Predicting Poverty

Paolo Verme and Aivin Vicquierra Solatorio

December 21, 2020

Contents

Objective

The objective of this paper is to establish:

- Whether standard regression analysis is superior or inferior to machine learning methods for predicting the poverty rate;
- Whether any particular method within the regression analysis or machine learning traditions is consistently superior to others for predicting the poverty rate

Poverty Measurement

- Poverty is typically measured with monetary indicators such as income, consumption or expenditure
- Censuses of the population that measure monetary indicators are extremely rare
- Poverty is more typically estimated using household surveys covering monetary indicators.
- Estimates come, therefore, with a standard error.
- Poverty predictions are designed to approximate poverty estimates that would result from surveys.
- Predictions come, therefore, with standard errors that contain survey estimation errors and model prediction errors

Poverty Predictions

- Poverty predictions are used in a variety of contexts such as
  - **Targeting** the poor when information on beneficiaries' poverty is missing
  - **Poverty mapping** using censuses when one wants to predict poverty at levels of geographical disaggregation where surveys are not representative
  - **Replace missing information** in surveys when observation units or observations on incomes is missing or poorly measured
  - **Cross-survey predictions.** Predict poverty in surveys that do not contain monetary indicators using other surveys that do contain this information
- *In all these cases the objective is to minimize the error between model poverty predictions and survey poverty estimates*

Two Traditions - Similarities

- There are two established traditions in the science of predictions:
  - Social science tradition: **Regression Analysis (RA)**
  - Computer science tradition: **Machine Learning Analysis (MLA)**
- These two traditions are converging. Computer scientists use regression analysis as part of ML instruments. Social scientists have, more recently, started to use ML methods that may or may not use regression analysis.
- RA and MLA rely on the same statistical foundations and both traditions may use Frequentist or Bayesian statistics.

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.*

Regression Analysis

- Predicting poverty in the social sciences with Regression Analysis has taken two different approaches:
  - **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 threshold to assign predictions to poor/non-poor status.
  - **Continuous Dependent Variable models** where the dependent variable is a monetary value of income, consumption or expenditure. 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.

Machine Learning Analysis

- Predicting poverty in computer science with Machine Learning Analysis has taken a multitude of approaches. Some of the most popular include:
  - **Regression Trees and Random Forest**
  - **LASSO, RIDGE, Elastic Nets**
  - **Ensemble**
  - **Deep learning and Neural Networks**
  - **Stochastic Gradient Boosting**
  - **Nearest Neighbor**
- We focus on ML models popular among economists

Machine Learning Methods Popular Among Economists

Table: Machine Learning Models Used by Economists

Paper	Regression Trees and Random Forest	LASSO, RIDGE, Elastic Nets	Ensemble	Deep Learning and Neural Networks	Spike and Slab	Stochastic Gradient Boosting	K-nearest neighbor
Lee et al (2010)							
Varian (2014)	X	X			X		
Einav and Levin, (2014)	X	X					
Abelson et al. (2014)	X						
Kleinberg et al. (2015)		X					
Blumenstock et al. (2015)		X	X				
Chalfin et al. (2016)		X				X	
Donaldson and Storeygard (2016)							
McBride and Nichols (2016)	X	X		X			
Neal et al. (2016)							
Mullainathan and Spiess (2017)	X	X	X				
Fitzpatrick et. al (2018)	X	X		X		X	X
Athey and Imbens (2019)	X	X		X			

Confusion Matrix

Whether we work with RA or MLA, the objective is to minimize the difference between the true estimated poverty rate and the predicted poverty rate. We want to maximize the number of observations that are correctly predicted as poor (true positives). A confusion matrix helps to clarify the problem. Note that the poverty rate is estimated as: $P = \frac{TP}{TN+FP+FN+TP} = \frac{TP}{Population}$.

Table: True and Predicted Poverty Confusion Matrix

		Poverty Estimates	
		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.

Predicting poverty Vs. Targeting the Poor

- Predicting the poverty rate is anonymous. We can predict the poverty rate with 100% accuracy but include among poor households some that are not poor while including among non-poor households some that are poor (check the confusion matrix. One can vary TP and any of the other three indicators and obtain the same poverty rate).
- Predicting household poverty status for the purpose of targeting is non-anonymous. A 100% accuracy requires that all households that are poor are correctly classified as poor.
- Predicting poverty status for the purpose of targeting is, therefore, much harder.

Metrics 1

- To rank methods for poverty predictions we need to define the metric to use for comparing models.
- There is a wide variety of objective functions that can be used for this purpose depending on what we wish to maximize or minimize
- The simplest case is minimizing the difference between the true poverty rate and the predicted poverty rate:

$$min(P_i - \widehat{P_i}) \tag{1}$$

Metrics 2

Depending on the final objective, there is a variety of metrics that can be used for comparing models' performance. In addition to the poverty difference, in this paper, we use the following:

- chi2, chi2lr
- Precision
- Accuracy
- F2
- Spearman R

Data

- Morocco 2007 Household Consumption Survey
- 7,062 Observations
- Nationally representative
- Poverty line = 3,719 per household per year
- Average HH income = 56,887
- For the sake of this experiment, we use an arbitrary poverty line to have roughly half of the observations below the poverty line. Poverty rate = 49.5%

Models

We compare the performance of several models:

1. The welfare model with continuous dependent variable (wcn);
2. The poverty model with dichotomous dependent variable (pct);
3. A Random forest model with continuous dependent variable (rcn);
4. A Random forest model with dichotomous dependent variable (rct);
5. A LASSO model with continuous dependent variable (lcn);
6. A LASSO model with dichotomous dependent variable (lct).
7. A Gradient Boosted Trees model with continuous dependent variable (gct)
8. A Gradient Boosted Trees model with categorical dependent variable (gcn)

In this way, we can compare the performance of econometric models

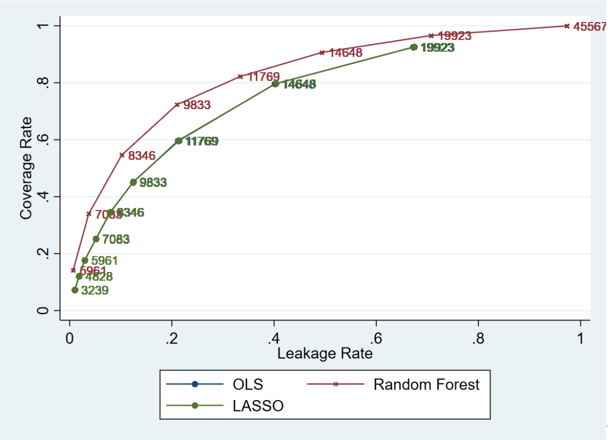
Models Comparison - Simple (Stata)

Table: Models Comparison - Simple (Stata)

	wcn	rcn	lcn	pct	rct	lct
TruePov	49.49	49.49	49.49	49.49	49.49	49.49
PredPov	41.72	49.46	41.66	48.46	51.85	48.34
Diff.	7.77	0.03	7.83	1.03	-2.36	1.15
Tstat. Diff.	11.82	0.07	11.92	1.6	-3.5	1.78
TN	2737	2737	2741	2539	2356	2541
FP	829	829	825	1027	1210	1025
FN	1379	832	1379	1101	1044	1107
TP	2117	2664	2117	2395	2452	2389
TPR=CR=Sens.	60.55	76.2	60.55	68.51	70.14	68.34
TNR=Spec.	76.75	76.75	76.86	71.2	66.07	71.26
FPR=IE=LR	23.25	23.25	23.14	28.8	33.93	28.74
FNR=ER=UR	39.45	23.8	39.45	31.49	29.86	31.66
chi2	1010.57	1980.31	1017.05	1114.38	926.91	1108.07
chi2lr	1038.45	2085.04	1045.34	1145.65	948.67	1138.97
Precision	71.86	76.27	71.96	69.99	66.96	69.98
Accuracy	68.73	76.48	68.79	69.87	68.08	69.81
F2	62.52	76.21	62.54	68.8	69.48	68.66

Stochastic Dominance (Continuous Models)

Figure: Models Comparison (Continuous Dep. Var.)

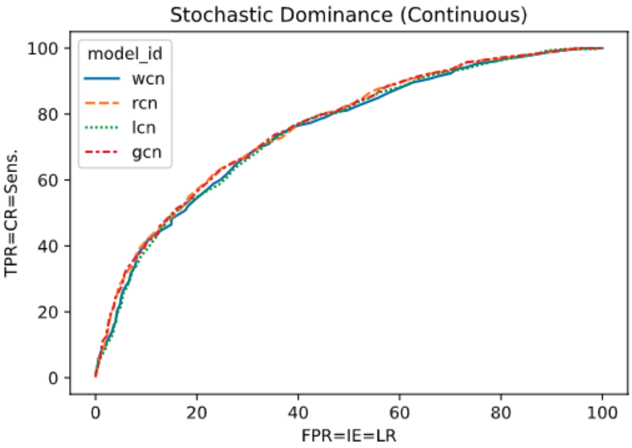


Models Comparison - Advanced (Python)

Table: Models Comparison - Advanced (Python)

	wcn	rcn	lcn	gcn	pct	rct	lct	gct
TruePov	49.49	49.49	49.49	49.49	49.49	49.49	49.49	49.49
PredPov	41.56	51.25	40.15	48.64	48.19	51.47	48.30	52.32
Diff.	7.93	-1.76	9.34	0.85	1.3	-1.98	1.19	-2.83
Tstat. Diff.	5.92	-1.31	6.97	0.64	0.98	-1.47	0.90	-2.13
TN	677	597	687	622	630	592	629	591
FP	215	295	205	270	262	300	263	301
FN	355	264	370	285	285	265	284	251
TP	519	610	504	589	589	609	590	623
TPR=CR=Sens.	59.38	69.79	57.67	67.39	67.39	69.68	67.51	71.28
TNR=Spec.	75.90	66.93	77.02	69.73	70.63	66.37	70.52	66.26
FPR=IE=LR	24.10	33.07	22.98	30.27	29.37	33.63	29.48	33.74
FNR=ER=UR	40.62	30.21	42.33	32.61	32.61	30.32	32.49	28.72
chi2	224.77	236.81	219.56	242.04	254.06	228.21	254.05	247.84
Precision	70.71	67.40	71.09	68.57	69.21	67.00	69.17	67.42
Accuracy	67.72	68.35	67.44	68.57	69.03	68.01	69.03	68.74
F2	61.35	69.30	59.93	67.62	67.75	69.13	67.83	70.48
SpearmanR	0.53	0.54	0.53	0.55	0.53	0.54	0.53	0.55

Stochastic dominance



Stochastic dominance - Continuous Models 1

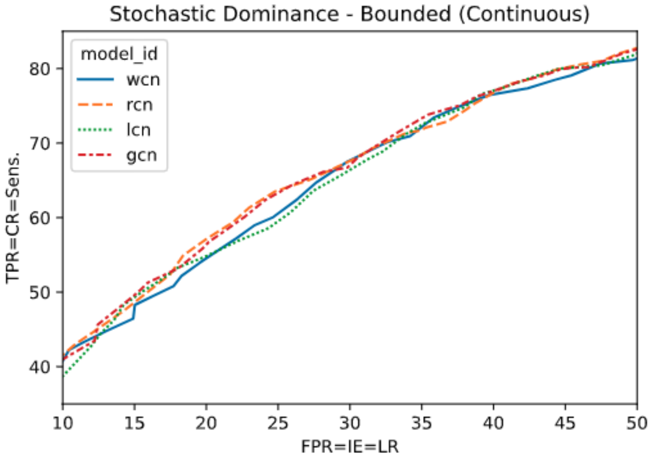
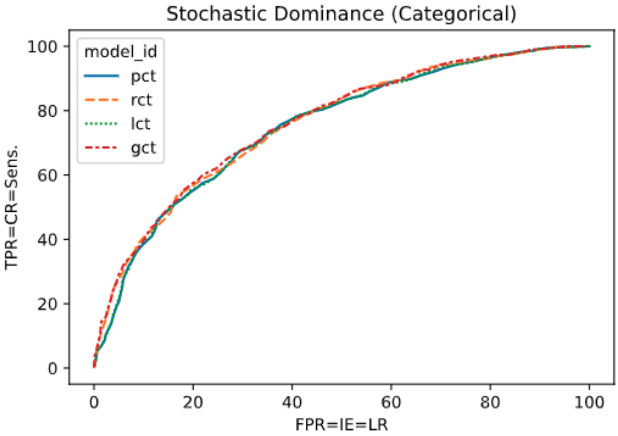


Figure: Stochastic dominance curves for continuous models

Stochastic dominance - Continuous Models 2



Stochastic dominance - Categorical Models 1

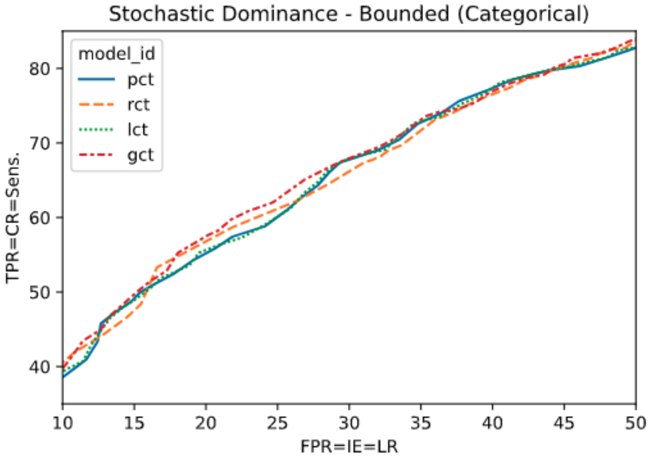


Figure: Stochastic dominance curves for categorical models

Stochastic dominance - Categorical Models 2

Table: Stochastic dominance curves area matrix

	gct	gcn	rcn	rct	lct	pct	lcn	wcn
gct	0.00	15.55	19.51	44.51	106.31	111.38	124.24	129.97
gcn	-15.55	0.00	3.98	28.89	90.54	95.51	108.58	114.67
rcn	-19.51	-3.98	0.00	24.99	86.74	91.75	104.74	110.72
rct	-44.51	-28.89	-24.99	0.00	61.93	66.81	79.91	85.83
lct	-106.31	-90.54	-86.74	-61.93	0.00	5.00	18.02	23.88
pct	-111.38	-95.51	-91.75	-66.81	-5.00	0.00	13.05	18.97
lcn	-124.24	-108.58	-104.74	-79.91	-18.02	-13.05	0.00	6.06
wcn	-129.97	-114.67	-110.72	-85.83	-23.88	-18.97	-6.06	0.00

Stochastic dominance

Table: Area between stochastic dominance curves of two models

Stochastic dominance

	gct	gcn	rcn	rct	lct	pct	lcn	wcn
gct		16	20	45	106	111	124	130
gcn			4	29	91	96	109	115
rcn				25	87	92	105	111
rct					62	67	80	86
lct						5	18	24
pct							13	19
lcn								6
wcn								

Note that the larger the area between two models, the more different the two in terms of general performance.

Conclusion - Research

- It is not obvious that categorical dependent variable models are superior or inferior to continuous dependent variable models
- It is not obvious that regression analysis models are superior or inferior to machine learning models
- Models ranking can change depending on the distribution of income and the poverty line
- When using new data, it is essential to test different models before opting for any particular model

Conclusion - Policy

- It is important to define the objective optimizing function for models' comparison based on the specific objectives that policy makers have.
- It is important to choose metrics for comparing models that take into account the objectives of the policy maker and the optimizing function considered.
- The optimizing function and the choice of metric are also influenced by normative criteria. Policy makers may have a preference for budget savings or for maximising the coverage of the poor. Each coverage/leakage trade-off has a different impact on budgets and budgets are important considerations of policy makers. Knowledge of these normative preferences are important for the selection of the right optimizing function, model and metric.