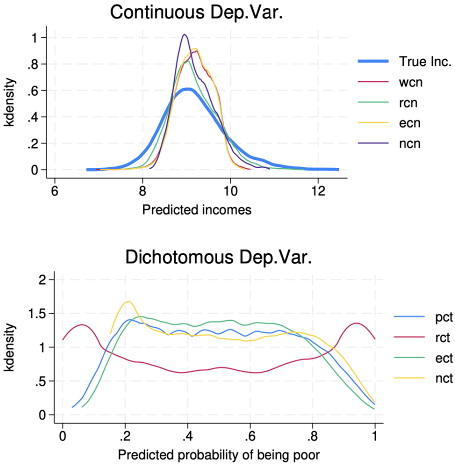
	wcn	wcn_r	rcn	rcn_r	ecn	ecn_r	ncn	ncn_r	pct	pct_r	rct	rct_r	ect	ect_r	nct	nct_r
Observations	7062	.	7062	.	7062	.	7062	.	7062	.	7062	.	7062	.	7062	.
TruePovRate	50	.	50	.	50	.	50	.	50	.	50	.	50	.	50	.
PredPoverty	43.09	.	50.06	.	43.16	.	49.63	.	49.11	.	49.92	.	49.45	.	50.48	.
Diff.(absmin)	6.91	8	.06	1	6.84	7	.37	3	.89	6	.08	2	.55	5	.48	4
Diff.(tstat)	10.61	.	-.12	.	10.51	.	.6	.	1.39	.	.2	.	.86	.	-.74	.
PrefTruePos(max)	67.83	8	83.08	2	67.94	7	72.95	3	70.76	4	87.66	1	70.68	5	70.58	6
PrefTrueNeg(max)	71.28	5	83.05	2	71.36	4	73.13	3	71.21	6	87.7	1	70.95	7	70.34	8
TruePos(max)	2212	8	2935	2	2218	7	2566	3	2475	6	3093	1	2481	5	2505	4
TrueNeg(max)	2700	4	2931	2	2701	3	2592	5	2538	6	3099	1	2520	7	2471	8
FalsePos(min)	831	4	600	2	830	3	939	5	993	6	432	1	1011	7	1060	8
FalseNeg(min)	1319	8	596	2	1313	7	965	3	1056	6	438	1	1050	5	1026	4
Leakage(min)	23.53	4	16.99	2	23.51	3	26.59	5	28.12	6	12.23	1	28.63	7	30.02	8
Undercoverage(min)	37.35	8	16.88	2	37.18	7	27.33	3	29.91	6	12.4	1	29.74	5	29.06	4
Sensitivity(max)	62.65	8	83.12	2	62.82	7	72.67	3	70.09	6	87.6	1	70.26	5	70.94	4
Specificity(max)	76.47	4	83.01	2	76.49	3	73.41	5	71.88	6	87.77	1	71.37	7	69.98	8
Precision(max)	72.69	5	83.03	2	72.77	4	73.21	3	71.37	6	87.74	1	71.05	7	70.27	8
Accuracy(max)	69.56	8	83.06	2	69.65	7	73.04	3	70.99	4	87.68	1	70.82	5	70.46	6

Legenda: wcn (Welfare - Continuous); rcn (Random Forest - Continuous); ecn (Elastic Net - Continuous); ncn (Neural Network - Continuous); pct (Poverty - Categorical); rct (Random Forest - Categorical); ect (Elastic Net - Categorical) and nct (Neural Network - Categorical). wcn-r refers to the rank position of the wcn model (horizontal ranking) with '1' indicating the top performing model and '8' the worse performing model. Similarly for other models. 'Diff' refers to the difference between the true and the predicted poverty rates. Leakage=FP/(FP+TN); Undercoverage=FN/(FN+TP); Sensitivity=TP/(FN+TP); Specificity=TN/(TN+FP); Precision=TP/(TP+FP); Accuracy=(TP+TN)/N. (min) (max) indicate whether the objective function is to be minimized or maximized. (absmin) indicates that the absolute value is to be minimized.

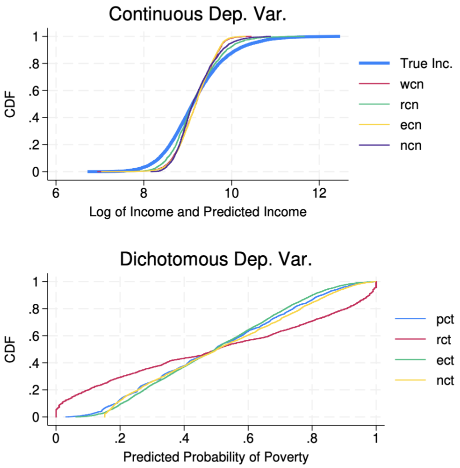
Missing Observations and Poverty Lines

	wcn	wcn_r	rcn	rcn_r	ecn	ecn_r	ncn	ncn_r	pct	pct_r	rct	rct_r	ect	ect_r	nct	nct_r
PovLine=5%	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MCAR95	4.8	1	4.8	2	4.8	3	4.7	7	4.8	4	4.8	5	4.8	6	4.7	8
MCAR75	4.1	2	4	3	4	4	3.7	6	3.8	5	4.2	1	3.7	7	3.7	8
MCAR50	2.9	2	2.7	4	2.8	3	2.5	6	2.6	5	3.2	1	2.5	7	2.5	8
MCAR25	1.9	2	1.6	4	1.8	3	1.2	7	1.5	5	2.7	1	1.3	6	1.2	8
MCAR5	1	1	.2	6	.9	3	1	2	.6	5	.7	4	.2	7	.2	8
MARpure	3.2	2	2.9	5	3.2	3	2.9	6	3	4	3.4	1	2.9	7	2.9	8
MAR_MNAR	3.9	2	3.9	3	3.9	4	3.9	5	3.9	6	4	1	3.9	7	3.9	8
MNARpure	4.2	2	4.2	3	4.2	4	4.2	5	4.2	6	4.3	1	4.2	7	4.2	8
Average	3.3	1.8	3	3.8	3.2	3.4	3	5.5	3	5	3.4	1.9	2.9	6.8	2.9	8
PovLine=25%	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MCAR95	24.5	3	24.6	2	24.4	5	23.7	8	24.4	6	24.8	1	24.4	7	24.5	4
MCAR75	21.9	3	22.3	2	21.7	4	18.7	7	21.7	5	23.4	1	21.4	6	18.7	8
MCAR50	18.2	3	20.1	2	17.6	5	13.5	8	17.6	6	21.9	1	16.7	7	18.1	4
MCAR25	15.2	3	16.7	2	14.2	5	6.2	8	14.6	4	20.5	1	12.9	6	12.6	7
MCAR5	10.5	3	5.6	7	9.4	4	8	6	11	2	13.6	1	4.3	8	8.3	5
MARpure	18.6	4	20.2	2	18.4	5	15.9	8	18.7	3	21.8	1	18	7	18.1	6
MAR_MNAR	19.7	4	20.1	2	19.7	5	19.9	3	19.7	6	20.8	1	19.7	7	19.7	8
MNARpure	21.4	3	21.8	2	21.4	4	21.3	8	21.4	5	22.2	1	21.4	6	21.4	7
Average	18.8	3.3	18.9	2.6	18.4	4.6	15.9	7	18.6	4.6	21.1	1	17.4	6.8	17.7	6.1
PovLine=50%	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MCAR95	49.6	6	49.9	1	49.6	7	47.5	8	49.9	2	50.1	3	49.9	4	49.9	5
MCAR75	47.8	6	49.4	4	47.8	7	37.5	8	49.6	1	49.6	2	49.5	3	49.2	5
MCAR50	45.7	6	48.8	4	45.5	7	29.5	8	49.2	3	49.7	1	49.7	2	47.3	5
MCAR25	43.6	6	47.8	5	43.6	7	12.5	8	49.3	2	50	1	49.1	3	50.9	4
MCAR5	44.3	6	49.2	2	44.9	5	35.1	8	52.1	4	50.2	1	51.8	3	44	7
MARpure	44.4	7	47.8	2	44.3	8	52.3	4	47.2	5	49.3	1	47	6	47.8	3
MAR_MNAR	42.3	6	43.3	2	42.2	7	41.7	8	43.1	4	44.5	1	43.2	3	43.1	5
MNARpure	44.4	6	45.5	2	44.4	7	42.6	8	44.5	4	46.2	1	44.5	5	44.8	3
Average	45.3	6.1	47.7	2.8	45.3	6.9	37.3	7.5	48.1	3.1	48.7	1.4	48.1	3.6	47.1	4.6
PovLine=75%	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MCAR95	75.3	4	75.1	3	75.4	5	76.2	8	75.4	6	75	1	75.5	7	75	2
MCAR75	76.6	3	76.3	2	76.7	4	81.2	8	77.1	5	75.9	1	77.6	7	77.1	6
MCAR50	78.4	3	77.5	2	78.7	4	87.5	8	79.5	5	76.7	1	80.7	7	79.7	6
MCAR25	80.8	3	79.5	2	81.3	5	93.7	8	82.7	6	78.7	1	84.7	7	81	4
MCAR5	89.7	4	87.9	3	90.8	5	91.7	7	87.4	2	84	1	91.1	6	98.7	8
MARpure	76.6	4	75.7	3	76.7	5	78.1	7	77.6	6	75.4	2	78.4	8	75.2	1
MAR_MNAR	71.8	3	71.3	7	72	2	69.2	8	71.4	4	71.4	5	72.5	1	71.4	6
MNARpure	71.1	8	71.3	5	71.3	6	78.7	7	72.8	3	71.8	4	73.7	1	73.5	2
Average	77.5	4	76.8	3.4	77.9	4.5	82	7.6	78	4.6	76.1	2	79.3	5.5	78.9	4.4

Predictions' Distributions



Predictions' Cumulative Distributions



Can OLS be Improved? Expanding Regressors

	wcn	wcn_r	rcn	rcn_r	ecn	ecn_r	ncn	ncn_r	pct	pct_r	rct	rct_r	ect	ect_r	nct	nct_r
Model1	43.2	7	50.1	1	43.2	8	48	6	49.2	5	50.1	2	49.4	3	50.6	4
Model2	43.2	6	50.1	1	43.2	7	43	8	49.2	4	50.3	2	49.4	3	54.5	5
Model3	42.3	8	51.9	3	43	7	45.4	6	50.7	1	53.5	5	51.1	2	52.5	4
Mobel4	39.6	2	39.6	3	39.6	4	38.4	8	39.6	5	39.6	6	39.6	7	41.8	1

Can OLS be Improved? OLS Error Adjusted

Poverty Line (%)	25	50	75
Continuous Dep. Var.			
OLS	12.1	43.2	82.1
OLS povimp (empirical)	25.0	47.4	73.2
Random Forest	16.6	49.7	81.0
Elastic Net	11.1	42.9	83.0
Neural Network	13.0	47.5	77.4
Categorical Dep. Var.			
Logit	11.4	51.9	84.8
Logit povimp (empirical)	25.5	51.0	75.7
Random Forest	20.2	50.9	79.7
Elastic Net	10.1	52.2	87.3
Neural Network	16.2	54.9	87.1

Can ML Models be Improved? Grid Search Parameters

	Grid range	Optimal parameters					
		Cont.	Cont.	Cont.	Cat.	Cat.	Cat.
Poverty Line (%)		*25*	*50*	*75*	*25*	*50*	*75*
Random Forest							
Iterations	50, 100, 200, 400	50	50	200	50	200	100
Number of Vars	1-12	6	9	9	6	3	11
Depth	3-8	8	6	5	7	7	6
Leaf size	5, 10, 50, 100	100	10	10	100	50	10
Elastic Net							
Alpha	0, 2, 4, 6, 8, 1	0	0.2	0	1	0.2	0.8
Lambda	50, 100, 200	50	50	50	100	100	50
Folds	5, 10, 20	5	10	5	5	5	5
Neural Network							
Layer 1	64, 128, 256	128	*128*	*256*	*64*	*128*	*64*
Layer 2	64, 128, 256	64	128	128	64	256	64
Learning Rate	.01, .001	0.01	0.001	0.01	0.01	0.001	0.001
Batch	20, 80	20	20	80	20	20	20
Epochs	50, 200	50	200	50.0	50	50	50

Can ML Models be Improved? Grid search results

Poverty Line (%)	Cont. *25*	Cont. *50*	Cont. *75*	Cat. *25*	Cat. *50*	Cat. *75*
Max						
Random forest	78.5	71.4	80.7	78.3	70.9	80.2
Elastic Net	77.9	69.5	79.5	78.1	70.4	79.2
Neural Network	78.2	70.9	80.2	78.6	71.0	80.3
Mean						
Random forest	77.7	69.9	79.0	77.7	69.9	78.2
Elastic Net	77.8	69.4	79.4	78.0	70.4	79.2
Neural Network	77.2	70.0	79.3	78.1	70.2	79.2
Std.Dev.						
Random forest	0.6	0.9	1.4	0.4	0.8	1.5
Elastic Net	0.1	0.0	0.0	0.0	0.0	0.1
Neural Network	0.6	0.6	0.5	0.5	0.9	1.3

Conclusions

- Ex-ante, it is not possible to know what the best prediction model is. With new data, it is important to test several models
- Prediction models can perform better or worse depending on the distribution of incomes and missing incomes, poverty line, and the objective function chosen
- With limited time and knowledge of ML models, random forest is the most accurate and flexible choice
- OLS error adjusted models used by cross-survey imputation specialists perform very well for estimating the poverty rate but do not estimate individual or household poverty
- With time and deep knowledge of ML models, any of the tested models with the exception of a simple OLS model can perform well