Determinants of the Armenian Household Poverty: An Econometric and a Machine Learning Approach

Submitted to

American University of Armenia

Manoogian Simone College of Business and Economics

In partial fulfillment of the requirements for the degree of BA in Business

By: Karen Gishyan

Supervisor: Dr. Aleksandr Grigoryan

Yerevan 2019

Abstract

The aim of the paper is to use combine econometric analysis and machine learning modeling to explain the multidimensional nature of the Armenian Household poverty. The multinomial logistic regression results show that there are monetary and socioeconomic variables affecting poverty. Food and Non-Food related purchases in dram, members of the household, settlement, which includes Yerevan, other urban and rural towns, income received from abroad, educational level of the head of the household and a few other variables have a significant effect on the poverty status. After measuring the direct variable impact, Neural Networks and Decision Tree models are constructed. All three models are fit on the same training data and later evaluated on the same testing data to find out how well they perform the task of classifying Poor and Very Poor Households. From the original data, less than 2 percent of the observations fall under Very Poor Category, so correct results for this class are the most prioritized. Neural Networks provide the best results in terms of correctly classifying the Poor and Very Poor Households from the testing data, followed by Decision Tree and Logistic Regression. As a main classification metric F1 score is taken.

Keywords: Household Poverty, Multinomial Logistic Regression, Neural Networks, Decision Tree

Acknowledgment

The fact that most of the literature on the economic research is based on econometric linear models inspired me to use machine learning modeling and approach to explain a socioeconomic phenomenon. The idea of doing the research on Armenian Household Poverty has been suggested by my supervisor Dr. Aleksandr Grigoryan, who has provided continuous guidance and support throughout the process. The remaining errors, if present, are mine.

Contents

List of Figures

List of Tables

Introduction

In the recent literature relating to socio-economic issues poverty reduction has been a key policy debate. The elaboration of policies for poverty relief requires a thorough knowledge of this phenomenon. In 2016, the poverty rate in Armenia was 29.4% compared to the 27.6% recorded in 2008 and the share of extremely poor was 1.8% as compared to 1.6% recorded in 2008. Armenia's administrative division consists of 10 marzes (regions) and the capital of Yerevan. The results of 2016 show that the poverty indicators in Shirak, Lori, Kotayk, Tavush and Armavir provinces are higher than the country average, and the highest poverty rate is in Shirak Region, where 46% of the population is below the poverty line ('Poverty Profile'). According to the same report, however, 62.4% of the poor in Armenia are urban residents. For these reasons I find it important to not only determine the factors affecting the poverty status in Armenia but to construct machine models that can go much beyond linear relationships and are able to classify the multi-label observations with higher accuracy. The estimation of the models is based on the Armenia Household Survey data from 2015-2017. The methodology starts with the multiclass logistic regression analysis, fit on the training data and coefficient interpretation for the significant variables. Then the obtained model is used to fit the testing data and the classifications metrics such as Recall, Precision, F1 score along with confusion matrix results are presented. The same training and testing data is used to build deep neural network and decision tree models. The classification metrics obtained from these models are compared to the one obtained from logistic regression. These section is followed by conclusion and discussion/recommendations.

Literature shows that most of the studies have used binary logistic regression models to find out the household poverty determinants. The use of machine learning methodology in combination with econometric interpretation will be a contribution to the existing literature.

Literature Review

Poverty is a mixture of economic and social aspects (Patlagean,1977, as cited in Jmaii, 2016) which must be studied simultaneously to find the efficient policy to fight against this scourge. The incidence of poverty is determined by multiple factors operating at micro (household) as well as macro (national) levels. (Rahman, 2013). Poverty reduction is a key objective of many nations around the world. Extreme poverty reduction is one of the eight Millennium Development Goals defined by the United Nations and all 191 member states in 2000. Initially aimed to be achieved in 2015, program has had certain success and aims to continue to meet its poverty reduction targets with the 2030 sustainable development agenda ('Millennium Development', 2015). According to the existing literature, we can distinguish two main forms of poverty. First of all, the monetary poverty which results from a lack of resources and leads to insufficient consumption. This approach is related to the economy of welfare since the monetary indicators define poverty according to an income deficiency or a too low consumption which reflect a lower standard of living (Townsend,1985, as cited in Jmaii, 2016). It is a widely used concept of classifying individuals according to their monetary resources and is usually referred to as a unidimensional index. The poor are those individuals or households whose income or consumption is below a given threshold (Ravallion,1998). This threshold is then defined by measuring the consumption of basket of goods and services which allows to achieve a minimum standard of living. The second concept of poverty, mostly referred to poverty of living conditions and initiated by Townsend (1979), is determined through a multidimensional index. This index is usually constructed by getting information about food consumption, education, working conditions etc. about a family by household surveys. The aim is to get an overall view of the living conditions to better capture the phenomena of poverty. This approach corresponds to the logic of Sen (1985) with his concept on individual capacities and it supports the idea that poverty reflects a lack of basic functional capabilities. This paper is not concerned with the construction of multidimensional index, rather it focuses on finding the variables which can define the multidimensional poverty.

For many economists, the one-dimensional study may appear to be more limited, less complete and therefore less relevant than a multidimensional study (Jmaii, 2016). If we consider income as a measure of a well-being, which it usually is, then univariate study can also be preferable. The analysis of poverty requires a definition of a poverty line to determine who is poor and who is not. Two widely used poverty line thresholds are absolute and relative poverty lines. The absolute poverty line is a constant threshold over time in terms of living standards, updated with price inflation only. The relative threshold measures both the evolution of inequality as well as poverty. The most commonly used threshold is the half of the median (or mean) income (or expenditure) per unit of consumption and per equivalent adult (Jmaii, 2016).

Sikander’s and Ahmed's (2008) study on Pakistan finds a high dependency with the size of the household having a positive impact on the household's probability of being poor. It has been demonstrated that the household size, and the dependency ratio have significant positive correlation with the household’s probability of being poor while the educational level of the households, age of the household head and landholding negatively affect the probability of being poor (Rahman, 2013). In South Africa, Maitra (2002, as cited in Rahman,2013) finds a significant effect of education of the household head on the poverty status of the household. Results show that the highest level of education attained by the household head significantly reduces the probability of the household being poor. In their studies, Bógale and Korf (2009) find that an increase in household size by one adult equivalent increases the probability of being extremely poor and moderately poor by 3.13 and 5.16 percent respectively and it lowers the likelihood that a household will fall under the category of slightly non-poor and non-poor by 0.49 and 7.79 percent respectively. Rahman (2013) demonstrates the following results in his findings. Households headed by younger persons are less likely to be poor than households headed by older persons. Female headed households are more likely to live in poverty than male headed households and larger households are more likely to live in poverty. (Alkire et al., 2015) have demonstrated that an increase of one year of education decreases the odds of being multi-dimensionally poor by 49%, ceteris paribus, whereas having a female household head increases the odds of being multi-dimensionally poor by 28%, ceteris paribus. Similarly, the odds of a household being multi-dimensionally poor decrease by 57% for households living in urban areas, ceteris paribus, and increase by 10% for each additional household member. The results of (Rodriguez et al, 2015) in their multinomial logistic regression analysis have shown that the higher the level of education of the head, the probability of chronic household poverty decreases by up to 20 percent. Increasing household size by one unit increases the probability of falling into chronic poverty by 3 percent while the probability of never being poor decreases by 2 percent. Living in a rural area increases the probability of being chronically poor by 3 percent and if the head of the household is a woman the probability of the household being chronically poor increases while the probability of never being poor decreases.

Background information and Data Description

The data is comprised of 18144 observations of Armenian Household Survey Data from years 2015 to 2017. Overall the dataset includes 60 variables. The important independent variables after collinearity check (De Veaux & Ungar) and significance check for both logistic regression models were identified and the descriptive statistics for these variables in the training data is presented.

Figure 1: Descriptive Statistics

	Non Food Purchases	Non Monetary Income	Food in Small Amount Per Month	Food Purchases	Present members of Household	Income received from abroad	Income from savings	Education Level of HH Head	Settlement
count	13,608.00	13,608.00	13,608.00	13,608.00	13,608.00	13,608.00	13,608.00	13,608.00	13,608.00
mean	76,551.95	12,320.94	2,504.73	49,548.78	3.43	16,438.71	3,951.42	5.94	1.08
std	128,736.12	21,978.14	1,320.62	34,796.32	1.83	48,760.32	22,254.94	1.47	0.81
min	0.00	0.00	27.80	0.00	1.00	0.00	0.00	1.00	0.00
25%	27,650.00	0.00	1,548.20	25,460.00	2.00	0.00	0.00	5.00	0.00
50%	50,200.00	2,898.79	2,260.61	41,410.00	3.00	0.00	0.00	5.00	1.00
75%	88,302.50	16,440.97	3,262.78	63,967.50	5.00	0.00	0.00	7.00	2.00
max	4,015,090.00	600,000.00	11,833.84	295,330.00	18.00	1,492,200.00	500,000.00	9.00	2.00

The variables are:

- Non-food purchased of household per month in dram
- Non-monetary income of household of household per month in dram
- Food in small amount of household per month in dram
- Food purchases of household per month in dram
- Present members of household
- Income from abroad- money received from relatives, living out of Armenia
- Income from savings
- Educational Level of head of the household- no primary, illiterate, no primary literate, primary, general, secondary, preliminary vocational, middle vocational, higher, post graduate
- Settlement- Yerevan, Other Urban, Rural

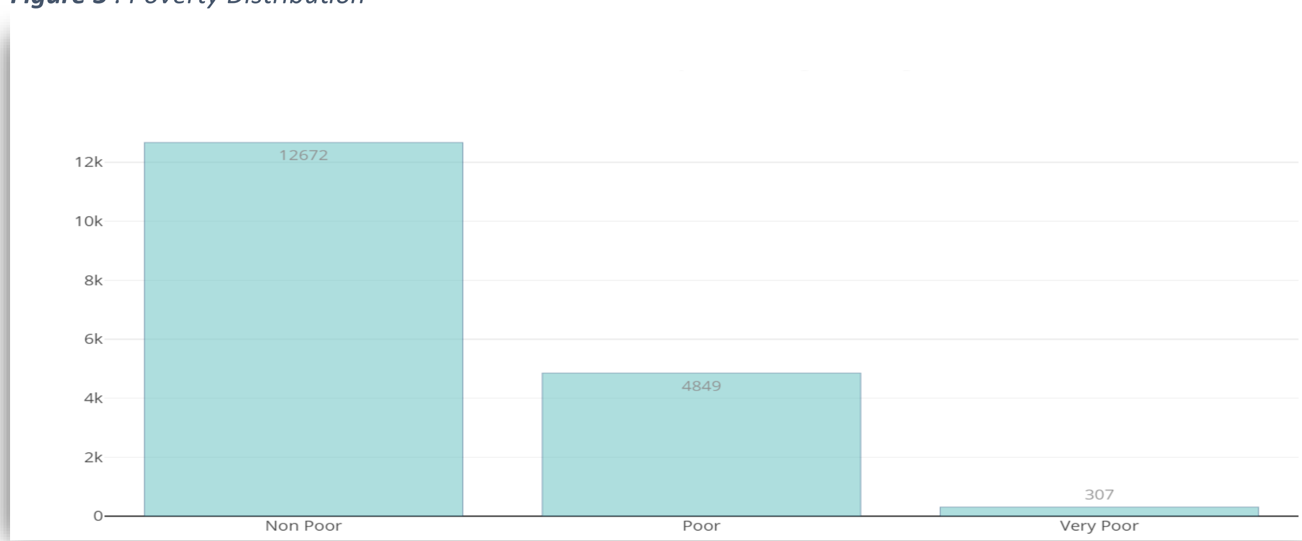
The correlation matrix shows that no high correlation is present in the dataset between the variables. The highest correlation is between the present members of the household and food purchased in small amounts per month and it is 59.5%. The next highest correlation is between present members of the household and food purchases variable. However, as they do not exceed 70% threshold, these variables are included in the model building.

Figure 2 : Correlation Matrix

	Non Food Purchases	Non Monetary Income	Food in Small Amount Per Month	Food Purchases	Present members of Household	Income received from abroad	Income from savings	Education Level of HH Head	Settlement
Non Food Purchases	1	0.0346383	0.165568	0.256822	0.173844	0.0160578	0.0418941	0.115248	-0.138423
Non Monetary Income	0.0346383	1	0.233391	-0.0301432	0.2643	-0.000201569	0.0468412	-0.106915	0.390087
Food in Small Amount Per Month	0.165568	0.233391	1	0.363117	0.595298	0.00598128	0.0766868	-0.0172397	0.120487
Food Purchases	0.256822	-0.0301432	0.363117	1	0.405723	0.0264125	0.0664225	0.14477	-0.182614
Present members of Household	0.173844	0.2643	0.595298	0.405723	1	-0.0627719	0.0624463	-0.0577104	0.0865005
Income received from abroad	0.0160578	-0.000201569	0.00598128	0.0264125	-0.0627719	1	-0.0406557	-0.0143512	0.0234386
Income from savings	0.0418941	0.0468412	0.0766868	0.0664225	0.0624463	-0.0406557	1	-0.048101	0.0545789
Education Level of HH Head	0.115248	-0.106915	-0.0172397	0.14477	-0.0577104	-0.0143512	-0.048101	1	-0.258091
Settlement	-0.138423	0.390087	0.120487	-0.182614	0.0865005	0.0234386	0.0545789	-0.258091	1

The dependent variable is Poverty. 71% of the observations belong to the Non-Poor category, around 27.1% to Poor category and around 1.72% to the Very Poor Category. The results are provided in the table. For the neural network model, alongside the variables presented above, five other variables, significant only for the $\ln\left(\frac{P_{(Poor)}}{P_{(Non\ Poor)}}\right)$ dependent variable model were included.

Figure 3 : Poverty Distribution



The observations for the network were normalized using Min-Max scaling method. The data was divided into training and testing sets, which were used for model building and validation. 75% of the observations were used for training the models, 25% for testing the models.

Models and Methodology

In order to study the relationship between the multiclass dependent categorical variable and the independent variables, logistic regression, deep neural networks and decision tree models are used. The goal is to explain the variables with their unit impact using logistic regression then find out the most optimal model among the three in terms of classification metrics.

We will start the analysis with a multinomial logistic regression. Let's note that here we do not assume the independent variables are normally distributed and the homoscedasticity is also not required. The independent variables linearly predict a logit transformation of the dependent variable while the equation in terms of probabilities is nonlinear. Probability (P) varies from 0 to 1, while the range of logit is from minus to plus infinity ('Logistic Regression'). Multinomial logistic regression, in other terms referred to as Softmax Regression, is used when the target variable has multiple classes. It gives the probability that the response variable takes on each of the possible classes.

Softmax function has the following form ('Softmax Regression').

$$Pr\left(Y^i = K \middle| x^i; \beta\right) = \frac{exp\left(\beta^{(k)T} x^{(i)}\right)}{\sum_{j=1}^{K} \exp(\beta^{(j)T} x^{(i)})}$$

The equivalent logit form for multiclass logistic regression can be written in the following way:

$$\ln \frac{\Pr(Y_i = 1)}{\Pr(Y_i = K)} = \beta_1 * X_i$$

.....

$$\ln \frac{\Pr(Y_i = K-1)}{\Pr(Y_i = K)} = \beta_{K-1} * X_i$$

where K is the number of possible target variable outcomes and also the pivot class, and K-1 is the number of independent binary logistic regression models built. In the binary case where K =2, softmax regression reduces to logistic regression, showing that it is a generalization of the logistic regression (‘Softmax Regression’).

The model provided in the table is a regularized multinomial logistic regression model, fit with an L1 regularization and with a 0.1 alpha term, which is the weight for the L1 penalty. As our dependent variable has three categories, Poor, Non-Poor and Very Poor, there will be two regression equations built (‘Logistic Regression’). Non-Poor is the reference class and there are two regression coefficients associated with each independent variable.

$$\ln\left(\frac{P_{(Poor)}}{P_{(Non\ Poor)}}\right) = \beta_{0Poor} + \beta_{Poor1}X_{1,i} + \beta_{2Poor}X_{2,i} + \cdots + \beta_{mPoor}X_{m,i}$$

$$\ln\left(\frac{p_{(Very\ Poor)}}{p_{(Non\ Poor)}}\right) = \beta_{0Very\ \ Poor} + \beta_{\ 1Very\ Poor}X_{1,i} + \beta_{2\ VeryPoor}X_{2,i} + \cdots + \beta_{m\ Very\ Poor}X_{m,i}$$

Table 1: Logistic Regression Model

Model: MNLogit Method: MLE Observations: 13608				Log-Likelihood: -3692.5 Pseudo R^2: 0.5354 LLR p-value: 0.000	
	Logit Coef.	**Std. Err.**	**Z**	**P> |z|**	**Odds Ratio Coef.**
const	0.1146	0.147	0.779	0.436	1.121453
Non Food Purchases	-9.516e-05	2.8e-06	-33.983	0.000***	0.999905

Non-Monetary Income	-6.674e-05	2.85e-06	-23.455	0.000***	0.999933
Food in Small Amount Per Month	-0.0001	3.34e-05	-4.364	0.000***	0.999854
Food Purchases	-3.013e-05	1.9e-06	-15.856	0.000***	0.999970
Present members of Household	1.7594	0.045	39.018	0.000***	5.809121
Income received from abroad	-3.59e-06	7.77e-07	-4.618	0.000***	0.999996
Income from savings	1.622e-05	2.88e-06	-5.677	0.000***	0.999984
Education Level of HH Head	-0.0864	0.021	-4.161	0.000***	0.917265
Settlement	-0.4286	0.045	-9.489	0.000***	0.651443
Poverty= Very Poor Variable	**Logit Coef.**	**Std. Err.**	**Z**	**P> |z|**	**Odds Ratio Coef.**
const	-0.0897	0.396	-0.226	0.821	0.914166
Non Food Purchases	-0.0002	7.8e-06	-23.851	0.000***	0.999814
Non-Monetary Income	-0.0001	9.06e-06	-14.552	0.000***	0.999868
Food in Small Amount Per Month	-0.0006	0.000	-5.494	0.000***	0.999420
Food Purchases	-7.147e-05	5.48e-06	-13.050	0.000***	0.999929
Present members of Household	2.6771	0.081	33.051	0.000***	14.542962
Income received from abroad	-6.47e-06	3.24e-06	-1.996	0.046**	0.999994

Income from savings	-3.004e-05	9.49e-06	-3.165	0.002***	0.999970
Education Level of HH Head	-0.3339	0.066	-5.032	0.000***	0.716154
Settlement	-0.2952	0.144	-2.047	0.041**	0.744351

The model assumes a linear relationship between the log odds of poverty variable and the independent variables, making the coefficient interpretation less intuitive. To be able to bypass this problem, we take an exponent of our initial coefficients, which allows us to explain the odds ratio ($\frac{P_{(Poor)}}{P_{(Non\ Poor)}}$) or ($\frac{p_{(Very\ Poor)}}{p_{(Non\ Poor)}}$) rather than the log odds ratio. If a variable has a negative coefficient, the respective odds ratio coefficient is less than 1, indicating a negative relationship with the Poverty variable. For explaining each variable coefficient, we need to consider the ceteris paribus effect. The results show that we if we increase non - food purchases of households by 1unit which is 1 dram in this case, the odds of being poor will change by 0.999905, or go down by 0.0095% and the odds of very poor will change by 0.999814 or decrease by 0.0186%, ceteris paribus. The impact of the variables is quite small because the unit is represented in one Armenian dram. Instead we could consider the Δchange in variables to be 1000 drams, which is a widely used form of currency in Armenia. Increasing non-monetary income of household by 1000 drams will decrease the odds of being poor by 6.7% and the odds of being very poor by 13.2%. This means a person will be less likely to be poor compared to non-poor by 6.7% and 13.2% less likely to be Very poor. The same logic applies to the other variables. If we increase the food purchased in small amounts of household by 1000 drams, the odds of being poor will decrease by 14.6% and the odds of being very poor by 58%. If we increase the food purchased per household per month in 1000 drams, the odds or chance of being Poor goes down by 3% and the odds of being very Poor goes down by 7.1%. It is visible that the variables which are related to money and purchasing have a bigger impact on the Very Poor Category. Income received from relatives living outside of Armenia has the following interpretation. Increasing income received from abroad by 1000 drams, decreases the odds of being poor by 0.4% and the odds of being very poor by 0.6%. The variable which overall has the biggest impact is the number of present members in the family. Increasing the present number of household members by 1 person increases the odds of being poor by 5.81 or by 481% and the odds of being very poor by 14.54 or by 1354%. The model also suggests that if income from savings goes up, a person by 1.6% is less likely to be poor and by 3% less likely to be very poor. Finally, if the education of the head of the family increases by one level, the odds of being poor decreases by 8.27% and the odds of being very Poor decreases by 28.38%. The model also says that the chance of a person who lives in an urban area rather than Yerevan is less likely to be poor by 34.86% and very poor by 25.56% and a person who lives in a rural area is less likely to be poor by 69.72% and very poor by 51.12%.

Table 2 : Logistic Regression Classification Metrics Report

	Classification Report				**Confusion Matrix**		
					Predicted Class		
	Precision	Recall	F1	Support	Non-Poor	Poor	Very Poor
Non-Poor	0.91	0.95	0.93	3524	0.95 (3361)	0.046(163)	0 (0)
Poor	0.75	0.65	0.7	958	0.35 (333)	0.65(622)	0.0031(3)
Very Poor	0.75	0.17	0.27	54	0 (0)	0.83 (45)	0.17 (9)
AVG Accuracy/Total	0.88	0.89	0.88	4536			

The classification report of the model enables us to assess the overall goodness of fit and the predictive power of the model and will be a common base for comparing Logistic Regression, Neural Networks and Decision Tree Models. The confusion matrix allows us to see how many observations from each category have been correctly classified and misclassified. The results are present in parenthesis alongside the proportions of classification. Accuracy, which is found by summing the correctly predicted diagonal elements and dividing by the total number of observations, is an appropriate measure for classifier evaluation when the target variable has balanced classes. The classes of poverty variable are imbalanced, for these reason we might consider using Precision, Recall or F1 for classifier comparison. Precision is defined as $\frac{True\ Positive}{True\ Positive+False\ Positive}$, while Recall is defined as $\frac{True\ Positive}{True\ Positive+False\ Negative}$ (Johari).

For a multilevel classification, true positives are the left to right diagonal elements of the confusion matrix, showing the actual classes which have been correctly classified. The total number of false negatives for a class is the sum of values in the corresponding row (excluding the true positive value) and the total number of false positives for a class is the sum of values in the corresponding column (excluding the true positive value). F1 is defined as the harmonic of precision and recall is an overall summary metric of this two measures: $F1 = 2 * \frac{P*R}{P+R}$. Support is the number of class occurrences in the testing dataset, thus is the same for all models. The intuition of Recall is the following: given that a Household actually belongs to a specific category (e.g. Very Poor), what proportion of the households has been correctly classified as Very Poor by our Model. The same logic applies to the other two categories as well. We see that 17% of the households that belonged to the Very Poor Category have been classified as Very Poor by our model. 65% of the Households which belonged to the Poor Category have been classified as Poor and 95% of the households which belonged to the Non-Poor Category have been classified as Non-Poor. We see that the model is highly accurate on the Non-Poor observations, and below 20% accurate for the Very Poor Category. This happens because out of 4536 testing observations, only 54 belong to the Very Poor Class, making the task of the model to correctly classify harder. And the intuition of sensitivity is the following. Given that our model has classified an observation as belonging to a specific category, what proportion of these classifications actually belonged to that category. In other words, out of all Very Poor Classification that the model made, 75% were correct, out of all observations classified as Poor, 75% were actually Poor and similarly 91% classified as Non-Poor were actually Non Poor. These two metrics might seem to represent the same thing but in reality they have different usages. Recall is the ability of a classifier to correctly find all the Very Poor instances while Precision is the ability not to label an instance as Non-Poor or Poor if it is actually Very Poor. After fitting the logistic regression model and obtaining the classification results, we move on to build a neural network model. The trained neural network, which is comprised of neurons at each layer, has 14 input variables. Neuron is a unit that takes the inputs and gives an output by a certain function. The function that does the following mapping is called an activation function ('Multi-Layer Neural Network').

$$h_{W,b} = f(W^T x) = f(\sum_{i=1}^{n} W_i X_i + b)$$

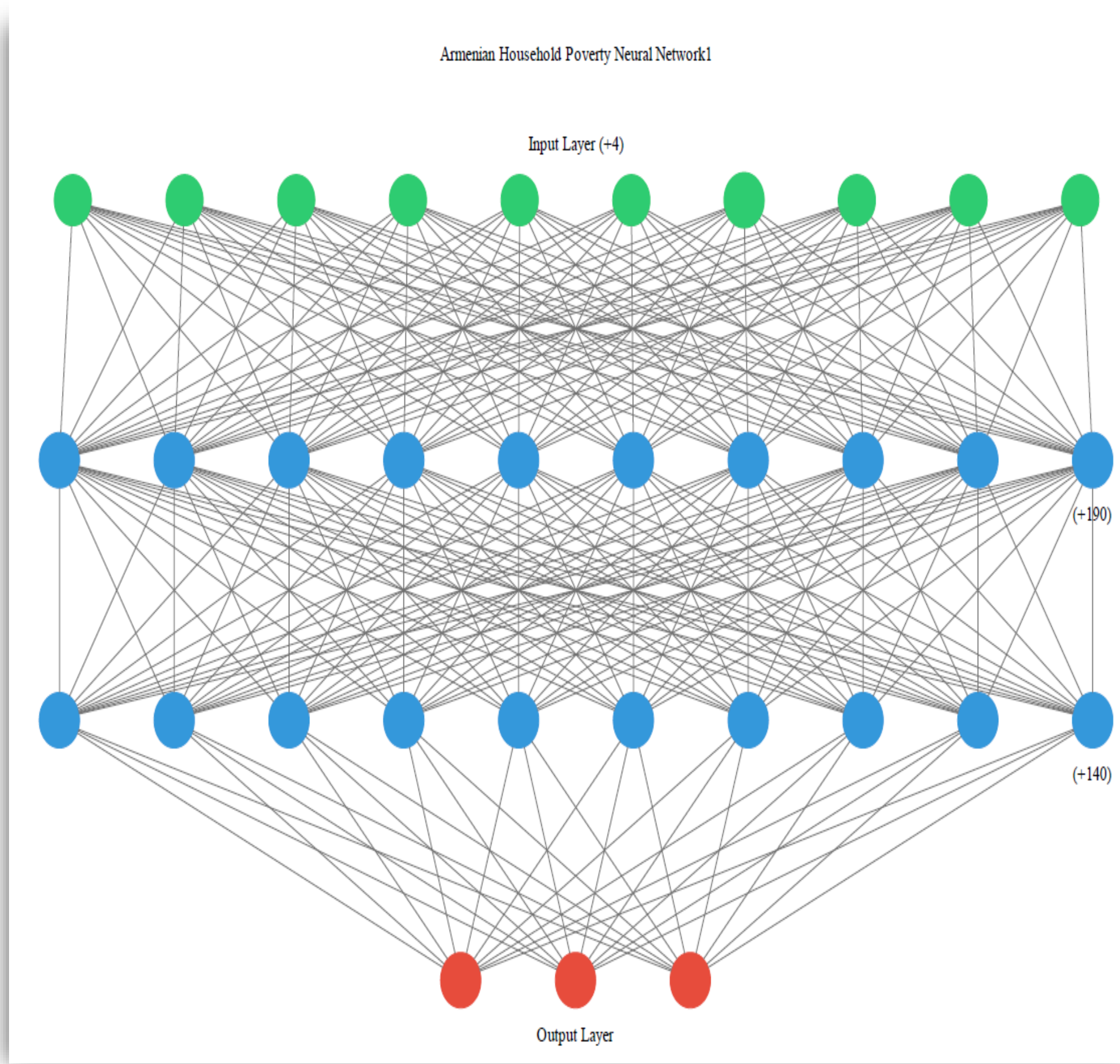
The first hidden layer of the network contains 200 neurons, the second one 150 neurons and rectified linear activation function (Relu) is used for these layers. Relu is given by:

$$f(z) = \max(0, z)$$

where z is the weighted sum of inputs including the bias term.

Softmax activation, introduced for logistic regression, is used as the activation function for the output layer. The model is minimizing a sparse categorical cross entropy loss function, the same loss function used for logistic regression, with 50 epochs and 100 batch size. The graph of the fitted neural network is presented.

Figure 4 : Visualized Neural Network



Softmax loss function ('Softmax Regression') which is a generalization of the binary logistic loss function and is otherwise known as a cross entropy loss function, has the following form.

$$J(\beta) = -\sum_{i=1}^{m}\sum_{k=1}^{K} 1\,\{y^i = k\} log \frac{exp\left(\beta^{(k)T} x^{(i)}\right)}{\sum_{j=1}^{K} \exp(\,\beta^{(j)T} x^{(i)})}$$

The graph of the testing and training losses is the following.

Figure 5: Neural Network Training and Testing Losses

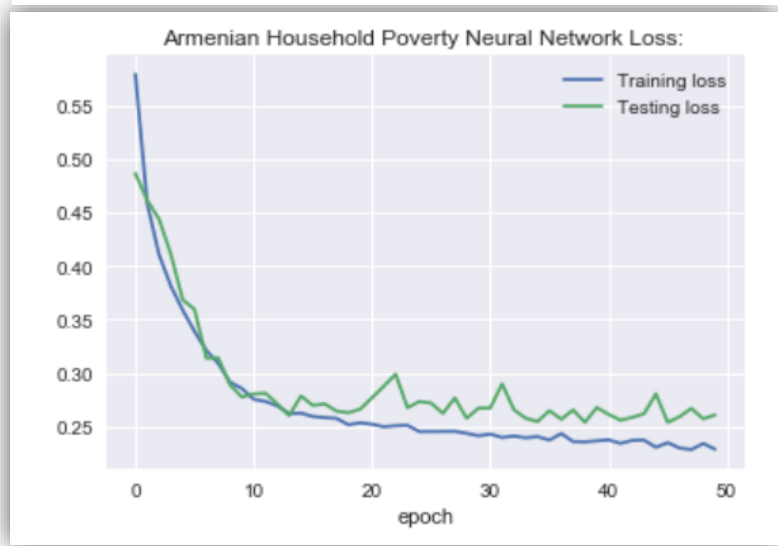


***Table 3** : Neural Networks Classification Metrics Report*

	Classification Report				Confusion Matrix		
					Predicted Class		
	Precision	Recall	F1	Support	Non-Poor	Poor	Very Poor
Non-Poor	0.93	0.95	0.94	3524	0.95(3336)	0.053 (187)	0.00028 (1)
Poor	0.77	0.69	0.73	958	0.27 (259)	0.69 (661)	0.04 (38)
Very Poor	0.49	0.70	0.58	54	0 (0)	0.3 (16)	0.7 (38)
AVG Accuracy/Total	0.89	0.89	0.89	4536			

For the Neural Networks model the Recall results are the following. Given that a person belongs to a Very Poor Category, the Model correctly classified 70% of them as Very Poor, for the Poor Category the percentage of correctly classified observations was 69% and for the Non-Poor Category, 95% of the observations who were Non Poor were correctly classified as Non-Poor. The Precision(specificity) shows that out of all observations that were classified as Very Poor 49% actually belonged to the Very Poor category, out of all observations classified as Poor 77% were actually Poor and from all Non-Poor classified observations 93% were actually Non Poor. In terms classifying the Very Poor category, the Neural Networks is highly outperforming the Logistic Regression Classifier. For the Decision Tree model, Gini impurity is chosen as the splitting criteria, and with maximum depth of 15.

$$I_G(p) = \sum_{i=1}^{N} p_i(1 - p_i) = 1 - \sum_{i=1}^{N} p_i^2$$

where: N-number of classes,

Pi-fraction of items labeled with class i in the set.

***Table 4** : Decision Tree Classification Metrics Report*

	Classification Report				**Confusion Matrix**		
					Predicted Class		
	Precision	Recall	F1	Support	Non-Poor	Poor	Very Poor
Non-Poor	0.93	0.92	0.92	3524	0.92 (3240)	0.08 (282)	0 (2)
Poor	0.69	0.7	0.69	958	0.27 (259)	0.7 (672)	0.003 (27)
Very Poor	0.47	0.48	0.48	54	0.02 (1)	0.5 (27)	0.48 (26)
AVG Accuracy/Total	0.87	0.87	0.87	4536			

The recall results state that from all of households being Very Poor 48% were correctly classified, for the actual Poor category the correctly classification was 70% and for actual Non Poor Category it was 92%. The results of precision state that from all very poor classified observations 47 % were actually very poor, out of all Poor classified observations 69% were actually poor and out of all Non Poor classified observations 93% were Non Poor.

To compare the overall predictive power of the models, we can look at the F1 score which incorporates both Recall and Precision. Neural Networks has the highest measures for all three categories and thus outperforms the other two models, though the score for the Very Poor Category is of the highest interest. Logistic regression has higher F1 scores in terms of Non-Poor and Poor Categories compared to the Decision Tree model, but the worst one out of three in terms of predicting the Very Poor category.

Discussion

The findings presented in the paper can be valuable in terms of policy development. We saw that besides the monetary and income variables, settlement, number of household members and head of the education turned out to be very significant. Educational level of the household head had the biggest role towards reducing household poverty, so the governments of Armenia should initiate policies or legislative changes to make education more accessible. The results demonstrate that larger households have higher chance of being poor. A policy solution could be to start educational trainings to inform families about the advantages of keeping the households small to avoid poverty risks. I believe doing the analysis on a dataset including pre 2015 observations would provide more insight, so this can be a future task to impellent. In terms of models, there is a tradeoff. Logistic regression provides is the most interpretative model, allowing to measure the direct impact of the variable on poverty, while decision tree and neural networks do a better job of correctly classifying a household as Poor or Very poor. For the Very Poor category, they do it with a much higher accuracy. If one is not much interested in model interpretation rather in a model providing the most accurate results in terms of categories, the Neural Networks is the one to choose. Though we should acknowledge that there are much more computational costs involved with running a deep learning model.

Conclusion

This paper, using 2015-2017 household data, aimed to find out the factors contributing towards the multidimensional poverty in Armenia and compare the predictive powers of logistic regression, neural networks, decision tree models using classification metrics. The results from logistic regression show that poverty status depends on both monetary and non-monetary factors. Increasing non-food related purchases, food related purchases, non-monetary income of households per month in dram and income from savings decreases the odds of being poor and very poor. Settlement variable is quite significant. Though a little surprising, the results show that outside of Yerevan a person has a less chance of being poor or very poor. Income received from relatives living outside of Armenia though small, but has an impact on the poverty status. Increasing income received from abroad by 1000 drams, decreases the odds of being poor by 0.4% and the odds of being very poor by 0.6%. Increasing the present number of household members by 1 person significantly increases the odds of being poor and very poor. If the education of the head of the family increases by one level, the odds of being poor decreases by 8.27% and the odds of being very Poor decreases by 28.38%. We are mostly interested in the models' predictive ability on the Poor and Very Poor categories. By looking at the F1 scores for the Poor and Very Poor Categories, which is a balanced measured between precision and recall, we see that the neural networks provide the best results. Logistic Regression does slightly well classifying the poor category compared to the decision tree, but provides the lowest F1 score in terms of Very Poor category (0.27) compared to Decision Tree's (0.48) and Neural Networks' (0.58) scores for the same category. Thus for interpreting the results logistic regression can be valuable, but we can see how much it lacks behind in terms of classifying the desired Very Poor Category.

Reference

Alkire, S., Foster, J. E., Seth, S., Santos, M. E., Roche, J. M., & Ballon, P. (March 2015). Multidimensional Poverty Measurement and Analysis: Chapter 10 – Some Regression Models for AF Measures. Oxford Poverty & Human Development Initiative (OPHI), n0. 91.

Bógale, A., and Korf, B., "Analysis of Poverty and its Covariates among Smallholder Farmers in the Eastern Hararghe Highlands of Ethiopia", 2009, paper presented at the International Association of Agricultural Economists Conference, August 16-22, Beijing, China

De Veaux, R., & Ungar, L. H. (n.d.). Multicollinearity: A tale of two nonparametric regressions. Princeton University/University of Pennsylvania. Retrieved from http://www.cis.upenn.edu/~ungar/Datamining/Publications/tale2.pdf

Jmaii. A. MICRO-ECONOMETRIC ANALYSIS OF POVERTY: EVIDENCE FROM TUNISIA. Statistics [stat]. University Tunis El Manar, 2016. English. fftel-01443104f

Johari, R. (n.d.). "Small" Data Classification. Retrieved from http://web.stanford.edu/~rjohari/teaching/notes/226_lecture8_prediction.pdf

Logistic Regression. (n.d.). NCSS. Retrieved from https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Logistic_Regression.pdf

Multi-Layer Neural Network. (n.d.). UFDL. Retrieved from http://ufldl.stanford.edu/tutorial/supervised/MultiLayerNeuralNetworks/

Maitra, P., "The Effect of Household Characteristics on Poverty and Living Standards in South Africa", Journal of Development Economics, 2002, Vol. 27, No. 1, pp. 75-83.

PART 1-ARMENIA: POVERTY PROFILE IN 2008-2016 (pp. 38-42, Publication). Yerevan: Statistical Committee of RA.

Patlagean, E. (1977). Pauvreté économique et pauvreté sociale à Byzance, 4e-7e siècles, Volume 48. Walter de Gruyter press.

Rahman, M. A. (2013). HOUSEHOLD CHARACTERISTICS AND POVERTY: A LOGISTIC REGRESSION ANALYSIS. *The Journal of Developing Areas, 47(1)*. University of Western Sydney, Australia.

Rodriguez, J. G., Ramos, J. F., García-Guerra, A. K., & Ramírez, G. M. (2015, April 6). The Dynamics of Poverty in Mexico: A Multinomial Logistic Regression Analysis. Universidad de Monterrey.

Ravallion, M. (1988). Expected poverty under risk-induced welfare variability. The Economic Journal 98 (393), 1171–1182.

Sen, A. (1985). Well-being, agency and freedom: the dewey lectures 1984. The journal of philosophy 82 (4), 169–221.

Sikander, M. U., and Ahmed, M., "Household Determinants of Poverty in Punjab: A Logistic Regression Analysis of MICS 2003-04 Data Set", 2008, paper presented at the 8th Global Conference on Business and Economics, Florence

Softmax Regression. (n.d.). UFDL. Retrieved from http://ufldl.stanford.edu/tutorial/supervised/SoftmaxRegression/

The Millennium Development Goals Report 2015. (2015, July). United Nations. Retrieved from https://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG 2015 rev (July 1).pdf

Townsend, P. (1979). Poverty in the United Kingdom: a survey of household resources and standards of living. Univ. of California Press

Townsend, P. (1985). A sociological approach to the measurement of poverty–a rejoinder to professor Amartya Sen. Oxford Economic Papers 37 (4), 659–668.

Appendix

***Figure 6** : Neural Network Training and Testing Accuracy*

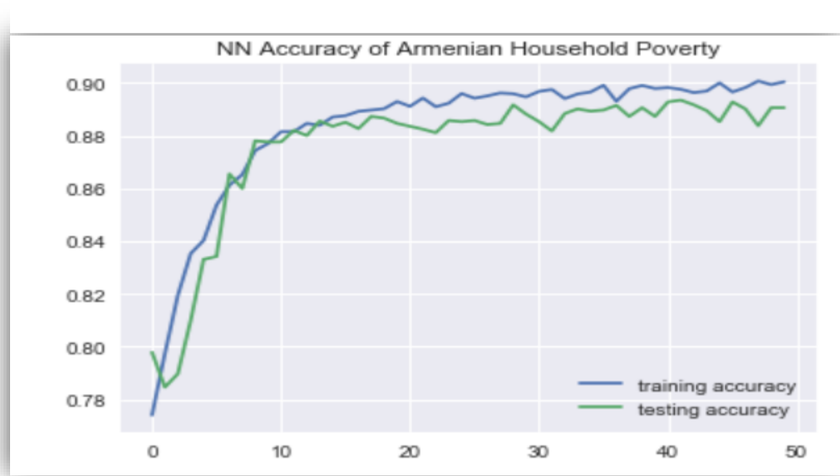


***Figure 7**: Logistic Regression Confusion Matrix*

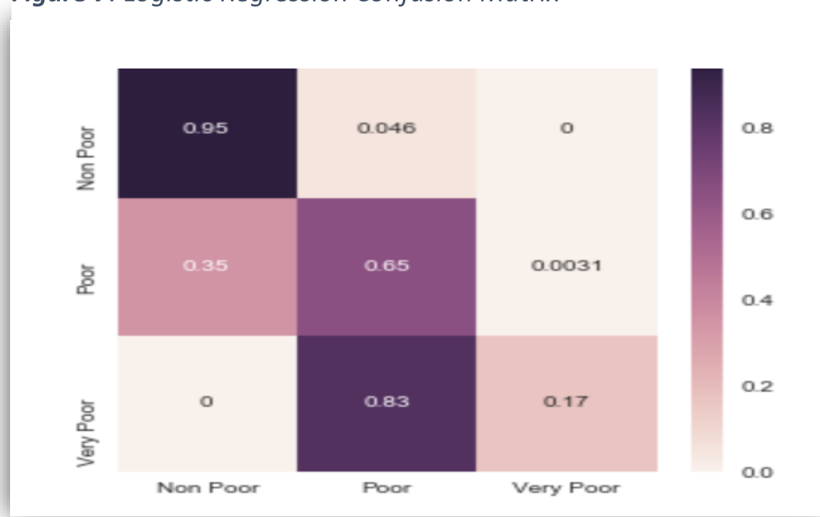
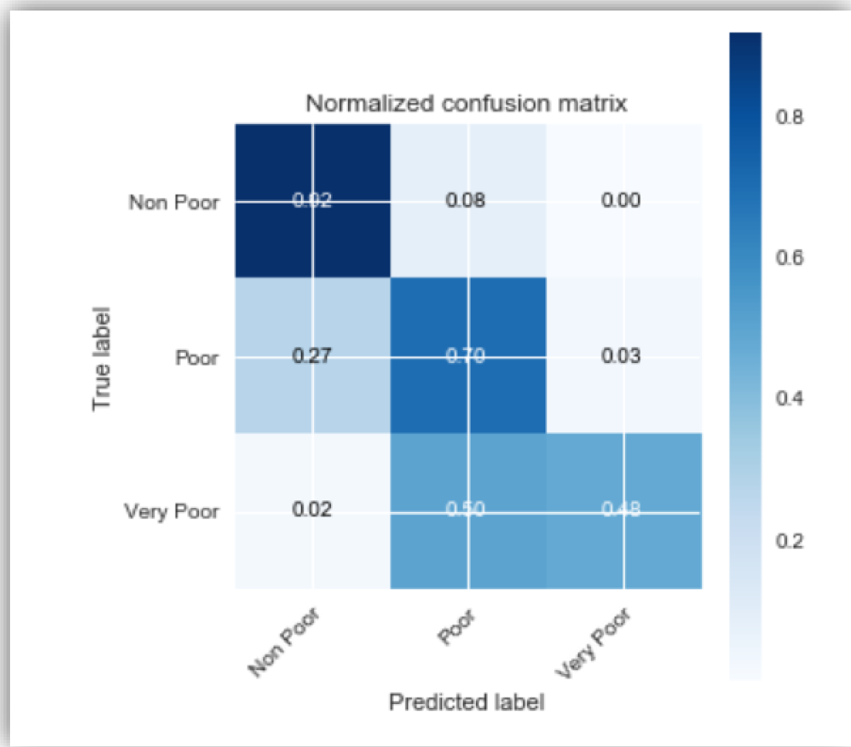


Figure 8 : Decision Tree Confusion Matrix



Endnote

I agree that the paper is posted for an open access to the AUA Community.