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 <u>causality</u>.
 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 <u>accuracy</u> 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 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 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



Step 1 - Modeling

$$W_i = \alpha + \beta_1 X_i + \eta_i + \epsilon_i \tag{1}$$

$$P_i = \delta + \gamma_1 X_i + \nu_i + \psi_i \tag{2}$$

where i is the unit of observation (usually a household or an individual, household for short), Wi = income, $P_i = \text{poor}$ 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, η_i and ν_i are random errors and ϵ_i and ψ_i are model fitting errors.

Step 2 - Prediction

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

$$\widehat{P}_i = \widehat{\gamma_1} X_i + \widetilde{\nu}_i + \widetilde{\psi}_i \tag{4}$$

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

Step 3 - Classification

if
$$\widehat{W}_i < z : i = poor$$

else: $i = nonpoor$ (5)

if
$$\widehat{P}_i > prob* : i = poor$$

else : $i = nonpoor$ (6)

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

Confusion Matrix

All prediction methods result in a confusion matrix:

Predicted Poverty

True Poverty

	Non-Poor = 0	Poor = 1
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

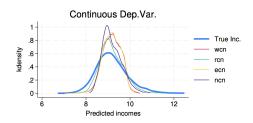
	wen	wcn_r	ren	rcn_r	ecn	ecn_r	ncn	ncn_r	pct	pct_r	rct	rct_r	ect	ect_r	nct	nct_r
Observations	7062		7062		7062		7062		7062		7062		7062		7062	
TruePovRate	50		50		50		50		50		50		50		50	
PredPoverty	43.09		50.06		43.16		49.63		49.11		49.92		49.45		50.48	
Diff.(absmin)	6.91	8	.06	1	6.84	7	.37	3	.89	6	.08	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: wen (Welfare - Continuous); ren (Random Forest - Continuous); ren (Elastic Net - Continuous); nen (Neural Network - Continuous); pet (Poverty - Categorical); ret (Random Forest - Categorical); etc (Elastic Net - Categorical) and net (Neural Network - Categorical). wen-r refers to the rank position of the wen 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+TP); Undercoverage=FN/(FN+TP); Sensitivity=FP/(FN+TP); Specificity=TN/(TN+FP); Precision=TP/(TP+FP). Accuracy=(TP-TN)N/N. (min) (max) indicates that the absolute value is to be minimized.

Missing Observations and Poverty Lines

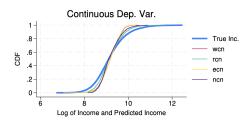
MNARpure 44.4 6 45.5 2 44.4 7 42.6 8 44.5 4 46.2 1 44.5 5 44.8 Average 45.3 6.1 47.2 28 45.3 6.9 37.3 75 48.1 31. 48.7 14.4 45.2 1 44.5 5 44.8 Average 45.3 6.1 47.5 28 45.3 6.9 37.3 75 48.1 31. 48.7 14.4 45.2 1 5 45.3 5 47.5 48.1 31. 48.7 14.4 45.2 1 5 45.3 5 47.5 48.1 31. 48.7 14.5 47.5 47.5 47.5 47.5 47.5 47.5 47.5 4		wen	wcn_r	ren	rcn_r	ecn	ecn_r	ncn	ncn_r	pct	pct_r	rct	rct_r	ect	ect_r	nct	nct_
MCARF5 4.1 2 4 8 3 4 4 3.7 6 3.8 5 4.2 1 3.7 7 3.7 MCARF5 MCARF5 19 2 1 6.7 4 1.8 3 12.5 7 1.3 1.5 1.5 1.2 1.3 1.5 1.5 1.2 1.3 1.5 1.2 1.3 1.5 1.2 1.3 1.5 1.2 1.3 1.5 1.2 1.3 1.5 1.2 1.3 1.5 1.2 1.3 1.5 1.2 1.3 1.5 1.2 1.3 1.5 1.2 1.3 1.5 1.2 1.3 1.5 1.2 1.3 1.5 1.2 1.3 1.5 1.2 1.3 1.5 1.2 1.3 1.5 1.2 1.3 1.5 1.3 1.5 1.3 1.5 1.3 1.3 1.3 1.2 1.3 1.3 1.3 1.3 1.3 1.3 1.3 1.3 1.3 1.3																	
MCAR60 29 2 27 7 4 2.8 3 2.5 6 2.6 5 3.2 1 2.5 7 2.5 MCAR62 1 1 2.6 4 1.8 3 1.2 7 1.5 5 5 2.7 1 1.3 6 1.2 MCAR63 1 1 2 6 6 9 3 1 1 2 6 6 5 7.7 4 2 7 2 MAP MAP MAR																	8
MCARES 1.9 2 1.6 4 1.8 3 1.2 7 1.5 5 2.7 1 1.3 6 1.2 MCARES MARDMER 3.2 2 2.9 6 5.3 1 2 2 6 6 5.0 4 3.4 1.2 29 7 2.9 MARDMER 3.2 2 2.9 5 3.2 3 2.9 6 3 3 4 3.4 1.2 29 7 2.9 MARDMER 4.2 2 4.2 3 3 4.5 3 5.5 3 5 5 3 5 6 4.1 1 3.9 7 3.9 6 MARDMAR 4.2 3 3 1.8 3 8 3.2 3 4.2 4 4 4.2 5 5 4.2 6 4.3 1 4.2 9 6 8 2.9 PoLine-25%													1	3.7			8
MCARS	CAR50				4		3		6		5		1	2.5	7		8
MARDMAR 32 2 29 9 5 32 3 29 6 3 4 3.4 1 2.9 7 2.9 MARDMAR MAR MAR MAR 42 2 4 4 2 5 5 42 6 4.1 1 3.9 7 3.9 MARDMAR 42 2 4 2 4 4 2 5 5 42 6 4.3 1 4.2 7 6 8.2 7 4.2 Average 3.3 1.8 3 8 32 3.4 3 5 5 5 3 5 6 4 1 3 4.9 7 6 8.2 7 4.2 Average 3.4 1 2 1.4 7 24.5 4 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1	CAR25	1.9	2	1.6	4	1.8	3	1.2	7	1.5	5	2.7	1	1.3	6	1.2	8
MARMAR 39 2 2 39 3 3 3.9 4 3.9 5 3.9 6 4 1 1 3.9 7 3.9 MARMARpure 31 8 3.1 8 3 3.8 3.2 3.4 3.2 5 4 3.2 5 4.2 6 4.3 1 4.2 7 4.2 5 MARAByre 4 4 5 2 2 4.4 6 5 2.3 4.4 6 2 4.8 1 2 4.4 7 24.5 MARAByre 5 4.5 6 4.7 8 4 8 4 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	CAR5								2	.6	5		4		7		8
MARDMRUR 42 2 42 3 42 3 42 4 42 5 44 27 6 4.3 1 4.2 7 4.2 PMARDMRUR 42 3 6 4.3 3 1.8 3.8 3.8 3.2 3.4 3 5 5 3 5 3 5 3 4.1 4.2 7 4.2 PMARDMRUR 42 3 6 4.3 1 4.2 7 4.2 PMARDMRUR 42 7 42.5 PMARDMRUR 44 7 47.5 PMARDMRUR 44 7 44.5 PM	ARpure	3.2	2	2.9	5	3.2	3	2.9	6	3	4	3.4	1	2.9	7	2.9	8
Average 3.3 1.8 3 3.8 3.2 3.4 3 5.5 3.5 5.3 5.9 3.4 1.9 2.9 6.8 2.9 MCAR75 24.5 3 24.6 2 24.4 5 23.7 8 24.4 6 24.8 1 24.4 7 24.5 MCAR76 24.5 3 23.2 2 21.7 4 18.7 7 21.7 5 23.4 1 24.4 7 24.5 MCAR80 18.2 3 20.1 2 17.6 5 13.5 8 17.6 6 21.9 1 16.7 7 18.1 MCAR80 18.2 3 20.1 2 17.6 5 13.5 8 17.6 6 21.9 1 16.7 7 18.1 MCAR80 18.2 3 20.1 2 17.6 5 62.8 14.6 4 20.5 1 12.9 6 12.6 MCAR81 10.5 3 5.6 7 9.4 4 8 8 14.6 4 20.5 1 12.9 6 12.6 MCAR81 10.5 3 5.6 7 9.4 4 8 8 18.7 3 21.8 1 19.7 7 19.7 MARAPure 18.6 3 21.8 2 21.4 4 21.3 8 18.7 3 21.8 1 19.7 7 19.7 MANAIpure 21.8 3 31.8 2 21.4 4 21.3 8 21.7 18.6 40.5 21.1 19.7 7 19.7 MCAR96 49.6 6 49.9 1 40.6 7 47.5 8 49.9 2 50.1 3 49.9 4 49.9 MCAR75 47.6 68.4 44.4 47.8 7 37.5 8 49.9 2 50.1 3 49.9 4 49.9 MCAR75 43.6 6 49.9 1 40.8 7 21.5 8 49.9 2 50.1 3 49.9 4 49.9 MCAR75 43.6 6 47.8 6 47.8 7 37.5 8 49.9 2 50.1 3 49.9 4 49.9 MCAR75 43.6 6 47.8 5 43.5 7 25.5 8 49.9 2 50.1 3 49.9 4 49.9 MCAR75 43.6 6 47.8 5 43.5 7 25.5 8 49.9 2 50.1 3 49.9 4 49.9 MCAR75 43.6 6 47.8 5 43.5 7 25.5 8 49.9 2 50.1 3 49.9 4 49.9 MCAR75 43.6 43.7 7 7 7 7 7 7 7 7 7	AR_MNAR	3.9	2	3.9	3	3.9	4	3.9	5	3.9	6	4	1	3.9	7	3.9	8
PorLine—25% MCAR75 245 3 246 2 244 5 23.7 8 244 6 248 1 244 7 245 MCAR75 219 3 223 2 217 4 18.7 7 21.7 5 23.4 1 21.4 6 18.7 MCAR76 129 3 223 2 217 4 18.7 7 21.7 5 23.4 1 21.4 6 18.7 MCAR75 15.2 3 16.7 2 14.2 5 6 2 8 14.6 4 20.5 1 12.9 6 12.6 MCAR75 15.2 3 16.7 2 14.2 5 6 2 8 14.6 4 20.5 1 12.9 6 12.6 MCAR75 15.2 3 16.7 2 14.2 5 6 2 8 14.6 4 20.5 1 12.9 6 12.6 MCAR85 15.2 3 16.7 2 18.4 5 11.5 9 8 18.7 3 21.8 1 18. 7 18.1 MAR MAR MAR MAR 18 3 3 12.8 2 2 14. 4 2 13. 8 2 14. 5 2 22.1 1 21.4 6 21.4 MAR DAR 18 8 3 3 18.9 2 6 18.4 6 15 12.0 8 2 11. 1 18. 7 18.1 MCAR25 16.8 3 3 18.9 2 6 18.4 6 15 12.0 8 2 11. 1 17. 7 19.7 MCAR95 49.6 6 49.9 1 49.6 15 40.6 15 2 2 2 1 2 1.4 6 21.4 MCAR5 41.8 3 3 18.8 4 45.5 7 25.5 8 49.9 2 5 11. 1 18. 7 49.9 MCAR75 47.8 6 49.9 1 49.6 7 47.5 8 49.9 2 5 10.1 3 49.9 4 49.9 MCAR75 48 6 49.4 4 47.8 7 37.5 8 49.9 2 5 10.1 3 49.9 4 49.9 MCAR75 48 6 49.8 4 47.8 7 37.5 8 49.9 2 5 10.1 3 49.9 4 49.9 MCAR75 48 6 49.8 4 47.8 7 25.8 49.9 2 5 10.1 3 49.9 4 49.9 MCAR5 41.3 6 4.3 6 4.3 6 4.5 7 25.5 8 49.9 2 5 10.1 3 49.9 4 49.9 MCAR75 48 6 48.3 6 4 47.8 7 3 5 45.6 4 4 4 5 5 5 2 2 2 1 2 1.4 6 5 12.4 MCAR95 41.3 6 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 4 5	NARpure	4.2	2	4.2	3	4.2	4	4.2	5	4.2	6	4.3	1	4.2	7	4.2	8
MCARGS 245 3 246 2 244 5 237 8 246 6 248 1 244 7 245 MCARGA 129 3 221 2 217 6 5 13.5 8 17 24.6 6 21.9 1 16.7 7 18.1 MCARGO 182 3 20.1 2 17.6 5 13.5 8 17.6 6 21.9 1 16.7 7 18.1 MCARGO 182 3 20.1 2 17.6 5 13.5 8 17.6 6 21.9 1 16.7 7 18.1 MCARGA 10.5 3 16.6 7 9.4 4 8 8 6 11 2 13.6 1 4.3 8 8.3 MCARGA 10.5 3 5.6 7 9.4 4 8 8 6 11 2 13.6 1 4.3 8 8.3 MARDHER 18.6 4 20.1 2 19.7 5 19.9 8 18.7 3 18.1 1 2 13.6 1 4.3 8 8.3 MARAMNAR 19.7 4 20.1 2 19.7 5 19.9 8 18.7 3 18.1 1 18.7 18.1 MARAMNAR 19.7 4 20.1 2 19.7 5 19.9 8 18.7 3 11.8 1 11.7 7 19.7 MCARGA 18.8 3.3 18.9 2.6 18.4 4.6 15.9 7 18.6 4.5 5 22.2 1 21.4 6 18.7 MCARGA 49.6 6 49.9 1 4.9 4.6 15.9 7 18.6 4.5 12.1 1 17.4 6.8 12.7 MCARGA 49.6 6 49.9 1 4.9 6 7 47.5 8 49.9 2 50.1 3 49.9 4 49.9 MCARTS 47.8 6 49.4 4 47.8 7 37.5 8 49.9 2 50.1 3 49.9 4 49.9 MCARTS 48.6 6 49.9 1 4.6 7 7 47.5 8 49.9 2 50.1 3 49.9 4 49.9 MCARTS 48.6 6 49.9 1 4.6 7 7 47.5 8 49.9 1 49.0 1 49.6 2 49.5 3 49.2 MCARTS 48.6 6 49.8 4 45.5 7 29.5 8 49.2 3 50.1 3 49.9 4 49.9 MCARTS 48.6 6 49.8 2 44.3 8 52.3 4 47.2 5 49.3 1 47.0 2 47.3 3 49.9 4 49.9 MCARTS 48.3 6 47.7 2.8 45.3 69.3 7.5 8 49.2 3 44.5 1 45.2 2 45.3 3 45.1 MARAMNAR 48.4 45.5 2 44.3 8 52.3 4 47.2 5 49.3 1 47.6 2 47.8 MARAMNAR 49.6 6 45.5 2 44.4 7 42.6 8 45.1 44.5 1 44.5 1 43.2 3 43.1 MXARDHER 44.4 6 45.5 2 44.4 7 42.6 8 43.1 44.5 1 43.2 3 43.1 MXARDHER 45.3 6.1 47.7 2.8 45.3 69.9 37.3 7.5 8 8.7 4 6 75 1 75.5 7 75.6 MCARTS 75.3 4 75.1 3 75.4 5 76.2 8 77.4 5 8 77.5 8 77.5 75.7 75.7 77.1 MCARGA 57.4 3 75.1 3 75.4 5 76.2 8 77.4 75.8 75.7 75.7 75.7 75.7 75.7 75.7 75.7	verage	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
MCAR95 24.5 3 24.6 2 24.4 5 23.7 8 24.4 6 24.8 1 24.4 7 24.5 MCAR50 18.2 3 20.1 2 17.6 5 13.5 8 17.6 6 21.9 1 16.7 7 18.1 MCAR50 18.2 3 20.1 2 17.6 5 13.5 8 17.6 6 21.9 1 16.7 7 18.1 MCAR50 18.2 3 20.1 2 17.6 5 62 8 14.6 4 20.5 1 29.6 12.6 MCAR51 10.5 3 56. 7 9.4 4 8 6 11. 2 13.6 1 43.3 8 8.3 MARMARD 18.6 4 20.1 2 19.7 5 19.9 8 18.7 3 18.1 1 18.7 18.1 MAR MAR 19.7 4 20.1 2 19.7 5 19.9 8 18.7 3 21.8 1 19.7 7 19.7 MARARJure 18.8 3.3 18.9 2.6 18.4 4.6 15.9 7 18.6 4.6 21.2 1 10.7 7 19.7 MCAR96 49.6 6 49.9 1 49.6 7 47.5 8 49.9 2 22.1 1 17.4 6.8 17.7 MCAR96 45.7 6 48.8 4 45.5 7 27.5 8 49.9 2 50.1 3 49.9 4 49.9 MCART5 47.8 6 47.8 7 7 27.5 8 49.2 3 3 49.7 4 49.9 MCART5 47.8 6 47.8 7 27.5 8 49.2 3 40.7 14.7 2 47.5 MCAR96 45.7 6 48.8 4 45.5 7 29.5 8 49.2 3 40.7 14.7 2 47.5 MCARD6 44.3 6 49.2 2 44.3 8 52.3 4 47.2 5 40.3 1 47.7 2 47.3 MARARMAR 44.5 6 47.8 7 37.5 8 49.2 3 40.7 1 47.7 2 47.3 MARARMAR 45.6 49.2 2 44.3 8 52.3 4 47.2 5 40.3 1 47.7 2 47.3 MARARMAR 45.5 6 47.7 2 47.5 6 37.3 7.5 8 43.1 4 4 4 5 7 4 MARARMAR 45.6 45.5 2 44.4 7 42.6 8 44.5 4 45.5 7 47.5 MCAR5 57.3 4 75.1 2 8 56.9 37.3 7.5 8 37.5 8 48.1 4 4 5 7 7 7 7 MCAR5 57.8 4 7 7 2 45.6 9 37.3 7.5 8 37.5 8 48.1 4 4 5 7 7 7 7 7 7 7 7 7	vLine=25%																
MCARGO 18.2 3 20.1 2 17.6 5 13.5 8 17.6 6 21.9 1 16.7 7 18.1 MCARGO 18.2 3 20.1 2 17.6 5 62.8 14.6 4 20.5 1 12.9 6 12.6 MCARS 10.5 3 5.6 7 9.4 4 8 8 6 11.1 2 13.6 1 43.8 8.3 MARpure 18.6 4 20.1 2 13.6 1 43.8 8.3 MAR MAR 18.4 8 19.7 4 20.1 2 13.6 1 43.8 8.3 MAR MAR 19.7 4 2 20.1 19.7 5 19.7 MAR 19.7 MAR 19.7 MAR 19.7 19.7 MAR 19.7 M	CAR95	24.5		24.6	2	24.4	5	23.7	8	24.4	6	24.8	1	24.4	7	24.5	4
MCARS 15.2 3 16.7 2 14.2 5 6.2 8 14.6 4 20.5 1 12.9 6 12.6 MCARS 10.5 3 5.6 7 9.4 4 8 6 11 2 3.6 1 1.8 7 8 MARDWR 18.6 4 20.2 2 18.4 5 15.9 8 18.7 3 21.8 1 18 7 7 19.7 MAR MNAR 19.7 4 20.2 2 18.4 5 15.9 8 18.7 3 21.8 1 18 7 7 19.7 MAR MNAR 18.8 3.3 18.8 2 21.4 4 21.3 8 21.4 5 22.2 1 21.4 6 21.4 MARDWR 21.4 3 21.8 2 21.4 4 21.3 8 21.4 5 22.2 1 21.4 6 21.4 MCARS 49.6 6 49.9 1 49.6 7 47.5 8 49.9 2 20.1 1 17 6 17.7 MCARDS 49.6 6 49.9 1 49.8 7 27.5 8 49.9 2 20.1 3 40.2 MCARTS 47.8 6 49.8 4 47.8 7 37.5 8 49.9 2 50.1 3 49.5 4 MCARTS 43.6 6 47.8 6 43.6 7 21.5 8 49.9 2 50.1 3 49.5 4 MCARS 44.3 6 47.8 6 47.8 7 27.5 8 49.9 2 50.1 3 40.2 MCARS 43.6 6 47.8 6 43.6 7 21.5 8 49.9 2 50.1 3 40.2 MARDWR 44.3 6 47.8 6 47.8 6 5 51.3 8 49.9 2 50.2 1 40.1 3 50.9 MAR MNAR 44.8 6 45.5 2 44.3 8 52.1 4 4 50.2 1 4 4 5 2 4 MARDWR 44.4 6 45.5 2 44.4 7 42.6 8 52.1 4 4 5 2 2 4 3 4 MARDWR 44.4 6 45.5 2 44.4 7 42.6 8 52.1 4 4 4 5 2 2 4 4 MARDWR 44.4 7 28 45.8 69 37.3 5 48.1 4 4 5 1 4 5 5 4 MCARS 57.3 4 7 7 8 5 6 5 7 7 5 5 5 5 7 7 7	CAR75	21.9	3	22.3	2	21.7	4	18.7	7	21.7	5	23.4	1	21.4	6	18.7	8
MCARS 10.5 3 5.6 7 9.4 4 8 6 11 2 1.3.6 1 4.3 8 8.3 MARDARDAR 18.6 4 20.1 2 18.4 5 15.9 8 1.87 18.1 1.8 7 18.1 MARMARDAR 19.7 4 20.1 2 18.4 4 21.9 3 19.7 6 20.8 1 19.7 7 19.7 MARMARDAR 19.7 4 20.1 2 18.4 4 21.3 6 20.8 1 19.7 7 19.7 MANARyure 18.8 3.3 18.9 2.6 18.4 4.6 15.9 7 18.6 4.6 21.1 6 21.4 6 11.4 4 20.2 21.1 4 6.8 17.7 9.4 4 9.9 1 49.6 7 47.5 8 49.9 2 50.1 3	CAR50	18.2	3	20.1	2	17.6	5	13.5	8	17.6	6	21.9	1	16.7	7	18.1	4
MCARS 10.5 3 5.6 7 9.4 4 8 6 11 2 13.6 1 4.3 8 8.3 MARAPure 18.6 4 20.1 2 19.7 5 19.9 3 19.7 6 20.8 1 19.7 7 19.7 MARAPure 18.8 3.3 18.9 2.6 18.4 4.6 15.9 7 18.6 4.6 21.2 1 1.1 17.4 6.8 17.7 MCARS 19.6 4.9 1 4.8 4.6 15.9 7 18.6 4.6 21.1 1 17.4 6.8 17.7 MCARS 47.8 6 49.9 1 49.6 7 47.5 8 49.9 2 50.1 3 49.9 4 49.9 MCARS 47.8 6 49.9 1 47.8 7 77.5 8 49.9 2 50.1 3 49.9 4 49.9 MCARS 45.7 47.8 6 48.8 4 47.8 7 7 29.5 8 49.2 3 49.5 2 49.5 3 MCARS 43.6 6 47.8 6 43.6 7 12.5 8 49.2 3 49.7 1 47.7 2 47.3 MCARS 43.6 43.6 49.2 2 44.3 8 52.3 4 47.2 5 49.3 1 47.6 5 MARDWR 44.4 7 47.8 2 44.3 8 52.3 4 47.2 5 49.3 1 47.6 6 47.8 MARDWR 44.5 6 47.5 2 44.3 8 52.3 4 47.2 5 49.3 1 47.6 6 47.8 MARDWR 44.5 6 47.5 2 44.3 8 52.3 4 47.2 5 49.3 1 47.6 6 47.8 MARDWR 44.5 6 45.5 2 44.4 7 42.6 8 43.1 4 45.5 1 44.5 1 44.5 1 MARAPURE 44.6 45.5 7 28.5 69.9 37.3 7.5 81.1 31.7 14.5 1 41.5 MCARS 75.3 4 75.1 2 8 76.7 4 8 12.2 8 75.4 4 6 7 7 7 7 7 MCARS 75.8 3 75.1 2 7 7 7 7 7 7 7 7 7	CAR25	15.2	3	16.7	2	14.2	5	6.2	8	14.6	4	20.5	1	12.9	6	12.6	7
MARMANAR 19.7 4 20.1 2 19.7 5 19.9 3 19.7 6 20.8 1 19.7 7 19.7 MARMARIMER 12.4 3 21.4 3 21.4 3 21.4 3 21.4 3 21.4 3 21.4 3 21.4 4 21.3 21.8 21.4 6 21.4 6 21.4 6 21.4 6 21.4 6 21.4 6 21.4 6 21.4 6 21.4 6 21.4 6 21.4 6 21.4 6 21.4 6 21.4 6 21.4 6 21.4 6 12.4 4 6 15.9 7 7 18.6 4.6 21.1 17.4 6.8 17.7 MCAR75 47.8 6 49.9 1 47.8 7 7.5 8 49.9 2 50.1 49.1 49.9 44.4 49.9 2			3		7		4		6		2	13.6	1		8		5
MARMANAR 19.7 4 20.1 2 19.7 5 19.9 3 19.7 6 20.8 1 19.7 7 19.7 MARMARJURG 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 POLILINES/GNZ V. L. 1.0 17.4 6.8 17.7 P. D. 2.0 2.1 1.0 17.4 6.8 17.7 P. D. V. V. 1.0 1.0 4.0 4.0 18.0 17.7 7.5 8 49.9 2 5.0 1.0 4.9.9 4.49.9 4.49.9 4.49.9 4.49.9 4.49.9 4.49.9 4.49.9 4.49.9 4.49.9 4.49.9 4.49.9 4.49.9 4.49.9 4.49.9 4.49.9 4.49.9 4.49.9	ARpure	18.6	4	20.2	2	18.4	5	15.9	8	18.7	3	21.8	1	18	7	18.1	6
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 MCAR5 49.6 6 49.9 1 49.6 7 47.5 8 49.9 2 50.1 3 49.9 4 49.9 MCAR5 47.8 6 49.8 4 47.8 7 37.5 8 49.9 2 50.1 3 49.9 4 49.9 MCAR5 47.8 6 48.8 4 45.5 7 29.5 8 49.2 3 49.7 1 49.7 2 47.3 MCAR5 43.6 6 47.8 5 43.6 7 12.5 8 49.3 2 50.1 1 49.7 2 47.3 MCAR5 44.3 6 47.8 5 43.6 7 12.5 8 49.3 2 50.1 1 49.7 2 47.3 MCAR5 44.3 6 47.8 5 43.6 7 12.5 8 49.3 2 50.1 1 49.7 2 47.3 MCAR5 44.3 6 47.8 47.8 47.8 48.5 4 45.1 4 50.2 1 51.8 3 44.8 MAR, MNAR MNAR 42.5 6 43.3 2 42.2 7 41.7 8 43.1 4 44.5 1 42.2 3 43.1 MNANAR with 44.4 6 45.5 2 44.4 7 42.6 8 43.1 4 46.2 1 44.5 5 44.8 MCAR5 43.6 47.7 2.8 43.3 69.9 37.3 7.5 48.1 31.8 47.1 MCAR5 75.3 4 75.1 3 75.4 5 76.2 8 75.4 6 75.1 75.5 7 77.1 MCAR5 48.4 3 75.5 2 76.7 4 81.2 8 79.5 5 76.7 1 8 77.7 MCAR5 48.4 3 75.5 2 76.7 4 81.2 8 79.5 5 76.7 1 8 77.7 MCAR5 50.8 3 79.5 2 81.3 5 93.7 8 82.7 6 78.7 1 84.7 7 81. MCAR5 48.7 47.8 47.8 47.8 47.8 47.8 47.8 47.8 MCAR5 50.8 3 79.5 2 81.3 5 93.7 8 82.7 6 78.7 1 84.7 7 81. MCAR5 50.8 3 79.5 2 81.3 5 93.7 8 82.7 6 75.4 2 74.4 8 75.2 MAR DWAR 70.8 3 76.7 5 76.1 7 77.1 77			4			19.7	5	19.9	3	19.7	6		1	19.7	7		8
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 MCAR5 49.6 6 49.9 1 49.6 7 47.5 8 49.9 2 50.1 3 49.9 4 49.9 MCAR5 47.8 6 49.8 4 47.8 7 37.5 8 49.9 2 50.1 3 49.9 4 49.9 MCAR5 47.8 6 48.8 4 45.5 7 29.5 8 49.2 3 49.7 1 49.7 2 47.3 MCAR5 43.6 6 47.8 5 43.6 7 12.5 8 49.3 2 50.1 1 49.7 2 47.3 MCAR5 44.3 6 47.8 5 43.6 7 12.5 8 49.3 2 50.1 1 49.7 2 47.3 MCAR5 44.3 6 47.8 5 43.6 7 12.5 8 49.3 2 50.1 1 49.7 2 47.3 MCAR5 44.3 6 47.8 47.8 47.8 48.5 4 45.1 4 50.2 1 51.8 3 44.8 MAR, MNAR MNAR 42.5 6 43.3 2 42.2 7 41.7 8 43.1 4 44.5 1 42.2 3 43.1 MNANAR with 44.4 6 45.5 2 44.4 7 42.6 8 43.1 4 46.2 1 44.5 5 44.8 MCAR5 43.6 47.7 2.8 43.3 69.9 37.3 7.5 48.1 31.8 47.1 MCAR5 75.3 4 75.1 3 75.4 5 76.2 8 75.4 6 75.1 75.5 7 77.1 MCAR5 48.4 3 75.5 2 76.7 4 81.2 8 79.5 5 76.7 1 8 77.7 MCAR5 48.4 3 75.5 2 76.7 4 81.2 8 79.5 5 76.7 1 8 77.7 MCAR5 50.8 3 79.5 2 81.3 5 93.7 8 82.7 6 78.7 1 84.7 7 81. MCAR5 48.7 47.8 47.8 47.8 47.8 47.8 47.8 47.8 MCAR5 50.8 3 79.5 2 81.3 5 93.7 8 82.7 6 78.7 1 84.7 7 81. MCAR5 50.8 3 79.5 2 81.3 5 93.7 8 82.7 6 75.4 2 74.4 8 75.2 MAR DWAR 70.8 3 76.7 5 76.1 7 77.1 77	NARpure	21.4	3	21.8	2	21.4	4	21.3	8	21.4	5	22.2	1	21.4	6	21.4	7
Pollines50%;											4.6				6.8		6.1
MCAR55 496 6 499 1 496 7 47.5 8 499 2 50.1 3 499 4 49.9 MCAR50 47.8 6 48.8 4 44.7 7 35.5 8 499 2 50.1 3 49.9 4 49.9 MCAR50 45.7 6 48.8 4 45.5 7 25.9 8 49.2 3 49.7 1 49.7 2 43.3 MCAR5 44.3 6 49.2 2 44.9 8 49.3 2 50.1 1 49.7 2 47.3 MCAR5 44.3 6 49.2 2 44.9 7 22.5 8 49.3 1 49.7 2 47.3 MCAR9 4 47.2 2 44.3 8 52.1 4 50.2 1 51.8 3 44 MARMAPure 44.4 7 <t></t>													-				
MCAR75 47.8 6 94.9 4 47.8 7 37.5 8 49.6 1 49.6 2 40.5 3 49.2 MCAR26 45.7 6 48.8 4 45.7 7 25.5 8 49.2 3 49.7 2 47.7 2 47.3 MCAR25 43.6 6 47.8 5 43.6 7 12.5 8 49.2 2 50.2 1 40.1 3 50.9 MARDWAR 44.3 6 42.7 44.3 8 52.1 4 45.0 1 41.5 1 47.6 47.8 MARDWAR 44.3 6 45.5 2 44.3 8 52.1 4 44.5 1 47.6 47.8 4 47.2 4 44.4 4 44.2 4 44.2 4 44.5 1 44.5 1 44.5 4 44.2 4 42.2 8																	5
MCAR50 45,7 6 48,8 4 45,5 7 29,5 8 49,2 3 49,7 1 40,7 2 47,3 MCAR5 44,3 6 49,2 2 44,9 7 12,5 8 49,3 2 50 1 40,1 3 50,9 MCAR5 44,3 6 49,2 2 44,9 5 35,1 8 52,1 4 50,2 1 51,8 3 4 MARMARMAR 42,3 6 43,3 2 42,2 7 41,7 8 43,1 4 44,5 1 43,2 3 43,1 MANARMAR 42,3 6 43,5 2 42,2 7 41,7 8 43,1 4 46,2 1 44,5 5 44,8 Average 43,5 6,1 47,7 2 44,5 6 75,4 48,7 4 48,2 1 48,2																	5
MCAR25 43.6 6 47.8 5 43.6 7 12.5 8 49.3 2 90 1 40.1 3 50.9 MCAR5 4.3 6 9 2 24.9 8 52.1 4 50.2 1 40.1 3 50.9 MARDWAR 4.3 6 42.8 44.3 8 52.3 4 47.2 5 49.3 1 47.6 6 47.8 MNARDWAR 44.3 6 45.5 2 44.4 7 42.6 8 44.5 4 45.2 3 42.2 3 4.3 4 47.2 4 46.2 1 44.5 1 44.5 5 44.8 4 4 46.2 1 44.5 1 44.5 5 44.8 4 4 4.2 4 6 45.5 4 48.2 4 46.2 1 44.5 5 44.8 4 46.2																	5
MCAR5 44.3 6 49.2 2 44.9 5 35.1 8 52.1 4 50.2 1 51.8 3 44 MARMARWAR 44.4 7 2 44.8 8 52.3 4 47.2 5 49.3 1 4 6.7 6 47.8 4 47.2 5 49.3 1 4 4.5 1 43.2 3 43.1 4 44.5 1 43.2 3 43.1 4 46.2 1 44.5 5 44.8 44.5 7 42.6 48.8 44.5 4 46.2 1 44.5 5 44.8 44.8 4 45.2 1 44.5 5 44.8 44.8 44.5 1 48.2 3 43.8 44.8 44.8 44.2 1 44.5 5 44.8 44.2 1 44.5 5 44.8 44.9 4 45.5 44.8 44.2 1																	4
MARDMAR 444 7 478 72 443 8 52.3 4 47.2 5 49.3 1 47 6 47.8 MARDMAR 423 6 43.3 2 42.7 41.7 8 43.1 44.5 1 43.2 3 43.1 MARDMAR 45.3 61 47.7 2.8 45.3 6.9 37.3 7.5 48.1 3.1 48.7 14.5 15.4 48.1 3.6 47.8 MARDMAR 57 53 4 75.1 3 75.4 8.1 3.1 48.7 14.5 15.6 14.8 MCAR95 75.3 4 75.1 3 75.4 8.1 3.1 48.7 14.5 15.7 7.7 1.1 MCAR95 75.3 4 75.1 3 75.4 8.1 3.1 48.7 14.5 15.7 7.7 1.1 MCAR95 81.8 3 79.5 2 81.3 5 87.5 6 8 75.4 6 75.1 75.5 7 75.1 MCAR95 81.8 3 79.5 2 81.3 5 87.5 8 75.5 6 75.1 18.7 7 77.1 MCAR95 81.8 3 79.5 2 81.3 5 93.7 8 82.7 6 78.7 1 84.7 7 81.1 MCAR95 81.7 4 87.5 8 8 8 87.5 8 8 87.5 8 8 8 87.5 8 8 8 87.5 8 8 8 87.5 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8			6												3		7
MARMANAR 42.3 6 43.3 2 42.2 7 41.7 8 43.1 4 45.5 1 42.2 3 43.1 MNARJurue 45.3 61 47.7 2.8 45.3 6.9 37.3 7.5 48.1 3.1 48.7 1.4 48.1 5 44.8 PorLine=7576 -																	3
MANARPINE 44.4 6 45.5 2 44.4 7 42.6 8 44.5 4 46.2 1 44.5 5 44.8 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 Pol-Line-75%									8		4						5
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 POPULINE=75%																	3
PoLine=75%																	4.6
MCAR55 75.3 4 75.1 3 75.4 5 76.2 8 75.4 6 75 1 75.5 7 75 MCAR50 76.6 3 76.3 2 76.7 4 81.2 8 77.1 5 75.9 1 77.6 7 77.1 MCAR50 78.4 3 77.5 2 78.7 4 87.5 8 79.5 5 76.7 1 80.7 7 79.7 MCAR25 80.8 3 79.5 2 81.3 5 93.7 8 82.7 6 78.7 1 84.7 7 81 MCAR50 80.8 3 79.5 2 81.3 5 93.7 8 82.7 6 78.7 1 84.7 7 81 MCAR50 80.8 3 79.5 2 81.3 5 93.7 8 82.7 6 78.7 1 84.7 7 81 MCAR50 80.8 3 79.5 2 81.3 5 93.7 8 82.7 6 78.7 1 84.7 7 81 MCAR50 80.8 3 79.5 2 81.3 5 93.7 8 82.7 6 78.7 1 84.7 7 81 MCAR50 80.8 1 87.9 3 90.8 5 91.7 7 87.4 2 84 1 91.1 6 98.7 81 MARDWAR 71.8 3 71.3 7 72.2 692.8 71.4 4 71.4 5 72.5 1 71.4																	
MCAR75 76.6 3 76.3 2 76.7 4 81.2 8 77.1 5 75.9 1 77.6 7 77.1 MCAR50 78.4 3 77.5 2 78.7 4 87.5 8 79.5 5 76.7 1 80.7 7 79.7 MCAR5 80.8 3 79.5 2 81.3 5 93.7 8 8.27 6 78.7 1 87.7 78.1 MCAR5 89.7 4 87.9 3 90.8 5 91.7 7 87.4 2 84 1 91.1 6 98.7 MARPUR 76.6 4 75.7 3 76.7 5 71.2 7 87.4 2 84 1 91.1 6 98.7 MARPUR 71.8 3 71.3 7 2 69.2 8 71.4 6 75.4 2 78.4 8 <td< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>2</td></td<>																	2
MCARS0 78.4 3 77.5 2 78.7 4 87.5 8 79.5 5 76.7 1 80.7 7 79.7 MCARS5 80.8 3 79.5 2 81.3 5 93.7 8 82.7 6 78.7 1 84.7 7 79.7 MCARS 89.7 4 87.9 3 90.8 5 91.7 7 87.4 2 84 1 91.1 6 98.7 MARpure 76.6 4 75.7 3 76.7 5 78.1 7 77.6 6 75.4 2 78.4 8 75.2 MAR MAR MAR 71.8 3 71.3 7 72 2 69.2 8 71.4 4 71.4 5 72.5 1 71.4																	6
MCAR25 80.8 3 79.5 2 81.3 5 93.7 8 82.7 6 78.7 1 84.7 7 81 MCAR5 89.7 4 87.9 3 90.8 5 91.7 7 87.4 2 84 1 91.1 6 98.7 MARpure 76.6 4 75.7 3 76.7 5 78.1 7 77.6 6 75.4 2 78.4 8 75.2 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
MCAR5 89.7 4 87.9 3 90.8 5 91.7 7 87.4 2 84 1 91.1 6 98.7 MARpure 76.6 4 75.7 3 76.7 5 78.1 7 77.6 6 75.4 2 78.4 8 75.2 MAR.MAR.MAR 71.8 3 71.3 7 72 2 69.2 8 71.4 4 71.4 5 7.25 17.5 7.25 17.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5																	4
MARpure 76.6 4 75.7 3 76.7 5 78.1 7 77.6 6 75.4 2 78.4 8 75.2 MAR.MNAR 71.8 3 71.3 7 72 2 69.2 8 71.4 4 71.4 5 72.5 1 71.4																	8
MAR_MNAR 71.8 3 71.3 7 72 2 69.2 8 71.4 4 71.4 5 72.5 1 71.4																	1
																	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

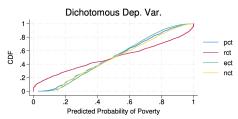
Predictions' Distributions





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

25	70	
23	50	75
12.1	43.2	82.1
25.0	47.4	73.2
16.6	49.7	81.0
11.1	42.9	83.0
13.0	47.5	77.4
11.4	51.9	84.8
25.5	51.0	75.7
20.2	50.9	79.7
10.1	52.2	87.3
16.2	54.9	87.1
	25.0 16.6 11.1 13.0 11.4 25.5 20.2 10.1	25.0 47.4 16.6 49.7 11.1 42.9 13.0 47.5 11.4 51.9 25.5 51.0 20.2 50.9 10.1 52.2

Can ML Models be Improved? Grid Search Parameters

	Grid range		Op	timal pa	aramete	ers	
		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

	Cont.	Cont.	Cont.	Cat.	Cat.	Cat.
Poverty Line (%)	25	50	75	25	50	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