ISSN: 2278-4632

## (UGC Care Group I Listed Journal) PREDICTION OF HOUSEHOLD EXPENDITURE ON THE BASIS OF HOUSEHOLD CHARACTERISTICS

 K.VELMURUGAN, Ph.D. Research Scholar, Department of Economics, Thiruvalluvar University, Vellore-632115-Tamil Nadu-India- email id: jayakumarp1988@gmail.com
Dr.C.SIVAKKOLUNDU, Assistant Professor in the Department of Economics, Thiruvalluvar

University, Vellore-632115-Tamil Nadu-India- email id: sivakozhundu.123@gmail.com

## ABSTRACT

There has been an increasing recognition that some aspects of household (HH) expenditures of potential importance for policymaker. It could potentially affect the effectiveness of policy intervention and may even lead to unintended consequences. The objective of this research is to predict three levels of household expenditures on the basis of household characteristics and to identify the significant household characteristics those can affect the household expenditures. For this purpose, data is obtained from Household Integrated Economic Survey (HIES) 2007-08 conducted by Federal Bureau of Statistics (FBS). In this study, dependent variables is household expenditures (low, medium and high), and independent variables are; household income, assets, number of earners, family size, region, highest education level of household, literacy and sex of household head. Multilayer Perceptron Neural Network has been used for data analysis. Results showed that all independent variables are playing a significant role to predict three levels of household expenditures. The order of importance of independent variables is HH income, HH assets, number of earners per HH, HH size, sex of HH head, region, literacy of HH head, and highest education level in household.

KEY WORDS: Household Expenditures, Multilayer Perceptron Neural Network

#### **INTRODUCTION**

Income is the consumption and savings opportunity acquired by an entity within a specified time frame, which is generally expressed in monetary terms. However, for households and individuals, "income is the sum of all the wages, salaries, profits, rents, interests' payments and other forms of earnings received in a given period of time". Household income is habitually a main determinant of household expenditure patterns. The differences between expenditure patterns are largely an indication of differences in income between household groups or individual households.

There are many factors which are influencing the preference on one hand and the income levels on the other hand, and they both are affecting the expenditure patterns. Expenditure patterns of household do not always differ spectacularly, particularly when controlling for income [Punt, et al. (2003)]. While analyzing the household income and consumption expenditures of household, we have to reflect on the sources of income. In Pakistan, there are many sources of income (such as wage and salaries, property owner occupied dwelling, livestock, crop, non-agricultural activities, social insurance benefits which are including gift and assistance, pension, domestic remittances, foreign remittances and other income).

There is a highly positive relationship between income share from rural nonagricultural activities and rising per capita income and a negative relationship between share of agricultural production and per capita income. As household income increases, the propensity to move away from agricultural production to the non-agricultural activities is a pattern similar to Engel's Law, which assumed that poorer households spend more share of their income on food than wealthier households [Akita (1999)]. Family size and dependency are highly correlated with expenditure and considered as important poverty predictors. Proportion of members greater than 65 and children are representing dependency. In rural areas, land, poultry, ownership of livestock, nonresidential and residential property are high positively related to household expenditure. Further, owner cultivator,

# Juni Khyat (UGC Care Group I Listed Journal)

# ISSN: 2278-4632 Vol-13, Issue-05, No.03, May: 2023

medium and large farmers (land ownership of 12.5 acres and greater) play a vital role in distinguishing between non-poor from poor. There are some factors which are creating disparity in the earnings including regional location factors, sex, sector of employment, marriage, and some other characteristics. Mostly youngest and oldest age groups contribute more to overall inequality in the earnings. [Nasir and Mahmood (1998)]. The nature of inequality has varied between income and consumption expenditure. The income inequalities is in the middle class households while consumption inequality is low within middle and lower income group as compared to upper expenditure class [Black low and ray (1999)].

# 2. OBJECTIVES OF STUDY

- To predict three level of household expenditures on the basis of HH characteristics
- .• To identify the significant household characteristics those can affect the HH expenditures

# **3. REVIEW OF LITERATURE**

Akita (2022) conducted a research to explore the factors underlying income inequality in Indonesia by using household data from 1987, 1990 and 1993 national socio economic survey. This was done by using Theil inequality decomposition technique. Results showed that was that inter provincial inequality has not been played an important role in inequality while education and gender inequality were a significant determinant of expenditure inequality. Household expenditure increases as the age of household head increase, but after age of 45 it again decreases. Household size also tended to increases with the age of household head but after the children become independent, again it became small. Income inequality within age group also had a tendency to increases with the age of household head. As household size increases, household expenditure increases while per capita expenditures tended to decrease. Agnes and John (2000) was conducted a research to judge against household's unitary models and collective models by using household data from Indonesia, Bangladesh, Ethiopia, and South Africa. They were presented measures of individual characteristics that were greatly correlated to the bargaining power, assessed at the time of marriage. In addition, the collective model predicted that intra household allocations reflecting the differences between "bargaining power" and preferences of household members within the household. In this study unitary model was rejected. Results suggested that assets which were controlled by women positively and significantly affected household expenditure allocations toward the subsequent generation, such as clothing and education of children. They also examined that parents had distinguishable preferences to their sons and daughters within and across all countries. Kirkpatrick and Tarasuk (2003).

# 4. DATA SOURCES AND ANALYSIS

The data for this study is taken from the Household Integrated Economics Survey (HIES) 2007-08 which conducted by Federal Bureau of Statistics (FBS), india. There are some missing cases related to variables; expenditure (low, medium and high), income, assets, number of rooms, number of earners, age of head, family size, proportion of members less than 5, proportion of out of school children, region of head, proportion of members greater than 65 of households. So we have used only eligible cases. Ordinal Logistic Regression and Multilayer Perceptron Neural Network have been used for data analysis.

# **5. RESULTS AND DISCUSSION**

Multi layer Perception Neural Network has been used for data analysis. Table A-1 represents the network summary analysis 1043(68.3%) cases assigned to the analysis

#### (UGC Care Group I Listed Journal)

# ISSN: 2278-4632 Vol-13, Issue-05, No.03, May: 2023

sample and 485(31.7%) cases to the holdout sample. Total no. of valid cases in this analysis are 1528(100%). Here model excluded 14104 cases from the analysis due to missing which are not match with the independent variables while total no of cases are 15632.

The network information, total no of units used by the model in input layer are13 and no of units in the hidden layers is 5 and units in output layer are 3. In this network model include total 8 independent variables in which 4 represent as factors and remaining 4 are covariates. Standardized rescaling method is used for input layers and softmax activation function used in output layers and the cross entropy error function used.

The category one(low expenditures) has high positive relationship with 1, 4 and 5 units of first hidden layer in this model and negative relation exists with 2 and 3 units of hidden layer. The second category (medium expenditures) show strong positive relationship with 2 and 5 units in the hidden layers and vary low positive relationship with 1 and 3 units of hidden layer and strong negative relationship with 4 unit of hidden layer. 3rd category (high expenditures) show positive relationship with 2 and 5 hidden layers units and having very low negative relationship with 3 unit of hidden layer this is same situation shown in table of parameter estimation of network model (table A-4).

The model summary of training sample and holdout sample which connected to error and incorrect prediction of the response variable household expenditures. Value of cross entropy error functions in this model is 348.09 which are very low corresponds to all sample data and this indicated that our estimated model is better. Training sample contains 14.1% incorrect prediction and in holdout incorrect prediction is 16.3%.

Summary of classification results of network model. Cells on the main diagonal showed the correct classifications and on off diagonal of the cross classifications are showed incorrect classifications of the response category. In training sample 189 cases are correctly classify in low category of expenditure who are previously fall in this category and 291 cases are correctly classify in medium expenditure category as they previously fall in this category and 71 cases incorrectly specify in other two category. 416 cases are correctly classified in high expenditure category and 40 cases incorrectly classify in other two categories. So overall correctly classify in low expenditure category and 20 cases are incorrectly classified in other two categories. 146 cases are correctly classified in other two categories. 146 cases are correctly classified in other two categories. 146 cases are correctly classified in other two categories. 146 cases are correctly classified in other two categories. 181 cases correctly classified in medium category and 40 cases incorrectly specify in other two categories. 181 cases correctly classify in high category and 19 cases incorrectly classify in other two category. Overall correctly classification in holdout sample for all three categories is 83.7%.

For categorical response variable, the predicted-by-observed chart demonstrate clustered of box plots with their predicted pseudo-probabilities for the both the training and testing samples. It is an alternative way to understand the correct classification for the purpose to check realistic importance of Neural Network. On the x-axis this showed the observed response categories, and the legend corresponding to the predicted categories. The part of box plot above the 0.5 on y-axis identified the correct predictions and the part below 0.5 represents incorrect predictions of response categories.

It show the predicted -by-observed chart most left box plot shows, for cases that have observed category for low expenditure but shows the predicted pseudoprobability of category for low in this category. By looking this chart we can say that all portion of box plot is above 0.5 this indicates that predicted pseudo probabilities are very high for low category and box plot for all remaining two categories are below 0.5 and incorrectly predicted in low category. In medium expenditure category the middle plot show all the portion of this box plot above the 0.5 so its predicted pseudo probability are high in medium category and all other categories of low and high expenditure are incorrectly classified in medium category. The next category of high expenditures the portion of last box plot is above 0.5 so it's predicted probability high and correctly classify in

#### (UGC Care Group I Listed Journal)

# high expenditure category and remaining two categories are incorrectly classified in high category of expenditure.

The receiver OC curve (ROC) and three curves; 1st for the low expenditures category, 2nd for the category of medium expenditures and 3rd category for high expenditures. The curves of these three categories are very close to upper left corner of the plot; which indicating that our model is better fit for given data and area covers for the category low expenditures is 0.976, area under the curve covers for the medium expenditures is .937and area covers for the category for high expenditures is 0.979 see more in table A-6. By looking this table we can say that mostly area covered closed to 1 which indicates perfect fit on 1.

The cumulative gains chart that is displayed the percentage of the overall numbers of cases in a given category is "gained" by targeting a percentage of the total number of cases. In this case, very first point on the curve against the low expenditures category is at (10%, 48%), which tell us that if you score a data set with the network and sort out all of the cases by predicted pseudo-probability of low expenditures category. This indicates that, the top 10% to contain approximately 48% of all of the cases that actually gained in category low expenditures. Figure A-5 shows the lift chart that is derived from the cumulative gain chart; the values on the y axis correspond to the ratio of the cumulative gain for each curve to the baseline. Thus, the lift at 10% for the category 48% / 10% = 4.8 it means approximately 4 cases correctly specify at top 10% of low expenditures.

The importance of independent variables. In this network model, 4 variables are playing very important role in predicting the expenditure categories of low, medium and high. This network model gives highest importance to Household (HH) income (100) than all others and second importance gives to household (HH) assets (72.3%) third important variable is no of earners per household (28.8%), forth importance give to household size (25.4%). In the light of all above results related to importance of predictor we conclude that all that play a significant in predicting the expenditures level.

## 6. CONCLUSION

Results shows that household income, assets, number of earners, family size, region and sex of household head are significant variables those can best discriminate the respondents among levels (low, medium and high) of household expenditure while highest education level and literacy are insignificant variables. These significant variables can help to predict the respondents among the levels of household expenditures.

Results from ordinal logistic regression shows that in first category of expenditures 75.3% cases, in second category, 78.8% cases while in last category 86.7% cases are correctly classified and overall correct classification for all three categories is 80.26%.

Neural network shows that four variables are playing most important role in predicting the expenditure categories of low, medium and high. The network has given highest importance to HH income rather than all others, second importance is given to HH assets, third important variable is number of earners per household, and forth importance is given to household size. Overall correct classification by the network for all three categories is 83.7%.

## REFERENCES

1. Agnes, R. Q. and John, A. M. (2000). Intrahousehold Allocation and Gender Relations: New Empirical Evidence from Four Developing Countries. International Food Policy Research Institute, FCND discussion paper 84, Washington DC.

2. Akita, T., Lukman, R.A. and Yamada, Y. (1999). Inequality in the Distribution of Household Expenditures in Indonesia: A Theil Decomposition Analysis. The Developing Economies, 37(2), 197-221.

3. Aslam, M. (2008). Gender and Household Education Expenditure in Pakistan. Applied Economics, 40(20), 2573-2592.

### ISSN: 2278-4632 Vol-13, Issue-05, No.03, May: 2023

#### (UGC Care Group I Listed Journal)

4. Blacklow, P. and Ray, R. (1999). A Comparison of Income and Expenditure Inequality Estimates: The Australian Evidence, 1975/76 to 1993/94. The Australian Economics Review, 33(4), 317-329.

5. Clement, M. (2011). Remittances and Household Expenditure Patterns in Tajikistan: A Propensity Score Matching Analysis. Cahiers du GREThA, 9, 1-25

6. Government of Pakistan (2009). Houshold Integrated Economic Survey 2007-08. Federal Bureau of Statistics, Islamabad, Pakistan.

7. Guzmán, J. C., Morrison, A. R. and Sjöblom, M. (2006). The Impact of Remittances and Gender on Household Expenditure Patterns: Evidence from Ghana. Gender and Development (PRMGE), World Bank. 8. Kirkpatrick, S. and Tarasuk, V. (2003). The Relationship between Low Income and Household Food Expenditure Patterns in Canada. Public Health Nutrition, 6(6), 589-597.

9. Lemay, P. (2006). Are Household Expenditure Decisions Cooperative? Evidence from Benin. CRED, University of Namur. Unpublished Paper.

10. Morrison, A.R., Schiff, M. and Sjoblom, M. (2008). The International Migration of Women. Co-Publication of World Bank and Palgrave Macmillan, 125-152.

11. Nasir, Z.M. and Mahmood, R. (1998). Personal Earnings Inequality in Pakistan: Findings from the HIES 1993-94. The Pakistan Development Review, 37(4), 781-792.

12. Punt, C., Pauw, K., Mohube, E., Gilimani, B., Rantho, L., Leaver, R., McDonald, S., Chant, L. and Valente, C. (2003). Household Expenditure Patterns in South Africa – 1995. PROVIDE Project Background Paper, 2, 1-16.

13. Rivera, J.J.M. and González, J.A. (2009). Effects of Remittances on Household Expenditure Patterns of Rural Mexico. Unpublished paper.