

# **RISK FACTORS AFFECTING POVERTY IN NEPAL: STATISTICAL MODELING APPROACH**



A THESIS SUBMITTED TO THE  
CENTRAL DEPARTMENT OF STATISTICS  
INSTITUTE OF SCIENCE AND TECHNOLOGY  
TRIBHUVAN UNIVERSITY  
NEPAL

FOR THE AWARD OF  
DOCTOR OF PHILOSOPHY  
IN STATISTICS

BY  
KRISHNA PRASAD ACHARYA

**July 2023**



**RISK FACTORS AFFECTING POVERTY IN NEPAL:  
STATISTICAL MODELING APPROACH**



A THESIS SUBMITTED TO THE  
**CENTRAL DEPARTMENT OF STATISTICS**  
**INSTITUTE OF SCIENCE AND TECHNOLOGY**  
**TRIBHUVAN UNIVERSITY**  
**NEPAL**

**FOR THE AWARD OF**  
**DOCTOR OF PHILOSOPHY**  
**IN STATISTICS**

BY  
**KRISHNA PRASAD ACHARYA**

**July 2023**



TRIBHUVAN UNIVERSITY  
Institute of Science and Technology  
**DEAN'S OFFICE**

Kirtipur, Kathmandu, Nepal

Reference No.:

**EXTERNAL EXAMINERS**

The Title of Ph.D. Thesis: **"Risk Factors Affecting Poverty in Nepal: Statistical Modeling Approach"**

Name of Candidate: **Krishna Prasad Acharya**

**External Examiners:**

- (1) Prof. Dr. Rabindra Kayastha  
School of Science  
Kathmandu University, NEPAL
- (2) Prof. Sada Nanda Dwivedi  
International Centre for Health Research (ICHR)  
RD Gardi Medical College  
Madhya Pradesh, INDIA
- (3) Prof. Dr. Deepak Sanjel  
Department of Mathematics and Statistics  
Minnesota State University  
Mankato, U.S.A.

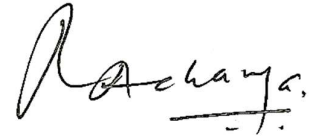
January 5, 2024

Dr. Surendra Kumar Gautam  
(Asst. Dean)

## DECLARATION

This thesis entitled “**Risk Factors Affecting Poverty in Nepal: Statistical Modeling Approach**” which is being submitted to the Central Department of Statistics, Institute of Science and Technology (IOST), Tribhuvan University, Nepal for the award of the degree of Doctor of Philosophy (Ph. D.), is a research work carried out by me under the supervision of Prof. Dr. Shankar Prasad Khanal, Central Department of Statistics, Tribhuvan University and co-supervised by Prof. Dr. Devendra Chhetry.

This research is original and has not been submitted earlier in part or full in this or any other form to any university or institute, here or elsewhere, for the award of any degree.

A handwritten signature in black ink, appearing to read 'Acharya', with a horizontal line underneath the name.


Krishna Prasad Acharya



## RECOMMENDATION

This is to recommended that **Krishna Prasad Acharya** has carried out research entitled “**Risk Factors Affecting Poverty in Nepal: Statistical Modeling Approach**” for the award of Doctor of Philosophy (Ph. D.) in **Statistics** under our supervision. To our knowledge, this work has not been submitted for any other degree.

He has fulfilled all the requirements laid down by the Institute of Science and Technology (IOST), Tribhuvan University, Kirtipur for the submission of the thesis for the award of Ph. D. degree.



**Dr. Shankar Prasad Khanal**  
**Supervisor**  
**Professor**  
Central Department of Statistics  
Tribhuvan University  
Kirtipur, Kathmandu, Nepal



**Dr. Devendra Chhetry**  
**Co-Supervisor**  
**Professor**  
Central Department of Statistics  
Tribhuvan University  
Kirtipur, Kathmandu, Nepal

**July 2023**



TRIBHUVAN UNIVERSITY  
CENTRAL DEPARTMENT OF STATISTICS  
OFFICE OF THE HEAD OF DEPARTMENT  
Kirtipur, Nepal

---

## LETTER OF APPROVAL

On the recommendation of Prof. Dr. Shankar Prasad Khanal and Prof. Dr. Devendra Chhetry, this Ph. D. thesis submitted by Krishna Prasad Acharya, entitled “**Risk Factors Affecting Poverty in Nepal: Statistical Modeling Approach**” is forwarded by Central Department Research Committee (CDRC) to the Dean, IOST, T. U..

**Dr. Gauri Shrestha**  
Professor,  
Head,  
Central Department of Statistics,  
Tribhuvan University  
Kirtipur, Kathmandu  
Nepal

## ACKNOWLEDGEMENTS

I wish to sincerely thank my Supervisor Prof. Dr. Shankar Prasad Khanal and co-supervisor, Prof. Dr. Devendra Chhetry for their guidance and support throughout my research. I thank them for providing me excellent guidance, brilliant ideas, scholarly advice and moral support during the course of my research. Their outstanding academic and research skills as well as thoughtful suggestions helped me excel my PhD work and writing research articles.

I am grateful to the Head of the Central Department of Statistics, Prof. Dr. Gauri Shrestha for her moral support, cooperation and encouragement. I am also grateful to senior faculties Prof. Dr. Srijan Lal Shrestha, Prof. Dr. Tika Ram Aryal, Prof. Dr. Chandramani Poudel and Prof. Dr. Ram Prasad Khatiwada, of Central Department of Statistics T. U., for their support and encouragement.

I am thankful to the Central Department Research Committee (CDRC) of Central Department of Statistics Tribhuvan University for its valuable comments and suggestions and to Prof. Dr. Kamal Deep Dhakal and Prof. Dr. Basanta Dhakal for their personal support and inspiration. I would like to thank Mr. Dipak Ratna Shakya, Associate Prof. of English, for English language editing.

Thanks are also due to Mr. Prabhat Upreti, Mr. Santosh Kumar Sah, Mr. Arun Kumar Yadav, Mr. Ishwori Prasad Banjade, Mr. Madhav Prasad Bhusal lecturers of the Central Department of Statistics and Dr. Rajendra Man Shrestha (Associate Professor of Padma Kanya Campus) for their valuable support.

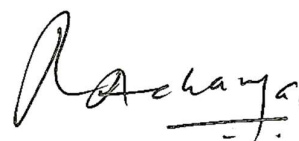
I would like to acknowledge to the University Grants Commission Nepal for providing me Ph. D. fellowship, and the then Central Bureau of Statistics (CBS) now National Statistics Office (NSO) for providing NLSS III data. My sincere thanks go to Prof. Leigh Blizzard, University of Tasmania who has provided STATA codes for calculation of H-L Chi-square in log-binomial model.

Special thanks are due to Dr. Hem Raj Regmi, Mr. Ram Hari Gahire, Mr. Dinesh Bhattarai, and Mr. Nanda Lal Sapkota of NSO and Mr. Ramesh Raj Paudel (Department of Archaeology, Government of Nepal) for providing secondary literature and other relevant information for this research. I also acknowledge Mr. Ramesh Maharjan for providing technical support in computer work for setting the format of thesis.



I likewise owe many thanks to my elder brothers Dr. Mahadev Sharma (Senior Scientist of Ministry of Natural Resources and Forestry, Ontario, Canada), Khaga Raj Acharya (Lecturer of Mathematics at Birendra Multiple Campus, Bharatpur), and nephew Er. Himal Acharya for their support and cooperation during the period of my research work.

Finally, I have no words to express my sincere gratitude to my late father Devi Prasad Acharya and mother Hari Kala Acharya who worked very hard for the wellbeing of their children and to educate them. I would also like to express my heartfelt thanks to my wife Kalpana Pandit and sons, Mr. Milan Acharya and Ganesh Acharya for their unlimited support and encouragement during my research work.

A handwritten signature in black ink, appearing to read 'Acharya' with a horizontal line underneath the 'ya' part.

Krishna Prasad Acharya

July 2023

## ABSTRACT

Poverty is one of the main problems of developing countries, like Nepal and its reduction is a central issue. The identification of its determinants to reduce the monetary poverty is one of the key issues. According to previous studies, log-binomial regression model (LBRM) is a good option to logistic regression model (LRM) for common outcomes, mostly used in the analysis of clinical and epidemiological data. However, the use of LBRM and the comparison with LRM for data on poverty has not been discussed yet. The objectives of this study are to identify the important risk factors, to compare the LRM and LBRM in identifying the risk factors and estimating their effects on poverty in Nepal, and to assess the stability of the model through bootstrapping method. The data used for the analysis is the cross sectional household level data (n = 5988) of Nepal Living Standard Survey 2010/11. All the data required for this study are not available in the provided household level data file of 5,988 households but are available in the individual level data file of 28,670 individuals. The individual level data are converted into household level data in order to generate the data on a number of variables, and merged into the main data file. With the support of rigorous review of literature and the availability of the variables in the dataset, seven possible independent variables have been considered for both the LRM and LBRM. They are: sex of household head (female / male), literacy status of household head (illiterate / literate), status of remittance recipient of household (no / yes), status of land ownership (no / yes), household with access to nearest market center (poor / better), number of children under 15 years (more than two / at most two), and number of literate members of working age population (WAP) (none / at least one). The response variable is household poverty (poor / non-poor). Implementing the stepwise forward and backward selection procedure with all these seven variables for the development of each final multiple regression model, only six variables except sex of household head has come out statistically significant at 5% level of significance. The LRM has yielded the odds ratio (OR) and LBRM has yielded risk ratio (RR) with 95% confidence interval estimate (CIE) for each covariate. Diagnostics of the model, the goodness of fit test, a risk assessment based on the presence of variables, and the stability of each model has been carried out. The classification and discrimination of the LRM has been also assessed. LRM and LBRM have been compared with respect to different criteria such as selection of covariates, effect size and its precision. The model's good fit test using  $H-L(\chi^2)$  and test of model's diagnostics criteria has also been compared. Further, the comparisons have also been

made in risk assessment on the basis of factors present in the model, stability of the model and convergence failure problem. The effect size in terms of OR and in RR of six factors in each final model namely illiterate household head (OR: 2.20, 95% CIE: 1.86 – 2.61,  $p < 0.001$ ; RR: 1.68, 95% CIE: 1.49 – 1.89,  $p < 0.001$ ), remittance non recipient household (OR: 1.90, 95% CIE: 1.64 – 2.20,  $p < 0.001$ ; RR: 1.45, 95% CIE: 1.33 – 1.59,  $p < 0.001$ ), household with no land holdings (OR: 1.53, 95% CIE: 1.31 – 1.78,  $p < 0.001$ ; RR: 1.22, 95% CIE: 1.11 – 1.34,  $p < 0.001$ ), household with poor access to market center (OR: 1.77, 95% CIE: 1.52 – 2.07,  $p < 0.001$ ; RR: 1.51, 95% CIE: 1.34 – 1.69,  $p < 0.001$ ), household having > 2 children aged under 15 (OR: 4.69, 95% CIE: 4.06 – 5.42,  $p < 0.001$ ; RR: 2.96, 95% CIE: 2.66 – 3.28,  $p < 0.001$ ) and household not having literate members of WAP (OR: 1.29, 95% CIE: 1.07 – 1.56,  $p < 0.001$ ; RR: 1.16, 95% CIE: 1.05 – 1.29,  $p < 0.001$ ) are significantly associated with the likelihood of poverty. For each covariate, the OR is overestimated than that of RR. There is narrower 95% CIE of RR than that of OR for each covariate. It shows that RR is more precise than OR. Greater elevation in risk in LRM compared to LBRM varies from 13% to 173%. In each model, there is no convergence issues have been countered, where both the models are equally stable as assessed by bootstrapping procedure. Almost all variables are repeated 100% times among 1000 times repetition. The visual assessments of diagnostics of each model are reasonably satisfactory. There is considerable acceptable discrimination of LRM (AUC: 0.78) and model correct classification values of 67.15%. The good fit of the model is satisfied by LRM [ $H-L(\chi^2)$  with 8 d.f.= 6.05,  $p = 0.53$ ] but not satisfied by LBRM [ $H-L(\chi^2)$  with 8 d.f.= 28.60,  $p = 0.0004$ ]. Since the LRM satisfied the majority of requirements of model performance instead of some limitations, this model seems to be better than the LBRM for this data set. Nevertheless, the LBRM is an option for the LRM since it has better accuracy and avoids overestimating effect size. The findings of this study is expected to be useful for researchers and policy makers in the relevant field.

**Keywords:** Odds ratio (OR), risk ratio (RR), diagnostics, elevation in risk, good fit, log-binomial, logistic, poverty

## LIST OF ACRONYMS AND ABBREVIATIONS

ADB	: Asian Development Bank
AIC	: Akaike Information Criterion
AUC	: Area under the Curve
BIC	: Bayesian Information Criterion
CBN	: Cost of Basic Needs
CBS	: Central Bureau of Statistics
CI	: Condition Index
CIE	: Confidence Interval Estimate
CN	: Condition Number
CQGs	: Consumption Quintile Groups
FGT	: Foster, Greer and Thorbecke
GDP	: Gross Domestic Product
GNP	: Gross National Product
H-L	: Hosmer and Lemashow
LBRM	: Log-binomial regression model
LL	: Log-likelihood
LRM	: Logistic regression model
LSMS	: Living Standard Measurement Survey
MoF	: Ministry of Fininance
MoH	: Ministry of Health
NDHS	: Nepal Demographic Health Survey
NLSS	: Nepal Living Standard Survey

NPC	: National Planning Commission
NRB	: Nepal Rastra Bank
OLS	: Ordinary Least Square
OR	: Odds Ratio
PCA	: Principle Component Analysis
PGI	: Poverty Gap Index
PR	: Prevalence Ratio
ROC	: Receiver Operating Characteristics
RR	: Risk Ratio or Relative Risk
SDGs	: Sustainable Development Goals
SE	: Standard Error
SES	: Socio-economis Status
SPSS	: Statistical Packages for Social Sciences
TFR	: Total Fertility Rate
UN	: United Nations
UNDP	: United Nations Development Program
UNFPA	: United Nations Population Fund
VIF	: Variance Inflation Factor
WAP	: Working Age Population
WB	: World Bank

## LIST OF SYMBOLS

$\alpha$	: Alpha
$b$	: Regression coefficient
$\beta$	: Beta
$\pi$	: Pai
$D$	: Deviance
$H$	: Hat matrix
$\Delta$	: Delta
$\chi^2$	: Chi-square
$<$	: Less than
$>$	: Greater than
$\geq$	: Greater than or equal to
$\leq$	: Less than or equal to
$\sum$	: Summation
$\wedge$	: Hat
$*$	: Star
$\ln$	: Logarithm

## LIST OF TABLES

	Page No.
<b>Table 1: Poverty Trends in Nepal at the National Level</b>	<b>9</b>
<b>Table 2: Poverty Trends in Nepal by Urban-Rural</b>	<b>10</b>
<b>Table 3: Poverty Trends in Nepal by Ecological Region</b>	<b>11</b>
<b>Table 4: Measures of Poverty by Social Group in 2010/11</b>	<b>12</b>
<b>Table 5: Estimated Head Count Ratio in % after 2010/11</b>	<b>13</b>
<b>Table 6: Poverty Incidence (%) of Households Grouped by Number of Children</b>	<b>50</b>
<b>Table 7: Selected Covariates and Response Variable with Group Formation and Coding Scheme</b>	<b>51</b>
<b>Table 8: Theoretical Classification Table Based on the Multiple LRM</b>	<b>63</b>
<b>Table 9: Layout of Computation of RR and OR</b>	<b>70</b>
<b>Table 10: Bootstrap Replication Matrix</b>	<b>76</b>
<b>Table 11: Nominal Per Capita Consumption and Income by Quintile</b>	<b>77</b>
<b>Table 12: Percentage Share of Each Quintile Group Income and Consumption</b>	<b>78</b>
<b>Table 13: Comparison of Mean Number of Three Population Groups across CQGs</b>	<b>79</b>
<b>Table 14: Comparison of Literacy Rate of WAP by Gender across CQGs</b>	<b>80</b>
<b>Table 15: Three Measures of Poverty for Fourteen Groups of Household Population</b>	<b>81</b>
<b>Table 16: Results of Bivariate Analysis (n = 5988)</b>	<b>82</b>
<b>Table 17: Regression Estimates of LRM with 95% CIE (n = 5988)</b>	<b>84</b>
<b>Table 18: Role of Remittance in Association with Independent Variables</b>	<b>85</b>
<b>Table 19: Correlation Matrix of Coefficients of LRM</b>	<b>87</b>
<b>Table 20: Collinearity Diagnostics</b>	<b>88</b>
<b>Table 21: Correct Classification Details of the Model</b>	<b>89</b>



<b>Table 22:</b>	<b>Correct Classification Values</b>	<b>89</b>
<b>Table 23:</b>	<b>Regression Coefficient for <math>\hat{y}</math> and <math>\hat{y}^2</math></b>	<b>91</b>
<b>Table 24:</b>	<b>Distribution of Households and OR on the Basis of Presence of Number of Factors</b>	<b>96</b>
<b>Table 25:</b>	<b>Results of the Bootstrap Resampling Procedure for LRM</b>	<b>98</b>
<b>Table 26:</b>	<b>Results of LBRM Considering One Variable at a Time (n = 5988)</b>	<b>99</b>
<b>Table 27:</b>	<b>Regression Estimates of LBRM with 95% CIE (n = 5988)</b>	<b>100</b>
<b>Table 28:</b>	<b>Distribution of Households and RR on the Basis of Presence of Number of Factors</b>	<b>103</b>
<b>Table 29:</b>	<b>Results of the Bootstrap Resampling Procedure for LBRM</b>	<b>104</b>
<b>Table 30:</b>	<b>Comparison of LRM and LBRM in terms of Different Parametrs (n = 5988)</b>	<b>106</b>
<b>Table 31:</b>	<b>Results of the Bootstrap Resampling Procedure for LRM vs. LBRM</b>	<b>109</b>
<b>Table 32:</b>	<b>Comparison of Models' Results</b>	<b>111</b>

## LIST OF FIGURES

	Page No.
<b>Figure 1: Schematic Diagram of Conceptual Framework</b>	<b>44</b>
<b>Figure 2: Comparison of Percentage Share of Food and Non-food Expenditure</b>	<b>79</b>
<b>Figure 3: Sensitivity / Specificity and Predicted Probability</b>	<b>90</b>
<b>Figure 4: ROC Curve</b>	<b>90</b>
<b>Figure 5: Plot of <math>\Delta\chi_j^2</math> and Estimated Probability from the Fitted Multiple LRM with Covariate Pattern J = 60</b>	<b>92</b>
<b>Figure 6: Plot of <math>\Delta D</math> and estimated probability from the fitted multiple LRM with covariate pattern J = 117</b>	<b>93</b>
<b>Figure 7: Plot of <math>\Delta\beta</math> and Estimated Probability from the Fitted Multiple LRM with Covariate Pattern J = 60</b>	<b>94</b>
<b>Figure 8: Plot of <math>\Delta\chi^2</math> and Predicted Probability of LRM with Size of the Symbol Proportional to <math>\Delta\beta</math>, Covariate Pattern J = 60</b>	<b>95</b>
<b>Figure 9: OR in Presence of Number of Risk Factors</b>	<b>97</b>
<b>Figure 10 (a): Leverage and Fitted Values of the LBRM</b>	<b>102</b>
<b>Figure 10 (b): Graph of <math>\Delta\chi^2</math> and predicted Values of LBRM with Plotting Symbol Proportional to Cook's Distance</b>	<b>103</b>
<b>Figure 11 (a): Graph of <math>\Delta\beta</math> and Estimated Probability (from LRM)</b>	<b>107</b>
<b>Figure 11 (b): Leverage and Fitted Values of LBRM</b>	<b>108</b>
<b>Figure 12 (a): Graph of <math>\Delta\chi^2</math> and Predicted Probabaility of LRM with Symbol Size Proportional to <math>\Delta\beta</math></b>	<b>108</b>
<b>Figure 12 (b): Graph of <math>\Delta\chi^2</math> and Predicted Probability of LBRM with Plotting Symbol Proportional to Cook's Distance</b>	<b>109</b>
<b>Figure 13: OR and RR in Presence of Number of Risk Factors</b>	<b>110</b>

# TABLE OF CONTENTS

	<b>Page No.</b>
External Examiners	ii
Declaration	iii
Recommendation	iv
Letter of Approval	v
Acknowledgements	vi
Abstract	viii
List of Acronyms and Abbreviations	x
List of Symbols	xii
List of Tables	xiii
List of Figures	xv
<b>CHAPTER 1</b>	
<b>1. INTRODUCTION</b>	<b>1</b>
1.1 Absolute Poverty	1
1.2 Refinements in Absolute Poverty	2
1.3 Contemporary Measures of Monetary Poverty	4
1.4 Poverty Measuring Practices in Nepal	7
1.5 Poverty Scenarios of Nepal	9
1.5.1 National Level Poverty Trends	9
1.5.2 Rural-Urban Poverty Trends	10
1.5.3 Regional Level Poverty Trends	10
1.5.4 Poverty Measures by Social Group	11
1.5.5 Head Count Ratio of Nepal after 2010/11	13
1.6 Statement of the Problem	13
1.7 Rationale of the Study	15
1.8 Objectives of the study	16
1.9 Research Questions and Hypotheses	16

1.10	Significance of the study	17
1.11	Limitation of the study	17
1.12	Chapter organization	18
<b>CHAPTER 2</b>		
<b>2.</b>	<b>LITERATURE REVIEW</b>	<b>19</b>
2.1	Understanding Rapid Decline of Poverty in Nepal	19
2.2	Economic Growth and Poverty Reduction	22
2.4	Statistical Methods/Models used in Poverty Analysis	40
2.5	Conceptual Framework	43
<b>CHAPTER 3</b>		
<b>3.</b>	<b>MATERIALS AND METHODS</b>	<b>45</b>
3.1	Data Source and Data File Preparation	45
3.2	Scheme of Data Analysis	45
3.3	Outcome Variable and Covariates	47
3.4	Dichotomization of Quantitative Variable	48
3.4.1	Dichotomization: Area of Land Holding	49
3.4.2	Dichotomization: Access to Nearest Market Center	49
3.4.3	Dichotomization: Number of Children	49
3.4.4	Dichotomization: Number of Literate Working Age Population (WAP)	50
3.5	Statistical Models	51
3.5.1	Selection of Variables for the Logistic Regression Model (LRM)	51
3.5.2	Logistic Regression Model (LRM) and its Fitting	52
3.5.3	Fitting of the Multiple Logistic Regression Model (LRM)	53
3.5.4	Test of Significance of the Fitted Model	54
3.5.5	Test of Significance of Individual Regression Coefficient	55
3.5.6	Confidence Interval for Regression Coefficient	56
3.5.7	Interpretations of Regression Coefficient	56

3.5.8	Coefficient of Determination ( $R^2$ ) in Logistic Regression	57
3.5.8.1	McFadden's $R^2$	58
3.5.8.2	Cox-Snell $R^2$	58
3.5.8.3	Nagelkerke $R^2$	59
3.5.9	Test of Goodness of Fit of the Model	59
3.5.9.1	Hosmer and Lemeshow Test	60
3.5.9.2	AIC and BIC Statistic	61
3.5.10	Classification and Discrimination of the Model	62
3.5.10.1	Description of Classification Table	62
3.5.10.1.1	Sensitivity, Specificity and Accuracy	63
3.5.10.1.2	Area under Receiver Operating Characteristic (ROC) Curve	64
3.5.11	Model Specification Test	64
3.5.12	Diagnostics of the Logistic Regression Model (LRM)	65
3.5.13	Assessment of Risk on the Basis of Factors Present in the Model	69
3.6	Log-binomial Regression Model (LBRM)	69
3.6.1	Log-binomial Regression Model (LBRM) and its Fitting	71
3.6.1.1	Maximum Likelihood Estimation for Log-binomial Regression Model (LBRM)	71
3.6.2	Interpretation of RR	72
3.6.3	Goodness-of-Fit Test of the Model	72
3.6.3.1	AIC and BIC for the Model	73
3.6.4	Diagnostics of the Log- binomial Model (LBRM)	73
3.6.5	Assessment of Risk for Different Factors	74
3.7	Comparison of the Models	74
3.7.1	Comparison of Models Based on Variable Selection	74
3.7.2	Comparison of Models Based on Individual Regression Coefficient	75
3.7.3	Comparison of Models Based on the Goodness of Fit and Diagnostic Criteria	75

3.7.4 Comparison of Models Based on the Robustness Criteria	76
3.8 Software Used for Statistical Analysis	76
<b>CHAPTER 4</b>	
<b>4. RESULTS AND DISCUSSION</b>	<b>77</b>
4.1 Analysis of Economic Characteristics	77
4.1.1 Per Capita Income and Expenditure	77
4.1.2 Percentage Share of Income and Expenditure	78
4.1.3 Share of Food and Non-food Expenditure	78
4.1.4 Analysis of Socio-demographic Characteristics	79
4.2 Poverty Indices	80
4.3 Association of Covariates with Response Variable	82
4.4 Results of Logistic Regression (LRM)	83
4.4.1 Assessment of Multicollinearity among Independent variables	87
4.4.2 Results of Sensitivity, Specificity and Correct Classification of the Model	88
4.4.3 ROC Curve for Model Discrimination	90
4.4.4 Results of Model Specification Test	91
4.4.5 Results of Diagnostics of the Fitted Multiple LRM	91
4.4.5.1 Plot of delta Chi-square and Estimated Probability	91
4.4.5.2 Plot of Changes in the Deviance ( $\Delta D$ ) and Estimated Probability	92
4.4.5.3 Graph of $\Delta\beta$ and Predicted Probability	93
4.4.5.4 Graph of $\Delta\chi^2$ and Predicted Probability with Symbol Size Proportional to $\Delta\beta$	94
4.4.6 Results of Risk Assessment on the basis of Factors Present in the Model	95
4.4.7 Stability of the Model	97
4.5 Results of Log-binomial Regression Model (LBRM)	99
4.5.1 Results of Diagnostics for the LBRM	101

4.5.2	Results of Risk Assessment on the basis of Factors Present in Log-binomial Model	103
4.5.3	Results of Stability of LBRM	104
4.5.6	Results of Comparison of Logistic and Log- binomial Regression Models	105
4.5.6.1	Results of Comparison with reference to Diagnostics	107
4.5.6.2	Results of Comparison Based on Stability of the Model	109
4.5.6.3	Results of Comparison Based on Risk Assessment	110
<b>CHAPTER 5</b>		
<b>5.</b>	<b>CONCLUSION AND RECOMMENDATIONS</b>	<b>114</b>
5.1	Conclusion	114
5.2	Recommendations	116
5.3	Further Study	117
<b>CHAPTER 6</b>		
<b>6.</b>	<b>SUMMARY</b>	<b>118</b>
<b>REFERENCES</b>		<b>122</b>
APPENDIX– A		139
APPENDIX– B		140
APPENDIX– C		143
APPENDIX– D		144



# CHAPTER 1

## 1. INTRODUCTION

Poverty is a multidimensional concept and has been defined, measured and practiced in several ways for the century. Broadly speaking, the developed concepts of poverty and their measurements can be categorized into two categories: monetary and non-monetary poverty. This study focuses on monetary poverty, specifically on the absolute poverty, then the historical development of the concept and measurement of absolute poverty with some scenarios of absolute poverty of Nepal over time and space. Finally this chapter presents statement of the problem, rationale, objectives, significance, limitation and chapter organization of this study.

### 1.1 Absolute Poverty

The definition and measurement of absolute poverty was first developed at the beginning of the 20<sup>th</sup> century by Charles Booth (1903) and Benjamin Seebhom Rowntree (1901). Charles Booth conducted study in London between 1886 and 1903 and the results of seminal work were published in multi-volume book, entitled *Life and Labor of the People in London*. Whereas Rowntree conducted study in York between 1899 and 1901, and the results of his seminal work were published in a book, entitled *Poverty, A Study of Town Life*. In both of the studies, poverty is defined as the lack of resources necessary to fulfill essential physical needs to an adequate standard (Niemi, 2011). Physical needs includes nutrition, shelter and clothing. The definition of Booth and Rowntree is based on the idea of ‘subsistence needs’ – what a person needs to survive. In order to operationalize the definition, they introduced the concept of poverty line - minimum amount of income/expenditure necessary for a family/individual to subsist. They conducted landmark surveys for collecting data on the living conditions of families from their respective study areas. Finally, based on survey data, they estimated poverty line as the cost of consumption baskets containing goods for subsistence needs.

The poverty line they estimated can be viewed as the minimal amount of money needed to keep a family/individual out of poverty and it demarcates the families/individuals into two groups, namely poor and non-poor. The proportion of poor they computed vary from one study to another in between 27.5 percent to 30.0 percent. Their works also

influenced government policy regarding poverty in the early 20<sup>th</sup> century. Their works helped initiate old age pension and free school meals for the poorest children.

There were several weaknesses in the methodology they adopted in the survey design as well as in the methodology they adopted in estimating poverty lines (Spicker, 1990). However, their basic strategy adopted in estimating poverty line and counting poor is pioneering work in the measurement of monetary poverty, and this strategy still prevails in estimating and monitoring national level poverty in developing and developed nations with some refinements.

## **1.2 Refinements in Absolute Poverty**

Some refinements in the definition of poverty, survey methodology, estimation of poverty line, and many more adopted by Booth and Rowntree were reckoned necessary due to several factors, such as the rise in living standards of populations, emergence of new ideas and availability of scientific tools and techniques. Some of the refinements are briefly discussed below.

In due course of time the concept of “basic needs” was introduced in the definition of absolute poverty instead of the concept of “subsistence needs”. The basic needs in addition to subsistence needs include basic facilities and services. For example, healthcare, sanitation and education are required for long-term physical well-being of peoples. This new concept was introduced by the International Labor Organization at the World Employment Conference in 1976. This concept in absolute poverty defines the absolute minimum resources necessary for long-term physical well-being, usually in terms of consumption goods.

Over time, various sampling techniques based on notion of probability sampling were developed that helped to select nationally and sub-nationally representative samples even with small sample size. These techniques enabled many worldwide agencies to conduct income/consumption surveys for estimating absolute poverty with less cost at the national and subnational level with more scientific and rigorous manner. As for example, Nepal since 1996 has been adopting two-stage stratified random sampling method in the selection of households for the Nepal Living Standard Survey (NLSS) whose one of the objectives is to provide estimates of contemporary measures of absolute poverty at the national and subnational levels.

In due course of time it was realized that in addition to data on income and consumption, data on relevant socio-economic, demographic and living standards would also be needed for better understanding of poverty. As a result, “effective questionnaires” were developed by research scholars (for example Grosh & Glewwe, 1998) for collecting the relevant data. This refinement enabled research scholars for establishing the linkage between poverty and development indicators using appropriate simple to complex statistical methods.

In addition to the simple measure of ‘proportion of poor’ some new measures such as intensity of poverty and severity of poverty were also developed in 1984 (to be discussed in later section of this chapter) and used by many developing countries. These two newly added measures helped academicians to understand poverty more in depth and policy makers to formulate policies for reducing poverty. Nepal has been reporting incidence, intensity and severity of poverty since 1996 (to be discussed in later section of this chapter).

The choice between household income or household consumption expenditure as a measure of household welfare in developing countries has almost been resolved. The better choice is household consumption expenditure since consumption can be a better indicator of lifetime welfare than income (World Bank Institute, 2005). This is because consumption remains fairly stable and households may be more able to recall what they have spent, rather than what they have earned.

The most difficult problem in the process of measurement of poverty is the estimation of absolute poverty line where a number of value judgments are embedded in this endeavor. In due course of time a number of methods for estimating absolute poverty line were developed. The most common method is the cost of basic needs (CBN) method. Nepal has been using this method since 1996. According to this method, absolute poverty line is the sum of *food poverty line* and *non-food poverty line*. These two poverty lines are estimated separately.

There seems to be have a great confusion among users or even scholars on the two definitions of poor. The first states that “*a person is said to be poor if his/her per capita expenditure falls below the poverty line*”. The second says that “*a person is said to be poor if his/her per capita income falls below the poverty line*”. These two definitions, in general, provide two different estimates of the proportion of poor in the same data.

In the context of Nepal, for example, the first definition provided 32 percent as poor while the second definition provided 36 percent as poor (NPC, 1978). According to the working paper 13 of UN (2017), the level of poverty in a nation is underestimated if the poverty is calculated from consumer expenditures as compared to income. Similarly, Slesnik (1993) found that consumption-based poverty indicators were much lower than those based on income. He recommended using consumption-based welfare measures rather than income-based ones in order to get more accurate identification of people who need help. In a developing country, like Nepal, first definition is preferred to the second, since the first definition is based on the standard of living.

In summary, in due course of time many refinements took place in order to make the measurements and analysis of poverty more scientific and cost-effective. As a result, many developing countries are currently measuring and monitoring poverty, and formulating suitable policies and programs for the reduction/eradication of *absolute* (aka *extreme* or *abject*) poverty.

### 1.3 Contemporary Measures of Monetary Poverty

The focus of this section is to review the theoretical development of the contemporary measures of poverty as well as their practical significances. Amartya Sen (1976) proposed both an axiomatic approach to poverty measurement and a specific index - widely known as Sen Index. Since then many research scholars felt that other new measures of poverty would be require in addition to the simple proportion of poor. As a result, a vast amount of theoretical literatures were developed on the measurement of poverty. In order to understand the contemporary measures of absolute poverty, let us assume that there are N individuals whose per capita consumption expenditure (y) arranged in ascending order as follows.

$$y_1 \leq y_2 \leq y_3 \leq \dots, \leq y_q < y_{q+1} \leq y_{q+2} \leq \dots \leq y_N \quad (1.1)$$

In the above set up (1.1), suppose z is poverty line satisfying the conditions  $y_q < z$  and  $y_{q+1} \geq z$ , which means there are exactly q individuals below the poverty line and (N – q) individuals above the poverty line. In this set up, headcount index (H) is algebraically defined as follows.

$$H = q/N. \quad (1.2)$$

Note that  $H$  is simply the proportion of individuals below the poverty line which is widely known as *head count ratio*, *incidence of poverty* or *poverty rate*, and its value ranges from 0 to 1. Three weaknesses of this measure are listed below.

1. It does not tell how poor the poor are. This means whether they are close to the poverty line or far below it.
2. It violates the **monotonicity axiom** of Sen. According to this axiom other things remaining the same, a reduction in income of someone below the poverty line must increase the poverty measure (Sen, 1976).
3. It also violates the **weak transfer axiom** of Sen. This indicates that, other things remaining the same, a transfer of income from a richer poor person to a poorer poor person must decrease the poverty measure (Sen, 1976).

In order to address these weaknesses of head count ratio, several new concepts were developed. A simple concept is the poverty gap of the  $i^{\text{th}}$  poor which is simply the shortfall in expenditure of the  $i^{\text{th}}$  poor for escaping out of poverty and algebraically it is defined by  $(z - y_i)$ . The average of shortfalls of all poor yields a new concept which is widely known as the *average poverty gap* and algebraically it is defined in (1.3) below.

$$\frac{1}{q} \sum_{i=1}^q (z - y_i) = (z - \bar{y}) \quad (1.3)$$

The average poverty gap measures the average cost per poor required to eliminate poverty and its value ranges in the interval  $(0, z)$ . Sen suggested to normalized the average poverty gap dividing the expression (1.3) by  $z$  in order to obtain a slightly different measure, called the *depth of poverty* ( $G$ ) where

$$G = \frac{1}{qz} \sum_{i=1}^q (z - y_i) = \frac{(z - \bar{y})}{z} \quad (1.4)$$

The depth of poverty is the average poverty gap expressed as a proportion of the poverty line. Average poverty gap as well as depth of poverty satisfies the **monotonicity axiom**; however it does not satisfy the **transform axiom**.

The three scholars Foster, Greer and Thorbecke introduced a class of poverty measures  $P(\alpha)$  in 1984 which is widely known as FGT measures, where  $\alpha$  is an index  $\geq 0$  and it is defined by

$$P(\alpha) = \frac{1}{N} \sum_{i=1}^q \left(1 - \frac{y_i}{z}\right)^\alpha \quad (1.5)$$

The three contemporary measures of poverty  $P(0)$ ,  $P(1)$  and  $P(2)$  are correspondingly known as *head count index*, *poverty gap index* and *squared poverty gap index*. The relationships between these three measures with head count ratio and depth of poverty are explored below.

The measure  $P(0)$  is algebraically defined in (1.6) is the same as head count ratio ( $H$ ) as can be seen below.

$$P(0) = \frac{1}{N} \sum_{i=1}^q \left(1 - \frac{y_i}{z}\right)^0 = \frac{q}{N} = H \quad (1.6)$$

The measure  $P(1)$  is algebraically defined in (1.7) which is just the ratio of total normalized poverty gap to  $N$  and the measure equals to the product of  $H$  and  $G$ . This measure satisfies the **monotonicity axiom** but not the **transfer axiom**.

$$P(1) = \frac{1}{N} \sum_{i=1}^q \left(1 - \frac{y_i}{z}\right) = \frac{q}{N} \frac{(z - \bar{y})}{z} = H \times G \quad (1.7)$$

The measure  $P(2)$  is algebraically defined in (1.8) which is just the ratio of the total of the squared normalized poverty gap to  $N$ . By squaring the normalized poverty gap for each individual,  $P(2)$  gives greater weight to those that fall far below the poverty line than those that are closer to it. The complex relationship between  $H$ ,  $G$  and the coefficient of variation ( $C$ ) of consumption distribution among the poor (Chaubey, 1995) can also be seen in (1.8). Here  $C$  is used to measure inequality in the distribution expenditure among poor. This measure satisfies the both **monotonicity** and **transform axiom**.

$$P(2) = \frac{1}{N} \sum_{i=1}^q \left(1 - \frac{y_i}{z}\right)^2 = H \times [G^2 + (1 - G)^2 \times C^2] \quad (1.8)$$

The three measures of poverty hold the following property:  $P(0) > P(1) > P(2) > 0$  and the strict inequalities become equality when  $H = 0$  or equivalently  $q = 0$ , in which case the concept of absolute poverty will become meaningless and the new concept of relative poverty will become relevant. Some practical significances are discussed below.

**Practical Significance:** The three measures  $P(0)$ ,  $P(1)$  and  $P(2)$  are computable using a household expenditure survey data. In their computations the most complex exercise is the estimation of poverty line. The computed three poverty measures at the national level can be decomposed for various sub-groups of the population defined by their place of residence, by their caste/ethnicity, and also by useful socio-economic and demographic factors. The computed measures can also be used to analyze poverty trends across time. All these facts widen the scope of poverty analysis. However, some remarks regarding three measures are in the following order.

1. The measure  $P(0)$  being a simple concept to understand and, therefore, it is widely used in political debate and public advocacy. But the measures  $P(1)$  and  $P(2)$  are not as simple concepts as  $P(0)$ , as they are less used in public domain. However, they are useful for policy maker, development planner and academician for understanding (in the sense that how far the per capita expenditure of the poor population is from the poverty line: higher is the value of  $P(1)$ , the more intense the poverty is said to be). Severity (being the average value of the square of depth of poverty for each individual, poorest people contribute relatively more to the index). Higher is the value of  $P(2)$ , the more severe the poverty of various sub-groups of the population.
2. Multiplying a country's  $P(1)$  by both the poverty line and the total number of population of the country, one can get the total amount of money needed to bring the poor in the population out of poverty and up to the poverty line, assuming perfect targeting of transfers.

#### **1.4 Poverty Measuring Practices in Nepal**

Nepal has been measuring poverty occasionally since 1978. The first measurement of poverty - based on a survey of 4,967 households - was published in 1978 by the National Planning Commission (NPC) of Nepal. According to the report, head count ratio was 36.2% (NPC, 1978). The second measurement of poverty - based on a survey of 5,323 households - was published by the Nepal Rastra Bank (NRB) in 1988. According to the report, the head count ratio was 41.4% (NRB, 1988). Due to various reasons (Chhetry, 2004), these head count ratios are not comparable. Moreover, NPC and NRB have defined those individuals as poor, each of whose per capita income falls below the poverty.



A more rational and scientific method for measuring monetary poverty is initiated in Nepal by the Central Bureau of Statistics (CBS) under the survey title - Nepal Living Standard Survey (NLSS). The survey methodology (both survey design and questionnaire design) and poverty line estimation procedure have been developed within the framework of the Living Standard Measurement Survey/Study (LSMS) program of the World Bank which makes the survey outcomes to a large extent comparable across the surveys. So far CBS has conducted three NLSS in the fiscal years 1995/96, 2003/04 and 2010/11 and for convenience they will correspondingly be designated by NLSS I, NLSS II and NLSS III in this study. Some salient features of these three surveys are as follows.

- NLSS I and NLSS II selected a nationally representative sample of households, using two-stage stratified random sampling method whereas NLSS III did the same using three-stage stratified random sampling method. All the surveys conduct face-to-face interview possibly with household heads for collecting data through questionnaires that covered a wide range of topics related to 'household welfare' such as - demography, consumption, income, education, health, employment, access to service centers, credit, remittance, housing conditions and so on including household income and expenditure. The number of enumerated households (or sample size) in each survey is presented in Appendix A1.
- Each survey uses the CBN method for estimating poverty line. The starting point for estimating the food poverty line is the estimation of the average per capita minimal calorie requirement per day to an individual for functioning. The estimated calorie requirement in each survey is presented in Appendix A2. The estimation procedure of both food and non-food poverty line of NLSS I is well documented (Chhetry, 2004). The poverty line of NLSS II is estimated just by updating prices for the same basic needs basket identified in NLSS I. For estimating the poverty line of NLSS III a *new basic needs basket* is identified in view of the change in food habit and consumption pattern of the poor in 2010/11 as compared to 1995/96 (CBS, 2011a). The estimated food and non-poverty lines in each survey are presented in Appendix A3.
- Each survey disseminates its survey methodology and estimation procedure of minimal calories requirement and poverty line in brief in reports. In addition, each

survey report disseminates the preliminary survey outcomes disaggregating by place of residence of respondents, consumption quintile groups etc. These preliminary outcomes provide impetus for more rigorous analysis of poverty.

## 1.5 Poverty Scenarios of Nepal

This section is devoted to present the three measures of monetary poverty – head count ratio, poverty gap and squared poverty gap – obtained from the three reports of NLSSs. While doing so, first they are presented for the national level and then for the sub-national level disaggregated by rural urban areas and by ecological regions. Finally they are presented for the year 2010/11 by disaggregating social groups.

### 1.5.1 National Level Poverty Trends

There has been considerable progress in Nepal in the reduction of poverty over the past one-and-half decade (1996 to 2011). The progress in poverty reduction could be observed even in a very unfavorable situation characterized by an armed conflict between the State and the Maoist (1995 to 2006), *Madhes Andolan* (2006 and 2007), frequent changes in government and sluggish economic growth. In 2006, the Peace Agreement was signed between the State and the Communist Party of Maoist. This ended the violent armed conflict. *Madhes Andolan* subsided after promulgation of the first Constituent Assembly which was held in 2008. The reduction in poverty in the mentioned period has been observed as follows:

The head count ratio had decreased from 42 percent in 1995/96 to 31 percent in 2003/05 and to 25 percent in 2010/11 (Table 1). Likewise, the other two measures of poverty had also decreased continually over the three survey years. There are several drivers responsible for the rapid decline in poverty measures under very unfavorable situation which has been discussed in Chapter II. Despite this fact, 1 in 4 persons still remained as poor.

**Table 1:** Poverty Trends in Nepal at the National Level

	1995/96	2003/04	2010/11
Head Count Ratio in %	41.8	30.9	25.2
Poverty Gap ×100	11.8	7.6	5.4
Squared Poverty Gap×100	4.7	2.7	1.8

Source: CBS (2011)

### 1.5.2 Rural-Urban Poverty Trends

Due to comparative advantages of urban population over rural population, poverty in the urban areas remained consistently below the rural areas (Table 2). Poverty in the rural areas has been continually decreasing over the three survey years. On the contrary, poverty in the urban areas had decreased from 1995/96 to 2003/04 and increased from 2003/04 to 2010/11.

**Table 2:** Poverty Trends in Nepal by Urban-Rural

Measures	Year	Urban	Rural
Head Count Ratio in %	1995/96	21.6	43.3
	2003/04	9.6	34.6
	2010/11	15.5	27.4
Poverty Gap $\times 100$	1995/96	6.5	12.1
	2003/04	2.2	8.5
	2010/11	3.2	6.0
Squared Poverty Gap $\times 100$	1995/96	2.7	4.8
	2003/04	0.7	3.1
	2010/11	1.0	2.0

Source: CBS (2005) and CBS (2011)

The cause of unstable result of increase in urban poverty, and decrease in rural poverty during the period 2003/04 and 2010/11 is the annexation of rural areas to the 58 number of urban centers after the Comprehensive Peace Agreement of 2006 and before the first Constituent Assembly Election held in 2008. This fact is clearly seen in the results of 2001 and 2011 population census: the number of urban centers was 58 in both censuses but the surface area of urban areas increased from 3,276 square kilometers in 2001 to 10,394 square kilometers in 2011, almost 3 fold-increased (for further information see Appendix A4).

### 1.5.3 Regional Level Poverty Trends

Economic opportunities, infrastructure developments, settlement patterns, climatic conditions and many more factors vary drastically across the three ecological regions, namely the terai, hill and the mountain regions. The population is expected to experience different levels of poverty across the three regions. In order to investigate the variations in poverty across the three regions the measures of poverty by ecological region have been presented in Table 3. In the terai and hill regions, the three measures

of poverty decreased continually over the three survey years. Note that the population of the terai region had in more comparatively advantages those of the hill region, but surprisingly the measures of poverty of the terai region was not as low as expected than those of the hill region.

**Table 3: Poverty Trends in Nepal by Ecological Region**

Measures	Year	Terai	Hill	Mountain
Head Count Ratio in %	1995/96	39.5	38.0	53.3
	2003/04	27.5	29.4	27.0
	2010/11	23.4	24.3	42.3
Poverty Gap ×100	1995/96	9.3	11.9	15.7
	2003/04	5.8	7.9	5.0
	2010/11	4.5	5.7	10.1
Squared Poverty Gap×100	1995/96	3.2	5.1	6.5
	2003/04	1.8	3.0	1.5
	2010/11	1.3	2.1	3.5

Source: CBS (2005), CBS (2011) and due to unavailability of the poverty gap and squared poverty gap for the years 1995/96 and 2003/04 in reports, and they were estimated by CBS expert

In the mountain region, the three measures of poverty had drastically decreased from 1995/96 to 2003/04 but they had sharply increased from 2003/04 to 2010/11. This may partly due to the migration of wealthy families from the mountain region to other regions and partly due to the annexation of more developed mountain areas to urban areas.

#### **1.5.4 Poverty Measures by Social Group**

In recent years caste/ethnicity has become a major social variable in understanding the process of social inclusion/exclusion and the level of socio-economic development/deprivation of the people in Nepal (Dahal, 2003). The CBS for the first time in Nepal collected data on ethnic/caste groups in the 1991 population census and has been continued in subsequent censuses. It identified only 60 ethnic/caste groups in the 1991 census. In the 2001 census the number increased to 100 caste/ethnic group (some reports reported this number to be 103 by including 3 ethnic/caste groups listed as ‘undefined’). In the 2011 census 125 ethnic/caste groups were identified. Since 1991

onwards other survey agencies also started to collect data on ethnic/caste groups of Nepal.

Attempts have also been made to categorize these large number of ethnic/caste groups into smaller number of groups for the sake of analysis of data to be more meaningful. As for example, Bennett et al. (2008) categorized the then 103 caste/ethnic groups into 7 groups and analyzed the Nepal Demographic and Health Survey (NDHS) data of 2006. The Country Partnership Strategy; Nepal, 2013 - 2017 of ADB categorized the 103 caste/ethnic groups into 11 groups and estimated the three measures of poverty for each of the 11 groups using the NLSS III data. UNDP (2014) categorized 125 caste/ethnic groups of 2011 into 11 category and measured Human Development Index and head count ratio of each of the 11 category. In this study, Bennett's 7 social groups with small modification (excluding 'Other') three measures of poverty were estimated and presented in Table4.

**Table 4:** Measures of Poverty by Social Group in 2010/11

Social Group	Head Count Ratio in %	Poverty Gap ×100	Squared Poverty Gap ×100
Brahaman/Chhetri	18.0	3.9	1.3
Terai/Madhese other castes	28.9	5.5	1.5
Dalit	41.8	9.9	3.5
Newar	10.3	2.1	0.7
Janajati	27.6	6.0	2.2
Muslim	20.2	3.4	0.9

Source: Computed from NLSS III (2010/11)

The head count ratio varied drastically across the six social groups. It varied from 10 percent for Newar to around 42 percent for Dalit. The poverty rates among the three groups, namely Newar, Brahaman/Chhetri and Muslim were below the national level of 25 percent whereas among the other three groups were above the national level. All these results clearly indicated that there was a high discrepancy in poverty rates among the social groups, which may be due to less commands over the resources and participation in the process of nation building programs and low capabilities of three social groups, whose poverty rate was below the national level.

An implication of such high discrepancy across social groups is that it is hard to meet the two goals of Sustainable Development by 2030. These two goals among 17 goals,

in this context, are Goal 10 and Goal 1. Goal 10 is to reduce inequality within and among countries and Goal 1 is to end poverty. These two are also inter-related in a way since all goals are integrated in the sense that action in one goal affects outcomes of the others .

### 1.5.5 Head Count Ratio of Nepal after 2010/11

After 2010/11 estimate of national level poverty rate based on nationwide survey, the national level poverty estimates based on nationally representative survey is still not available due to various reasons such as 2015 disasterous earthquake and Covid 19. However, the fourth NLSS is still in progress. Despite this fact the Ministry of Finance has published from time to time the Economic Survey estimated head count ratios for internal use as in Table5.

**Table 5:** Estimated Head Count Ratio in % after 2010/11

2011/12	2012/13	2013/14 to 2016/17	2017/18
24.4	23.8	NA	18.7

Source: Economic Survey (2012/13 to 2017/18)

The latest estimate of head count ratio made by the National Planning Commission was 16.7 percent for the year 2019/20 (MoF, 2020).

### 1.6 Statement of the Problem

It is well known that poverty is affected by many socio-economic and demographic factors (or indicators) as well as factors related to living standard of households. A vast number of research literatures dealing with the problem of identification of factors that affect poverty are available in numerous journals and the Internet (reviewed in Chapter II). The findings of these research literatures have succeeded in establishing the linkage between poverty and indicators of development using a wide variety of statistical methods ranging from simple to rigorous (will review in Chapter III). These linkages are useful for academicians in understanding the issues of poverty more in depth and developing more theoretical works. They are useful to policy makers and development planners for formulating appropriate policies and preparing action plans for poverty alleviation program.

The factors that affect poverty are not static; they are dynamic in the sense that they change over time and space due to the change in the level of development as well as

living standards of the people over time and space. As a result, from time to time country specific research directing towards the problem of identification of factors affect poverty is essential, and in such endeavor academicians can play dominant role.

It has been observed that some of these identified factors aggravate poverty when the value of a factor increases: a classical example of such factor is *household size*. While some factors alleviate poverty when the value of a factor increases, a classical example of such factor is *farm size*. If the real intention of the government is to reduce poverty, then the foremost work of the government is to identify factors that affect poverty and depending upon the identified factors address the issue of poverty reduction through appropriate policies and action plans.

Some practical as well as technical problems may arise in the identification of such factors. One major practical problem is the identification of *policy-driven factors* - such factors that can help policy makers to formulate meaningful policies as well as can help development planners to prepare pragmatic action plans for a poverty alleviation program. The notion of policy-driven factor is elaborated below with an example.

Numerous research studies has identified 'household size' as an important factor affecting poverty (Chapter II). For an academician, this finding could be useful for the sake of knowledge. But how can a policy maker address the issue of poverty reduction based on the information that household size is an important factor that aggravates household poverty? He/she will be able to address the needs of which group (s) of household members, since a group of household members in a society comprises of three groups of population: *children*, *working-age group population* and *elders*, and their needs to be addressed vary across the three groups. For example, a major need of children is their quality of education, a major need of working-age group population is gainful employment, and a major need of elders is social security. As a result, instead of household size it would be more logical to identify which group of population that are more prominent factors affecting poverty or in other words identify policy-driven factors.

This study, therefore, first focusses on the problem of identification of policy-driven several factors based on their practical significance in Nepal in Chapter II and Chapter III. Then this study tests the identified policy-driven factors for their statistical significance through advanced statistical models in Chapter IV.



## 1.7 Rationale of the Study

The existing literature clearly shows that very few rigorous studies have been carried out in identifying the linkage of poverty with other demographic and socio-economic indicators using the NLSS III cross-section data. This is a long standing gap in the analysis of NLSS III data in Nepal. This study attempts to fulfill this gap by incorporating policy-driven indicators.

Whatever studies have been carried out so far in analyzing the NLSS and other surveys data has been found to be using demographic and socio-economic indicators but not necessarily from the perspective of policy-driven point. The most popularly used household level demographic and socio-economic indicators used as risk factors in poverty analysis are household size, number of children in household, dependency ratios, sex and age of household head, literacy status of household head, area of land holding, remittance, income/consumption expenditure, access to service centers etc.

As far as the use of rigorous statistical models in poverty analysis is concerned, the most popular models are multiple linear regression, logistic regression, multinomial logistic regression, quantile regression. Among these, the most widely used model is the logistic regression which utilizes the odds ratio (OR) for establishing the linkage between poverty and the covariates included in the model.

The logistic regression model is abbreviated as LRM and the log-binomial regression model is abbreviated as LBRM from here and onwards writing of this document.

Another regression model, namely, log-binomial regression will be very useful to measure the relative risk of poverty (risk ratio) but not the odds ratio. Generally the log-binomial regression model (LBRM) is used for the analysis based on the cohort type of studies having binary outcomes. However, the LBRM can also be considered as an options to the logistic regression model (LRM) even if for the cross-sectional data (Barros & Hiraakata, 2003) when the occurrence of the event of interest is frequent i.e. 10% and more (Greenland, 1987; McNutt et al., 2003; Katz, 2006; Viera, 2008; Ranganathan et al., 2015 and Gallis & Turner, 2019). The risk of developing poverty due to considered predictor variable while applying the logistic regression, is generally measured through the odds ratio (OR) while the same can be measured through the risk ratio (RR) by using log-binomial regression. Some authors prefer OR generated through

logistic regression model (Walter, 1998; Olkin, 1998; and Cook, 2002) whereas other prefer RR which is generated through Log-binomial regression (Sackett et al., 1996; De Andrade & Carabin, 2011 and Gallis & Turner, 2019). The applications of log-binomial regression model are mostly found to be discussed with reference to the epidemiological data. With the best of the researcher knowledge based on extensive review of literature, till date, the identification of risk factors of poverty and their estimates have been found mostly only by using logistic regression model. The use of LBRM is almost rare in measuring the risk of poverty for different risk factors. Also, no study has been found comparing LRM and LBRM in poverty data in terms of estimates, goodness of fit, variables to be selected, and stability, etc. of the model through rigorous statistical procedures. Keeping in view to address this research gap, this study was planned, and is expected to be very useful for statistical view point and policy point of view.

### **1.8 Objectives of the study**

The general objective of this study is to identify the most important factors associated with poverty of Nepal. The specific objectives are as follow:

1. To identify the important risk factors of poverty of Nepal
2. To compare logistic regression model and log-binomial regression model in identifying the risk factors and estimating their effects on poverty
3. To assess the stability of the model through bootstrapping method

### **1.9 Research Questions and Hypotheses**

This research has examined risk factors affecting poverty in Nepal. It has used the monetary approach and theoretical framework developed by Rowntree and contributed by others. The research questions that the thesis focuses on are;

1. What are the risk factors affecting poverty in Nepal?
2. Which model (logistic or log-binomial) is appropriate for identifying the factors associated with poverty in poverty analysis?

Based on the extensive review of literature, discussions with the experts who have long experience on the relevant study and also based on the research questions and objectives of the study and to guide this investigation the following research hypothesis:

1. All the seven household level covariates as presented in the Conceptual Framework (Chapter 2) demonstrate significant effects on household poverty in Nepal in both models.
2. Both models fit the data significantly.
3. Both the fitted models are stable with the final set of covariates associated with poverty using bootstrap resampling procedure.

### **1.10 Significance of the study**

Some of the significances of this study are as follows.

1. For the first time in Nepal, this study introduces as well as emphasizes on the use of policy-driven socio-economic and demographic indicators instead of the usual ones for establishing the linkage between poverty and the policy-driven indicators with rigorous treatment of NLSS III data.
2. The study provides a framework for poverty data analysts, particularly the data of the forthcoming NLSS IV as well as data of other small surveys usually conducted by research scholars and other stakeholders
3. This study provides pragmatic significance for understanding of poverty among policy makers, academicians as well as other stakeholders in future.
4. This study attempts to compare empirically the LRM and LBRM in poverty data which is expected to be helpful to encourage researchers to apply log-binomial regression model as an alternative in social science data satisfying the required conditions.

### **1.11 Limitation of the study**

Some limitations of this study are as follows.

1. The findings of this study depict the poverty scenarios of 2011, which is more than a decade old. This is mainly due to unavailability of nationwide survey data.
2. There might be other relevant independent quantitative and qualitative variables associated with poverty which cannot be incorporated to analyze in a single framework in this study because of data problem and lack of data relevant indicators.

## **1.12 Chapter organization**

This study contains six chapters. The first chapter is the introduction that includes development of concept of poverty, measurement as well as refinements of poverty, statement of the problem, rationale, objective, research questions, hypotheses, significance, limitation of the study and chapter organization. Similarly, chapter two outlines the rigorous review of literature for poverty scenario, factors associated with poverty, statistical methods/models used in poverty analysis. Conceptual framework is also included in this chapter.

The data and methods of the present study are thoroughly described in chapter three. It includes data source and data file preparation, scheme of data analysis, selection of variables, dichotomization of quantitative variables and theoretical aspects of two statistical models. This chapter also describes the coefficient of determination, good fit, and diagnostics of the fitted model. The theoretical aspects of assessment of presence of risk factors and the stability of each fitted model are also discussed. Further, it also shows how the comparison of two models based on variable selection, individual regression coefficient, goodness of fit and diagnostics criteria and robustness are incorporated.

Chapter four reports the estimates of poverty in Nepal. It also presents an elementary data analysis in terms of consumption quintile groups (CQGs). Further, this chapter explains the results of poverty indices and statistical analysis of the risk factors of poverty in Nepal. Results and discussions based on LRM and LBRM are presented in details. Moreover, the results of comparison of both models with respect to effect size and its precision, good fit, diagnostics, stability, and assessment of risk on the basis factors present in the model are explained.

Chapter five includes the conclusion, recommendation and the indication for further research works in this study area. Chapter six is the summary. Finally, references and appendices are included at the end.

## **CHAPTER 2**

### **2. LITERATURE REVIEW**

One of the major thrusts of development programs in many developing countries is to reduce absolute poverty which is found to be interlinked with numerous socio-economic, demographic and many more factors. A vast amount of research works have been carried out for understanding such interlinks. The main objective of this chapter is to review some relevant literatures and examine their relevancy in the context of Nepal empirically. This chapter is broadly divided into 4 units. First unit deals with the review of fundamental understanding of rapid decline in poverty in the context of Nepal. Second unit focuses on the review on linkage between economic growth and poverty. The review of different studies focusing on identification of important factors associated with poverty is discussed in detail in the third unit of this chapter. In the fourth unit, the review focusing on the statistical models used to identify the factors associated with poverty are discussed. Furthermore, the review on the statistical models used to identify the factors associated with binary outcome other than poverty data are also explored keeping in mind for the possible application of such statistical models in the poverty data of Nepal. Finally, the schematic diagram of conceptual framework of this research work is also included.

#### **2.1 Understanding Rapid Decline of Poverty in Nepal**

Nepal has succeeded in reducing absolute poverty from 42% in 1995/96 to 25% in 2010/11, showing on an average, 3.46% declining rate of poverty reduction per annum in very unfavorable situations as mentioned in Chapter I. The reduction of absolute poverty has made such remarkable progress responsible to a number of factors. The substantial rise in individual remittances obtained from overseas, increase in labor incomes, and progress in household demographics are three major drivers identified by the World Bank (Uematsu et al., 2016), and they respectively contributed 27%, 52%, and 15% to the reduction in poverty between 1996 to 2011. A large volume of Nepali people went to abroad for the sake of better employment during the same period. The exact number of out-migrants is not known but the absent population are also reported in the population censuses of Nepal. However, the following two evidences justify the fact that the outmigration is huge in Nepal. The reported absentee population in the past

two censuses was more than double: 762 thousands in 2001 to 1.92 million in 2011 (CBS, 2014). The remittance receiving households in the past one-and-half decade was more than double: 23.4% in 1995/96 to 55.8% in 2010/11 (CBS, 2011c). These out-migrants brought many visible impacts in the socio-economic and demographic sectors of Nepal as illustrated below empirically.

Two noticeable impacts of remittance have been observed. The first one is at micro-level and the second one is at macro-level, and they are given below.

- (i) The nominal average amount of remittances per recipient household had increased 5.3 folds over the past one-and-half decade: NRs 15,160 in 1996 to NRs 80,436 in 2011 (CBS, 2011b).
- (ii) The percentage share of remittances in GDP increased from 1.8 in 1996 (MoF, 2005) to 18.5 in 2011 (MoF, 2012).

Rise in labor incomes can be seen from the following evidences. First, the mean daily wage rate in agriculture sector increased from NRs 40 in 1995/96 to NRs 170 in 2010/11, and in non-agriculture sector it increased from NRs 74 to NRs 263. Second, partly due to the spillover effects of remittance and partly due to the rise in labor incomes, household income increased by a factor of 4.6 times over the past one-and-half decade. For example, the nominal average household income increased from NRs 43,732 in 1996 to NRs 202,374 in 2011 (CBS, 2011c).

Improvement in household demographics can be seen from the following three evidences.

- a) The average annual growth rate was 2.25% during the census period of 1991-2000 (CBS, 2014). It was 1.35% during the census period of 2001 to 2011 (CBS, 2014). Evidently, there is considerable decrease in the value comparing in these different census periods.
- b) The total fertility rate (TFR) in 1996 was reported to be 4.6 births per women (MoH, 2011). It was 2.6 births per women in 2011 (MoH, 2011). There is clear indication of decreasing the TFR from 1996 to 2011.
- c) The percentage of children under the age of 15 was 42.4 in 1996 (CBS, 2011c). It was 36.7 in 2011 (CBS, 2011c). It also shows the declining scenario of the percentage of children under the age of 15.

The following improvements can be observed because of above discussed three evidences.

1. Increase in Household Income: The nominal average household income increased from NRs 43,732 in 1995/96 to NRs 202,374 in 2010/11 with high inequality, e. g. mean household income of the richest group was 3.6 times higher than that of the poorest group and the mean per capita income of the richest group was 5.9 times higher than that of the poorest group (CBS, 2011b).
2. Increase in Wage Rate: The mean daily wage rate in agriculture sector increased from NRs 40 in 1995/96 to NRs 170 in 2010/11. During the same period, the mean daily wage rate in non-agriculture sector had increased from NRs 74 to NRs 263. This raise in wage rate occurred due to the shortage of skill labors in Nepal.
3. Decline in Population Growth Rate: The average annual inter-census population growth rate of more than 2% during the period of 1961 to 2001 of Nepal suddenly declined to 1.35% during the inter-census period 2001 to 2011 (UNFPA, 2017). This drastic decline in population growth was due to substantial outmigration of population for foreign employment and decline in fertility.
4. Decline in Fertility: The TFR of Nepal declined from 4.6 births per woman in 1996 to 2.6 births per woman in 2011 (MoH, 2011) with high disparity across the wealth quintile groups. For example, the total fertility rate of the lowest quintile group was 4.1 births per woman while that of the highest quintile group was 1.5 births per woman.
5. Increase in Literacy Rate: The literacy rate among the population aged 6 years and above increased from 37.8% in 1995/96 to 60.9% in 2010/11 with gender disparity of 72% among males and 51% among females (CBS, 2011b). The report also showed that there is an alarming disparity between the poor and rich group in the 6+ literacy rate: 79% among the richest quintile population and 45% among the poorest quintile population.

## 2.2 Economic Growth and Poverty Reduction

Development has long been conceptualized in terms of economic growth. An important macro-economic parameter, is an increase in the production of economic goods and services of a nation, compared from one period of time to another. Economists usually measure economic growth in terms of growth rate of gross domestic product (GDP) or gross national product (GNP). Assuming that economic growth will invariably take care of poverty reduction. As a result, many developing countries introduced economic growth as the main agenda in the development plans in the past. In adequate course of time international communities realized that economic growth alone did not automatically reduce poverty. For example, Brazil with very rapid and sustained economic growth continued with high level of poverty in the past. As a result, a series of dialogues started among eminent scholars in the past for the refinement of the notion of development and several concepts of development had emerged, and one of them is the *pro-poor growth* which need to be materialized primarily through national policies to stimulate economic growth for the benefit of the poor people.

Numerous studies have been conducted to examine the relationship between economic growth and the occurrence of poverty incidence across nations and historical periods (Ravallion & Chen, 1997; Adams, 2003). According to estimates, a 1% rise in the rate of per capita earnings growth can result in a reduction of the number of persons living in poverty of up to 2%, provided that the mechanism of income change is distribution-neutral in character. However, inequality has a tendency to fluctuate, and while some nations have achieved excellent growth rates while reducing poverty, others have been able to do so while experiencing relatively low growth.

The adverse impacts of inequalities in the process of development, poverty reduction, social cohesion and stability have been well documented in the literature; for example: one of the most crucial concerns in development is inequality, which is seen unfair from most philosophical views. Evidence shows that inequality is harmful to overall well-being, social stability, economic progress, and prosperity. Depending on the goal, inequality can be described in a variety of “spaces” or dimensions, such as income, assets, capabilities, satisfaction, or opportunities (Stewart & Samman, 2013).



### 2.3 Factors Associated with Poverty

Okojie (2002) examined the nexus between household head gender, education, and poverty from 1980 to 1996. Data used in their study came from four national consumer expenditure surveys collected by the Federal Office of Statistics in Nigeria in 1980, 1985, and 1996. The head count ratio as well as the gap and severity of poverty were calculated using the FGT index. Theil's index and Gini coefficients were used to examine inequality trends. Two models namely the ordinary least square regression and the multiple logistic regression were used for all survey years. For all survey years, sensitivity, specificity and correctly predicted value were also tested. In order to test the goodness of fit of the model, the pseudo  $R^2$  also reported for all surveys. For the multiple linear regression model, the outcome variable was log (household per capita expenditure) and for LRM, the poverty status (poor / non-poor) was used as dependent variable. The associations between gender, poverty, and other household factors, such as education, were investigated using multivariate analysis for all families as well as for subgroups of male and female headed households, respectively.

Chhetry (2005) applied LRM for comparative analysis of absolute poor and non-poor. The separate LRM was run for each of three different belts (Terai, Hill and Mountain). The study summarized that the poor households of Nepal were not only disadvantaged by low income but were also severely disadvantaged by socio-demographic factors as well as access to reproductive health care. Excessively large number of children, high fertility, and high child-dependency at household level were the major demographic disadvantages of the poor. Low level of literacy and relatively high gender inequalities among children as well as among adults were the major social disadvantages of the poor.

Yusuf et al. (2008) attempted to examine the poverty situation of urban agricultural households from a variety of perspectives. The study was conducted in the city of Ibadan Sub-Saharan Africa with 200 agricultural households. The data were analyzed using poverty indices and LRM. According to this study, agricultural farmers were reported to have the highest rates of poverty. The developed LRM revealed that age of household head ( $\beta = -0.08$ ,  $p < 0.05$ ), educational attainment of household head ( $\beta = -0.41$ ,  $p < 0.01$ ), years of work experience in farming ( $\beta = -0.09$ ,  $p < 0.05$ ), types of agricultural employment in crop farming ( $\beta = 1.54$ , reference category = otherwise,  $p$

$< 0.10$ ) and dependency ratio ( $\beta = 0.75, p < 0.01$ ) were associated with the household poverty. The value of  $R^2$  reported was 0.35.

Sikander and Ahmed (2008) analyzed Multiple Indicator Cluster Survey (MICS) during 2003-2004 in Pakistan with almost 31,000 households. The study was performed by using LRM model considering poverty status (poor vs. non-poor) as outcome variable. On the basis of a threshold of Pakistan rupees (PKR) 848.798 and a daily caloric intake of 2350 calories for per capita monthly expenditure, various households were categorized as poor or non-poor. The findings indicated that variations in the probability of being poor were strongly explained by the age, education, and sex of the household heads. Additionally, those with access to remittances and agricultural land were more likely to escape the cycle of poverty. The probability of becoming the poor household groups was positively affected by the dependency ratio and bigger family size. The job sector also contributed significantly to the explanation of cross-regional and geographic variations in the determinants of poverty.

The study conducted by Adepoju and Oluoha (2008) at Obefemi-Owode LGA in Ogun State examined the impact of access to microcredit on the poverty status of rural households in study area. The information gathered from 94 randomly chosen households in the research region were considered for analysis. The FGT poverty indices and the LRM were used to analyze the data utilized in their study. Age of household head ( $\beta = 0.16, p < 0.10$ ), family size ( $\beta = 0.12, p < 0.01$ ), education of the household head ( $\beta = -0.16, p < 0.05$  for secondary education,  $\beta = -0.28, p < 0.01$  for tertiary education, reference category = no formal education), access to credit ( $\beta = -0.14, p < 0.01$ ) and employment of the household heads ( $\beta = 0.16, p < 0.10$  for primary occupation, reference category = non-farming) were shown to be major characteristics that affected poverty status in the study area.

Onu and Abayomi (2009) examined poverty amongst households living in Yola metropolis of Adamawa State, Nigeria. In this study, 120 respondents were selected and the FGT poverty indices tool was used to calculate three measures of poverty indices. Results from this study showed that the study area had a high rate of poverty. The incidence of poverty, poverty gap and severity poverty of female headed households were  $P_0 = 47.7, P_1 = 0.42, P_2 = 0.22$ , respectively. These values for male were

$P_0 = 44.4$ ,  $P_1 = 0.26$ ,  $P_2 = 0.08$ , respectively. Similarly, the incidence of poverty, poverty gap and severity poverty of illiterate household head were  $P_0 = 100.0$ ,  $P_1 = 0.50$ ,  $P_2 = 0.26$ , respectively. These values for older farmers of age 60 years and above were  $P_0 = 100.0$ ,  $P_1 = 0.49$ ,  $P_2 = 0.25$ , respectively.

Achia et al. (2010) examined the determinants of poverty in Kenya. The data for this study were from the Demographic and Health Surveys (DHS) Kenya. The principal component analysis (PCA) was also used. This PCA was used to create an asset index which gave the social economic status of each household. A logistic regression was estimated to check the association of different variables with socio-economic status (SES) (i.e. is poor and non-poor). According to the findings of this study, age, religion, location, education and ethnicity of the household head were the set of demographic variables that raised the likelihood of poverty. Size of the household was statistically associated with social economic status when examined as a univariate model, but it was not statistically associated when added to the multivariate analysis.

Sakuhuni et al. (2011) investigated empirical examination of economic factors that contributed to poverty in Zimbabwe using cross-section data for 2005. Based on this data, a multiple linear regression model with per capita consumption as response variable and a number of economic and demographic factors as the explanatory variables were estimated. Age of household head ( $\beta = -0.15$ ,  $p < 0.01$ ), gender (male) ( $\beta = 0.71$ ,  $p < 0.01$ ), working in the private sector ( $\beta = 1.58$ ,  $p < 0.01$ ), number of income sources ( $\beta = 0.87$ ,  $p < 0.01$ ) were statistically significant ( $\alpha = 0.01$ ). The value of  $R^2$  reported was 0.69.

Akerele and Adewuyi (2011) centered their study on assessment of household poverty and welfare among households in Ekiti State, Nigeria. A total of 80 households were selected to analyze poverty using multiple linear regression analysis throughout the study. Their results explored that educational levels of household head ( $\beta = 0.07$ ,  $p < 0.10$ ) and spouse ( $\beta = 0.16$ ,  $p < 0.10$ ), gender of household head ( $\beta = 0.14$ ,  $p < 0.01$ , male = 0) and dependency ratio ( $\beta = -0.20$ ,  $p < 0.05$ ) were the factors that impose significant influence on the household welfare. For the goodness of fit of the model, the value of  $R^2$  was 0.68.

Javed and Asif (2011) examined the relationship between male-headed households, female-headed households, and the variables that influence the likelihood of poverty in two Tehsils of the District of Faisalabad. In total, eighty samples were chosen. Multiple linear regressions and the multiple logistic regressions analysis were used to find the relationship between households headed by men and women and the variables that influence a probability of falling into poverty. Their results from multiple linear regression revealed that the households monthly income was significantly influenced by the households occupation ( $\beta = 0.15$ ,  $p < 0.01$ ), number of children ( $\beta = -0.46$ ,  $p < 0.01$ ), secondary earners ( $\beta = 0.07$ ,  $p < 0.01$ ), educational attainment of the household head in years ( $\beta = 0.02$ ,  $p < 0.01$ ). Whereas findings of the LRM, factors influencing poverty included family consumption, family size, headship status, and family income.

Ennin et. al. (2011) used information from three rounds of Ghana Living Standards Survey (GLSS3, 1991/92; GLSS4, 1998/99; and GLSS5, 2005/06) to analyze determinants of poverty. A nationally representative sample of 4552 (first round), 5998 (second round), and 8687 (third round) households were chosen. Among them, around 200, 300, and 580 enumeration areas were used for the first, second, and third rounds, respectively. Separate LRM was run for each of three different samples. They concluded that poorer households were those with larger sizes of households, heads who were illiterate and employed mostly in agriculture and who lived in rural localities and the savanna zone.

Gounder (2012) analyzed factors determining household expenditure and poverty in Fiji using the data from 2002–2003 household survey. An ordinary least square (OLS) model was used to identify the factors (household characteristics) associated with poverty. A probit model was also estimated for the robustness of the determinants of poverty. The dependent variable of this model was the household poverty status whereas the log of household per capita expenditure was the dependent variable of OLS model. Six regression models were performed separately for central, western, northern, eastern, rural and urban regions of the country. It was also reported that higher levels of education, policies that support agricultural growth in rural areas, and the real location of labor to the formal sector of the economy were likely to be effective in reducing poverty at the household level.

Issahaku and George (2012) estimated and examined different socio-economic factors that determine poverty in the Kwabre East District of the Ashanti Region of Ghana. The research was performed in such a way that a semi-structured questionnaire was constructed to different 208 households that were selected through random sampling. The time period when the data was collected was 2009/2010 farming season. Weighted least squares multiple regressions was used mainly to estimate the determinants of poverty throughout the study. The dependent variable in this study was a logarithm of per capita household consumption of household and the independent that were household age characteristics, household educational characteristics, household employment characteristics, access of household to basic facilities, household assets, household vulnerability, location, household remittance, tenure system and household access to capital. This study exhibited that the number of children aged 6-12 years in a household and the distance of household dwelling to the nearest portable water source impacted negatively on the welfare of households. In addition, the households that were female headed were found to be prone to be poor. On the other hand, different factors like the number of household members in skilled jobs, value of home assets, and access to micro-credit were found to be enhancing the well-being at security of households.

Sekwati et al. (2012) examined different household characteristics in Botswana that contributed to poverty. Data from Gaborone's baseline urban food security survey were used to develop a LRM, which was then used to analyze the demographic, social, and economic aspects that contributed to household poverty. Results of this regression showed that there was a strong and positive correlation between household size ( $\beta = 0.37$ ,  $p < 0.001$ ) and total consumption expenditure ( $\beta = 0.94$ ,  $p < 0.001$ ) with household poverty.  $R^2$  in this case was 0.43.

Osohole et al. (2012) used data from 2003–2004 National Living Standard Survey (NLSS), which included 19,158 households. In order to identify the determinants relevant to poverty of Nigerian households, a LRM was used. Results of the LRM indicated that the most important variables of poverty in the study area were household size ( $\beta = 0.30$ ,  $p < 0.001$ ) and the highest level of education of the household head ( $\beta = -0.25$ ,  $p < 0.001$ ). Other factors included the gender of household head ( $\beta = -0.26$ ,  $p < 0.001$ ), age in years of household head ( $\beta = -0.01$ ,  $p < 0.001$ ), level of education ( $\beta = 0.04$ ,  $p < 0.001$ ) and employment ( $\beta = 0.01$ ,  $p < 0.001$ ) of the father, employment of

the mother ( $\beta = -0.01$ ,  $p < 0.001$ ), and occupation group ( $\beta = 0.06$ ,  $p < 0.001$ ) of the households head were also significant ( $\alpha = 0.05$ ).

Rusnak (2012) explored factors that increased the chance of poverty. The source of data in this study was unidentifiable unitary data from household budget research carried out by Central Statistical Office (CSO) in 2008. The study applied LRM and the explanatory variables statistically significant ( $\alpha = 0.05$ ) were the location and the size of the household and number of children under 14 years in the household.

Sekhampu (2013) used household level data and examined factors that contributed to household poverty in 283 female-headed households in South African Township of Bophelong. Based on these data, a LRM was developed with the dependent variable as poverty status and the explanatory variables as a set of demographic factors, including household size, age, educational level, employment status and marital status of household head. Their findings showed that various factors including household size (OR: 1.44,  $p < 0.01$ , 95% CI: 1.15 - 1.80), the age (OR: 0.95,  $p < 0.01$ , 95% CI: 0.93 - 0.98) and employment status of the household head (OR: 0.14,  $p < 0.01$ , 95% CI: 0.07 - 0.29) significantly contributed to variations in the probability of being poor.

Salami and Atiman (2013) assessed different determining factors of poverty among households in Adamawa North Senatorial. A total of 400 sample households were identified and ordinary least square (OLS) model was used to identify the determinants of poverty. Results of this study revealed that household occupation, energy use, and educational levels had positive effects and the estimate coefficients were 0.36, 0.25 and 0.07, respectively. On the other hand, dependency ratio, water sources, and inadequate nutrition had negative effects with coefficients -0.20, -0.11, and -0.19, respectively. The value of  $R^2$  was 0.68.

Dudek and Lisicka (2013) analyzed the influence of the variables on poverty among the households of the employees in Poland. The data were drawn from the Household Budget Survey (HBS-2011) carried out by the Central Statistical Office (CSO). The LRM was used to determine the households at risk of poverty using 18441 households. Their results revealed that working in manual jobs, living in a rural region, household size, living in cities or medium-sized townships in the central region and having at least

a secondary education were the factors that were associated with the household poverty status.

Thapa et al. (2013) used LRM to identify the determinants of poverty. The sample size used for the study was 279 households from 6 districts of western region of Nepal. They reported that literacy status of household head ( $\beta = -0.94$ , reference category = illiterate,  $p < 0.01$ ), family size ( $\beta = 0.02$ ,  $p < 0.10$ ), occupation of household head ( $\beta = -1.09$ , reference category = agriculture,  $p < 0.01$ ), size of land holding ( $\beta = 0.96$  for 4-10 *ropani*,  $p < 0.01$ ,  $\beta = 1.96$  for 10-20 *ropani*,  $p < 0.05$ , reference category = less than 2 *ropani*), females involvement in service ( $\beta = -0.77$ , reference category = yes,  $p < 0.05$ ), social involvement ( $\beta = 1.28$ , reference category = yes,  $p < 0.01$ ), and *Dalits* ( $\beta = 1.36$ ,  $p < 0.05$ ) were significantly associated with the rural poverty. The value of Cox and Snell  $R^2$  and Nagelkerke  $R^2$  were 0.25 and 0.34 respectively.

Omogbee et al. (2013) examined the characteristics of 244 farmers and their effects on poverty in Nigeria's Delta South Senatorial region. Their study concluded that poverty status of farmers was significantly associated with gender issues ( $\beta = 0.57$ , reference category = male), level of education ( $\beta = 0.25$ ), size of the farm ( $\beta = -0.34$ ) with odds ratio of 1.78, 1.28 and 0.71, respectively, using LRM.

Sanusi et al. (2013) examined different socio-economic factors affecting the poverty status of farming households in Ikorodu Local Government Area (LGA) of Lagos State, Nigeria. The primary data for this study was obtained from 120 respondents using a multi-stage sampling method. Foster, Greer, Thorbecke (FGT) indices and logistic regression analysis were used in this study. LRM was used to assess the factors of farmer household poverty in the study area. Poverty status was used as a dependent variable whereas age, marital status, gender, and previous farming experience of the household head, off farm activities, farm size, household size, marketable surplus, farm produce consumed, educational qualification of household head and farm produce given as gift were taken as the independent variables. They concluded that the household head's education level, farming experience, household size significantly influenced the chances of household's poverty.

Khudri and Chowdhury (2013) evaluated the poverty status of households and determined important factors of poverty in Bangladesh. They used Bangladesh Demographics and Health Survey (BDHS-2007) data of sample size 10,400 and LRM to identify key factors of poverty in Bangladesh. Socio-economic Status (SES) was the dependent variable in this study. A number of demographic factors, including marital status, the type of residence, ownership of land usable for agriculture, greatest level of education, and work status of household head, were found to be important predictors of poverty. The variables significant in their regression were: place of residence (OR: 0.19,  $p < 0.05$ , 95% CI: 0.17 - 0.22, rural = reference category), employment status (OR: 0.94,  $p < 0.05$ , 95% CI: 0.68 - 0.99, no = reference category), owned agricultural property (OR: 0.56,  $p < 0.05$ , 95% CI: 0.50 - 0.61, no = reference category), level of education (OR: 0.48,  $p < 0.05$  with 95% CI: 0.43 - 0.53, for primary education, no education = reference category), OR: 0.21,  $p < 0.05$ , 95% CI: 0.19 - 0.24, for secondary education, no education = reference category) and OR: 0.05,  $p < 0.05$ , 95% CI: 0.04 - 0.07, for higher education, no education = reference category) and administrative division (OR: 1.89,  $p < 0.05$ , 95% CI: 1.61 - 2.22, for Barisal division, Dhaka division = reference category), OR: 1.45,  $p < 0.05$ , 95% CI: 1.25 - 1.68, for Rajshahi division, Dhaka division = reference category).

Leekoi et al. (2014) identified risk factors that affected rural households and examined their relationships to socioeconomic factors in Thailand's Pattani Province. They selected 600 households and used LRM in this study. The outcome of logistic regression indicated that household size ( $\beta = 1.17$ ,  $p < 0.001$ ), location of homes ( $\beta = 0.96$ ,  $p = 0.002$ ), and the sex of household heads ( $\beta = 2.19$ ,  $p < 0.001$ ) had significant impact on risk exposure.

Balarabe (2014) investigated poverty determinants in Kano Metropolis, Nigeria. Probit model was used to identify the factors of poverty. Primary dataset of about 120 households was used in their study. Their results indicated that education had a negative relationship with poverty while other explanatory variables such as households headed by women in the residence ( $\beta = 0.76$ ,  $p < 0.05$ ), people with no job in a given residence ( $\beta = 0.35$ ,  $p < 0.10$ ), people with no high school diploma ( $\beta = -0.63$ ,  $p < 0.10$ ) and households whose heads moved to another state ( $\beta = 0.39$ ,  $p < 0.10$ ) were found to be associated with poverty.



Deressa and Sharma (2014) used the most recent Household Income, Consumption and Expenditure Survey (HICES) 2010–11 to investigate the effects of socioeconomic and demographic factors of households on household poverty in Ethiopia using LRM. They categorized different households as poor or not poor depending on the per capita expenditure of 3781 Birr. They concluded that people who own agricultural land (OR: 0.74,  $p < 0.001$ ) were more likely to escape poverty. They also concluded that households with female heads (OR: 1.60  $p < 0.001$ ), big families (OR: 2.77,  $p < 0.001$ ), households having in rural area (OR: 5.23,  $p < 0.001$ ) and high dependency ratios (OR: 1.21,  $p < 0.001$ ) were poorer than their counterparts. The good fit of the model was also tested by the value of  $R^2$  (0.40), Hosmer & Lemeshow  $\chi^2$  test ( $\hat{C} = 2.22$ ,  $p = 0.97$ ), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) and compared full and null model. AIC for null and full models were 32855.56 and 23463.28, respectively. Similarly, BIC for null and full model were 32859.01 and 23502.39, respectively. For the predictive power of the model, sensitivity (55.3%), specificity (90.10%) and correctly classified value (80.53%) were also reported.

Adekoya (2014) scrutinized the status of poverty of agricultural households in the Nigerian state of Ogun. The study was conducted using LRM involving 117 sample sizes of farm households. The results of the analysis showed that large households, farm households headed by uneducated people ( $\beta = -1.90$ ,  $p < 0.01$ ), households without credit access ( $\beta = -0.15$ ,  $p < 0.05$ ), farming experience ( $\beta = -0.58$ ,  $p < 0.05$ ), sex ( $\beta = -1.16$ ,  $p < 0.05$ ), number of adult in households ( $\beta = -0.14$ ,  $p < 0.05$ ), farm size ( $\beta = -0.68$ ,  $p < 0.01$ ) and households with other non-farm income ( $\beta = -0.03$ ,  $p < 0.01$ ) were statistically significant in explaining the variation in household poverty status. The value of  $R^2_{Logistic}$  for this regression was 0.22.

Edoumiekumo et al. (2014) assessed income poverty in Nigeria's south-south geopolitical region using 2888 samples collected by National Living Standard Survey (2009–2010). The FGT poverty indices and LRM were used in this study. They reported that gender (OR: 0.67,  $p < 0.01$ , reference = male) occupation (OR: 3.74,  $p < 0.01$ , reference = others), size of household (OR: 1.76,  $p < 0.01$ ), education expenditure per capita (OR: 0.99,  $p < 0.01$ ), health expenditure per capita (OR: 0.99,  $p < 0.01$ ) and share of food expenditure (OR: 0.02,  $p < 0.01$ ) were the major significant factors that were related to poverty.

Myftaraj et al. (2014) conducted a study to examine poverty in household per capita consumption. They used 3600 households from rural and urban area of Albania to measure monetary poverty. They identified potential factors of household poverty using logistic regression analysis and used the household level and individual level characteristics as independent factors (literacy status, education level, household size, location, geographic division, employment status, sex) and the poverty status as dependent variable. Their results revealed that educational level, employment status and location (rural/urban) of the household heads, more than two children in a house, larger households size, access to public services, age of the households head were significant factors in explaining the variation in households poverty status.

Makame and Mzee (2014) used 4293 households of Zanzibar Household Budget Survey (ZHBS) (2004/2005 and 2009/2010) to assess the factors associated with poverty using LRM. The dependent variable of this study was the poverty status of household and the independent variables were set of social and demographic variables (household size, age of household head, sex of household head, type of residence, educational attainment, and dependent status of household head, employment status, farming, fishing, and administrative location). They reported that social and demographic factors were important in explaining poverty. In addition, the chance of poverty was significantly associated with household size (OR: 1.39,  $p < 0.001$ , 95% CI: 1.34 - 1.46), residence (OR: 0.62, reference = rural,  $p < 0.001$ , 95% CI: 0.44 - 0.87) and basic educational attainment (OR: 0.58,  $p < 0.001$ , 95% CI: 0.44 - 0.78 for primary and OR: 0.40,  $p < 0.001$ , 95% CI: 0.29 - 0.57 for secondary). They also reported that all of Pemba's districts were at high risk of poverty. The value of pseudo  $R^2$  they reported was 0.19.

Khafaj and Nurja (2014) used logistic and linear log regression models to identify the most important factors that affect poverty. The sample size of this study was 3600 that was collected by Living Standards Measurement Study (LSMS) (2008) in Albania. The poverty status and household consumption expenditure per capita were dependent variable in the logistic and linear log regression models. Household size, level of education, sex, and residence of household head were used as independent variables. The findings of both models (linear and logistic regression models) showed that level of education, sex, and location of household head were most significantly related to

response variables. The value of  $R^2$  based on Cox & Snell and Nagelkerke were 0.11 and 0.22, respectively.

Spaho (2014) conducted a study to assess different determining factors of poverty in Albania based on the household level data. A direct interview with 215 randomly chosen households in both urban and rural areas was conducted in November 2013. Based on the data gathered for the study, two regression analysis were estimated: a logistic regression analysis having poverty status as the response variable and a log-linear model with per person monthly consumption as the response variable. A group of demographic and economic factors, including the age, level of education, employment status, and place of location of household head, and household size were used as independent variables. The pseudo  $R^2$  reported was 0.21.

Tuyen (2015) used 1800 households to analyze socio-economic determinants of household income among ethnic minorities in the North-West Mountains, Vietnam. In this study, community and household factors that affected ethnic minorities' household income were examined. He found that a large percentage of sample families strongly depended on agricultural activity. Factors affecting household income per capita were investigated using multiple linear regression model. He reported that there was a strong positive relationship between household income and non-farm employment.

Habyarimana et al. (2015) created an asset index by using principle component analysis (PCA) method to assess poverty status. Data collected by Rwanda Demographic and Health Survey (RDHS) (2010) from 12540 households (2009 from urban and 10531 from rural areas) was used to determine socioeconomic status (SES) of households using LRM. Results of this study showed that the household size (OR: 1.09,  $p < 0.001$ ), location of the residence (OR: 0.79,  $p < 0.001$ , rural = reference category), age (OR:1.01,  $p < 0.001$ ) and education (OR: 6.48,  $p = 0.002$  for secondary education, higher education = reference category), OR: 24.42,  $p < 0.001$ , for primary education, higher education = reference category) and OR: 41.97,  $p < 0.001$ , for no education, higher education = reference category) of the household head were the most important predictors of households poverty in Rwanda. Hosmer and Lemeshow  $\chi^2$  test was reported in this study (H-L  $\chi^2 = 7.33$ ,  $p = 0.502$ ).

Yusuf et al. (2015) randomly sampled 210 households head from Moa ward in the Mkinga district of Tanga region in Tanzania to assess the factors of poverty. They used ordinal LRM to analyze the data and found that about 93% of participants in the study area were poor. The response variable of this study were poor, moderately poor and non-poor households and the independent variables were sex of the household head, land ownership of the household and the size of the farm.

Khan et al. (2015) had attempted to identify the determinants of household poverty in a district of Pakistan (n = 600 households) using binary logit model to estimate the risk of rural household poverty. Their findings indicated that socioeconomic empowerment ( $\beta = -0.07$ ,  $p < 0.05$ ), only agriculture employment ( $\beta = -1.63$ ,  $p < 0.10$ ), remittance recipient households ( $\beta = -2.28$ ,  $p < 0.05$ ), female to male ratio ( $\beta = 0.41$ ,  $p < 0.10$ ) and household size ( $\beta = 0.25$ ,  $p < 0.05$ ) were associated with rural household poverty. The value of  $R^2$  of Nagelkerke and Cox & Snell were 0.42 and 0.27, respectively.

Farah (2015) identified variables that had relative impact on household poverty. The data used in this study was obtained from 17,142 households collected by Demographic and Health Surveys (DHS) in Bangladesh. A LRM was estimated using the socio-economic status as the response variable and a set of demographic variables as the explanatory variables. The age (OR: 1.47,  $p < 0.001$ ), type of residence (OR: 7.43,  $p < 0.001$  urban = reference category), education (OR: 0.82,  $p < 0.001$ ) and region (OR: 0.96,  $p < 0.001$ ) of the household head, the household size (OR: 1.21,  $p < 0.001$ ), land ownership (OR: 3.14,  $p < 0.001$ ) and condition (OR: 1,  $p < 0.001$ ) of the household were demographic variables that influenced the probability that a household would get poor.

Margwa et al. (2015) conducted a study exploring the level of poverty among the households of Adamawa State in Mubi region, Nigeria. They selected 160 households to collect the data and the data was analyzed using FGT indices. Logistic regression analysis was used to identify households' poverty level. They determined the relationship between different factors affecting the poverty (age of respondents, respondent income, household size and time taken to reach health center) and household poverty status. Results of this analysis showed a significant association between respondents' age ( $\beta = -0.07$ ,  $p < 0.05$ ), incomes ( $\beta = -0.29$ ,  $p < 0.10$ ), size of the

household ( $\beta = 0.33, p < 0.05$ ), level of education ( $\beta = 0.42, p < 0.05$ ) and travel times to reach health centers ( $\beta = 0.02, p < 0.05$ ).

Majeed and Malik (2015) investigated household and individual characteristics as the factors that influenced poverty in Pakistan. Education, experience, gender, age, and employment status of the household head were taken into account as various individual characteristics of the household. Whereas size of household head, provincial location status, regional location status, and remittance receiving status were used as various characteristics of the household. The poverty status was used as dependent variable. The LRM was applied to find and examine the impact of explanatory variables on the likelihood that the household would become poor. The following factors were statistically significant in explaining the variation in household poverty. Household size (OR: 1.22,  $p < 0.01$ ), head of household's age (OR: 1.04,  $p < 0.01$ ), male headed households (OR: 1.82,  $p < 0.01$ ), receiving remittances (OR: 0.58,  $p < 0.01$ ), levels of education (OR: 0.78,  $p < 0.10$  for primary, OR: 0.46,  $p < 0.01$  for middle, OR: 0.36,  $p < 0.01$  for matric, OR: 0.13,  $p < 0.01$  for inter, OR: 0.01,  $p < 0.01$  for bachelor, OR: 0.11,  $p < 0.01$  for professional) and place of living in the province (OR: 1.94,  $p < 0.01$  for Punjab, OR: 1.31,  $p < 0.01$  for Sindh and OR: 1.91,  $p < 0.01$  for KPK, Balochistan = reference category), school starting age (OR: 0.99,  $p < 0.10$ ) and in agriculture employment status (OR: 0.73,  $p < 0.01$ ).

Maloma (2016) used a Survey questionnaire as an instrument to collect the data based on a sample of 300 households in Bophelong town in Gauteng province, South Africa during the second half of 2013. To analyze the poverty status, different independent variables were taken. Sex of household head, status of the employment, level of education, household income and age of the household head were the factors included in the study. The data was analyzed using a LRM. It was concluded that the age ( $\beta = -0.03, p < 0.10$ ), employment status ( $\beta = -1.27, p < 0.05$ ), and the household head educational level ( $\beta = -1.16, p < 0.05$ ) were negatively associated with poverty status.

Mohammed (2017) examined how urban poverty in Ethiopia's Southern Nations, Nationalities, and Peoples' Region (SNNPR) was measured and what caused it. The data used in this study was obtained from 5015 urban households surveyed by SNNPR. The primary objectives of this study were to measure urban poverty and identify its root causes using logistic regression. The variables; marital status and educational

attainment of the household head, size of household, overall dependency, saving habit and energy source were statistically significant in identifying the causes of urban poverty. The values of pseudo  $R^2$  was 0.11 and Hosmer and Lemeshow ( $\chi^2$ ) test was 11.86 ( $p = 0.16$ ).

Kona et al. (2018) used logistic regression to identify and generalize the impact of the several factors that determine poverty. Altogether 120 respondents were used in this study. Explanatory variables such as the age and sex of household head, household size, education level, and job status of women were used to estimate dependent variable (socio-economic status). Results of this analysis revealed that several factors, including sex of household head, highest educational attainment of family members, age of household head and employment of women were significantly associated with household poverty.

Imam et al. (2018) explored factors determining poverty in rural areas of Bangladesh using the data of nationally representative Household Income and Expenditure Survey (HIES) 2010. They used 7840 rural households to determine the key variables that contribute to poverty. The LRM was used to identify significant factors associated with poverty as well as to capture and assess the unobserved variability between communities. Two poverty lines (lower and upper poverty lines) were used in the analysis. Independent variables used for lower poverty line which were as follows: the sex, age and education of household head, sex ratio, household size, type of house, land ownership, access to electricity, livestock, and other assets. Similarly, the dependent variable for the lower poverty line was household poverty status. Age (OR: 0.94,  $p < 0.01$ ) and education (OR: 0.66 for class I-V and OR: 0.54 for class above VI,  $p < 0.01$ , reference = no education) of the household head, family size (OR: 2.02,  $p < 0.01$ ), household types (OR: 2.65 for *kacha* and OR: 3.52 for *Jhupri*, reference = *pacca* and *semi pacca*,  $p < 0.01$ ), number of dependents (OR: 1.45,  $p < 0.01$ ), per capita income (OR: 0.51 for 1000-2000TK, OR = 0.24 for 2000-3000TK and OR: 0.09 for 3000TK and above OR: 0.09,  $p < 0.01$ , reference = 0-1000TK,), ownership of land by the household (OR: 0.69 for 50-100 decimal, OR: 0.42 for 100-200 decimal and OR: 0.23 for 200 decimal and above,  $p < 0.01$ ), access to electricity (OR: 0.38, reference = no,  $p < 0.01$ ), ownership of non-agricultural assets by the household (OR: 0.43,

reference = no,  $p < 0.01$ ), and the proportion of male (OR: 1.20,  $p < 0.05$ ) and female (OR: 1.66,  $p < 0.01$ ) all had significant association with poverty.

Mamo and Abiso (2018) evaluated variables that influenced rural household's poverty levels in five districts of the Gamo Gofa zone, Southern Regional State of Ethiopia. A cross-sectional study was carried out using 4092 households. A household was considered to be poor if its welfare fall below the poverty level, and non-poor if it was above the poverty line. A LRM was applied to analyze the data. The following variables had a significant impact on the poverty status of households in the study area: dependency ratio (OR: 1.00,  $p < 0.001$ ), use of improved tools (OR: 1.560,  $p < 0.001$ ), household size in adult equivalent scale (AES) (OR: 1.85,  $p < 0.001$ ), saving habit (1.51,  $p < 0.001$ ), access to loan (OR: 1.55,  $p < 0.001$ ), resource base (OR: 2.20,  $p < 0.001$ ), land ownership of household in hecter (OR: 0.78,  $p < 0.001$ ), household labor availability (OR: 1.38,  $p < 0.001$ ), number of farm animals in tropical livestock unit (TLU) (OR: 0.86,  $p < 0.001$ ), use of agricultural inputs (OR: 3.99,  $p < 0.001$ ), and market access (OR: 2.11,  $p < 0.001$ ). The value of Hosmer and Lemeshow ( $\chi^2$ ) test with p-value was H-L  $\chi^2 = 7.22$ ,  $p = 0.51$ . The classification of the fitted LRM was estimated using sensitivity (84.6%), specificity (76.3%) and correctly predicted (81.1%).

Abrar-ul-Haq (2018) analysed the data of 600 households in rural Pakistan. He also constructed household empowerment index. The household empowerment not only provided new insights into the analysis of poverty in emerging nations but also into the monitoring and spatial comparing of household empowerment. Altogether 42 variables were used for the construction of Household Empowerment Indicator (HEMI) which were considered to be helpful in the study of poverty analysis. These variables were incorporated under the major three pillars namely – economic empowerment, social empowerment and political empowerment. The study revealed that household empowerment significantly reduced monetary poverty.

Teka et al. (2019) examined poverty and its factors and income disparity in Ethiopia's pastoral and agro-pastoral communities. The Gini coefficients, FGT indices and LRM were used to analyze 2295 households from zones 1 and 2 of Afar area. At 1% level, the following factors were found statistically significant with household poverty status: access to loan, household size, mobility, participation in non-pastoral/farm employment

by household members, and sex of household head. In addition to 1%, the following variables were also significant at 5% level: the Productive Safety Net Program (PSNP) membership, household head participation in local institutions, remittances, and market distance all showed statistical significance. 10% level of significance included literacy. The value of  $R^2$  in this case was 0.21.

Baser and Kaynakci (2019) examined poverty and its numerous causes in smallholder farms in Turkey's central district of Hatay province. A questionnaire was used to obtain the required information from 73 small farmers. Poverty was assessed using the poverty incidence and Poverty Gap Index (PGI). A LRM was used to investigate the causes of poverty. The dependent variable was poverty status and independent variables were the age, sex and literacy status of household head, social security, household size, land size, membership in farmer groups, and retirement status. They found that household size ( $\beta = 1.48$ ,  $p < 0.01$ ), retirement status ( $\beta = -2.88$ ,  $p < 0.05$ , reference category = no), social security status ( $\beta = -1.77$ ,  $p < 0.10$ , reference category = not having social security) and land size ( $\beta = -0.73$ ,  $p < 0.05$ ) were shown to be the most significant factors influencing poverty in smallholders farms. . The value of pseudo  $R^2$  reported in this study was 0.42.

Eyasu (2020) examined the severity and main causes of poverty at various spending quantiles using 350 households in North-Western Ethiopia. FGT index and quantile regression model were used for determining rural households' poverty. The size of the family as a whole and the household head's poor health were found to be factors that could raise poverty in rural families and lower their standard of life.

Shaga et al. (2021) identified factors that contributed to rural poverty in Ethiopia's Sodo Zuria Woreda of the Wolaita zone. To determine the level of rural household poverty, 152 rural households were selected. FGT indices and a logistic regression were used to determine rural household poverty. A total of fifteen explanatory variables (7 numeric and 8 categorical) were used in the analysis. Eight explanatory factors were statistically significant. The key elements, such as gender and age of household head, size of the family, level of education, land size, the number of livestock, use of technology, and saving habit, were all important in explaining the poverty of rural households. Family size was positively correlated but others were negatively associated with



poverty status. The predictive power of the model, sensitivity (95.7%), specificity (91.7%) and correct classified value (94.1%) were also reported.

A few other Nepalese studies on the issue of poverty are presented as follows: Wagle (2014) mainly focused on study of changing pattern of inequalities between different caste/ethnic groups in Nepal based on NLSS I to NLSS III data set; Patel (2012) just reviewed the data related to caste/ethnicity including poverty rates mentioning the source as CBS 2011; Adhikari (2016) assessed poverty dynamics and found chronic poverty based on the NLSS II and NLSS III dataset, and was concentrated on Dalit and some other ethnic groups.

Pant (2017) examined the factors that affect remittance receipt and how it affects household expenditure and child wellbeing in Nepal based on the NLSS III dataset. Devkota (2014) estimated the impact of migrant's remittances on poverty and inequality using the Probit model to calculate poverty types of household based on the NLSS III dataset. Thapa and Acharya (2017) focused households receiving remittances tended to spend on household expenditure based on NLSS III dataset. Lamichhane et. al. (2014) compared the poverty profile between people with and without disabilities in Nepal using NLSS III dataset. All these Nepalese studies were not focused to identify the factors associated with poverty except the study reported by Thapa et al. (2013). However, Thapa et al. (2013) attempted to identify the factors associated with poverty but the used data were not nationally representative. Based on these extensive review of literature, there is a research gap for identifying the most important factors associated with poverty using nationally representative data of Nepal. In this context, this study is an attempt for the same using a nationally representative data of NLSS III.

In general, based on the extensive review of literature, various characteristics or factors are reported to be associated with poverty. It is crucial to keep in mind that poverty is a socio-economic conditions caused by different factors and is not only associated with income and consumption expenditure. The significant variables associated with poverty as reported by the extensive review of literatures which can be kept broadly under three different characteristics such as: demographic characteristics (age of the households head, sex of household head, household size, number of children) , socio-economic characteristics (literacy status of the households head, educational level of the households head, land ownership of the households, remittance status of the

households, number of literate members in a house, and community characteristics (status of access to nearest market center, access to the nearest health center).

Based on these review of literature, and considering the availability of the variables in the Nepal Living Standard Survey data file of 2010/11, the following variables are considered as possible candidate variables from which the final risk factors are identified through suitable statistical modeling: Sex of household head, literacy status of household head, remittance receiving status of household, land ownership status of household, household with access to nearest market centre, number of literate members of working age population (WAP), and status of household with number of children under 15 years (Acharya et al, 2022a). The details about the reasons of categorization of quantitative variables and their coding schemes, etc. are expalind in detail in Chapter three of this thesis.

#### **2.4 Statistical Methods/Models used in Poverty Analysis**

Many researchers have used the Foster, Greer and Thorbecke (FGT) poverty indices as descriptive statistics (Yusuf et al., 2008; Adepoju & Oluoha, 2008; Onu & Abayomi, 2009; Sanusi et al., 2013; Edoumiekumo et al., 2014; Margwa et al., 2015; Teka et al., 2019; Baser & Kaynakci, 2019; Eyasu, 2020; Shaga et al., 2021). On the other hand, a few researchers have used multiple linear regression to model per capita income or consumption expenditure (Sakuhuni et al., 2011; Gounder et al., 2011; Akerele & Adewuyi, 2011; Salami & Atiman, 2013; Tuyen, 2015). In another case, poverty status has been modeled using probit regression (Gounder, 2012; Balarabe, 2014), and almost all studies have found to be using logistic regression (Chhetry, 2005; Yusuf et al., 2008; Achia et al., 2010; Ennin et al., 2011; Sekwati et al., 2012; Osowole et al., 2012; Sekhampu, 2012; Dudek & Lisicka, 2013; Thapa et al., 2013; Omoregbee et al., 2013; Sanusi et al., 2013; Khudri & Chowdhury, 2013; Leekoi et al., 2014; Deressa & Sharma, 2014, Adetayo, 2014; Edoumiekumo et al., 2014; Myftaraj et al., 2014; Makame & Mzee, 2014; Xhataj & Nurja, 2014; Spaho, 2015; Habyarimana et al., 2015; Khan et al., 2015; Farah, 2015; Margwa et al., 2015; Majeed & Malik, 2016; Maloma, 2017; Mohammed, 2018; Kona et al., 2018; Immam et al., 2018; Mamo & Abiso, 2019; Teka et al., 2019; Baser & Kaynakci, 2019; Shaga et al., 2021).

To evaluate the fitted LRM, some have used coefficient of determination ( $R^2$ ) as a measure of goodness of fit (Yusuf et al., 2008; Sekwati et al., 2012; Thapa et al., 2013; Adetayo, 2014; Xhataj & Nurja, 2014; Khan et al., 2015) whereas others have used pseudo  $R^2$  as a measure of goodness of fit (Okojie, 2002; Makame & Mzee, 2014; Spaho, 2014, Mohammed, 2017; Baser & Kaynakci, 2019). Moreover, pseudo  $R^2$  as well as Hosmer and Lomeshow ( $\chi^2$ ) (goodness of fit test) have also been used for model evaluation (Deressa & Sharma, 2014; Habyarimana et al., 2015; Mohammed, 2017; Mamo & Abiso, 2018). Nonetheless, none of the reviewed studies have used regression diagnostics of the used LRM.

Most of the studies have been found to be using LRM in identifying the factors associated with poverty, and to quantify the effect of each factor on poverty. LRM computes the regression coefficient and odds ratio (OR) of each independent variable. The association of dependent variable (having only two levels) with independent variables can also be analyzed by using LBRM. However, the use of LBRM in identifying the factors associated with poverty is found to be almost rare in the previous literature with the best of our knowledge. Most of the applications of LBRM have been found reported in epidemiological or clinical studies, but not to analyze in poverty data. This model also facilitates the assessment of the effects of independent variables on dependent variable. Similarly, the LBRM computes the regression coefficient and the risk ratio or relative risk (RR) or prevalence ratio (PR) of each independent variable. Generally, RR or PR have been reported in cohort studies. However, it can also be reported in cross sectional data (Barros and Hirakata, 2003) if the outcome of the event interest is common i.e.  $\geq 10$  (Greenland, 1987; McNutt et al., 2003; Katz, 2006; Viera, 2008; Ranganathan et al., 2015 and Gallis & Turner, 2019).

Barros and Hirakata (2003) used the information that came from a population-based survey to measure the association of maternal smoking with explanatory variables and quantify the effect of these independent covariates on maternal smoking in Pelotas, Southern Brazil. Since the prevalence ratio is more understandable and easier to convey to non-specialists than the odds ratio, they used LBRM to directly predict the prevalence ratio rather than logistic regression for analyzing binary outcomes. Similarly, Lumley et al. (2006) used a sample that consisted of 6,814 men and women that were Caucasian, African-American, Hispanic, or Chinese-American to analyze

relative risk of coronary of Multi-Ethnic study of Atherosclerosis. They concluded that relative risk should be estimated rather than odds ratio. In another study, Coutinho et al. (2008) used the data collected between May 2003 and April 2005 from a cross-sectional epidemiological study on 2072 senior citizens in Sao Paulo, southeast of Brazil to estimate the association between depressive episodes and self-rated poor health for empirically comparing the Cox, log-binomial, Poisson and logistic regressions. They concluded that the LBRM produced unbiased prevalence ratio. The confidence interval of the regression coefficient yielded by log-binomial regression is reported narrower than the confidence interval yielded by LRM (Deddens & Petersen, 2008; Barr et al., 2016). However, if the outcome is very prevalent and the confounding variables are continuous, there might be problems with model convergence. The failure convergence of log-binomial regression was also reported by Williamson et al. (2013). Yelland et al. (2011) realized that the relative risk was a clinically important predictor of treatment impact on binary outcomes. They used LBRM to estimate relative risk of randomized controlled trials of fish oil supplements versus placebo for preterm infants. However, this model frequently suffered from convergence. Therefore, they recommended trying log-binomial model first and using other alternative in the case of convergence problem. Similarly, Espelt et al. (2017) used cross-sectional data from the Survey on Health, Ageing, and Retirement in Europe (SHARE) which included 41,263 participants from 16 European countries to examine the difference between in prevalence ratios (PR) and odds ratios (OR) of hazardous drinker between men and women. In another study, Schwendinger et al. (2021) reported that associations and effects due to their simplicity of interpretability should be assessed using estimated values of relative risks. Relative risk can be directly inferred using estimated values of regression coefficients of LBRM. These results have indicated that log-binomial model is one of the options model for analyzing cross-sectional data with binary outcomes.

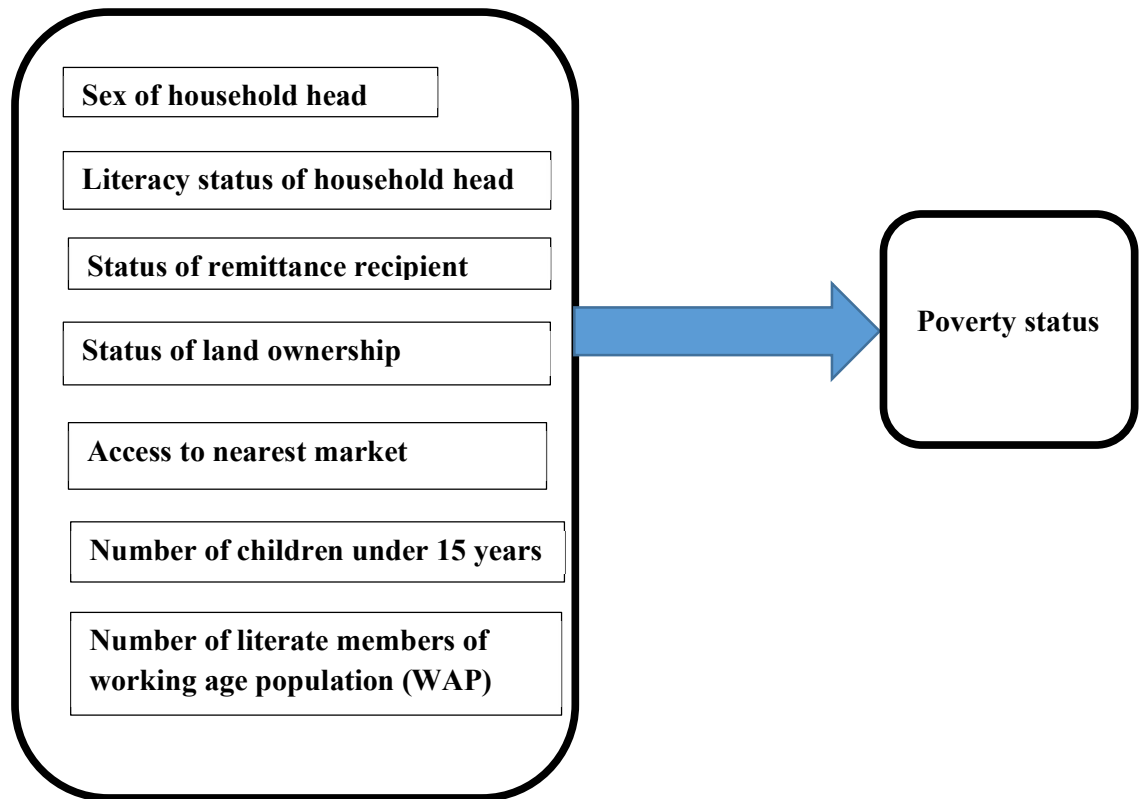
Wacholder (1986) initially suggested a straightforward method of directly assessing risk ratio (RR) for examining the relationship between independent factors and the binary outcome variable. Later, Barros and Hirakata (2003) argued that in cross-sectional investigations with common number of desired event, the odds ratio (OR) usually overestimates the risk ratio (RR). The LRM is a generalized linear model (GLM) with logit link and binomial probability distribution, which results in the odds ratio (OR). The LBRM is also a GLM with log link and binomial probability

distribution, which results in the risk ratio (RR). Their descriptions and uses have also been highlighted by Robbins et al. (2002) and McNutt et al. (2003). Blizzard and Hosmer (2006) recommended the method for examining the model's good fit, some diagnostics of the LBRM. They performed the test of model's good fit and regression diagnostics with real data of Tasmanian Infant Health Study (TIHS). LBRM model was found good fitted and satisfied the regression diagnostics (plots based on leverage values, and Cook's distance) through graphical assessment. There is standard practice for converting the OR into RR. However, Robbins et al. (2002) made it clearly evident that the confidence intervals generated by these converted approaches were incorrect.

In summary, the extensive review of literature has clearly indicated that the study for examining the independent factors in relation to household poverty specially using the dataset representing Nepal has not been reported so far. The frequent use of logistic regression as a statistical model is found to be reported in identifying the factors associated with poverty. The use of LBRM can be found as an alternative to LRM to quantify the effects of independent covariates mostly in the epidemiological and in clinical research even for cross-sectional as well as for cohort data. The studies related to the comparison of log-binomial and LRM are also reported using epidemiological and clinical data but not for the data of social sciences such as poverty. Keeping in view of these research gaps, this study is an attempt to identify the factors associated with poverty using suitable regression model. Attempts have been made to compare the models with respect to effect size and its precision and good fit of the model. The comparison has also been attempted with respect to regression diagnostics and the issue of convergence of the empirically developed model.

## **2.5 Conceptual Framework**

On the basis of reviewed empirical studies, and on the basis of theoretical considerations, conceptual framework for this study has been developed and presented through the schematic diagram (Figure 1).



**Figure 1:** Schematic Diagram of Conceptual Framework

## CHAPTER 3

### 3. MATERIALS AND METHODS

The main objective of this chapter is to deal with the materials and methods applied in this study. More specifically, this chapter discusses on the available secondary data and preparation of data file, scheme of data analysis, selection of variables, dichotomization of quantitative covariates and theoretical aspects of two statistical models with their diagnostics and stability. Further, theoretical features of comparison of the two models with respect to different criteria are also dealt with.

#### 3.1 Data Source and Data File Preparation

The main source of data for this study is the available data on the socio-economic and demographic variables of 5,988 households and 28,670 individuals of the NLSS III. The individual level data are converted into household level data to have data on a number of variables such as - the number of children (0-14 years of old), working-age members (15 to 64 years old) and elders (65+ years old) by gender and literacy status (literate/illiterate) within each household. All these household level data including the available data on the variables poverty status (poor/non-poor), household weight and individual weight are compiled in an SPSS data file. The household and individual weights have used to estimate weighted and un-weighted poverty rates of household and individual at the national, sub-national and desired social groups. Before finalizing the above data, a number of meetings held with CBS experts in order to clarify and resolve number of issues that arose during the preliminary phase of analysis of provided data.

#### 3.2 Scheme of Data Analysis

A set of poverty profiles presented in Tables 1 to 4 of Chapter I of this study clearly shows that poverty is not a state of static reality but a dynamic reality. This works in the sense that poverty varies across the time, space and population groups stratified by caste/ethnicity. Moreover, poverty rate also varies across the population groups stratified by other variables such as employment sector, educational level, sex and age of the household head, and household size, number of children within household, area of land holding by household (Tables 1.4.1 to 1.4.6 of CBS, 2005). This type of poverty

profiles is useful for policy makers for identifying locations and group of population where the poverty reduction initiatives are most urgently required.

The other set of poverty profile deals with the comparison of population or household level socio-economic and demographic characteristics between poor and non-poor groups. One major problem behind such profile is on the definition of 'poor' group which may not be agreeable to all scholars. This problem can be avoided as per the suggestion of Miller and Roby (1967) by comparing the socio-economic and demographic characteristics across the *consumption quintile groups* which are basically the five equal groups of individuals ordered from the poorest to the richest depending upon their level of per capita consumption. If  $G_1, G_2, G_3, G_4$  and  $G_5$  are five consumption quintile groups then each group includes 20% of total population and satisfies the following ordering relation:

$$G_1 < G_2 < G_3 < G_4 < G_5$$

where the symbol "<" is a group ordering in the sense that the per capita consumption of any individual belonging to a lower consumption quintile group is smaller than that of any individual belonging to a higher consumption quintile group. Note that  $G_1$  group includes all those individuals who are poorest among the poor and  $G_5$  includes all those who are richest among the rich based on the measure of poverty level. In literature of poverty analysis,  $G_1$  is called the poorest group and  $G_5$  is called the richest group.

In the present study, SPSS syntax file was created to generate the consumption quintile groups in the data file using the available data on the two variables - per capita consumption expenditure per household and household size (Appendix B1). The number of individuals and households within each quintile group are presented in Appendix B2. This type of poverty profiles are useful for policy makers as well as academicians for understanding type of initiatives for the needy people as well as for understanding the inequalities or disparities in socio-economic and demographic indicators.

After preparing relevant poverty profile, data analyses for generating statistical results of two theoretical regression models, namely the LRM and the LBRM as per objective of the present study are carried out. Theoretical aspects of these two models have been



discussed later in this chapter, since before the discussion of theoretical model it is more appropriate to discuss about the outcome variable as dependent variable and covariates as independent variables for both models.

### **3.3 Outcome Variable and Covariates**

The outcome variable for both the model is household poverty status with two possible traits 'poor' and 'non-poor'. After extensive review of literatures (Chapter II), efforts have been made to select some of the most suitable variables in the context of Nepal, from the NLSS III data file, as covariates. Some of the efforts made in this endeavor have been discussed in Chapter IV. Finally, in consultation based on extensive review on relevant field and in consultants meeting with relevant experts, the following seven household level variables are identified as covariates for both models.

1. Sex of household head with two possible outcomes - 'male' and 'female'
2. Literacy status of the household head with two possible outcomes - 'literate' and 'illiterate'
3. Status of remittance receiving households with two possible outcomes - 'receiving' and non-receiving
4. Area of land holding in hectare
5. Access to the nearest market in minutes of time taken to reach by any mode of transportation
6. Number of children under 15
7. Number of literate members of working-age population (WAP)

The last covariate in this study is considered as a proxy measure of *human capital within each household*, since very few literatures analyze the impact of human capital flight on poverty in Nepal. During literature review, no study is found using all the seven covariates for estimating odds ratios or relative risks of each covariate. However, many of these covariates are directly or indirectly related to a list of 42 variables selected by Abrar ul haq et al. (2018) for constructing the Household Empowerment Index (HEI) for rural households of Pakistan.

The last four covariates as mentioned above are quantitative variables. As a result, the presence or absence of outliers in each variable is thoroughly investigated in Appendices B3 and B4. The conclusion drawn through the investigation is that substantial number of outliers presented in each quantitative variable (see Appendix

B5). In the presence of such substantial number of outliers, the results produced by the two models may have several problems, such as stability and convergence of models unless some adjustment been made on each of the quantitative variables.

### **3.4 Dichotomization of Quantitative Variable**

Transforming a quantitative variable into a dichotomous variable using an appropriate threshold value is not a new practice in statistical applications, particularly in poverty analysis. For example, in the measurement of *poverty incidence* (percentage of poor population), a quantitative variable ‘per capita consumption expenditure’ is transferred into a dichotomous variable ‘poverty status’ using a threshold value ‘poverty line’ which demarcates individuals into poor and non-poor group based on their level of per capita consumption expenditure.

Scholars generally recommend avoiding dichotomization of quantitative variables due to several reasons such as loss of explanatory information, loss of statistical power, and so on (Maccallum et al., 2002). However, scholars also write that in a rare situation when there are two distinct taxonomy classes underlying the quantitative variable, dichotomization is possible with rigorous justification such as taxometric analysis.

Theoretically it is reasonable to argue that each of the four quantitative covariates selected for the models of this study possesses two distinct taxonomy groups, namely disadvantage and advantage group that corresponds to poor and non-poor group, respectively. With this argument the four quantitative covariates are dichotomized each with appropriate threshold value. The key objectives of dichotomizing each of the four quantitative variables are to provide significant comparisons between the two disjoint and exhaustive groups of households (Acharya et al., 2022b). These two are disadvantaged and advantaged groups (Acharya et al., 2022b).

It is possible to obtain the significant results from the quantitative variables too, but it would be a difficult task for the policy makers to differentiate the vulnerable or targeting group. Therefore, it is logical to dichotomize numeric variables, so that the interpretation becomes meaningful especially from the policy perspective discussed in Chapter II. Below is an explanation of the dichotomization procedure of four numeric variables. The dichotmization is performed selecting a threshold value with rationale.

#### **3.4.1 Dichotomization: *Area of Land Holding***

Possession of land has a variety of advantages such as shelter, crop production, protection against natural disasters or shocks, and socio-political prestige in a society (Kousar et. al., 2015). In view of this fact the above said covariate was dichotomized with a threshold value of 0 hectare which demarcates the households into two groups: having no land (disadvantaged group) and having land (advantaged group). The available NLSS III data shows that the distributions of disadvantaged and advantaged group are correspondingly 28.8 and 71.2 percent, and the poor households within group are correspondingly 27.0 and 15.1 percent.

#### **3.4.2 Dichotomization: Access to Nearest Market Center**

Taylor et al. (2009) reported that a poor household cannot learn about or adopt new technologies, market and its production, receive inputs, sell labor, obtain credit, insure against risks, or purchase consumption goods at affordable prices without strong access to markets. Realizing such importance of access to markets, the above said covariate is dichotomized with a threshold value of 30 minutes. This demarcates the households into two groups: beyond the reach of 30 minutes of market (disadvantaged group) and within the reach of 30 minutes of market (advantaged group). This threshold value of 30 minutes has been taken from CBS reports. The distributions of disadvantaged (by having poor access) and advantaged (by having better access) are correspondingly 48.0 and 52.0 percent, and the poor households within group are correspondingly 26.0 and 11.6 percent.

#### **3.4.3 Dichotomization: *Number of Children***

The rationale for choosing threshold value is presented in Table 6. This shows the incidence of poverty is lower than the national average for every households' group with  $\leq 2$  children. It is higher for each group of households with more than two children. The above said covariate is dichotomized by a threshold value of 2 children which demarcates the households into two groups: households with  $> 2$  children (disadvantaged group) and households with  $\leq 2$  children (advantaged group). The distributions of disadvantaged and group of households are correspondingly 73.8 and 26.2 percent, and the poor households within group are correspondingly 40.1 and 10.9 percent.

**Table 6: Poverty Incidence (%) of Households Grouped by Number of Children**

	Household grouped by the number of children						National
	0	1	2	3	4	5+	
Within group poverty incidence	5.9	11.6	19.6	33.5	42.3	55.7	25.2
Within group poverty incidence	13.5			41.4			

Source: Acharya et al. (2022a)

The threshold value of 2 children is consistent with the 2011 total fertility rate of 2.3 births per woman (NDHS 2016). It is also consistent with respondents' responses to mean ideal number of children 2.2 and 2.3 for currently married women and men respectively (NDHS 2016).

#### **3.4.4 Dichotomization: Number of Literate Working Age Population (WAP)**

Out-migration of educated WAP to remit money back at home or to settle abroad is a well known problem in Nepal. In order to investigate the impact of such out-migration on poverty, a household level numeric variable “number of literate members of WAP” is selected. Then, converted it into dichotomous variable by grouping the households into two groups. one group of households each has no literate member of WAP. The other group of households each has at least one literate member of WAP. The rationale behind grouping is as follows: households without no literate working-age member (disadvantaged group) have more difficulty to fight against poverty than those with at least one literate member (advantaged group). The percentages of households in the former and later groups are 19.3 and 80.7 percent, respectively, and the percentages of poor households are 30.8 and 15.6 percent, respectively.

All the independent variables selected for the study are binary. The binary coding scheme of the selected seven covariates and the response variable is presented in Table 7.

**Table 7:** Selected Covariates and Response Variable with Group Formation and Coding Scheme

Response variable and covariates	Group coding schemes
Response Variable	
Household poverty status	Poor = 1 and Non-poor = 0
Covariates:	
1. Sex of household head	Female = 1 and Male = 0
2. Literacy status of household head	Illiterate = 1 and Literate = 0
3. Status of remittance recipient household	No = 1, Yes = 0
4. Status of land ownership	No = 1, Yes = 0
5. Access to nearest market center	Poor access = 1 and Better access = 0
6. Number of children under 15 years	More than two = 1 and At most two = 0
7. Number of literate members of working age population (WAP)	None = 1 and At least one = 0

Source: Defined based on NLSS III data

### 3.5 Statistical Models

The association between poverty status and each dichotomized independent variable is assessed using the Chi-square test. The Phi-coefficient is used to calculate the association and amount of effect size of each test. To determine the risk factors affecting the poverty which are associated with poverty status, the multiple LRM and LBRM are estimated since the response variable is binary. The goodness of fit, diagnostic, risk assessment and stability of each model are also assessed. The analysis for both the statistics model are exclusively based on the unweighted data file.

#### 3.5.1 Selection of Variables for the Logistic Regression Model (LRM)

Based on extensive review of literature, seven covariates associated with the poverty were selected. Since the response variable of the study is poverty status of a household, LRM has been used to investigate the relationship between dependent variable and independent variables. The Chi-square test has first been used to determine the relationship between each of the seven proposed covariates and the response variable. The phi-coefficient is also used to calculate the effect size of each Chi-square test. Finally, to determine potential independent variables which are statistically significant with response variable, both stepwise forward and backward selection procedure are implemented for the development of LRM.

### 3.5.2 Logistic Regression Model (LRM) and its Fitting

In order to identify the effects of covariates on the outcome variable, appropriate regression model which suits for the given data structure is applied. When the outcome of interest is dichotomous and the independent covariates may be of categorical or continuous, logistic regression or simply a logistic regression model (LRM) is used. In this situation of having binary outcome, many distribution functions have been suggested and some of them were discussed by Cox and Snell (1989). The rationale behind using logistic distribution is flexible and easily defined mathematical function and its ease of meaningful interpretation (Hosmer and Lemeshow, 2000). In this thesis work, the outcome of interest is the poverty status (poor vs. non-poor), for which the category 'poor' is coded by 1 and 'non-poor' is coded by 0. From the list of different independent variables as identified through extensive review of literature and series of discussions with the experts in the relevant field, the candidate variables for multivariable LRM are finalized through bivariate analysis using Chi-square test.

Based on the fundamental notion of applying linear regression, let us use the notation  $\pi(x) = E(Y | x)$ , which represents the conditional mean of Y given x when the logistic distribution is applied.

Let us consider  $p$  independent covariates  $x_1, x_2, \dots, x_p$  and the specific form of the LRM is given by

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}} \quad (3.1)$$

The logit transformation in terms of  $\pi(x)$  is as follows.

$$g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (3.2)$$

Equation (3.2) may have followed many desirable properties of linear regression model.

The ratio term defined in equation (3.2) is called the odds. Computing the odds is a commonly used technique of interpreting probabilities (Fleiss, et al., 2003).

In LRM, the conditional mean of the regression equation is defined in such a way that the formulation will be bounded between 0 and 1. Another very important component in the LRM is that the error term is distributed as binomial distribution in contrary to the normal distribution in linear regression model. Let us explain this initially considering the fundamental notion of linear regression model, and then slowly transferring into LRM as follows.

The researcher expresses the regression model as defined earlier  $y = E(Y|x) + \varepsilon = \pi(x) + \varepsilon$ . The outcome variable  $y$  will have only two values 0 and 1. When  $y = 1$ ,  $\varepsilon = 1 - \pi(x)$  with probability  $\pi(x)$ ; when  $y = 0$ ,  $\varepsilon = -\pi(x)$  with probability  $1 - \pi(x)$ . This indicates that  $\varepsilon$  has a distribution with mean zero and variance of  $\pi(x)[1 - \pi(x)]$ . For detail, please see Hosmer and Lemeshow (2000).

### 3.5.3 Fitting of the Multiple Logistic Regression Model (LRM)

Let us suppose that there are  $n$  independent observations  $(x_i, y), i = 1, 2, \dots, n$ , and  $p$  independent covariates which are used in the regression model (3.1). Fitting of the multiple LRM demands the estimates of  $(p+1)$  number of regression coefficients  $(\beta_0, \beta_1, \dots, \beta_p)$  including the intercept in the model. In order to estimate the regression coefficients, based on the principle of maximum likelihood, the log likelihood function will be generated and then  $(p+1)$  likelihood equations are developed differentiating the log likelihood function with respect to the considered  $(p+1)$  regression coefficients. The likelihood equations may be expressed in the following form. Please see Hosmer and Lemeshow (2000); Neter et al. (1996); Afifi (2004); Kleinbaum (2010) for detail explanations about the likelihood functions.

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0 \quad (3.3)$$

$$\sum_{i=1}^n x_{ij} [y_i - \pi(x_i)] = 0 \quad (3.4)$$

where  $j = 1, 2, \dots, p$  stands for  $p$  number of independent variables.

The maximum likelihood equations (3.3) and (3.4) can be solved by using the computer intensive program such as iterative weighted least square method, and it cannot be solved through manual procedure. Finally the estimates of  $(p+1)$  coefficients will be

obtained. McCullagh and Nelder (1989) explained in detail about the methods adopted by different programs used for solving maximum likelihood function.

### 3.5.4 Test of Significance of the Fitted Model

After running the regression model, it is generally checked whether the independent variables used in the model are significantly associated with the outcome or not. The assessment of the significance of independent variables with outcome variable is done through the F-test using analysis of variance in the case of linear regression through the comparison of observed and predicted values under two models. One model is the model without covariates and another is the model with the covariates. In linear regression model, when no variable is considered  $\hat{\beta}_0 = \bar{y}$  i.e. the mean of the response variable. Similar approach can also be adopted in LRM while assessing the significance of the relation of independent variables with the outcome variable but not directly. In regression model, the comparison of observed and predicted values is done based on likelihood functions. For this purpose, the deviance  $D$  based on the likelihood function (McCullagh & Nelder, 1989) is computed and is as follows.

$$D = -2 \ln \left[ \frac{\text{likelihood of the fitted model}}{\text{likelihood of the saturated model}} \right] \quad (3.5)$$

where, the saturated model is that model which contains as many parameters as there are data points. The quantity inside the large brackets is known as the likelihood ratio. The likelihood of the saturated model is equal to 1 (Hosmer & Lemeshow, 2000) since outcome variable are either 0 or 1.

In order to assess the overall significance for  $p$  coefficients of independent variables, the change in the value of deviance ( $D$ ) with and without independent variables is computed as follows.

$$G = D(\text{model without variables}) - D(\text{model with the variables}) \quad (3.6)$$

This  $G$  statistic can be expressed in terms of log likelihood as follows

$$G = -2 \ln \left[ \frac{\text{likelihood without variables}}{\text{likelihood with variables}} \right] \quad (3.7)$$



Equation (3.7) can be further simplified as follows.

$$G = -2[\ln(L_0) - \ln(L_f)] \quad (3.8)$$

where,  $L_f$  stands for the value of the likelihood function for the full model and  $L_0$  stands for the value of likelihood function for the null model (i.e. only having intercept).

The statistics  $G$  defined in (3.8) follows the Chi-square distribution with  $p$  degrees of freedom ( $G \sim \chi^2_{(p)}$ ).

The overall significance of  $p$  regression coefficients of independent variables in the multiple LRM has been assessed through the  $\chi^2$ . The null hypothesis for this is :  $\beta_j = 0$ , for  $j=1,2,\dots,p$ . The alternative hypothesis is: at least one slope coefficient for the covariate is different from zero. This test is also termed as the Omnibus test for the test of significance of the overall significance of  $p$  independent covariates in the fitted model. The significance is assessed at 5% level of significance.

### 3.5.5 Test of Significance of Individual Regression Coefficient

In order to test the significance of individual regression coefficient ( $\beta_j$ ), Wald test and Score test have been suggested. Rao (1973) had explained the assumptions and other details about these tests. The behavior of Wald test and its anomalous behavior for failing to reject the null hypothesis even for having the significant regression coefficient was discussed by Hauck and Donner (1977). Later, Jennings (1986) also studied the adequacy of inference based on Wald test and concluded in the similar direction as indicated by Hauck and Donner (1997).

However, Wald test (Wald, 1943) is easily available while running logistic regression in most of the statistical software such as STATA, IBM SPSS, etc. In this research work also, Wald test has been applied to test the significance of individual regression coefficient ( $\beta_j$ ), and is computed as follows.

$$W_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (3.9)$$

This Wald Statistic  $W_j$  follows normal distribution, and it works under the null hypothesis that  $\beta_j = 0$  vs.  $\beta_j \neq 0$ . The regression coefficient is considered significant at 5% level of significance.

### 3.5.6 Confidence Interval for Regression Coefficient

The confidence interval estimation for the slope coefficient  $\beta_j$  and intercept  $\beta_0$  are based on Wald test and formulated adopting the as usual statistical theory. The limits of a  $100(1-\alpha)\%$  confidence interval for the slope  $\beta_j$  and the intercept  $\beta_0$  are formulated in equation (3.10) and (3.11) respectively as follows.

$$\hat{\beta}_j \pm Z_{1-\alpha/2} SE(\hat{\beta}_j) \quad (3.10)$$

$$\hat{\beta}_0 \pm Z_{1-\alpha/2} SE(\hat{\beta}_0) \quad (3.11)$$

where  $Z_{1-\alpha/2}$  is the value of standard normal variate for given level of significance ( $\alpha$ ) and  $SE(\cdot)$  estimator of standard error of respective parameters based on the fitted regression model

### 3.5.7 Interpretations of Regression Coefficient

The slope coefficients ( $\beta_j$ ) in logistic regression are generally interpreted in terms of log odds which may not be easy to understand and not much meaningful with reference to problem specific situations. The range of log odds varies from  $(-\infty$  to  $+\infty)$ . The LRM also yields the odds ratios shortly denoted by (OR) which is computed as  $OR = \exp(\beta_j)$ , for  $j = 1, 2, \dots, p$ . Odds ratio lies between  $(0$  to  $\infty)$ . Generally OR measures the effect of independent variable on outcome variable in terms of risk in the defined risk category as compared to the reference category in the case of categorical independent variable, and measures the risk per unit change in the case of independent continuous variable used in the regression model.

The interpretation of OR for categorical independent variables is as follows:

- i. If  $OR = 1$ , there is no risk between the comparing groups.

- ii. If  $OR > 1$ , the risk is increased with reference to the reference category or reference group
- iii. If  $OR < 1$ , the risk is decreased with reference to the reference category or reference group which indicates that the factor is protective.

For continuous independent covariate  $X$  (say), slope coefficient  $b = \exp(b)$  is interpreted as the ratio of the odds with value  $(x+1)$  with respect to the odds with value  $x$ . Therefore,  $\exp(b)$  is the incremental odds ratio corresponding to an increase of one unit in the variable  $x$ , keeping effects of all other  $x$  variables constant.

The limits of a  $100(1 - \alpha)\%$  confidence interval for odds ratio (OR) for the independent variables in the model is based on the limits of confidence interval of regression coefficient ( $\beta_j$ ). The exponentiation of limits of confidence interval of regression coefficient ( $\beta_j$ ) yields the confidence interval for OR, the expression for which is as shown below.

$$\exp\left[\hat{\beta}_j \pm Z_{1-\alpha/2} SE(\hat{\beta}_j)\right] \quad (3.12)$$

In this thesis work, the odds ratio (OR), standard error of OR for each independent variable of the fitted LRM with confidence interval along with p-value is computed and reported.

### 3.5.8 Coefficient of Determination ( $R^2$ ) in Logistic Regression

The role and the use of coefficient of determination or coefficient of multiple determination ( $R^2$ ) while analyzing the data through regression analysis, for the first time, established by Rao (1973). In linear regression model, the coefficient of determination can be measured as follows.

$$R^2 = \frac{SSR}{SST} = \frac{\text{Sum of square due to regression}}{\text{Total sum of square}} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (3.13)$$

The coefficient of determination measures the portion of the total variation in outcome variable explained by the variation of independent variables.

There are different techniques so far proposed by the statisticians in order to compute  $R^2$  in LRM. In this regard, Mittlbock and Schemper (1996) had reviewed twelve different methods of measuring explained variation in LRM. Furthermore, Menard (2000) also explained other methods for measuring the same in the context of logistic regression. In practice, it could not be found any consensus for only one method for computing  $R^2$  while applying LRM. Nonetheless, most of the statistical software such as STATA, SAS, SPSS have used either  $R^2$  proposed by McFadden (1974) or Cox-Snell  $R^2$  (Cox & Snell, 1989) or suggested by Nagelkerke (1991). The computational formula to compute  $R^2$  using each of these techniques are briefly explained as follows.

### 3.5.8.1 McFadden's $R^2$

In LRM, the value of  $R^2$  can be computed by using the expression suggested by McFadden (1974) is as follows.

$$R^2_{(MF)} = 1 - \frac{\ln(L_f)}{\ln(L_0)} \quad (3.14)$$

where, 'ln' stands for log,  $L_f$  stands for the value of the likelihood function for the full model and  $L_0$  stands for the value of likelihood function for the null model (i.e. only having intercept).

The McFadden's  $R^2$  is also known as log-likelihood ratio  $R^2$  since the formula is exclusively based on the ratio of the value of log likelihood functions.

Statistical software STATA generally yields the MCFadden's  $R^2$  by default with the name of Pseudo  $R^2$ .

### 3.5.8.2 Cox-Snell $R^2$

The value of  $R^2$  in LRM can also be computed by the method proposed by Cox and Snell (1989) known as Cox-Snell  $R^2$ . It is computed also based on the log-likelihood values as follows.

$$R^2_{(C-S)} = 1 - \left\{ \frac{L_0}{L_f} \right\}^{2/n} \quad (3.15)$$

where  $L_0$  stands for the value of likelihood function for the null model and  $L_f$  stands for the value of likelihood function for the full model,  $n$  is the sample size.

### 3.5.8.3 Nagelkerke $R^2$

This is another method of computing  $R^2$  in LRM which had been proposed by Nagelkerke (1991). This method is the modified method of Cox-Snell  $R^2$ , and can be computed by using the following expression.

$$R^2_{(N)} = \frac{R^2_{(C-S)}}{1 - (L_0)^{2/n}} = \frac{\left[ 1 - \left\{ \frac{L_0}{L_f} \right\}^{2/n} \right]}{\left[ 1 - (L_0)^{2/n} \right]} \quad (3.16)$$

where,  $L_0$  and  $L_f$  represents the value of likelihood function of null and full model respectively.

Statistical software IBM SPSS and SAS both provides Nagelkerke  $R^2$  while running the LRM.

There may be some reasons to prefer one coefficient of determination over another. Nonetheless, Menard (2000) had indicated that McFadden's  $R^2$  is to be preferable than others in logistic regression based on satisfying the majority of criteria proposed by Kvalseth (1985) along with its reasonable interpretations relatively. In this thesis work, the value of McFadden's  $R^2$  i.e Pseudo  $R^2$  is computed and reported.

### 3.5.9 Test of Goodness of Fit of the Model

After fitting the model, it is necessary to evaluate the goodness of fit of the model i.e. to assess whether the fitted model is effectively able to describe the outcome variable. In order to assess the goodness of fit of the multiple logistic regressions model the following two tests have been performed.

- i. Hosmer & Lemeshow test
- ii. AIC and BIC Statistic

### 3.5.9.1 Hosmer and Lemeshow Test

In order to examine the goodness of fit of the fitted LRM, one of the most popular and widely used test is the Hosmer and Lemeshow  $\chi^2$  test (Hosmer & Lemeshow, 1980; Lemeshow & Hosmer, 1982; Hosmer et al., 1988). This test compares the observed and expected values generally dividing into 10 groups. Hence, the subjects are grouped into 'g' groups. Each group containing  $\frac{n}{10}$  subjects. The number of groups 'g' is generally about 10. It can also be less than 10, for fewer subjects. Ideally, the first group contains  $n_1' = \frac{n}{10}$  subjects having the smallest estimated success probabilities. The second group contains  $n_2' = \frac{n}{10}$  subjects having the second smallest estimated success probabilities, and so on. The success probabilities are obtained from the fitted model.

The outcome variable y takes the value 1 and 0. Therefore, for y = 1, estimates of expected values are found by summing the estimated probabilities over all subjects in a group. In similar fashion, for the outcome variable taking zero i.e. y = 0, the estimated expected values are found by summing over all subjects in the group with complement of estimated probability i.e. one minus the estimated probability. For this, the Pearson Chi-square statistic was computed from g×2 table of observed and expected frequencies.

The goodness of fit statistic can be computed as follows.

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} \quad (3.17)$$

where  $n'_k$  is the total number of subjects in the  $k^{th}$  group

$O_k$  is the observed number of responses among the  $c_k$  covariate patterns in the  $k^{th}$  decile, and is defined mathematically as follows.

$$O_k = \sum_{j=1}^{c_k} y_j \quad (3.18)$$

The average estimated probability  $\bar{\pi}_k$  is computed as follows

$$\bar{\pi}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{\pi}_j}{n_k} \quad (3.19)$$

where,

$\hat{\pi}_j$  is the estimated probability for  $j$  covariate pattern

$m_j$  is the number of subjects with  $x = x_j$ , for  $j = 1, 2, 3, \dots, J$  associated with covariate.

The goodness of fit statistic is approximated with Chi-Square distribution with  $(g-2)$  degrees of freedom. This test works under the null hypothesis that the good fit of the model is not violated.

### 3.5.9.2 AIC and BIC Statistic

For examining the relative quality or the performance of the fitted LRM, the Akaike information criterion (AIC) has been computed. The computation formula for AIC statistic (Akaike, 1974) is as follows.

$$AIC = -2LL + 2k \quad (3.20)$$

where  $LL$  is the maximum log-likelihood of the fitted model

$k$  is the number of parameters estimated in the model

Another statistic based on Bayesian approach termed as Bayesian information criterion (BIC) has also been generated based on the finally fitted multiple LRM. The mathematical formulation of BIC statistic (Schwarz, 1978) is as follows.

$$BIC = -2LL + k \ln(N) \quad (3.21)$$

where  $LL$  is the maximum log-likelihood of the fitted model

$k$  is the number of parameters estimated in the model

$N$  is the sample size used in developing the model

Lesser the both AIC and BIC, better will be the model performance.

### **3.5.10 Classification and Discrimination of the Model**

After fitting the multiple LRM, whether the model is correctly classified or not can be assessed using sensitivity, specificity and accuracy through classification table. Further, the discrimination of the events and non-events in the model needs to be examined. It can be performed using the analysis of Area under Receiver Operating Characteristics (ROC) curve. In the following sections, firstly, the description of classification table and secondly the technique of discrimination analysis will be discussed as follows.

#### **3.5.10.1 Description of Classification Table**

An intuitive way of summarizing the results of fitted LRM is through the classification table. This table yields sensitivity, specificity and the overall rate of correct classification (accuracy) of the regression model. The classification table consists of cross-classification of the outcome variable ( $y$ ) with a dichotomous variable. The values for this table are generated through the estimated logistic probabilities. To generate the dichotomous variable, one has to define the cut point  $c$  (say) and compare each estimated probability with  $c$ . If estimated probability is greater than  $c$ , the generated dichotomous variable would be denoted by 1 and otherwise by 0. In practice, the most commonly used value of cut point  $c$  is 50% i.e. 0.5. If one is interested to identify the optimal cut point for the purpose of classification, a cut point might be chosen which maximizes both sensitivity and specificity. Such cut point can be achieved through the graph of probability cutoff in X-axis and sensitivity/specificity in the Y-axis. The crossing point of sensitivity and specificity curve yields the cutoff for this, and has been implemented accordingly in this thesis work. The classification table shows a comparison of the number of successes ( $y = 1$ ) predicted by the LRM compared to the number actually observed and similarly the number of failures ( $y = 0$ ) predicted by the LRM compared to the number actually observed. In this data set, the success represents the poor household and failures represent the non-poor household. We have four possible outcomes. The four cell values of classification are explained in Table 8.



**Table 8:** Theoretical Classification Table Based on the Multiple LRM

	Predicted “0” (Predicted Negative)	Predicted “1” (Predicted Positive)	Total
Observed “0” (Observed Negative)	<b>A</b> (TN)	<b>B</b> (FP)	ON
Observed “1” (Observed Positive)	<b>C</b> (FN)	<b>D</b> (TP)	OP
Total	PN	PP	Tot

- i. TN= Number of cases predicted as “0” and were observed “0” (True Negative)
  - ii. FP = Number of cases predicted as “1” but were observed “0” (False Positive)
  - iii. FN = Number of cases predicted as “0” but were observed “1” (False Negative)
  - iv. TP = Number of cases predicted as “1” and were observed “1” (True Positive)
- where,

$$PP = \text{predicted positive} = TP + FP,$$

$$PN = \text{predicted negative} = FN + TN,$$

$$OP = \text{observed positive} = TP + FN,$$

$$ON = \text{observed negative} = FP + TN, \text{ and}$$

$$\text{Tot} = \text{the total sample size} = TP + FP + FN + TN$$

### 3.5.10.1.1 Sensitivity, Specificity and Accuracy

Sensitivity represents the proportion of all actual positive cases correctly predicted as positive. It is computed by  $TP / (TP+FP)$ . It is also known as the true positive rate (TPR). In this research work, sensitivity explains the proportion of all poor cases correctly predicted as poor.

Specificity is the proportion of all actual negative cases correctly predicted as negative. It is computed by  $TN / (TN+FP)$ . It is also known as the true negative rate (TNR). In the context of this research work, specificity is the proportion of actual non-poor cases predicted as non - poor.

Accuracy is proportion of all cases correctly predicted as Negative and Positive cases computed by  $(TN + TP) / (TN+FP+FN+TP)$ . It is also known as correct classification rate. The accuracy in this data set represents the proportion of cases correctly predicted as poor and non-poor cases.

Note that  $1 - \text{Specificity} = \text{FPR}$ ,  $1 - \text{Sensitivity} = \text{FNR}$ ,  $1 - \text{Accuracy} = \text{Error rate or misclassification rate}$

### **3.5.10.1.2 Area under Receiver Operating Characteristic (ROC) Curve**

In classification table, sensitivity, specificity and correct classification rate were computed based on single cut point. In order to have better description of classification accuracy can be enhanced through area under the Receiver Operating Characteristic (ROC) curve. This curve is plotted keeping false signal i.e. (1-specificity) in X-axis and true signal i.e. sensitivity in Y-axis for an entire range of possible cut points. The area under the curve ranges from 0 to 1 which measures the ability of the fitted LRM to discriminate between those subjects who experience the outcome of interest and those subjects who do not experience the outcome of interest. In this thesis work, area under the ROC curve will be able to measure the ability of the fitted LRM to discriminate poor and non-poor households. The ROC curve has been plotted based on the final multiple LRM of this poverty data.

A useful statistic which can be computed from ROC curve is the area under the curve (AUC), whose value practically ranges from 0.5 to 1.0. Higher the value of AUC, there is better discrimination of the fitted regression model. As a general rule (Hosmer & Lemeshow, 2000), the range of ROC is considered as follows:

- i. If  $\text{ROC} = 0.5$ , there is no discrimination
- ii. If  $0.70 \leq \text{ROC} < 0.8$ , acceptable discrimination
- iii. If  $0.8 \leq \text{ROC} < 0.9$ , excellent discrimination
- iv. If  $\text{ROC} \geq 0.9$ , outstanding discrimination

### **3.5.11 Model Specification Test**

It is logical to check whether the finally fitted LRM may need more independent variables or not. For the assurance of this, the model specification test can be done. It can be done by regressing the original outcome variable on predicted value of the model and the square of the predicted value as independent variables. The null hypothesis for this is that there is no specification error. If the regression coefficient for square of the predicted value is not significant (i.e. p-value  $> 0.05$ ) at 5% level of significance, then we fail to reject the null hypothesis. This indicates that the fitted

multiple LRM is correctly specified in this respect. This test has been attempted in this research work to assess the model specification in this respect.

### 3.5.12 Diagnostics of the Logistic Regression Model (LRM)

With the fundamental concept of regression model, the residual sum of squares is one of the major components. One of the major assumptions of linear regression is that the error variance does not depend on the conditional mean  $E(Y_j | x_j)$ . Contrary to this, in LRM, the error terms follow binomial distribution instead of normal distribution, and consequently the error variance is a function of the conditional mean i.e.

$$Var(Y_j | x_j) = m_j E(Y_j | x_j) [1 - E(Y_j | x_j)] = m_j \pi(x_j) [1 - \pi(x_j)] \quad (3.22)$$

The Pearson residual  $r(y_j, \hat{\pi}_j)$  or shortly  $r_j$  is expressed as:

$$r_j = \frac{(y_j - m_j \hat{\pi}_j)}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}} \quad (3.23)$$

where  $\pi_j$  is the estimated probability for  $j$  covariate pattern.

Let us also define the deviance residual  $d_j$  as follows.

$$d_j = \pm \left\{ 2 \left[ y_j \ln \left( \frac{y_j}{m_j \hat{\pi}_j} \right) + (m_j - y_j) \ln \left( \frac{m_j - y_j}{m_j (1 - \hat{\pi}_j)} \right) \right] \right\}^{1/2} \quad (3.24)$$

where the sign (+) and (-) is the same as the sign of  $(y_j - m_j \hat{\pi}_j)$

In both the expressions (3.23) and (3.24), the denominator is the approximate estimate of standard errors of residuals. Under such scenario, if the LRM is correct, each quantity defined in (3.23) and (3.24) has mean approximately equal to zero and variance approximately equal to one. However Pearson residuals do not have variance equal to 1 unless they are further standardized (Pregibon, 1981). This is true only when  $m_j$  is sufficiently large to justify that the normal distribution gives adequate approximation to the binomial distribution, a condition obtained under m-asymptotic (Hosmesr & Lemeshow, 2000).

The standardized Pearson residual denoted by  $r_{sj}$  for covariate pattern  $x_j$  can be defined as follows.

$$r_{sj} = \frac{r_j}{\sqrt{1-h_j}} \quad (3.25)$$

where  $h_j$  is the  $j^{\text{th}}$  diagonal element of the hat matrix  $H$  derived by Pregibon (1981) as a linear approximation to the fitted values for LRM. Now,  $h_j$  may be defined as follows.

$$h_j = m_j \hat{\pi}(x_j) [1 - \hat{\pi}(x_j)] X_j' (X'VX)^{-1} X_j = V_j \times b_j \quad (3.26)$$

$$\text{where } b_j = X_j' (X'VX)^{-1} X_j \quad (3.27)$$

And  $X_j' = (1, x_{1j}, x_{2j}, \dots, x_{pj})$  is the vector of covariate values in the  $j^{\text{th}}$  covariate pattern.

Pregibon (1981) also suggested another influence statistic  $\Delta \hat{\beta}_j$  based on covariance matrix  $\hat{\beta}$  which is defined as follows.

$$\Delta \hat{\beta}_j = \frac{r_{sj}^2 h_j}{1 - h_j} \quad (3.28)$$

This statistic is also called delta beta statistic, and also mathematically denoted shortly by  $(\Delta \hat{\beta})$  without using the suffix  $j$ , where  $\Delta$  stands for the difference. Delta beta measures the changes in the regression coefficients for every covariate pattern. This is measured when we were to eliminate that pattern. In order to identify the individuals with relatively large influence on the estimated regression coefficients, scatter plot can be used. The plot is made considering  $(\Delta \hat{\beta})$  in the y-axis and the estimated probability based on the finally fitted LRM in the x-axis. The pattern of the influence of estimated regression coefficient can be assessed through the scatter plot.

On the use of same approximation, it can be shown that the decrease in the value of Pearson Chi-square statistic due to deletion of the subjects with covariate pattern  $x_j$  is given by:

$$\Delta\chi_j^2 = \frac{r_j^2}{1-h_j} = r_{sj}^2 \quad (3.29)$$

This statistic is also known as delta chi-square ( $\Delta\chi^2$ ). This generally measures the effects of patterns on the fit of the model. When excluding the patterns, the changes to the overall chi-square statistic can be measured and examined through scatter diagram. It can be assessed through the graph of the estimated probability on the horizontal axis and the delta-chi-square on the vertical axis. It would be helpful to detect the patterns of the scatter plot. Further, the same plot can be made allowing the size of the symbol to depend on the corresponding ( $\Delta\hat{\beta}$ ) value which would be more influential plot.

A similar quantity can also be obtained for the change in the deviance which is given by

$$\Delta D_j = d_j^2 + \frac{r_j^2 h_j}{(1-h_j)} \quad (3.30)$$

If  $r_j^2$  is replaced by  $d_j^2$ , then equation (3.30) becomes

$$\Delta D_j = \frac{d_j^2}{(1-h_j)} \quad (3.31)$$

The statistic of changes in the deviance ( $\Delta D_j$ ) i.e. delta deviance works in the similar direction as done by delta chi-square. In this case also, the pattern of the change in the deviance with the estimated probability based on the LRM has been assessed through the scatter plot keeping change in the deviance in the vertical axis and the estimated probability of the LRM in the horizontal axis. If the points either fall on the top left or on the top right corners of the scatter plot, it indicates that the covariate pattern of these subjects are poorly fit. Actually this is based on the distance from the balance of the data plotted which can be assessed through visual assessment and numerical values too.

Different statistical software performs the diagnostics of LRM differently. In STATA software, all the residuals and the diagnostics statistics are computed based on covariate pattern not on observations. Then it finally retains the size of the original data. Hence all subjects in a particular covariate pattern have the same covariate value and the diagnostic statistics. However, each subject has individual outcome. On the other hand,

performing diagnostics of LRM, it performs based on the data structure (Hosmer & Lemeshow, 2000). Lack of routine methods and lack of uniform techniques in the statistical software for diagnostics of fitted LRM, it is not straight forward to report this diagnostic statistics or that or all possible. In this context, Hosmer and Lemeshow (2000) have indicated that there is no substitution for experience for effective use of diagnostic statistics for LRM. Number of different types of graphical techniques has been suggested for the diagnostics of LRM (Hosmer and Lemeshow, 2000; Pregibon, 1981; Landwehret et al., 1984; Fowlkes, 1987).

Hosmer and Lemeshow (2000) have suggested seven diagnostic statistics grouped into three such as (i)  $(r_j, d_j, h_j)$  (ii) derived measures of the effect of each covariate pattern on the fit of the model,  $(r_{sj}, \Delta\chi_j^2, \Delta D_j)$  and (iii) derived measure of the effect of each covariate pattern on the value of the estimated parameters,  $(\Delta\hat{\beta}_j)$ .

Some of the graphical presentations, which can generally be made based on these diagnostics statistics are listed as follows.

- i. Plot of  $(\Delta\chi_j^2)$  versus estimated logistic probability  $(\hat{\pi}_j)$
- ii. Plot of  $(\Delta D_j)$  versus estimated logistic probability  $(\hat{\pi}_j)$
- iii. Plot of  $(\Delta\hat{\beta}_j)$  versus estimated logistic probability  $(\hat{\pi}_j)$
- iv. Plot of  $(\Delta\chi_j^2)$  versus  $(h_j)$
- v. Plot of  $(\Delta D_j)$  versus  $(h_j)$
- vi. Plot of  $(\Delta\hat{\beta}_j)$  versus  $(h_j)$

Among different graphical presentations mentioned above and others for the diagnostics of the fitted LRM, first (i) to (iii) are considered more meaningful and suggested to report (Hosmer & Lemeshow, 2000).

In this thesis work, the final multiple LRM consists of six independent covariates, each categorical having only two levels. Considering these all theoretical as well as practical aspects in this regard, these first four (listed above) plots have been attempted. The visual assessment of patterns of the distribution of data points will be helpful to assess whether the covariate patterns are poorly fit or not in most of these plots.

### **3.5.13 Assessment of Risk on the Basis of Factors Present in the Model**

There are six factors namely household grouped by the number of literate persons of working age group (at least one versus none), household grouped by the number of children under 15 years (at most 2 versus at least 3), literacy status of household head (male versus female), household grouped by the area of land holding (with land versus without land), remittance receiving status of household (receiver versus not receiver) and access to market center ( $\leq 30$  min versus  $> 30$  min). The distribution of households (with percentage) on the basis of presence of any one factor, any two factors, up to presence of all six factors are presented. Finally, the odds ratios for each situation i.e. presence of any one factor, any two factors, etc. are computed by regressing the same outcome variable (poverty: Poor vs. non-poor) with this newly generated indicator variable ( $x_i$ , for  $i = 0, 1, 2, 3, 4, 5, 6$ ), where 0 stands for none of the factors present, 1 stands for presence of one factor, and so on. Computation of odds ratio (OR) for the presence of number of risk factors would be helpful for policy point of view in the relevant area.

### **3.6 Log-binomial Regression Model (LBRM)**

The most popular modeling approach for examining association between predictors and binary outcomes is LRM. LRM has historically been used to assess binary outcomes that estimates odds ratio. However, recent literature reveals an increasing preference for assessing relative risk ratio (RR) rather than odds ratio (OR). The odds ratio has been characterized as being incomprehensible (Lee, 1994), but the risk ratio (RR) is assumed to be easier for understanding (De Andrade & Carabin, 2011; Schechtman, 2002). The relative risk (RR) often referred to as the prevalence ratio (PR) or risk ratio (RR) is easier to interpret and express, especially to non-epidemiologists, which is alternative techniques (Barros & Hirakata, 2003). Skov et al. (1998) and Wacholder (1986) have suggested to use log-binomial regression to calculate relative risks. The LBRM assumes a log link and the logistic regression model is a logit link (Janani et al., 2015). Espelt et al. (2017) recommended using RR rather than OR for analyzing cross-sectional data sets with binary response variable.

Regarding the question of reporting, there is still academic disagreement over whether “OR” or “RR” is preferable. Some authors prefer OR, whereas others prefer RR. Cook

(2002), Newman (2001), Walter (1998), and Olkin (1998) preferred odds ratio (OR), and Gallis and Turner (2019), De Andrade and Carabin (2011), and Sackett, et al. (1996), preferred risk ratio (RR) as it is simple to understand. Lee (1994) has also indicated that OR is incomprehensible. It has been recommended by Williamson et al. (2013) to apply relative risk in epidemiological research whenever practical and also advocated for RR. If the event of interest under study is rare, or normally measured as less than 10%, the odds ratios and risk ratios are closer (Viera, 2008; Greenland et al., 1986; Greenland & Thomas, 1982). Odds ratio cannot approximate risk ratio if the event of interest under study is common i.e.  $\geq 10\%$  (Gallis & Turner, 2019; Ranganathan et al., 2015; Viera, 2008; Katz, 2006; McNutt et al., 2003; Greenland, 1987). In the study data set (NLSS III), 18.5% of households still live in poverty. This poverty rate is greater than 10%, and hence the computation of RR is also justifiable in the analysis for poverty data of Nepal.

Let us consider a contingency table having exposure status (independent variable) in association with the event status (occurrence of event and non-occurrence of event) in Table (9). Each group whether exposed or not is represented in rows and the outcome status in the column.

**Table 9:** Layout of Computation of RR and OR

Exposure Status	Events	Non-Events	Total	Probability of occurrence of risk
Exposed	$n_{11}$	$n_{12}$	$T_{10}$	$p_1 = n_{11} / T_{10}$
Not exposed (reference group)	$n_{21}$	$n_{22}$	$T_{20}$	$p_2 = n_{21} / T_{20}$
Total	$T_{01}$	$T_{02}$	n	

where  $p_1$  is the probability of occurrence of event in the exposed group, and  $p_2$  is the probability of occurrence of event in unexposed group. The risk ratio (RR) and the odds ratio (OR), are computed as follows.

$$RR = p_1 / p_2$$

The OR is the ratio of two odds, and it can be written as

$$OR = p_1 / (1 - p_1) / p_2 / (1 - p_2)$$



In order to examine the link between a number of explanatory variables and the binary response variable, Wacholder (1986) initially proposed a simple way of directly evaluating risk ratios (RR). Later, Barros and Hirakata (2003) found that in cross-sectional data with common number of events, the OR frequently overestimates the RR. The risk ratio (RR) is generated via the LBRM, which is based on a generalized linear model with a log link and a binomial probability distribution. Both Robbins et al. (2002) and McNutt et al. (2003) focused on the descriptions and applications of these models. Blizzard and Hosmer (2006) recommended diagnostics for the LBRM and goodness of fit tests. There are accepted techniques for converting OR to RR. However, the confidence intervals generated by these converted techniques were inaccurate (Robbins et al., 2002). Additionally, it has also been shown that a number of implementations of the LBRM do not converge (Williamson et al., 2013).

### 3.6.1 Log-binomial Regression Model (LBRM) and its Fitting

LBRM is a special form of the generalized linear model, and its link function is log link. Let  $X_1, X_2, \dots, X_p$  are the p number of covariates which are associated with response variable, then the LBRM can be written as:

$$\log \pi = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p, \quad (3.32)$$

where  $\pi = \text{Pr ob}[Y = 1 | X] = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)$  for binary outcome Y.

$\beta_1, \beta_2, \dots, \beta_p$  are the regression coefficients for covariate  $X_1, X_2, \dots, X_p$  and  $\beta_0$  is the constant term of the model.

Initially, the simple LBRM has been used with these seven independent covariates taking one variable at a time to identify the candidate variables for the multiple LBRM. With these candidate variables, both stepwise forward and backward selection method as used in LRM, have been performed in LBRM. Then the final multiple LBRM has been estimated with significant covariates.

#### 3.6.1.1 Maximum Likelihood Estimation for Log-binomial Regression Model (LBRM)

Maximum likelihood estimation is the basis of the common estimation process for all generalized linear models. Estimates of the (p+1) number of regression coefficients

$(\beta_0, \beta_1, \dots, \beta_p)$ , including the intercept in the model are to be estimated. The  $(p+1)$  regression coefficients are estimated by using the following likelihood function (Blizzard & Hosmer, 2006).

$$l(\beta) = \sum_{i=1}^n [y_i \log \pi_i + (1-y_i) \log(1-\pi_i)] \quad (3.33)$$

where,  $\pi_i = e^{\left(\sum_{j=0}^p \beta_j x_{ij}\right)}$

### 3.6.2 Interpretation of RR

The value of RR is interpreted as follows.

- i. If  $RR = 1$ , then there is no risk between exposed group and not exposed group (reference group).
- ii. If  $RR > 1$ , then the risk is more in exposed group as compared to the not exposed group (reference group).
- iii. If  $RR < 1$ , then the risk is less in exposed group as compared to the not exposed group (reference group).

### 3.6.3 Goodness-of-Fit Test of the Model

The approach based on deciles of risk test is also used to assess the goodness of fit of LBRM as used in LRM. It is to compare the observed outcome frequencies to model-based estimates of those frequencies within groups based on the rank-ordered fitted values. Blizzard and Hosmer (2006) suggested goodness-of-fit test of the LBRM is

$$\hat{c}_k = \sum_{j=0}^1 \sum_{k=1}^{10} \frac{(o_{jk} - \hat{e}_{jk})^2}{e_{jk}} \quad (3.34)$$

where,  $C_k$  is a set containing the indices of the subjects in the  $k^{\text{th}}$  deciles of risk

$$o_{1k} = \sum_{i \in C_k} y_i, o_{0k} = \sum_{i \in C_k} 1 - y_i,$$

$$\hat{e}_{1k} = \sum_{i \in C_k} \hat{\pi}(x_i) \text{ and } \hat{e}_{0k} = \sum_{i \in C_k} 1 - \hat{\pi}(x_i)$$

Hosmer and Lemeshow (1980) showed in the LBRM that the number of groups minus two is used to compute the degrees of freedom for this test.

### 3.6.3.1 AIC and BIC for the Model

In order to measure the performance of the fitted models, the value of AIC and BIC for each model are used. The Akaike Information Criterion (AIC) and the Schwarz Bayesian Information Criterion (BIC) as measures of good fit to determine which model is the most appropriate and best fit. The model with the lowest AIC or BIC, better will be the model performance. Akaike's Information Criterion (AIC) is defined as

$$AIC = -2LL + 2K \quad (3.35)$$

The formula for the Bayesian Information Criterion (BIC) is

$$BIC = -2LL + \log(N)K \quad (3.36)$$

where  $K$  = the number of parameters in the model (including the intercept),  $N$  = sample size, and  $LL$  is the log likelihood estimate (Christensen & Angeles, 2018).

### 3.6.4 Diagnostics of the Log- binomial Model (LBRM)

For the fitted LBRM, two scatter plots are mostly used among the various diagnostic techniques described in the statistics literature. They are: (i) Scatter plot of leverage values keeping in Y-axis and the predicted values keeping in X-axis. (ii) Scatter plot of change in  $\chi^2$  ( $\Delta\chi^2$ ) in Y- axis and model predicted values with plotting symbol proportional to Cook's distance in X- axis.

According to McCullagh and Nelder (1989), the leverage values are the diagonal

elements of the hat matrix,  $H$ , from the regression,  $z_i = x_i'\hat{\beta} + \frac{y - \hat{\pi}(x_i)}{\hat{\pi}(x_i)}$  is

$$H = \hat{W}X(X'\hat{W}X)^{-1}X' \quad (3.37)$$

where  $\hat{W} = \text{diag}(\hat{w}_i)$  and  $X$  is the  $n(p+1)$  data matrix. The individual leverage values are:

$$h_i = \hat{w}_i x_i' (X'\hat{W}X)^{-1} x_i$$

The formula of the Cook's distance measure of influence is

$$\Delta\hat{\beta}_i = \Delta\chi_i^2 \frac{h_i}{(1-h_i)} \quad (3.38)$$

where,  $\Delta\chi_i^2 = \hat{r}_{is}^2$  and  $\hat{r}_{is} = \frac{y_i - \hat{\pi}(x_i)}{\sqrt{\hat{\pi}(x_i)(1-\hat{\pi}(x_i))(1-h_i)}}$

The above explained two scatter plots are considered more meaningful and recommended graphical presentations for the diagnostics of the fitted LBRM (Blizzard & Hosmer, 2006). In this research work, these two graphical methods are performed for the diagnostic assessment of LBRM.

### 3.6.5 Assessment of Risk for Different Factors

When the fitting of final multiple LBRM is completed, the risk assessment of household poverty has been done. It has been done taking into consideration of number of risk factors present. This has been performed by running the LBRM again with same outcome variable (poverty status: poor vs. non-poor) and newly generated independent variable ( $x_i, i = 1, 2, \dots, p$ ). This independent variable  $x_i$  represents the presence of number of risk factors. The risk ratios are assessed based on the presence of number of risk factors in the final model as did in LRM.

## 3.7 Comparison of the Models

LRM and the LBRM has been compared based on different criteria. The criteria are variable selection, effect size, precision of effect size for each covariate (using confidence interval), fit of the model. Additionally, model's diagnostics, model's stability (using bootstrapping procedure), risk assessment, and model convergence issue are also compared between the two models (Acharya et al., 2022c).

### 3.7.1 Comparison of Models Based on Variable Selection

Based on the extensive review of literature, same set of independent variables are considered for both LRM and LBRM to check their association with response variable. In order to select the candidate variables for multiple LRM, Chi-square test has been used as a bivariate analysis. Simple LBRM has been used as a bivariate analysis to select the candidate variables for multiple LBRM. Stepwise forward and backward

selection procedure each is adopted for the final selection of variables in multiple LRM and in multiple LBRM. The models are compared with respect to which variables are selected in each model building process.

### 3.7.2 Comparison of Models Based on Individual Regression Coefficient

After fitting both the models, effect size of each covariate of each model has been compared to check whether they are statistically significant at 5% level of significance or not. Magnitude of effect size of those significant variables in terms of RR (obtained from LBRM) and OR (obtained from LRM) are compared. Precision of the effect size of each independent covariate of both models are compared on the basis of the confidence interval estimate (CIE). In addition, the comparisons of the covariate has also been done by the quantity of risk elevation which is computed considering the difference between the OR and RR. The formula for the computation of risk elevation is as follows (Acharya et al., 2022c).

$$\text{Elevation risk (\%)} = [(OR-1) - (RR-1)] \times 100$$

### 3.7.3 Comparison of Models Based on the Goodness of Fit and Diagnostic Criteria

The Hosmer and Lemeshow (H-L) goodness of fit test and AIC& BIC have been used to assess the goodness of fit test for both LRM & LBRM. The results of H-L test and AIC and BIC values used are also compared.

In order to assess the model diagnostics, the following graphs are made for multiple LRM:

- (i)  $\Delta\beta$  and predicted probability
- (ii)  $\Delta\chi^2$  and predicted probability with symbol size proportional to  $\Delta\beta$

The following graphs are generated for assessing the diagnostics of multiple BLRM:

- (i) Leverage in Y-axis and the model predicted probability in X-axis
- (ii)  $\Delta\chi^2$  in Y-axis and values of fitted LBRM with plotting symbol proportional to Cook's distance in X-axis.

The above mentioned scatter plots of model diagnostics are compared to assess the patterns of influential points.

### 3.7.4 Comparison of Models Based on the Robustness Criteria

As with the Cox regression model, the stability of the proposed model has been examined using the bootstrapping technique (Chen & George, 1985; Altman & Anderson, 1989; Saurbrei & Schumacher, 1992). The model is run taking ‘M’ number of bootstrap replications with replacement for each model. If the repetition of a variable is at least 50% in bootstrap replication at selection level of 5%, then that variable can be considered as a strong variable to explain the outcome variable (Khanal et al., 2019). Same approach has used to assesses the robustness of the estimated each LRM and LBRM. The replication frequencies of each independent variable for each bootstrap replication are presented in bootstrap replication matrix as shown in Table10.

**Table 10:** Bootstrap Replication Matrix

Bootstrap replication	Covariates included in the model				
	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	.....	X <sub>p</sub>
1	*	*	*	.....	*
2	*	*	*	.....	*
3	*	*	*	.....	*
.	.	.	.	.....	.
.	.	.	.	.....	.
.	.	.	.	.....	.
.	.	.	.	.....	.
M	*	*	*	*	*
Total					

\*: A covariate included in the model through bootstrap replication.

X<sub>i</sub>, i=1,2,.....,p: Covariates include in the model

### 3.8 Software Used for Statistical Analysis

All the statistical analysis has been carried out using STATA version 13.0 Stata Crop LP, College Station, Texas, USA, and IBM SPSS Version 20 except for bootstrapping procedure. Bootstrapping replication procedure has been performed by using R software.

## CHAPTER 4

### 4. RESULTS AND DISCUSSION

This chapter presents the results of the main data analysis of this study with discussions. The first few sections are devoted to the analysis of socio-economic and demographic characteristics across the consumption quintile groups (hereafter refer to as quintile group), defined in Chapter III, with a view to discuss some disadvantages of the poor more vividly. Then the results of bivariate analysis and multiple LRM and LBRM along with their goodness of fit, diagnostics, stability of the model are discussed in detail. In order to select a good performer model among these two models, the comparisons are made with respect to different comparison parameters such as effect size, precision of the effect size, etc. are discussed in detail.

#### 4.1 Analysis of Economic Characteristics

This section first deals with the analysis of the *per capita income and consumption expenditure (hereafter refer to as expenditure)* across the quintile group, second the *share income and expenditure* of each quintile group, and finally compares the *share of food and non-food expenditure* of each quintile group.

##### 4.1.1 Per Capita Income and Expenditure

The nominal per capita income and expenditure in rupees of each quintile group is displayed in Table 11. Comparison between the two economic indicators within each quintile group reveals a surprising result - *per capita expenditure is greater than per capita income within each of the bottom three quintile groups!* One reason of such surprising result, as pointed out by Brewer and O’Dea in 2012, is due to a substantial number of households of lower quintile groups had fulfilled their basic needs by borrowing or using their own savings during the reference period of the survey.

**Table 11:** Nominal Per Capita Annual Consumption expenditure and Annual Income by Quintile

	Poorest	Second	Third	Fourth	Richest	Overall
Income in NRs	8,498	16,294	25,329	41,138	117,063	41,659
Consumption Expenditure in NRs	13,168	19,317	26,253	36,962	78,504	34,829

Source: CBS (2011), Table 10.2 and Table 11.5

Obviously, the low level of income and consequently low level of expenditure of the poorest group relative to the richest group is a major disadvantage of the poorest group. A crude measure of inequality is the ratio of the per capita income/expenditure of the 20% richest group to that of the 20% poorest group of individuals. This ratio in terms of per capita income is almost 14 implying that per capita income of the richest group is 14 times higher than that of the poorest group. Similarly, the ratio in per capita expenditure is around 6.

#### 4.1.2 Percentage Share of Income and Expenditure

The percentage shares of income and expenditure of each quintile are presented in Table 12. In terms of the percentage share of total income, the share of lowest two quintiles or 40% poor individuals is around 12% while that of highest two quintile or 40% of rich individuals share around 76%. Likewise, in terms of the percentage share of total expenditure, the shares of corresponding turn out to be around 19% and 66% respectively.

**Table 12:** Percentage Share of Each Quintile Group Annual Income and Annual Consumption expenditure

	Poorest	Second	Third	Fourth	Richest	Total
% Share of income	4.1	7.8	12.2	19.7	56.2	100.0
% Share of consumption expenditure	7.6	11.1	15.0	21.3	45.0	100.0

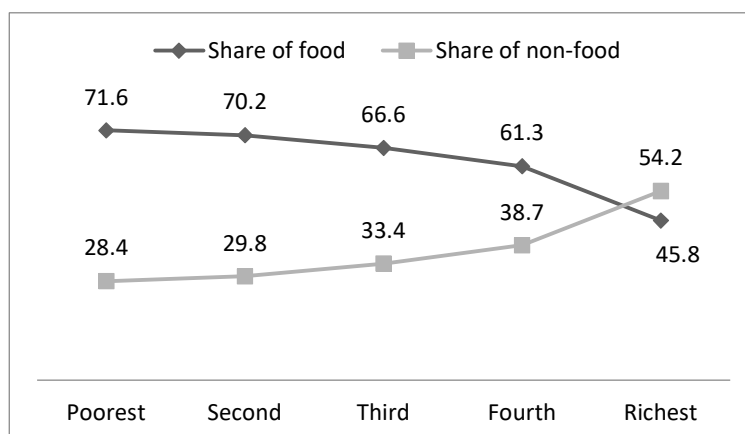
Source: CBS (2011), Table 10.2 and Table 11.5

A more rigorous measure of inequality is Gini-coefficient. The value of Gini-coefficient in income turned out to be 0.464.

#### 4.1.3 Share of Food and Non-food Expenditure

The comparison of the share of food and non-food expenditure to total expenditure provides more serious disadvantage of the poorest group as compared to the richest group. The available percentage share of food and non-food expenditure in NLSS III report (CBS, 2011, Table 10.4) is displayed in Figure 2.





**Figure 2:** Comparison of Percentage Share of Food and Non-food Expenditure

The figure clearly shows that the share of food expenditure continually decreases from around 72% of the poorest group to around 46% of the richest group. One disadvantage of the poorest group relative to the richest group is very high expenditure on food and consequently very low expenditure on non-food which among others includes expenditure on education and health of children. Inability to invest as much as it should be on education and health of children is another disadvantage of the poorest group relative to the richest group.

#### 4.1.4 Analysis of Socio-demographic Characteristics

It is important to note that a group of household comprises of three mutually exclusive groups of population – children, working-age population (WAP) and elders – and the major needs of these three groups of population are to be addressed by policy makers or to be taken care by development workers are entirely different. In view of this argument the average number of children, WAP and elders within each quintile is investigated in order to get more insight (Table13).

**Table 13:** Comparison of Mean Number of Three Population Groups across CQGs

Population Group	Poorest	Second	Third	Fourth	Richest
Children (under 15 years)	2.90	2.19	1.71	1.34	0.87
Working-age population (15 to 64 years)	2.92	2.96	2.92	2.78	2.62
Elders (65 or over years)	0.25	0.25	0.26	0.25	0.21
Child Dependency Ratio (%) = 100*children/WAP	99.32	73.99	58.56	48.20	33.21

Source: Computed from NLSS III data

Table 13 discloses an important fact that large number of WAP in a household does not necessarily improve poverty level of the household, since average number of WAP/household of the poorest group is 2.92 and that of the richest group is 2.62. The earning capability of WAP matters more than the number of WAP and the earning capability to a large extent depends upon their education. With this premise Table 14 displays the literacy rate of WAP (proxy measure of the human capital) within each quintile group by sex.

**Table 14:** Comparison of Literacy Rate of WAP by Gender across CQGs

	Poorest	Second	Third	Fourth	Richest	Ratio of richest to poorest
Male	61.3	70.6	76.6	85.4	90.6	1.5
Female	31.5	41.6	50.9	59.1	72.0	2.3
Overall across CQGs	44.1	54.5	62.1	70.7	80.4	1.8

Source: Computed from NLSS III data

The male and female literacy rates of WAP are in increasing trend from the poorest to the richest quintile group. Another disadvantage of the poorest group compared to the richest group is the low literacy rate among male and female WAP. The ratios of the literacy rate of the richest group to the literacy rate of the poorest group have been computed and presented in Table 14 for male, female and overall. The ratio for female is 2.3 implying that the female literacy rate of the richest group is 2.3 times higher than that of the poorest group. The low female literacy rate among WAP, particularly of the poorest group, is a serious concern since due to outmigration females' dominance over the WAP is high within each quintile group (Table 14).

## 4.2 Poverty Indices

The results of the poverty profile among individuals by socio-economic and demographic characteristics in the study area is presented in Table 15.

**Table 15: Three Measures of Poverty for Fourteen Groups of Household Population**

Variables	Head Count Index (P(0))×100	Poverty Gap Index P(1)×100	Square Poverty Gap Index P(2)×100
Sex of the household head:			
Male	25.6	5.5	1.8
Female	23.7	5.1	1.7
Literacy status of the household head:			
Literate	16.7	3.2	0.9
Illiterate	24.1	5.6	2.0
Status of remittance recipient:			
Yes	13.3	2.5	0.7
No	28.4	6.4	2.2
Status of land ownership:			
With land	21.4	4.5	1.4
Without land	32.9	7.5	2.6
Access to nearest market center:			
Having better access	16.3	3.3	0.9
Having poor access	32.1	7.1	2.5
Number of children under 15:			
At most 2	13.5	2.4	0.7
At least 3	41.4	9.6	3.3
Number of literate members of WAP:			
At least one	21.5	4.3	1.4
None	41.7	10.4	3.9
National level of estimates	25.2	5.4	1.8

Source: Computed from NLSS III

Three measures of poverty for fourteen groups of household population in Nepal are shown in Table 15. The poverty head count index is the highest (41.7%) in the none of the literate member of WAP in a house followed by the number of children under 15 more than two in a household (41.4%), without land (32.9%), having poor access to the nearest market center (32.1%), status of the household not receiving remittance (28.4%), Illiterate household head (24.1 %) and female headed households (23.7 %). The overall result indicates that all kinds of poverty indices is higher in disadvantaged group than in advantaged group. Except female headed households and illiterate household heads, all kinds of poverty indices is higher in disadvantaged group than in national level while all kinds of poverty indices is lower in advantaged group than in national level.

Comparison of two economic variables within each quintile group has shown that per capita expenditure is higher than income in each of the lowest three quintile categories. One main disadvantage of the lowest group is its low level of income. As a result, this

group has low level of expenditure compared to the richest group. In fact, the poorest group spends a lot more percentage of their income on food than the richest group. Therefore, the poorest group has a lot less money on non-food items like children's health and education. It makes more sense to consider the number of children per household to analyze poverty in a country like Nepal than using household size and the child dependence ratio. All types of poverty indicators(head count index, poverty gap, square poverty) are greater among disadvantaged groups than at the national level with the exception of female-headed households and literate household head.

### 4.3 Association of Covariates with Response Variable

In order to investigate the association between each of the seven covariates and response variable, chi-square test is carried out. The results of Chi-square test with its effect size, measured by Phi-coefficient, as well as descriptive statistics are presented in Table 16.

**Table 16:** Results of Bivariate Analysis (n = 5988)

Description of covariates	Percentage distribution of households	Association of covariates with poverty			Phi-coefficient
		% of poor households within category	Chi-square value	p - value	
Sex of household head:					
Male	73.3	18.9	1.7	0.193	-0.02
Female	26.7	17.4			
Literacy status of household head:					
Literate	60.2	12.2	240.7	<0.001	0.20
Illiterate	39.8	28.1			
Status of remittance recipient:					
Yes	53.1	15.7	35.7	<0.001	0.08
No	46.9	21.7			
Status of land ownership:					
Yes	71.2	15.1	114.9	<0.001	0.14
No	28.8	27.0			
Access to nearest market center:					
Better	52.0	11.6	206.7	<0.001	0.19
Poor	48.0	26.0			
Number of children under 15:					
At most two	73.8	10.9	653.0	<0.001	0.33
More than two	26.2	40.1			
Number of literate members of WAP:					
At least one	80.7	15.6	142.0	<0.001	0.15
None	19.3	30.8			

Source: Acharya et al. (2022a)

Seven covariates, except sex of household head, are found statistically significant with response variable ( $p < 0.05$ ). Interestingly, sex of the household head has not become significant. Similar result is observed by Spaho (2014). One possible reason for being the sex of household head insignificant might be that majority of young Nepalese males have migrated to foreign countries after 1990 (Umatsu et. al., 2016) which subsequently lead females to be the household head. As a result, the male headed household is more likely to be poorer compared to female headed households. Similar results was found by Edoumiekumo et. al. (2014). The minimum effect size of Chi-square test (0.08) is found for the covariate “remittance receiving status of household” and the maximum effect size (0.33) is noted for “number of children under 15 years”. One of the possible reasons for having minimum effect size for remittance might be the reduction of poverty in Nepal is not only caused by the heavy inflows of remittance but also by increment in the labor wages and production of growth-led agricultural products. One reason for having maximum effect size for children might be the population of children after 2011 seems to be decreasing but the majority of children born before 2011 were dependent and still needed some time for them to mature and become self-dependent.

#### **4.4 Results of Logistic Regression (LRM)**

From the bivariate analysis six covariates have been found significant, and they are along with sex of household head considered as candidates for multiple LRM. The stepwise selection procedure has been adopted for finalizing the model. The final multiple LRM retains with six variables except sex of household head. The results of estimated multiple LRM including the value of beta, odds ratio (OR), standard error (SE), p-value and 95% confidence interval estimate (CIE) of OR are shown in Table 17. The estimated model is significant ( $p < 0.001$ ) at 5% level of significance. The model significance is examined by using the Omnibus test. Fit of the model is well as shown by H-L test ( $p = 0.5340$ ). The value of McFadden pseudo  $R^2$  is 0.16.

**Table 17: Regression Estimates of LRM with 95% CIE (n = 5988)**

Independent factors	<i>b</i>	OR	S.E.	P -value	95% CIE for OR
Literacy status of household head:					
Literate		1.00			
Illiterate	0.79	2.20	0.09	<0.001	(1.86, 2.61)
Status of remittance recipient:					
Yes		1.00			
No	0.64	1.90	0.08	<0.001	(1.64, 2.20)
Status of land ownership:					
Yes		1.00			
No	0.43	1.53	0.08	<0.001	(1.31, 1.78)
Access to nearest market center:					
Better		1.00			
Poor	0.57	1.77	0.08	<0.001	(1.52, 2.07)
Number of children under 15:					
At most two		1.00			
More than two	1.55	4.69	0.07	<0.001	(4.06, 5.42)
Number of literate members of WAP:					
At least one		1.00			
None	0.25	1.29	0.10	<0.001	(1.07, 1.56)
Constant	-3.27	0.04	0.09	<0.001	
Log likelihood(null model ) = -2869.408, Log likelihood(full model): -2399.922 LR Chi-square with 6 d.f. = 938.97 (p < 0.001), Pseudo R <sup>2</sup> = 0.1636 AIC = 4813.844, BIC = 4860.727 Goodness of Fit of the Model: Hosmer and Lemeshow goodness of fit test : Chi-square with 8 d. f. = 6.05 , p = 0.5340					

Source: Results adopted from Acharya et al. (2022a)

Each regression coefficient has a positive sign, which means that each of the disadvantaged groups is more likely to be poorer than its counter parts. This is explained as follows.

In Nepal, household head is generally needs to be accountable for managing all of the household's resources. Illiterate household heads are normally more likely to work in low-paying jobs, have weaker negotiating positions, and engage in fewer other forms of economic activity. As a result, the household income is lower and the poverty level of the household will rise. Keeping the effects of all other factors constant, it is found that household which are led by illiterate member are 2.2 times more risk (OR: 2.20;

95% CIE: 1.86 - 2.61) of having poor compared to its counterpart. Results of Teka et al. (2019), Imam et al. (2018), and Botha (2010) corroborate the finding.

If all other factors are constant, households who do not receive remittance are 1.9 times more likely to be poorer than those households who receive remittance (OR: 1.90; 95% CIE: 1.64 - 2.20). Similar results are obtained from the study conducted in Pakistan. According to Majeed and Malik (2015), households receiving remittances has a 43% lower probability of being poor (OR = 0.57) than those not receiving remittances. The present study findings have agreed with those of Abrar ul Haq et al. (2018) and R. E. A. Khan et al (2015). The findings of the Chi-square test used in this study to assess the association between remittance and the remaining covariates are shown in Table 18. Interestingly, all five of the disadvantaged groups-aside from the group of households with more than two children have a much greater percentage of households receiving remittance than their counterparts. However, the odds ratio for the chance of disadvantaged households groups being poor remains higher than 1 compared to that of their advantaged groups.

**Table 18:** Role of Remittance in Association with Independent Variables

		% of households receiving remittance	Chi-square value	p-value
Literacy status of household head	Literate	49.0	59.8	< 0.001
	Illiterate	59.2		
Status of land ownership	Yes	49.8	62.8	< 0.001
	No	61.1		
Access to nearest market center	Better	49.3	37.5	< 0.001
	Poor	57.2		
Number of children under 15	At most 2	53.2	0.1	0.739
	More than 2	52.7		
Number of literate members of WAP	At least one	51.2	34.5	< 0.001
	None	60.8		

Source: Acharya et al. (2022a)

The availability and accessibility of various resources, such as work possibilities, land's availability, and loan's accessibility, are crucial for escaping rural poverty in Nepal. The social standing of a person is closely correlated with possessing or not having land. The ability of a household to obtain loans, launch enterprises, and rent land is typically severely limited. As a result, these households ability to engage in economic activity is

hampered, leading to an increase in their degree of poverty. Keeping the effects of all other variables constant, our analysis has clearly shown that the group of families without land is 1.5 times more likely to be poorer than the group of households with land (OR: 1.53; 95% CIE: 1.31 - 1.78).

It is exceedingly challenging for farmers and smallholders to sell their goods and obtain financing in rural areas of Nepal if the market center is far away. This remains a problem if the highways and feeder roads are not improved. Household poverty is significantly impacted by the issue of post harvest food loss caused by a lack of cold storage facilities and poor infrastructure (Shively & Thapa, 2017). The risk of poverty is 1.8 times (OR: 1.77; 95% CIE: 1.52 - 2.07) more in the household with poor market access in comparison to the household with better market access keeping the effect of other variables constant. This finding is comparable to that made public by Mamo and Abiso (2018).

In each household in a country like Nepal, children are dependents. Households with more children need more money for food, clothes, health care, and education. The level of household poverty will consequently rise. In relation to this, the research analysis found that, after controlling the effects of all other factors, the odds of having children poor is 4.7 (OR: 4.69; 95% CIE: 4.06 - 5.42) times higher for a family having > 2 children compared to the family having  $\leq 2$  children. This conclusion is consistent with that of Myftaraj et al. (2014), who found that families with two children had a 20% lower probability of being poor (OR: 0.8). However, after adding one more child (i.e. three children) in a household, the likelihood of poverty (OR: 1.03) increases.

A household having illiterate members of working age will not have good awareness about a better job, market demands, and consequently less possibility of getting the job. The members of such households will also be less familiar with the most recent information and technology. Because of such constraints, their involvements in social and economic activities become very narrowed. In this regard, the findings of this study has indicated that, after controlling the effects of all other factors, the risk of poverty is 1.3 (OR: 1.29; 95% CIE: 1.07 - 1.56) times more in the households having no literate members with reference to those households having at least one literate member. Mamo and Abiso (2018) showed a similar outcome in rural Ethiopian communities (OR: 1.4). According to Omoregbee et al. (2013), in Nigeria, farmers with less education were 1.3



times more likely to live in poverty than those with higher education. Another Pakistani research found that, compared to households with illiterate earners, adding an additional educated earner of any level considerably lowers the likelihood of the household being poor by 11% (OR: 0.89) (Majeed & Malik, 2015).

#### 4.4.1 Assessment of Multicollinearity among Independent variables

In order to assess the multicollinearity among independent variables in Ordinary Least Square (OLS) regression, the Variation Inflation Factor (VIF) is generally used. For such scenario, the correlation matrix of the Pearson's correlation coefficient of independent variables are also reported in practice, and the simple correlation between explanatory variables less than 0.8 or 0.9 is considered as a general rule of thumb for not indicating multicollinearity (Farrar & Galuber, 1967). The application of VIF and the simple correlations, each is generally reported for the continuous explanatory variables. However, the data set of this research work consists of six categorical variables, for which the use of VIF may not be logical. There is informal approach of reporting the correlation matrix of regression coefficients of explanatory variables, which has been attempted in this research work and presented in Table 19.

**Table 19:** Correlation Matrix of Coefficients of LRM

Indicators	Literacy status of household head	Status of remittance recipient	Status of land ownership	Access to nearest market center	Number of children under 15	Number of literate members of WAP
Literacy status of household head	1.00					
Status of remittance recipient	0.09	1.00				
Status of land ownership	-0.02	0.09	1.00			
Access to nearest market center	-0.08	0.05	-0.23	1.00		
Number of children under 15	0.04	0.04	-0.04	-0.07	1.00	
Number of literate members of WAP	-0.51	0.04	0.04	-0.06	-0.01	1.00

Source: Computed from NLSS III

The correlation matrix has been generated using STATA (a statistical software) codes. By observing the correlation matrix of the regression coefficients generated through the fitted LRM, there is not much high correlation. It indicates that there is no serious multicollinearity. It has also attempted to compute the Condition Index (CI) and Condition Number (CN) without considering the regression constant. Majority of values of CI are less than 2 and only one value of CI is less than 3 and the value of CN is 2.02 which is depicted in Table 20.

Table 20: Collinearity Diagnostics

Eigen Value	Condition Index
1.85	1.00
1.15	1.27
1.00	1.36
0.85	1.47
0.69	1.63
0.45	2.02
Condition Number:	2.02

Source: Computed from NLSS III

There is weak dependencies of the explanatory variables if the value of CI is around 5 or 10 (Belsley et al.,1980). Further, when the the value of CN is less than 100, then it is considered as non harmful multicollinearity or weak multicollinearity (Callaghan & Chen, 2008). Based on the results of correlation coefficient of regression coefficients, condition indices and condition number, there is clear indication of not having the problem of multicollinearity of the independent variables in the developed model.

#### 4.4.2 Results of Sensitivity, Specificity and Correct Classification of the Model

The classification of the fitted LRM has been estimated using sensitivity, specificity, and accuracy. The results of classification is presented in Table21.

**Table 21:** Correct Classification Details of the Model

Observed		Predicted			
		Household Poverty Status		Total	Percentage Correct
		Poor	Non-poor		
Household	Poor	822	1680	2502	74.12
Poverty Status	Non-poor	287	3199	3486	65.57
	Total	1109	4879	5988	67.15

Source: Computed from data of NLSS III

The algorithm properly identified 822 households that it predicted as poor but 287 households are really not poor out of the 1109 households that does include a poor person. In a similar manner, the algorithm accurately identified 3199 households out of 3484 that are predicted to be non-poor but 822 households were really poor.

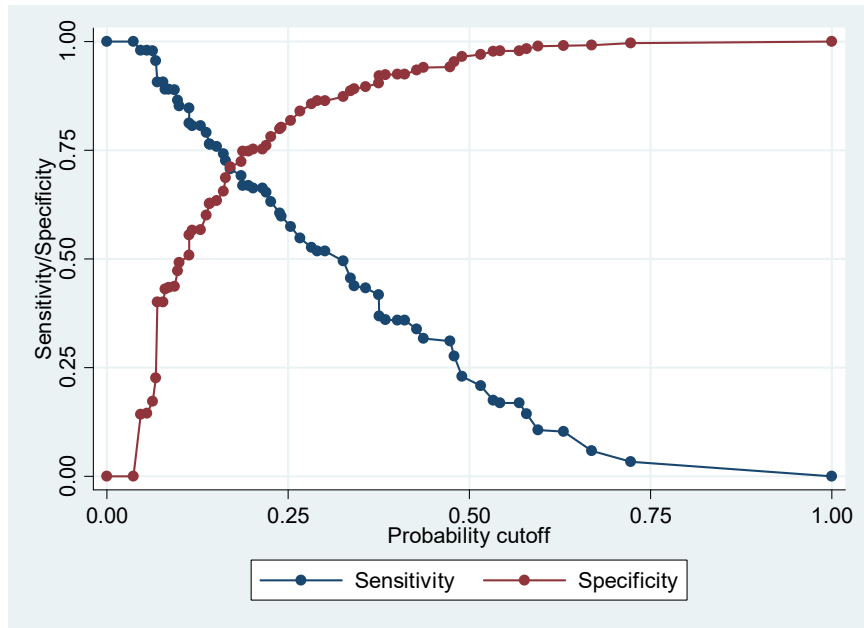
The values of sensitivity and specificity of prediction are 74.12% and 65.57%, respectively. According to the regression classification, a LRM is able to accurately predict 67.15 % of the responses.

The crossing point of sensitivity and specificity curve is shown in the cutoff point (Figure 3). The cutoff has been taken as 0.16 instead of 0.5 based on the curve drawn from this data set. The results related to correct classification is provided in Table 22. The model's correctly predicted for poor and non poor are 74.12% and 65.57% respectively.

**Table 22:** Correct Classification Values

Cutoff	Sensitivity	Specificity	Correct Classification
0.16	74.12%	65.57%	67.15%

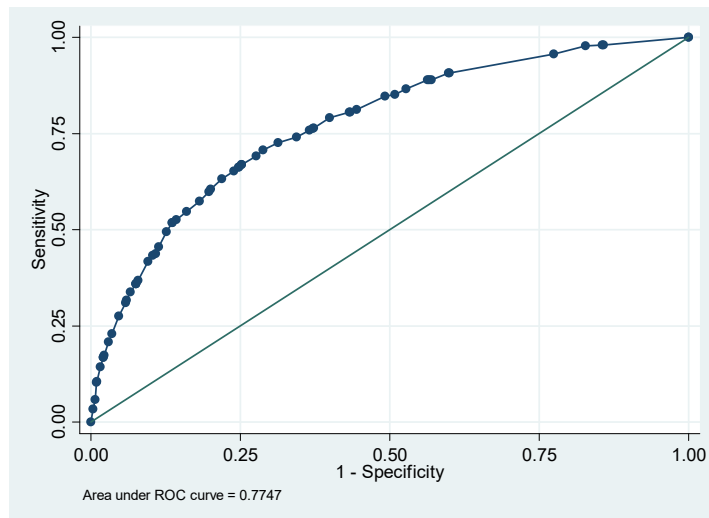
Source: Acharya et al. (2022a)



**Figure 3:** Sensitivity / Specificity and Predicted Probability

#### 4.4.3 ROC Curve for Model Discrimination

The capacity of the model's discrimination has been evaluated with the help of ROC curve as shown in figure 4. The area under the curve (AUC) is 0.78. This value is generally considered as the acceptable discrimination of the model (Hosmer & Lemeshow, 2000).



**Figure 4:** ROC Curve

#### 4.4.4 Results of Model Specification Test

In this test, a new regression model has been made for which the dependent variable is same as used in LRM. The independent variables for this is the model estimated value ( $\hat{Y}$ ) and the square of the estimated value ( $\hat{Y}^2$ ). If the regression coefficient of square of the estimated value is not statistically significant at  $\alpha = 0.05$ , then the fitted model is correctly specified.

**Table 23:** Regression Coefficient for  $\hat{y}$  and  $\hat{y}^2$

	Coefficient	S. E.	Z	p-value	95% CIE for regression coefficient
$\hat{y}$	0.97	0.09	11.26	$\leq 0.001$	(0.80, 1.14)
$\hat{y}^2$	-0.01	0.03	-0.39	0.696	(-0.08, 0.05)
Constant	-0.01	0.06	-0.10	0.923	(-0.12, 0.11)

Source: Acharya et al. (2022a)

The regression coefficient for  $\hat{y}^2$  is not statistically significant ( $p = 0.696$ ) at 5% level of significance as shown in Table 23. This indicates that the model is correctly specified.

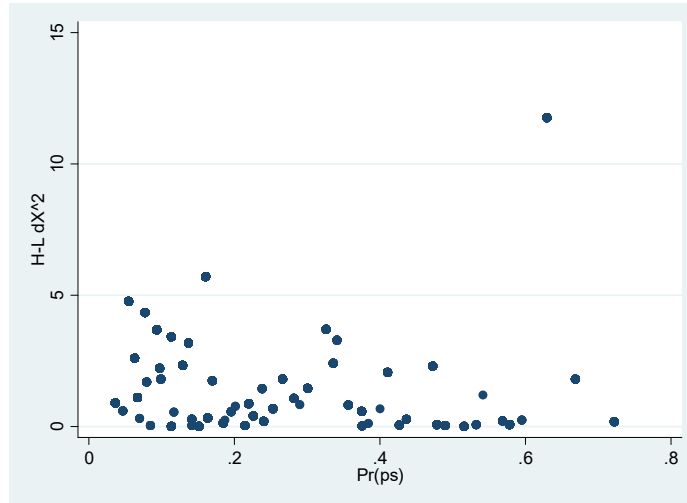
#### 4.4.5 Results of Diagnostics of the Fitted Multiple LRM

The model diagnostics is assessed graphically by the following four plots.

- i. Graph of  $\Delta\chi_j^2$  in Y-axis and estimated probability in X-axis
- ii. Graph of  $\Delta D$  in Y-axis and estimated probability in X axis
- iii. Graph of  $\Delta\beta$  in Y-axis and estimated probability in X axis
- iv. Graph of  $\Delta\chi^2$  in Y axis and estimated probability with symbol size proportional to  $\Delta\beta$  in X axis

##### 4.4.5.1 Plot of delta Chi-square and Estimated Probability

Few data points (4 data points) seem to be having unusual compared to most of the data points as shown in Figure 5.

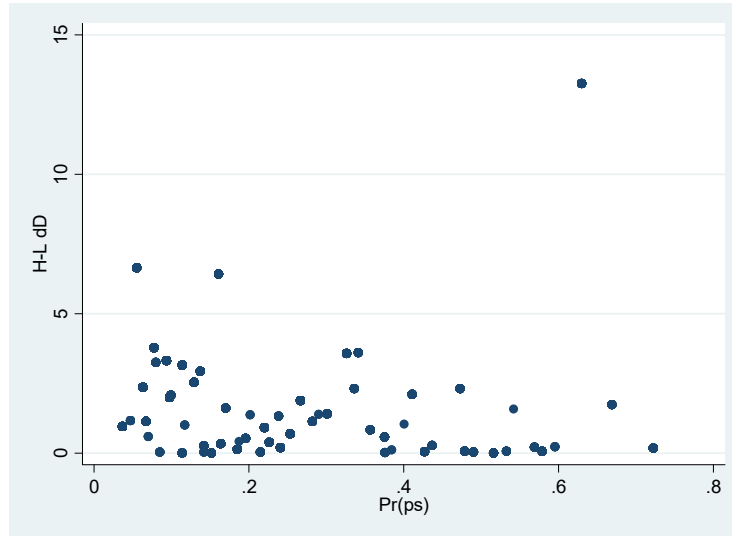


**Figure 5:** Plot of  $\Delta\chi_j^2$  and Estimated Probability from the Fitted Multiple LRM with Covariate Pattern  $J = 60$

The 4 data points are having the value of  $\Delta\chi_j^2$  greater than 4. The value of  $\Delta\chi_j^2$  equals to 4 is considered as the crude approximation to the 95<sup>th</sup> percentile of the distribution of  $\Delta\chi_j^2$  and  $\Delta D$  under  $m$ -asymptotic as each of them would be distributed as  $\chi^2$  distribution with 1 degrees of freedom. Among these 4 data points, only one seems to be very far away. Hence apart from these minor variations noted in the scatter plot, the plot shows that the multiple logistic regression fits reasonably well.

#### 4.4.5.2 Plot of Changes in the Deviance ( $\Delta D$ ) and Estimated Probability

In this plot (Figure 6) also similar patterns can be observed and be seen in the plot of delta chi-square (Figure 5).

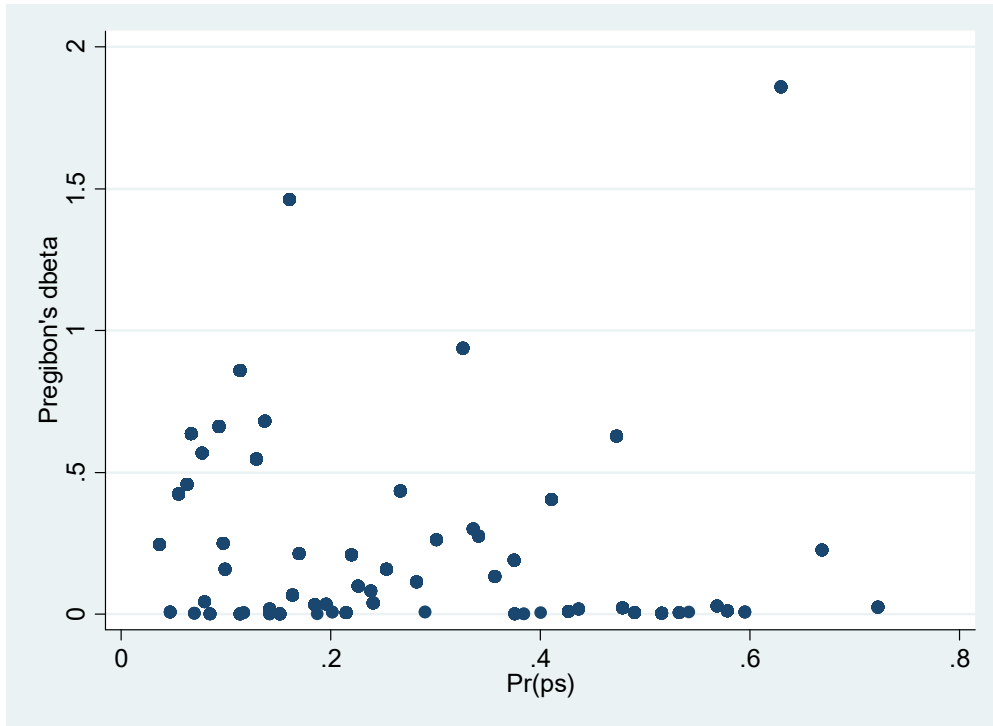


**Figure 6:** Plot of  $\Delta D$  and estimated probability from the fitted multiple LRM with covariate pattern  $J = 117$

Very few covariate patterns have indicated the poor fit of the model. It is argued that the fit of the model is overall reasonably good with respect to this delta deviance plot except for some covariate patterns.

#### 4.4.5.3 Graph of $\Delta\beta$ and Predicted Probability

The influence diagnostic  $\Delta\beta$  has been plotted with an estimated probability resulted from LRM (Figure 7). It is obvious that two data points are sticking out from the rest of the data. In this plot the values of  $\Delta\beta$  of two data points are greater than 1. This indicates that these individual covariate patterns are influential in determining the estimated values of regression coefficients (Hosmer and Lemeshow, 2000). Overall, other individual covariate patterns do not have much influence in the estimated values of regression coefficient in overall.

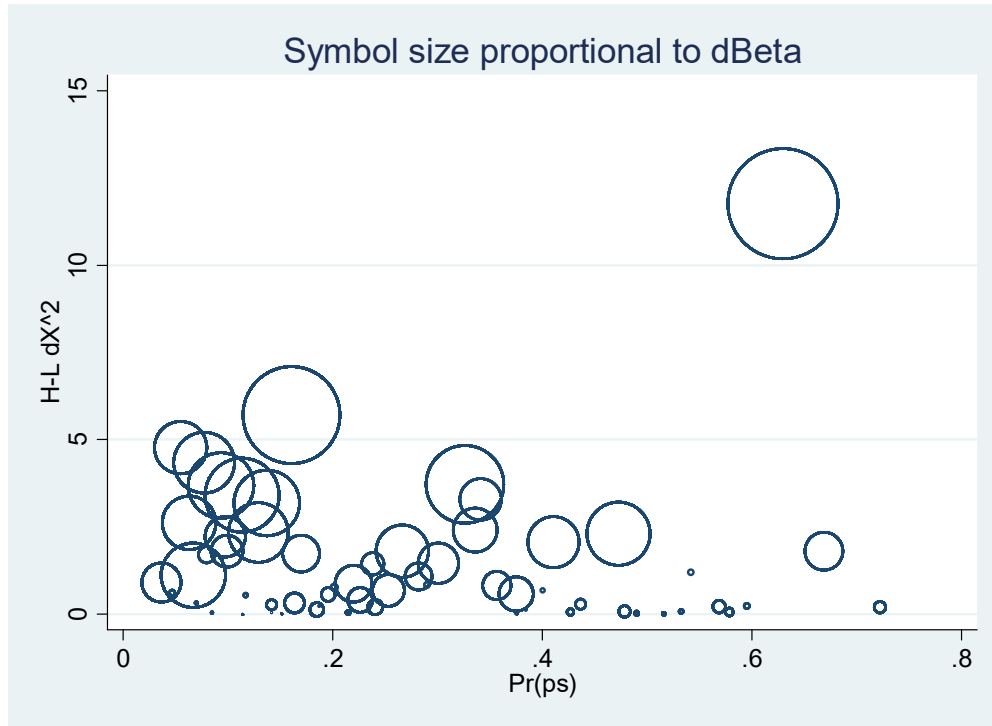


**Figure 7:** Plot of  $\Delta\beta$  and Estimated Probability from the Fitted Multiple LRM with Covariate Pattern  $J = 60$

**4.4.5.4 Graph of  $\Delta\chi^2$  and Predicted Probability with Symbol Size Proportional to  $\Delta\beta$**

Figure 8 displays the scatter plot of  $\Delta\chi^2$  against the probability based on the fit of LRM with the size of the symbol proportional to  $\Delta\beta$ . In this plot, symbol size proportional to  $\Delta\beta$  determines the influence of the covariate patterns on the overall fit of the model. There are only a few extremely big circles in the plot. The value of  $\Delta\chi^2$  is small for all these circles except for one. This shows clearly that only one covariate patterns has an impact on the regression coefficients and  $\Delta\chi^2$ .





**Figure 8:** Plot of  $\Delta\chi^2$  and Predicted Probability of LRM with Size of the Symbol Proportional to  $\Delta\beta$ , Covariate Pattern  $J = 60$

Overall, assumptions related to this diagnostic measure and the plots have not been flagrantly violated. Just a relatively small number of covariate patterns has been seen in four plots (Figures 5 to 8). The value of  $\Delta\chi^2$ ,  $\Delta D$  each is not much greater. Just two covariate patterns with values of  $\Delta\beta$  greater than one are identified. Therefore, it can be concluded that the fitting of the model using this data of Nepal seems to be reasonably good on the assessment of regression diagnosticis prospect (Acharya et al., 2022a).

#### 4.4.6 Results of Risk Assessment on the basis of Factors Present in the Model

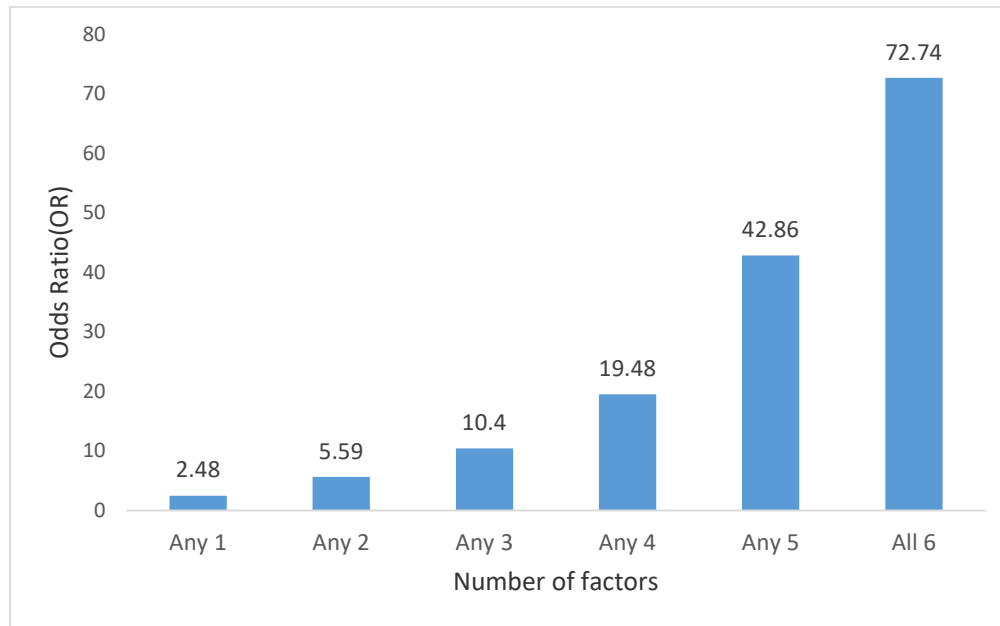
The distribution of houheholds corresponding to presence of different risk factors and the risk of a household being poor (in terms of OR with 95% CIE) are shown in Table 24.

**Table 24:** Distribution of Households and OR on the Basis of Presence of Number of Factors

Number of factors	Number of households(%)	% of households	OR (95% CIE)
None	714	11.9	-
Any one	1626	27.2	2.48 (1.56 – 3.95)
Any two	1431	23.9	5.59 (3.57 – 8.76)
Any three	1191	19.9	10.40 (6.67 – 16.22)
Any four	732	12.2	19.48 (12.43 – 30.55)
Any five	241	4.0	42.86 (26.12 – 70.35)
All six	53	0.9	72.74 (35.27 – 150.03)
Total	5988	100	-

Computed from data of NLSS III

Of the total households, 12% households were found not presenting any factor. Remaining 88% of the households are associated with having either of any factor. The highest number of households (27%) are associated with any one factor. Nonetheless, considerable number of households are associated with either 2 factors (24%) or three factors (20%) or four factors (12%). The odds ratio of any one factor is found to be 2.5 and increased accordingly. The highest odds ratio of all six factors present is seen to be 72.7. The confidence interval of OR seems to be wider gradually as the number of presence of risk factors increases in the households. The wideness in the confidence interval is expected since the number of households are gradually decreasing as the presence of number of risk factors increases. Nonetheless, there is not much wider confidence interval of OR for presence of any one, any two and any three risk factors for each of them, the sample size is also relatively larger, as shown in Table 24.



**Figure 9: OR in Presence of Number of Risk Factors**

The likelihood of risk of household poverty increases as the number of factors increases as shown in Table 28 and Figure 9. When one risk factor presents, the risk of poverty is approximately 2.5 times more in comparisons of not presenting any one factor. If three factors are present, then OR goes up to 10.4, and it increases as the number of factors increases.

#### 4.4.7 Stability of the Model

Bootstrapping resampling of the model has been run 1000 times with the final set of variables that are independent, to evaluate high repetition of each factor in each final model. The main goal of this approach is to determine the significance of each factor in each final model to examine how frequently each variable appeared in each model. Naturally, the more frequently a variable occurs, the more significant it is in the model. As a result, the fitted model would be expected to be more stable. Among all the six independent covarites, one variable namely number of literate members of WAP appeared 97.4% times and other remaining five variables appear 100% times (Table 25). This indicates that each covariate in the final model appeared to be almost equally important to affect the household poverty. On the basis of the assessment of the repeatability of the variable in the model, examined through bootstrapping procedure, has evidently indicated that the developed LRM with six independent covariates is stable.

**Table 25:** Results of the Bootstrap Resampling Procedure for LRM

Variable	Out of 1000 total bootstrapping repetition (%)
Literacy status of household head	1000 (100%)
Remittance receiving status of household	1000 (100%)
Status of land ownership	1000 (100%)
Access to nearest market center	1000 (100%)
Number of children under 15 years	1000 (100%)
Number of literate members of (WAP)	974 (97.4%)

Now, summarizing the results based on the fitting of the LRM, its diagnostic criteria and stability of the model, based on extensive literature review, 7 variables are identified as candidate covariates potentially affecting the poverty in Nepal. These variables were: (sex of household head, literacy status of household head, remittance receiving status of household, land ownership status of household, access to nearest market center, number of children under 15 years, number of literate members of WAP. Chi-square test as a bivariate analysis is used to confirm whether these variables are associated with poverty. All variables except the sex of household are statistically significant in explaining the variation in poverty in both bivariate as well as in the multiple LRM finalized through stepwise (backward and forward) selection method. As a result, poverty is expressed in terms of these 6 variables in this study. The final developed multiple LRM with six factors is statistically significant as shown by omnibus test, and the fit is good that is confirmed by H-L Chi square test. Similarly, odds ratio (OR) of each of these 6 variables is also found to be statistically significant at 5 % level of significance with reasonably narrower 95% confidence interval. The odds of having poverty is the highest (OR: 4.69; 95% CIE: 4.06 - 5.42) for the variable “number of children under 15”. It is the smallest (OR: 1.29; 95% CIE: 1.07 - 1.56) for the variable “number of literate member of WAP”. The model has reasonably satisfied the diagnostic criteria assessed through graphical approaches. According to the model specification test, the model is correctly specified. Similarly, the household poverty increases with the presence of number of risk factors increases. Running bootstrap resampling 1000 times, only one variable has appeared 97.4% of the time but all other variables are present all the time (100%). This has indicated that almost all 6 variables

are equally important in the model, and the fitted multiple logistic regression model is found to be stable.

#### 4.5 Results of Log-binomial Regression Model (LBRM)

Several studies have advocated that the use of LBRM as a useful model and as an alternative to the LRM for binary outcome with common number of event of outcome. However, the use of LBRM has not been found in social science set up data. Since the household poverty based on NLSS III data is 18.5%, the use of LBRM as an alternative to the logistic model is justified. The results of fitting of the LBRM with the goodness of fit test and other diagnostics are discussed in the following sections.

In total, there are seven independent variables in association with response variable. The LBRM considering one independent variable at a time has been used to determine the potential covariates for the final LBRM. The results consisting of the relative risk (RR), p-value, and 95% CIE for each covariate are shown in Table 26.

**Table 26:** Results of LBRM Considering One Variable at a Time (n = 5988)

Independent characteristics	RR	p-value	95% CIE
Sex of household head:			
Male	1.00		
Female	0.92	0.195	(0.82 1.04)
Literacy status of household head:			
Literate	1.00		
Illiterate	2.31	< 0.001	(2.07 2.57)
Status of remittance receipt:			
Yes	1.00		
No	1.38	< 0.001	(1.24 1.54)
Status of land ownership:			
Yes	1.00		
No	1.79	< 0.001	(1.61 1.99)
Access to nearest market center:			
Better	1.00		
Poor	2.25	< 0.001	(2.00 2.52)
Number of children under 15:			
≤ 2	1.00		
> 2	3.68	< 0.001	(3.32 4.09)
Number of literate members of WAP:			
≥ 1	1.00		
0	1.97	< 0.001	(1.77 2.20)

Source: Results adopted from Acharya et al. (2022c)

Only six of these seven independent variables are statistically significant ( $\alpha = 0.05$ ) (Table 26) except one variable - sex of the household head. As a result, these six variables are regarded as potential covariates for developing multiple LBRM.

**Table 27:** Regression Estimates of LBRM with 95% CIE (n = 5988)

Independent characteristics	RR	S.E.	p-value	95% CIE
Literacy status of household head:				
Literate	1.00			
Illiterate	1.68	0.1006	< 0.001	(1.49 1.89)
Status of remittance recipient:				
Yes	1.00			
No	1.45	0.0685	< 0.001	(1.33 1.59)
Status of land ownership:				
Yes	1.00			
No	1.22	0.0594	< 0.001	(1.11 1.34)
Access to nearest market center:				
Better	1.00			
Poor	1.51	0.0888	< 0.001	(1.34 1.69)
Number of children under 15:				
$\leq 2$	1.00			
$> 2$	2.96	0.1590	< 0.001	(2.66 3.28)
Number of literate members of WAP:				
$\geq 1$	1.00			
0	1.16	0.0606	< 0.001	(1.05 1.29)
Constant	0.05	0.0034	< 0.001	(0.05 0.06)
Log likelihood(null model) = - 4068.888				
Log likelihood(full model) = - 2412.336				
AIC = 0.808				
BIC = - 47195.150				
H-L ( $\chi^2$ ) with 8 d.f = 28.602, p = 0.0004				

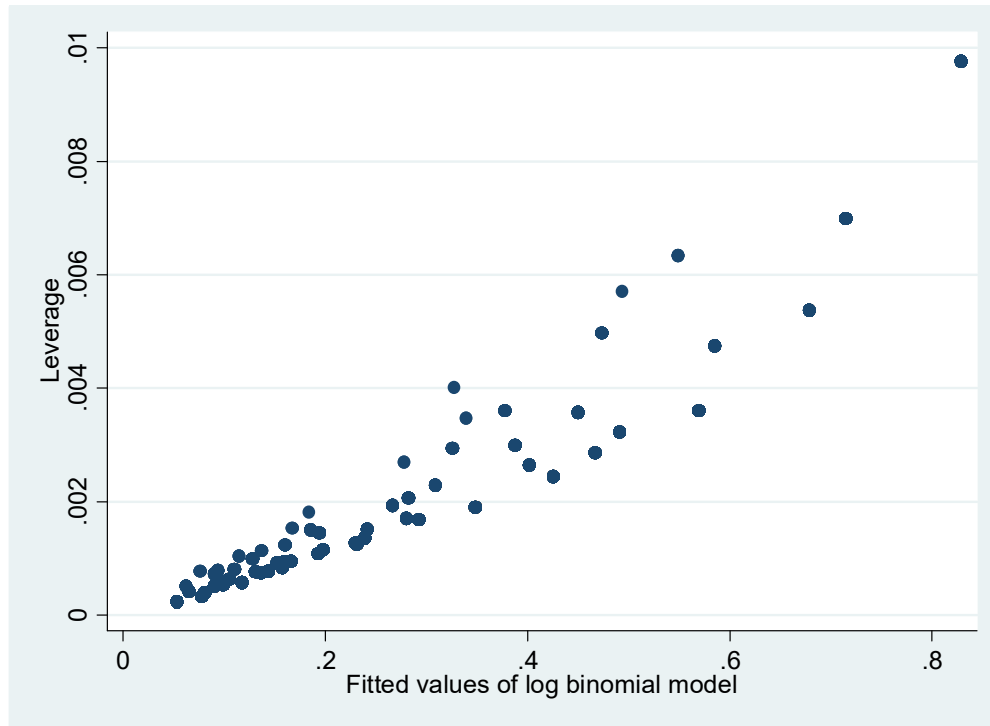
Source: Results, adopted from Acharya et al. (2022c).

Among the final six independent variables, household with  $> 2$  children under 15 has been determined to be the greatest risk associated with poverty (RR: 2.96; 95% CIE: 2.66 - 3.28). This is followed by households with an illiterate household head (RR: 1.68; 95% CIE: 1.49 – 1.89), poor access to the market center (RR: 1.51; 95% CIE: 1.34 – 1.69), households not receiving remittances (RR: 1.45; 95% CIE: 1.33 – 1.59), and households without land (RR: 1.22; 95% CIE: 1.11 – 1.34). On the other hand, households without a single literate member of WAP have the lowest risk of being poor

(RR: 1.16; 95% CIE: 1.05 - 1.29) compared to those households which have more than one literate member. According to this, households without a single member who is literate and of working age are 1.16 times more likely to be poor than households with at least one such member. The model's goodness of fit, as measured by the H-L ( $\chi^2$ ) test with 8 degrees of freedom, is violated ( $p= 0.0004$ ) (Table 27). The finding of this study regarding the goodness of fit of the model is not in the similar direction of the study findings reported by (Blizzard & Hosmer, 2006). In their findings based on the real data of Tasmania, there was good fitting of the LRM with the value of H-L Chi-square test 3.80 with 8 degrees of freedom ( $p = 0.92$ ). The considerable violation for good fitting of the model as assessed by H-L test in the study data set (NLSS III) might have been because of having all independent covariates of categorical with two levels (Acharya et al., 2022c). However in the model developed by Blizzard and Hosmer (2006), among two independent variables, one variable was of continuous type.

#### **4.5.1 Results of Diagnostics for the LBRM**

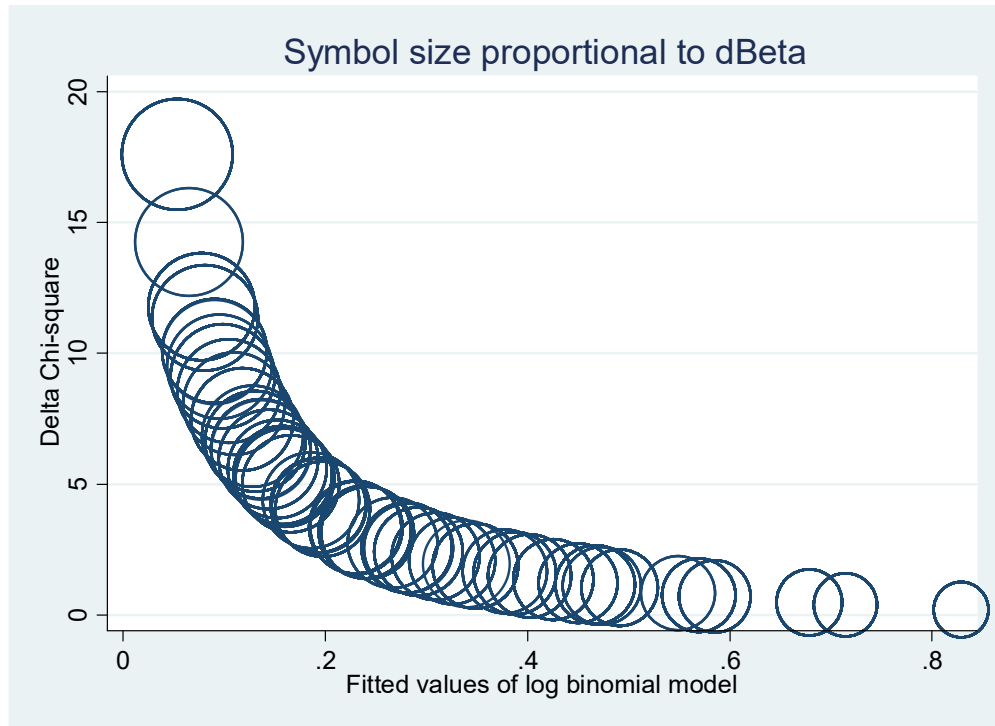
Figure 10 (a) shows a visualization of the leverage values on the y-axis and the model-fit values on the x-axis. One of the data values appears to have more leverage than the others in the upper right hand corner. The majority of leverage values in this dataset are less than 0.008, and the highest leverage value is also below 0.01. As a result, all leverage values are below 0.08 (Blizzard & Hosmer, 2006), showing there is not any violation of the model's leverage-based diagnostics.



**Figure 10 (a):** Leverage and Fitted Values of the LBRM

The diagnostic of the model has also been attempted graphically keeping  $\Delta\chi^2$  in Y-axis the predicted probabilities in the X-axis with plotting symbol proportional to Cook's distance (Figure 10 (b)). There are four data points with poor fit i.e.  $\Delta\chi^2 > 10$ . The four circles are seen to be larger than other circles. This finding seems to be in better direction as compared to the findings reported by Blizzard and Hosmer(2006). In their data set, there were 22 poorly fit subjects ( $\Delta\chi^2 > 10$ ). In the study based on this thesis work, three data points are located in the right lower corner, one of which is far from the other two. Blizzard and Hosmer (2006) has reported that two subjects' circles were larger than the others. The findings regarding the diagnosis of the LBRM based on this criteria has reasonably indicated that the model diagnostics has not been violated. Hence, based on these two diagnostic plots there are not major significant violations on the diagnostics of the fitted LBRM.





**Figure 10 (b):** Graph of  $\Delta\chi^2$  and predicted Values of LBRM with Plotting Symbol Proportional to Cook’s Distance

#### 4.5.2 Results of Risk Assessment on the basis of Factors Present in Log-binomial Model

The number and percentage distribution based on the presence of various risk factors as well as the likelihood that a household would be poor (in terms of risk ratio with 95% CIE) are presented in Table 28.

**Table 28:** Distribution of Households and RR on the Basis of Presence of Number of Factors

Number of factors	Number of households(%)	% of households	RR (95% CIE)
None	714	11.9	1.00
Any one	1626	27.2	2.38 (1.52 – 3.71)
Any two	1431	23.9	4.9 (3.19 – 7.53)
Any three	1191	19.9	8.07 (5.28 – 12.13)
Any four	732	12.2	12.41 (8.14 – 18.92)
Any five	241	4.0	18.72 (12.23 – 28.64)
All six	53	0.9	22.66 (14.48 – 35.46)
Total	5988	100	-

Source: Computed from data of NLSS III

The discussion about the distribution of households based on the presence of risk factors has also already been performed in section 4.4.6. The value of risk ratio of any one factor has been found to be 2.4 and increased in accordance. The value of risk ratio 18.7 is determined to be the greatest among the six variables. As more risk factors are present in the households, the 95% confidence interval for the RR has appeared gradually wider. Since households are gradually becoming less as the number of risk factors increases, a wider confidence interval is to be estimated as indicated in Table 28.

### 4.5.3 Results of Stability of LBRM

Among the six independent variables in the final model, number of literate members of WAP has occurred 97.4% times and other variables have occurred 100% times (Table 29) while performing the bootstrap resample procedure for 1000 times. This shows that each of the six covariates on household poverty seems to be almost equally important. Based on the bootstrapping procedure, analysis of the repeatability of the variable of the model, it is clear that the developed LBRM with six independent variables is stable.

**Table 29:** Results of the Bootstrap Resampling Procedure for LBRM

Variables	Out of 1000 total bootstrapping repetition (%)
Literacy status of household head	1000 (100%)
Remittance receiving status of household	1000 (100%)
Status of land ownership	1000 (100%)
Access to nearest market center	1000 (100%)
Number of children under 15 years	1000 (100%)
Number of literate members of WAP	974 (97.4%)

The results of the LBRM is summarized as follows.

The same seven independent candidate variables are considered for multiple LBRM as did in the case of LRM. Out of these seven independent variables, only six variables are come out statistically significant at  $\alpha=0.05$ , except for sex of household head, which is selected through stepwise backward and forward selection procedure. The good fit of the model has been assessed through H-L Chi-square test. As mentioned earlier, the

RR is the highest (RR: 2.96; 95% CIE: 2.66 - 3.28) for number of children under 15 and it is the lowest (RR: 1.16; 95% CIE: 1.05 - 1.29) for the variable number of literate members of WAP. Two graphs are made for model diagnostics for log-binomial. One is the fitted values where all the leverages are lower than 0.08. This implies that there is no violation in the leverage diagnostics. The other one is the graph of  $\Delta\chi^2$  against the predicted probability with plotting symbol proportional to Cook's distance. This plot also does not show any violation of model assumptions. The bootstrap resampling has also indicated that all variables are almost equally important indicating the model is stable as is in the case for LRM. However, the H-L Chi square test has showed that the good fit is not satisfied for LBRM.

#### **4.5.6 Results of Comparison of Logistic and Log- binomial Regression Models**

At the beginning, the same set of seven variables are used in both the LRM and LBRM building processes. With the exception of variable “sex of household head,” both models have identified other six factors statistically significant. LRM has showed a high effect size and has a larger 95% confidence interval estimation than LBRM when comparing the effect size for each variable. In this study, the OR of LRM varies from 1.29 of the variable “number of literate members of WAP” to 4.69 of the variable “number of children under 15 years”. The RR of LBRM varies from 1.16 of the variable “number of literate members of WAP” to 2.96 of the variables “number of children under 15 years.” The variable whose OR is minimum, and there is also minimum RR for the same variable, and similar pattern is also observed in the case of maximum OR and maximum RR. The OR is overestimated than RR for each independent covariate (Table 30). Similar results are found in other studies (Barros & Hiraakata, 2003; Blizzard & Hosmer, 2006; Diaz-Quijano, 2012; Espelt et al., 2017). The greater elevation of risk for variable estimates(logistic regression vs. log-binomial regression) varies from 13% to 173% (Table 30). The largest (173%) greater risk elevation is observed for ‘number of children under 15’. The smallest (13%) elevations of risk are observed for covariates ‘number of literate persons of WAP’.

**Table 30:** Comparison of LRM and LBRM in terms of Different Parametrs (n = 5988)

Independent characteristics	LRM		LBRM		Elevation in risk (%)
	OR(95% CIE)	Width of interval	RR(95% CIE)	Width of interval	
Literacy status of household head:					
Literate	1.00		1.00		
Illiterate	2.20(1.86 2.61)	0.75	1.68(1.49 1.89)	0.4	52
Status of remittance receipt:					
Yes	1.00		1.00		
No	1.90(1.64 2.20)	0.56	1.45(1.33 1.59)	0.26	45
Status of land ownership:					
Yes	1.00		1.00		
No	1.53(1.31 1.78)	0.47	1.22(1.11 1.34)	0.23	31
Access to market:					
Better	1.00		1.00		
Poor	1.77(1.52 2.07)	0.55	1.51(1.34 1.69)	0.35	26
Number of children under 15:					
≤ 2	1.00		1.00		
> 2	4.69(4.06 5.42)	1.36	2.96(2.66 3.28)	0.62	173
Number of literate members of WAP:					
≥ 1	1.00		1.00		
0	1.29(1.07 1.56)	0.49	1.16(1.05 1.29)	0.24	13
H-L ( $\chi^2$ ) with 8 d.f	6.05, p = 0.534		28.602, p = 0.0004		
AIC	4813.844		0.808		
BIC	4860.727		- 47195.150		

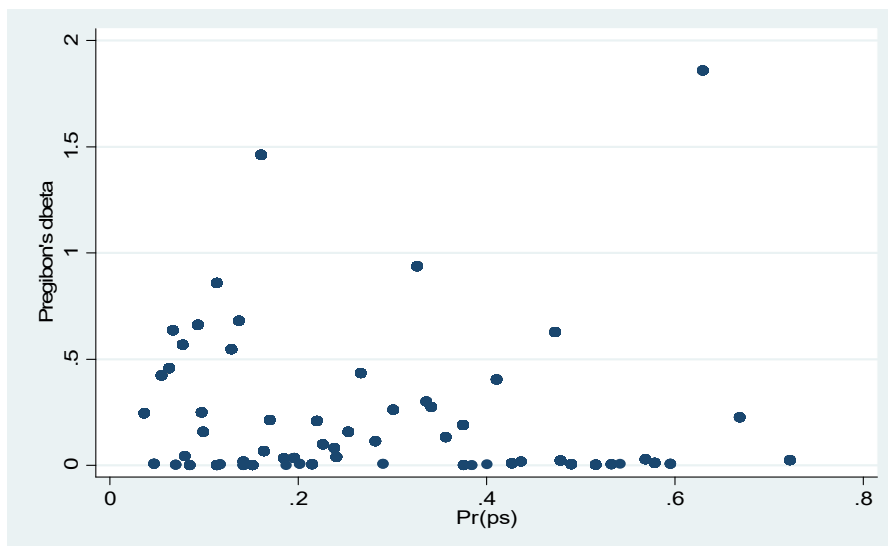
Source: Results adopted from Acharya et al. (2022a and 2022c)

The good fit of the LBRM as measured by the H-L ( $\chi^2$ ) test (8 d.f.) is not satisfied (p = 0.0004). However, this test for the LRM is satisfied (p = 0.534). But the good fit test for both LBRM and LRM is reported satisfied by the study of Blizzard and Hosmer (2006).

The value of AIC is larger (AIC: 4813.84) for LRM than for LBRM (AIC: 0.81). On the basis of comparison of AIC, LBRM seem to be better comparatively. The BIC value is positive in LRM (BIC: 4860.73) whereas it is negative in LBRM (BIC: -47195.15). While comparing BIC value, in terms of magnitude alone, it seems that there is smaller value in LRM compared to LBRM (Table 30). There is no problem of convergence in each logistic and LBRM indicating that there is not such misbehaving characteristics. The findings of the research based on this convergence issue is similar to the findings of Barros and Hirakata (2003) and Blizzard and Hosmer (2006).

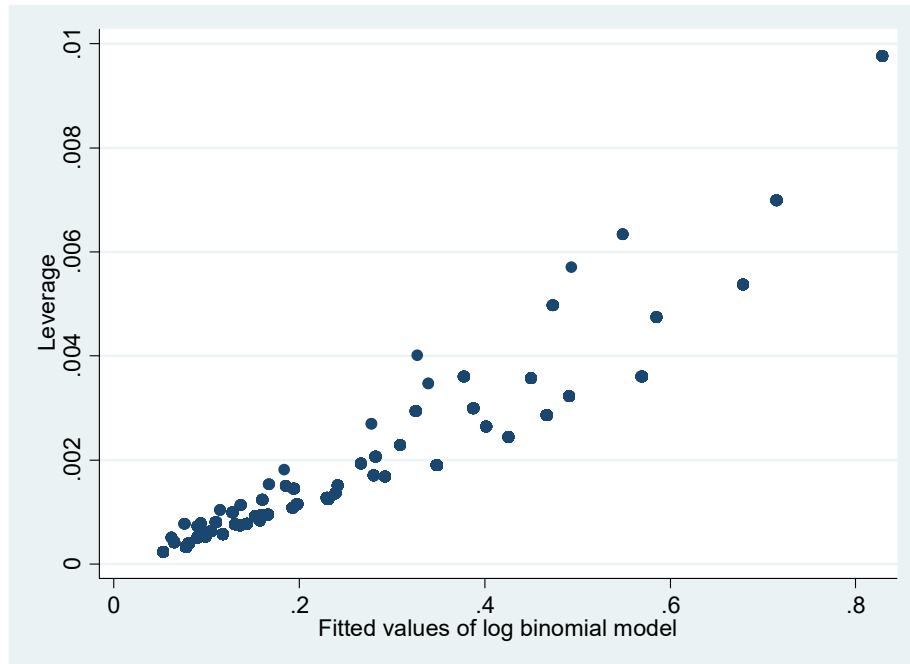
#### 4.5.6.1 Results of Comparison with reference to Diagnostics

The diagnostics of the LRM has been examined by the graph of  $\Delta\beta$  and estimated probability (Figure 11a). The same is also examined in LBRM by the graph of leverage and estimated probability (Figure 11b). Figure 11(a) does not violate the diagnostic criteria except 2 data points greater than 1. Figure 11(b) also does not violate the diagnostics criteria except one covariate. Hence, both the graphs have reasonably satisfied the diagnostics evaluations.



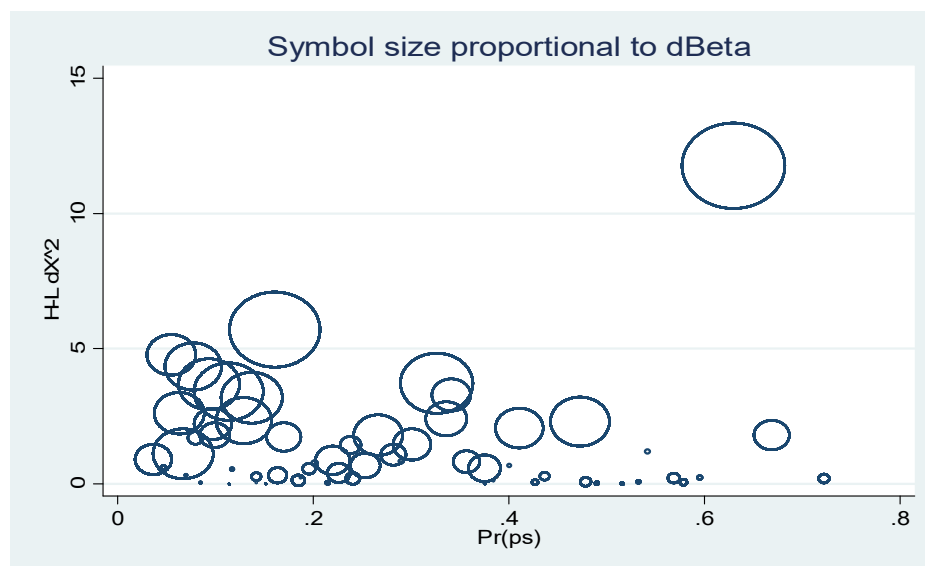
**Figure 11 (a):** Graph of  $\Delta\beta$  and Estimated Probability (from LRM)

[Source: Acharya et al. (2022)]

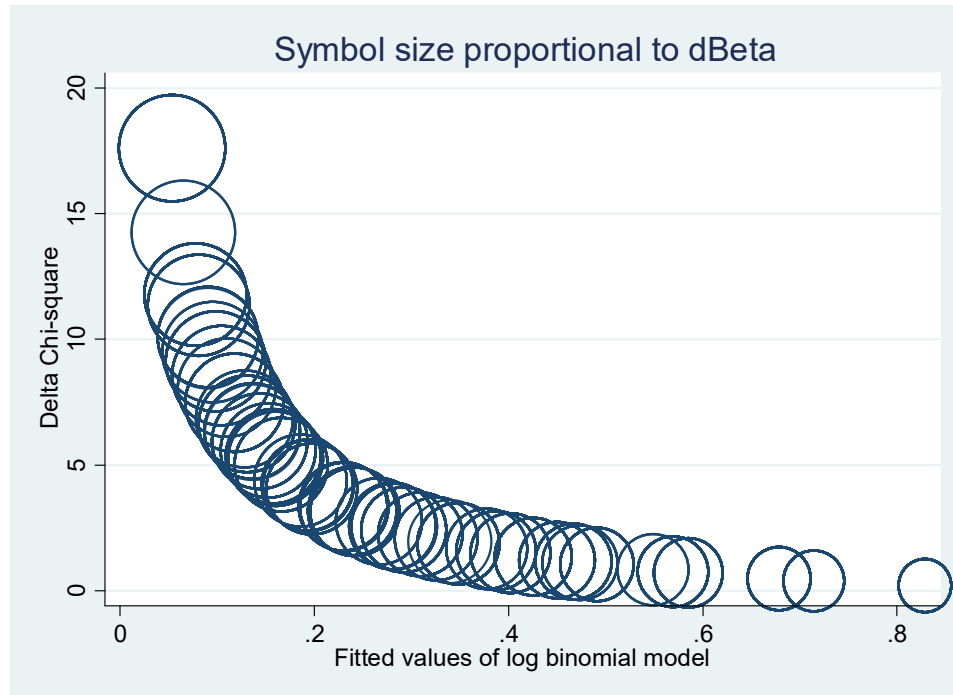


**Figure 11 (b):** Leverage and Fitted Values of LBRM

Figure 12(a) represents for LRM and Figure 12(b) BLRM. Based on the visual assessment of these two graphs, there is not so much violations of the diagnostics criteria. In the case of regression diagnostics, these results are corroborated by the study of Blizzard and Hosmer (2006).



**Figure 12 (a):** Graph of  $\Delta\chi^2$  and Predicted Probabaility of LRM with Symbol Size Proportional to  $\Delta\beta$  [(Source: Acharya et al. (2022c)]



**Figure 12 (b):** Graph of  $\Delta\chi^2$  and Predicted Probability of LBRM with Plotting Symbol Proportional to Cook’s Distance [Source: Acharya et al. (2022c)]

#### 4.5.6.2 Results of Comparison Based on Stability of the Model

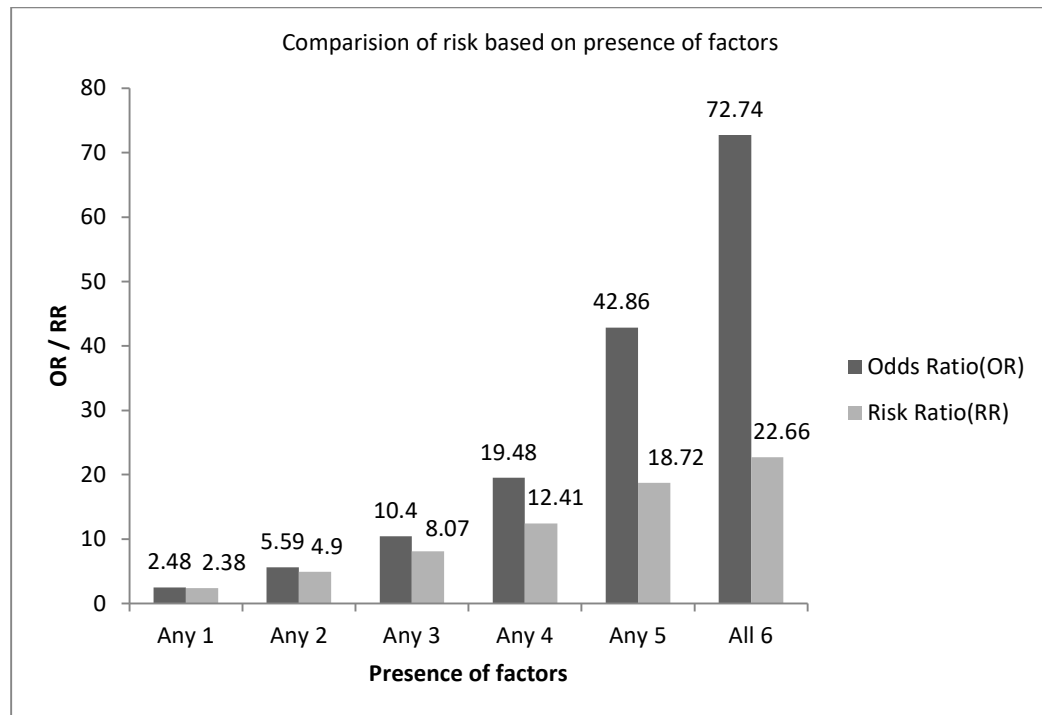
Only one variable, “number of literate members of WAP”, repeated 97.4% of the time in each model. The rest are found to be repeating 100% of the time (Table 31) out of 1000 bootstrap resampling procedure. As a result, both models meet the stability criteria. This has indicated that the variables played almost equally important role in both the models.

**Table 31:** Results of the Bootstrap Resampling Procedure for LRM vs. LBRM

Variables	LRM	LBRM
	Out of 1000 total bootstrapping repetition (%)	Out of 1000 total bootstrapping repetition (%)
Literacy status of household head	1000 (100%)	1000 (100%)
Remittance receiving status of household	1000 (100%)	1000 (100%)
Status of land ownership	1000 (100%)	1000 (100%)
Access to nearest market center	1000 (100%)	1000 (100%)
Number of children under 15 years	1000 (100%)	1000 (100%)
Number of literate members of WAP	974 (97.4%)	974 (97.4%)

### 4.5.6.3 Results of Comparison Based on Risk Assessment

Risk assessment has been individually carried out by running LRM and LBRM model considering the occurrence of the risk factors.



**Figure13:** OR and RR in Presence of Number of Risk Factors

As the number of variables increases in each model, the likelihood of households having poor increases continuously (Figure13). However, the LRM overestimates the risk of each factor compared to LBRM.

Summary of effect size with 95% CIE of each independent covariate developed by LRM and LBRM, AIC, BIC, statistical tests related to goodness fit, results of regression diagnostics and results based on robustness criteria have been presented side by side in Table 32.



**Table 32:** Comparison of Models' Results

Independent variables	LRM		LBRM	
	OR (95% CIE)	Out of 1000 total bootstrapping repetition (%)	RR (95% CIE)	Out of 1000 total bootstrapping repetition (%)
Literacy status of household head: Literate Illiterate	1.00 2.20 (1.86, 2.61)	1000 (100)	1.00 1.68 (1.49, 1.89)	1000 (100)
Status of remittance recipient: Yes No	1.00 1.90 (1.64, 2.20)	1000 (100)	1.00 1.45 (1.33, 1.59)	1000 (100)
Status of land ownership: Yes No	1.00 1.53 (1.31, 1.78)	1000 (100)	1.00 1.22 (1.11, 1.34)	1000 (100)
Access to nearest market center: Better Poor	1.00 1.77 (1.52, 2.07)	1000 (100)	1.00 1.51 (1.34, 1.69)	1000 (100)
Number of children under 15: ≤ 2 > 2	1.00 4.69 (4.06, 5.42)	1000 (100)	1.00 2.96 (2.66, 3.28)	1000 (100)
Number of literate members of WAP: ≥ 1 0	1.00 1.29 (1.07, 1.56)	974 (97.4)	1.00 1.16 (1.05, 1.29)	974 (97.4)
H-L ( $\chi^2$ ) with 8 d.f.	6.05, p = 0.5340		28.602, p = 0.0004	
AIC	4813.844		0.808	
BIC	4860.727		-47195.150	
Model diagnostics are reasonably satisfied in each model	<ul style="list-style-type: none"> <li>• <math>\Delta \chi^2</math> vs. estimated probability</li> <li>• <math>\Delta D</math> vs. estimated probability</li> <li>• <math>\Delta \beta</math> vs. estimated probability</li> <li>• <math>\Delta \chi^2</math> vs. estimated probability with symbol size proportional to <math>\Delta \beta</math></li> </ul>		<ul style="list-style-type: none"> <li>• Graph of Leverage and fitted values</li> <li>• Graph of <math>\Delta \chi^2</math> and predicted values with plotting symbol proportional to Cook's distance</li> </ul>	

Based on the above presented results and discussion of fitting of LRM, LBRM and their comparison, the following summary has been made.

The same set of variables are used to compare LRM and LBRM. All six variables except the sex of household head are statistically significant in both models. When calculating the OR and RR for 6 significant variables, the effect size and 95% confidence intervals for LRM are higher and wider, respectively, as compared to LBRM. The greater elevation of effect size for each covariate is found in LRM compared to that of LBRM. The greater elevation is noted from 13% to 173% (Acharya et al., 2022c). The variables for which effect size and CIE are higher for the LRM, are also higher for the LBRM. Similarly, the variable for which effect size and CIE are lower for the logistic regression model, are also lower for the LBRM. The value of AIC is less in LBRM compared to LRM (0.808 vs. 4813.84). However, the BIC is larger with negative sign in LBRM than that of LRM (-47195.15 vs. 4860.73). H-L Chi square test is satisfied for the LRM, but not for the LBRM.

Graphs of leverage vs. fitted values, and  $\Delta\chi^2$  vs. fitted values with plotting symbol proportional to Cook's distance are satisfactory for LBRM. Similarly, graphs of  $\Delta\beta$  vs predicted probability, and  $\Delta\chi^2$  vs predicted probability with symbol size proportional to  $\Delta\beta$  are also reasonably satisfied for LRM. The poverty of a house increases continuously with the increase of presence of number of variables for both models. In the case of variable's importance, literate member of working age population has occurred 97.4% and the other variables have appeared 100% times in 1000 times bootstrap resampling. This shows that almost all the variables are equally important for each model, and both models seem to be stable.

As an overall finding of this research work, the individual poverty profile of socio-economic and demographic characteristics shows that all measures of poverty are higher in the disadvantaged group than in advantaged groups. All types of poverty (head count index, poverty gap index and square poverty gap index) indicators are greater among disadvantaged groups than at the national level, with the exception of female and illiterate head of the households.

In terms of the factors chosen for the final model, the diagnostics of the fitted model, the stability of the model, the LRM and LBRM run in the same way. The effect size is overestimated by the LRM which also has a larger CIE of effect size than the LBRM. In comparison to the LRM, LBRM's AIC value is lower. In a comparison using

estimates, estimate accuracy, and AIC, the LBRM outperforms the LRM. The good fit of the model is satisfied by the LRM but not by the LBRM as assessed by H-L Chi-square test. Based on a comprehensive comparison that takes into account the model's good fit, the LRM is better than the LBRM. Comparing these two models based on the different diagnostic criteria, both models reasonably satisfy these criteria. There is not the issue of convergence encountered in both the models. Both models satisfy the stability criteria assessed by using bootstrap resampling procedure. On the basis of the overall comparison of these two models based on different criteria including the good fit of the model, the LRM slightly overscores to explain this poverty data of Nepal compared to the LBRM. However, the LBRM can also be one of the important options to the LRM for cross-sectional poverty data with considerable number of outcome of event.

## CHAPTER 5

### 5. CONCLUSION AND RECOMMENDATIONS

This chapter is divided into three sections namely conclusion, recommendations and further study. The first section focuses on the conclusion of the entire research findings and discussions. The second section devotes on highlighting the recommendations based on the research's findings, and the third section, is devoted on indicating the prospective paths for future research in this area.

#### 5.1 Conclusion

Using the sample data of 5,988 households of NLSS III, this study has developed a LRM and LBRM. The aim of this study is to identify important determinants influencing household poverty. The constructed LRM and LBRM with six variables (literacy status of household head, remittance receiving status of household, land ownership status of household, access to nearest market center, number of children under 15 years, number of literate members of WAP) are identified which are statistically significant (at  $\alpha = 0.05\%$ ) with household poverty status. Based on the model used in regression coefficient of correlation coefficient, condition indices, and condition number, there is no significant relationship between independent variables.

From NLSS I to NLSS III, the poverty of Nepal has been in the decreasing trend. In 1996, the poverty of Nepal was 41.8%, similarly, in 2004 it was 30.9%, and in 2011, it came down to 25.2%. Similarly, the poverty rate has decreased in 2004 in comparison to that in 1996. Similarly, the poverty rate in 2011 increased 15.5% as compared to the poverty rate in 2004 (9.6%) for the urban areas of Nepal. Whereas, the poverty rate of Nepal in rural areas have been on the regular decreasing trend 43.3% in 1996 and 34.6%, in 2004 and, 27.4% in 2011. Similarly, the poverty rate of the terai region and the hilly region has been on the continuously decreasing trend, whereas the poverty rate of mountain region has decreased in 2004 as compared to 1996, while in comparison to 2004, there has been increase in the poverty rate in 2011. When referring to the rate of poverty, the Newar community has the least poverty rate with only 10.3%, whereas the highest poverty has been found in the Dalit which is 41.8%. Furthermore, Nepal's household poverty rate for the un-weight cases is 18.5%, while for the weight cases it

is 20.0%. As far as poverty indices are concerned, all kinds of poverty indices are higher in disadvantaged group than in national level except female headed households and illiterate household heads. Those households are considered disadvantaged group if a household headed by female or a household having more than two childrens or household not having at least one literate members or household not having literate household head or household not having any land or household not having better access to the nearest market centre or household not receiving remittance.

The proportion of food expenditure for the poorest group in comparison to the richest group is very high. While, the proportion of non-food expenditure of the poorest group is nearly half than the proportion of richest group.

Results from the developed regression models have indicated that if at least one of the household member is literate that can help to reduce poverty. Therefore, building human capital should be a focus of policies and programs to reduce poverty, especially in households with low levels of education. In addition, remittances play a significant role in reducing poverty. This is due to the fact that as the number of remittance beneficiaries grows, household income rises and poverty decreases.

More than two children in a house makes poverty worse. If the children are not provided with a sufficient education, the intergenerational poverty cycle (a vicious cycle of poverty) may occur in poor households. Lack of land of a household is also a factor contributing to increased household poverty. In addition, the possibility of household poverty surges if there is no easy access of market center.

As far as the LRM and the LBRM are concerned, the effect size of each covariate is overestimated by the LRM compared to LBRM. The LRM also has a larger confidence interval and higher AIC value than the LBRM. Based on the estimates of parameters, precision, and AIC, the LBRM outperforms the LRM in this study. The LRM reasonably satisfies the diagnostics criteria assessed by (i) plots of  $\Delta\beta$  vs. model predicted probability, and (ii)  $\Delta\chi^2$  vs. model predicted probability with symbol size proportional to  $\Delta\beta$ . In a similar fashion, the LBRM also reasonably satisfies the model's diagnostic criteria evaluated graphically through (i) leverage versus predicted value of LBRM and (ii) graph of  $\Delta\chi^2$  versus values of fitted LBRM with plotting symbol proportional to Cook's distance. Both the models do not encounter the model

convergence issues. In both the models, each independent variable in the final model are observed to be almost equally important which is justified through the bootstrapping replication procedure. The likelihood of poor households increases gradually as the presence of number of risk factors increases, according to the risk assessment based on factors included in both the models. As far as the stability of the model is concerned, the same independent variables are almost equally important in both the LRM and the LBRM which is assessed by using bootstrap resampling procedure. On other hand, the good fit assessed by H-L Chi-square test is satisfied by the LRM but not by the LBRM. This implies that, the LRM seems to be comparatively outperformed to the LBRM. Nevertheless, the LBRM is also a reasonable substitute for the LRM even for cross sectional data with considerable number of event of outcome due to its higher precision, lesser AIC, and reduced tendency to overestimated magnitude of effect size. Based on variable selection, effect size, precision of effect size, and AIC, LBRM is better than the LRM not only for clinical and epidemiological data but also for cross-sectional data of poverty. The LRM is better than the LBRM when referring to good fit examined by H-L Chi- square test. Even having the overestimation of effect size and the wider precision of the effect size, when incorporating goodness of fit with other criterias, the LRM can be considered comparatively better than the LBRM.

## **5.2 Recommendations**

Following are the major recommendations based on the findings of this research work.

- This study largely has succeeded in demonstrating the dichotomization scheme which distinguishes the poor into two groups, “advantaged group” and “disadvantaged group,” so that the concerned authorities concentrate the poverty reducing programs towards the disadvantaged group of people.
- Since those households with more than two childrens are at higher risk of poverty, the government has taken initiatives to support for children’s education specially for those families having more than 2 children.
- The findings has clearly indicated that those households either headed by literate member or having at least one literate members of WAP has less risk of poverty. The poverty reducing policies and programs need to be focused for increase the human capital for those households having weaker human capital.

- Remittance receiving households are in better position with respect to poverty. The government of Nepal needs to create friendly environment to invest remittance returnees money in a secure way and to utilize their learned skills in the relevant field.
- The concerned authority must address the issues faced by households who have not any land by adopting reasonable procedures, and develop policies and initiatives to help them. These actions help to reduce landless household's poverty.
- The initiative for improving the market accessibility by creating infrastructures is essential. These infrastructures are to increase road networks, transport networks, cold storage centers, and electricity especially in the rural parts of the country. Easy access of the nearest market centre helps to reduce the household poverty.
- When the event of outcome of interest is frequent for bivariate response variable, the LBRM can also be an important alternative to the LRM even for cross section data related to social sciences such as household poverty. Researchers are recommended to apply the LBRM cautiously under such scenario for social sciences data considering it as one of the useful options to the LRM which is stable, yields precise effect size and the effect size not over estimated.

### **5.3 Further Study**

Future studies can be planned that include other more relevant factors using upcoming NLSS IV data, and one can also plan through primary data considering the standard poverty line as provided by the relevant government authority. Considering other relevant variables, the household empowerment indices can also be constructed. Since this analysis is based on the NLSS III data, it does not permit to make separate analysis for considering the seven provinces of the country. Based on these indices, the factors affecting poverty in different provinces can also be planned using the same statistical framework. Besides them, new studies can be planned to capture other community characteristics related to poverty. The variables identified in this study along with other variables can also be applied using multilevel modeling in a single statistical framework. There are different literatures proposed for assessing the stability of the model. Here it has been done using bootstrapping method. Augustin et al. (2005) has proposed the model selection procedure based on different variables using Bayesian approach, which can also be considered as the possible future research work.

## CHAPTER 6

### 6. SUMMARY

This research work has been carried out to study the risk factors affecting poverty in Nepal. The study has been designed with three fold objectives: (i) to identify the important risk factors of poverty of Nepal, (ii) to compare logistic regression and log-binomial regression in identifying the risk factors and estimating their effects on poverty, and (iii) to assess the stability of the model through bootstrapping method. This study has been exclusively based on secondary data i.e. Nepal Living Standard Survey-2011 (NLSS III). The secondary data has been obtained from the Central Bureau of Statistics (CBS), government of Nepal. This data consists of two separate data files. One data file is of 5988 households and another data file is of 28,670 individuals. The analysis is solely based on the different household characteristics. In order to develop the model, not all variables are readily available in the household data. Therefore, the individual level data has been converted into household level data to have data on a number of variables such as - the number of children (0-14 years of old), working-age members (15 to 64 years old) and elders (65+ years old) by gender and literacy status (literate/illiterate) within each household. Finally the entire analysis has been based on this newly generated household data of 5988 households with different socio-economic, demographic variables along with the response variable household poverty status (poor vs. non-poor) without considering weights.

Based on the extensive review of literature performed for this thesis work, and taking into account of the availability of the variables in the Nepal Living Standard Survey data file of 2010/11, the candidate independent variables have been finalized to be applied in the models, where the outcome of interest is the household poverty. The candidate independent variables are sex of household head (female / male), literacy status of household head (illiterate / literate), status of remittance recipient of household (no / yes), land ownership status of household (no / yes), household with access to nearest market centre (poor / better), number of children under 15 years (more than two / at most two) and number of literate members of working age population (WAP) (none / at least one).



The rigorous review of literature has also clearly indicated that the factors associated with the poverty is generally identified using the LRM. To examine the effect of independent variable over the dependent variable, the LBRM is used as an alternative of the LRM mostly in the clinical and epidemiological studies but not found in the poverty studies. Similarly, the comparison between the LBRM and the LRM has been found in the cohort as well as cross-sectional data of clinical and epidemiological research but not in the analysis of poverty data.

The association of each of these 7 covariates with household poverty as a bivariate analysis has been performed by using  $\chi^2$  test. The effect size of each  $\chi^2$  test has been assessed by using Phi-coefficient. Only six variables have come out significant with poverty at  $\alpha = 0.05$  except sex of household head. However, all these seven variables are considered as candidate variables for each the LRM and LBRM. In each model, the final variables have been selected using both stepwise forward and stepwise backward selection procedure. Except for the sex of the household head, which has been chosen using a stepwise backward and forward selection technique, only six of these seven independent variables are found to be statistically significant at the 5% level of significant in each model.

The overall significance of the fitted logistic regression model has been assessed by using Omnibus test, assessment of classification and discrimination of the model by using sensitivity and specificity, and the discrimination of the model by AUC. The value of pseudo  $R^2$  has also been computed based on the fitted LRM. The good fit of each model (the LRM and LBRM) has been evaluated using H-L  $\chi^2$  test. Akaike Information Criteria (AIC), Bayesian information criterion (BIC) for each model has been computed. Diagnostics of the model, risk assessment based on the presence of factors in the model and stability of the model through bootstrap resampling procedure are also attempted.

The estimated multiple LRM is statistically significant ( $p < 0.001$ ) as shown by the omnibus test. There is no discernible association between the independent variables assessed by using the correlation matrix of regression coefficients, condition indices, and condition number. The regression model fits well as assessed by H-L  $\chi^2$  test with 8 d.f. ( $p = 0.51$ ). The value of McFadden pseudo  $R^2$  is 0.16. Among the six covariates, the likelihood of risk of household being poor is the highest (OR: 4.7; 95% CIE:

4.06 – 5.42) for the household having more than two children compared to that household having less than or equal to two children. The likelihood of risk of household being poor is the least (OR: 1.3; 95% CIE: 1.07 – 1.56) for the household not having single literate member of WAP compared to that of household having at least one literate member of WAP. The correct classification of the fitted logistic regression model is 67.15% at the cut off point (crossing point of sensitivity and specificity) of 0.16. The value of AUC is 0.78 which can be considered as acceptable discrimination of the model. Different diagnostics plots (discussed in chapter 4) revealed that the fit of the model is reasonably good for this poverty data. The risk of poverty based on the presence of factors are assessed using logistic regression model. As the number of presence of risk factors increases, the risk of household being poor also increases. All variables are almost equally important and the model is stable assessed by bootstrapping method.

As stated earlier, all same six independent variables are statistically significant in the final multiple LBRM too. As previously noted, the number of children under 15 have the greatest RR (RR: 2.96; 95% CIE: 2.66 - 3.28) while the variable number of literate WAP members have the lowest RR (RR: 1.16; 95% CIE: 1.05 - 1.29). The graph of leverages and fitted values of the LBRM, and the graph of delta Chi-square and values of fitted the LBRM with plotting symbol proportional to Cook's distance have used for examining the model's diagnostics. Results have shown that all leverages are less than 0.08. This indicates that the leverage diagnostics are acceptable. There are 4 data points whose values of delta Chi-square greater than 10, and 3 data points are in the lower right corner, and one of which is from these two are far away even from these two data points. Based on the previously published relevant literature, there is not so much violations on the diagnostics of the fitted LBRM. With these two graphical assessments of diagnostics of the model, the LBRM does not violate the diagnostics criteria. The model is stable according to the bootstrap resampling, which also shows that all variables are almost equally important. However, the good fit for the LBRM is not satisfactory according to the H-L Chi-square test. In this model too, the risk of poverty increases as the number of presence of risk factors increases.

The final estimated multiple LRM and multiple LBRM have been compared with respect to variable selection, estimate and its precision, and good fit of the model. Further, the comparison have been made with the model's diagnostics, stability and the

problem of convergence. Both models have finally selected the same set of six variables. The effect size for each independent covariate measured in OR in LRM have been overestimated than that of LBRM measured in RR. The OR varies from 1.29 to 4.69 in the LRM, and RR varies from 1.16 to 2.96 in the LBRM. The greater elevation of risk varies from 13% to 173% for LRM for independent variables compared to the LBRM. The precision of the effect size is evaluated by 95% CIE for each variable for the LBRM which is narrower than that for the LRM. Evidently, the precision of the effect size generated through the LBRM is better than that generated by LRM. The good fit of the LRM has satisfied its criteria but the LBRM is not satisfactory as assessed by H-L ( $\chi^2$ ) test. The value of AIC is smaller for the LBRM (0.808) than for the LRM (4813.844). There is no convergence problem in fitting both models. Both models have satisfied the diagnostics criterion. While assessing the stability of the model using bootstrapping procedure, among six variables, five are repeated 100% times and remaining one got repeated 97% of times for both models. This signifies that both finally fitted models could be considered as stable.

When good fit of the model is also taken into account among other comparison parameters, the LRM is considered as a better choice inspite of having overestimated effect size with a bit wider confidence interval of estimation. In a cross-sectional poverty data, the LBRM is also seem to be a good option to the LRM based on selection of variables, effect size, precision of effect size, and value of AIC.

## REFERENCES

- Abrar-ul-Haq, M., Jali, M. R. M., & Islam, G. M. N. (2018). The development of household empowerment index among rural household of Pakistan. *Pertanika Journal of Social Sciences & Humanities*, **26**(2). <http://www.pertanika.upm.edu.my/>
- Acharya, K.P., Khanal, S.P., & Chhetry, D. (2022a). Factors Affecting Poverty in Nepal - A Binary Logistic Regression Model Study. *Pertanika Journal of Social Science and Humanities*, **30**(2). DOI: <https://doi.org/10.47836/pjssh.30.2.12>
- Acharya, K. P., Khanal, S. P., & Chhetry, D. (2022b). Dichotomization of quantitative variables in poverty analysis. *BIBECHANA*, **19**(1-2): 142-149. DOI: <https://doi.org/10.3126/bibechana.v19i1-2.46407>
- Acharya, K. P., Khanal, S. P., & Chhetry, D. (2022c). On the use of logistic regression model and its comparison with log-binomial regression model in the analysis of poverty data of Nepal. *Nep. J. Stat*, **6**: 63-79. DOI: [10.3126/njs.v6i01.50806](https://doi.org/10.3126/njs.v6i01.50806)
- Achia, T. N., Wangombe, A., & Khadioli, N. (2010). A logistic regression model to identify key determinants of poverty using demographic and health survey data. *European Journal of Social Sciences*, **13**(1): 38-45. *Acta Academica*, **42**(4). <https://www.researchgate.net/publication/287486539>
- Adams, R. H. (2003). *Economic growth, inequality and poverty: Findings from a new data set, 2972*. World Bank Publications. <https://doi.org/10.1002/sim.1956>
- Adekoya, A. O. (2014). Analysis of farm households poverty status in Ogun states, Nigeria. *Asian Economic and Financial Review*, **4**(3): 325-340.
- Adepoju, A. O., & Oluoha, K. (2008). Rural households' access to microcredit and poverty status in Obafemi-Owode local government area of Ogun State, Nigeria. *Journal of Rural Economics and Development*, **17**: 62-71.
- Adhikari, S. R. (2016). Poverty dynamics in Nepal between 2004 and 2011: An analysis of hybrid dataset. *NRB Economic Review*, **28**(1): 29-40.

- Afifi, A., Virginia, A. C., & Susanne, M. (2004). *Computer Aided Multivariate Analysis*. New York, Chapman & Hall/CRC.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**: 716–723.
- Akerele, D., & Adewuyi, S. A. (2011). Analysis of poverty profiles and socioeconomic determinants of welfare among urban households of Ekiti State, Nigeria. *Current Research Journal of Social Sciences*, **3**(1): 1-7.
- Altman, DG & Anderson, PK (1989). Bootstrap investigation of the stability of a Cox regression model. *Statistics in Medicine*, **8**(7): 771-783. DOI: 10.1002/sim.4780080702
- Augustin, N., Sauerbrei, W., & Schumacher, M. (2005). The practical utility of incorporating model selection uncertainty into prognostic models for survival data. *Statistical Modelling*, **5**(2): 95-118. DOI: 10.1191/1471082X05st089oa
- Balarabe, I. I. (2014). Empirical investigation of the determinants of poverty in Kano Metropolis, Nigeria. *Journal of Economics and Sustainable Development*, **5**(14): 14-18.
- Barr, Margo L., Clark, Robert, & Steel, David G.(2016). Examining associations in cross-sectional studies, National Institute for Applied Statistics Research Australia, University of Wollongong, Working Paper. <https://ro.uow.edu.au/niasrawp/35> Accessed on 17th Dec., 2022
- Barros, A. J. D., & Hirakata, V.N. (2003). Alternatives for logistic regression in cross-sectional studies: An empirical comparison of models that directly estimate the prevalence ratio. *BioMed Central Medical Research Methodology*, **3**(21). DOI: <https://doi.org/10.1186/1471-2288-3-21>
- Baser, U., & Kaynakci, C. (2019). Determinants of poverty among smallholder farms in central district of Hatay Province, Turkey. *Journal of International Environmental Application and Science*, **14**(4): 145-151.

- Bennett, L. (2008). *Caste, ethnic, and regional identity in Nepal: further analysis of the 2006 Nepal Demographic and Health Survey*. Population Division, Ministry of Health and Population, Government of Nepal.
- Besley D. A., Kuh, E. & Welsch R. E.(1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York John Wiley
- Blizzard, L. & Hosmer D.W. (2006). Parameter estimation and goodness-of-fit in log binomial regression. *Biometrical Journal*, **48**: 5–22. DOI: <https://doi.org/10.1002/bimj.200410165>
- Booth, Charles, (1903). *Life and Labour of the People in London*. Final Volume, London, the Macmillan Co.
- Botha, F. (2010). The impact of educational attainment on household poverty in South Africa.
- Brewer, M, & O’Dea, C. 2012, Measuring living standards with income and consumption: evidence from the UK, IFS Working Paper W12/12.
- Callaghan, K. J., & Chen, J. (2008). Revisiting the collinear data problem: An assessment of estimator'ill-conditioning'in linear regression. *Practical Assessment, Research, and Evaluation*, **13**(1). DOI: <https://doi.org/10.7275/c8aa-bj67>
- Central Bureau of Statistics. (2005). *Poverty Trends in Nepal (1995-96 and 2003-04)*. Kathmandu: Central Bureau of Statistics, National Planning Commission.
- Central Bureau of Statistics. (2011a). *Nepal Living Standard Survey (2010/11)*. Poverty in Nepal (A brief report based on NLSS III), Central Bureau of Statistics. National Planning Commission Secretariat, Government of Nepal.
- Central Bureau of Statistics. (2011b). *Nepal Living Standard Survey (2010/11)*. Statistical Report, Volume One, Central Bureau of Statistics, National Planning Commission Secretariat, Government of Nepal. [https://cbs.gov.np/wp-content/uploads/2018/12/Statistical\\_Report\\_Vol1.pdf](https://cbs.gov.np/wp-content/uploads/2018/12/Statistical_Report_Vol1.pdf)
- Central Bureau of Statistics. (2011c). *Nepal Living Standard Survey (2010/11)*. Statistical Report, Volume Two, Central Bureau of Statistics, National Planning

Commission Secretariat, Government of Nepal. [https://time.com/wp-content/uploads/2015/05/statistical\\_report\\_vol2.pdf](https://time.com/wp-content/uploads/2015/05/statistical_report_vol2.pdf)

Central Bureau of Statistics. (2014). *Population Monograph of Nepal 2014: Population Dynamics*, 1. <https://mohp.gov.np/downloads/Population%20Monograph%20V01.pdf>

Chaubey. P. K. 1995. 'Poverty Measurement: Issues, Approaches and Indices'. New Age International Publishers

Chen, CH, & George, SL (1985). The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. *Statistics in Medicine*, 4(1): 39-46. DOI: 10.1002/sim.4780040107

Chhetry, D. (2004). Practices of poverty measurement and poverty profile of Nepal.

Chhetry D.2005. Poverty and Reproductive Health: Linkages and Consequences, unpublished report submitted to UNFPA 2005.

Christensen, W., & Angeles, L. (2018). Model Selection Using Information Criteria (Made Easy in SAS?). *University of California, Los Angeles*.

Cook, T. D. (2002). Advanced statistics: Up with odds ratios! A case for odds ratios when outcomes are common. *Academic Emergency Medicine* 9: 1430–1434. DOI: 10.1111/j.1553-2712.2002.tb01616.x

Coutinho, L., Scazufca, M., & Menezes, P. R. (2008). Methods for estimating prevalence ratios in cross-sectional studies. *Revista de saude publica*, 42: 992-998.

Cox, D.R. & E.J. Snell (1989) *Analysis of Binary Data*. Second Edition. Chapman & Hall.

Dahal, D. R. (2003). Social composition of the population: caste/ethnicity and religion in Nepal. *Population monograph of Nepal*, 1: 87-135.

- Deddens, J. A., & Petersen, M. R. (2008). Approaches for estimating prevalence ratios. *Occupational and environmental medicine*, **65**(7): 501-506.  
<http://dx.doi.org/10.1136/oem.2007.034777>
- De Andrade, B. B., & Carabin, H. (2011). On the estimation of relative risks via log binomial regression. *Revista Brasileira de Biometria*, **29**(1): 15.
- Deressa, T. K., & Sharma, M. K. (2014). Determinant of poverty in Ethiopia. *Ethiopian Journal of Economics*, **23**(1): 113-130.
- Diaz-Quijano, F. A. (2012). A simple method for estimating relative risk using logistic regression. *BMC medical research methodology*, **12**(1): 1-6. DOI: DOI: 10.1186/1471-2288-12-14
- Devkota, J. (2014, March). Impact of migrants' remittances on poverty and inequality in Nepal. In *Forum of International Development Studies* , **44**: 36-53).
- Dudek, H., & Lisicka, I. (2013). Determinants of poverty–binary logit model with interaction terms approach. *Econometrics. Ekonometria. Advances in Applied Data Analytics*, **3** (41): 65-77.
- Edoumiekumo, S. G., Karimo, T. M., & Tombofa, S. S. (2014). Determinants of households' income poverty in the South-South Geopolitical Zone of Nigeria. *Journal of Studies in Social Sciences*, **9**(1): 101-115.
- Ennin, C. C., Nyarko, P. K., Agyeman, A., Mettle, F. O., & Nortey, E. N. N. (2011). Trend analysis of determinants of poverty in Ghana: Logit approach. *Research Journal of Mathematics and Statistics*, **3**(1): 20-27.
- Espelt, A., Mari-Dell'Olmo, M., Penelo, E., & Bosque-Prous, M. (2017). Applied Prevalence Ratio estimation with different Regression models: An example from a cross-national study on substance use research. *Adicciones*, **29**(2): 105-112. DOI: 10.20882/adicciones.823
- Eyasu, A. M. (2020). Determinants of poverty in rural households: Evidence from North-Western Ethiopia. *Cogent food & agriculture*, **6**(1): 1823652.  
<https://doi.org/10.1080/23311932.2020.1823652>



- Farah, N. (2015). Impact of household and demographic characteristics on poverty in Bangladesh: A logistic regression analysis. *Eastern Illinois University*.  
[https://thekeep.eiu.edu/lib\\_awards\\_2015\\_docs/3/](https://thekeep.eiu.edu/lib_awards_2015_docs/3/)
- Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics*, 92-107.  
<https://doi.org/10.2307/1937887>
- Fleiss, J.L., Levin, B., & Paik, N.C. (2003). *Statistical Methods for Rates and Proportions*. Third Edition. New York, Wiley.
- Fowlkes, E.B. (1987). Some diagnostics for binary regression via smoothing. *Biometrika*, **74**: 503-515.
- Gallis, J.A., & Turner, E.L. (2019). Relative Measures of Association for Binary Outcomes: Challenges and Recommendations for the Global Health Researcher. *Annals of Global Health*, **85**(1): 137, 1–12. DOI:  
<https://doi.org/10.5334/aogh.2581>
- Gounder, N. (2012). The determinants of household consumption and poverty in Fiji. *Griffith University*.
- Greenland, S. (1987). Interpretation and choice of effect measures in epidemiologic analyses. *American Journal of Epidemiology*, **125**(5): 761–768.
- Greenland, S., & Thomas, D. C. (1982). On the need for the rare disease assumption in case-control studies. *American Journal of Epidemiology*, **116**(3): 547–553.
- Greenland, S., Thomas, D.C., & Morgenstern, H. (1986). The rare-disease assumption revisited: A critique of “estimators of relative risk for case-control studies”. *American Journal of Epidemiology*, **124**(6): 869–883.
- Grosh, M. E., & Glewwe, P. (1998). Data Watch: The World Bank's Living Standards Measurement Study Household Surveys Margaret E. Grosh and Paul Glewwe. *Journal of Economic Perspectives*, **12**(1): 187-196.

- Habyarimana, F., Zewotir, T., & Ramroop, S. (2015). Analysis of demographic and health survey to measure poverty of household in Rwanda. *African population studies*, **29**(1): 1472-1482.
- Hauck, W. W., & Donner, A. (1977), "Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association*, **72**: 851-853.
- Hosmer & Lemeshow (2000). *Applied Logistic Regression*. Second Edition. John Wiley & Sons.
- Hosmer, D. W., Jr., & S. A. Lemeshow (1980). Goodness-of-fit tests for the multiple logistic regression model. *Communications in Statistics*, **A9**: 1043–1069.
- Hosmer, D. W., S. A. Lemeshow, & J. Klar (1988). Goodness-of-fit testing for the logistic regression model when the estimated probabilities are small. *Biometrical Journal*, **30**: 911–924.
- Imam, M. F., Islam, M. A., & Hossain, M. J. (2018). Factors affecting poverty in rural Bangladesh: An analysis using multilevel modelling. *Journal of the Bangladesh Agricultural University*, **16**(1): 123-130. [https:// doi: 10.3329/jbau.v16i1.36493](https://doi.org/10.3329/jbau.v16i1.36493)
- Issahaku, H., & Anawart, G. (2012). Determinants of Poverty in the Kwabre East District of the Ashanti Region of Ghana. doi:10.5707/cjsocsci.2012.5.2.22.31
- Janani, L., Mansournia, M. A., Nourijeylani, K., Mahmoodi, M., & Mohammad, K. (2015). Statistical issues in estimation of adjusted risk ratio in prospective studies. *Archives of Iranian medicine*, **18**(10)
- Javed, Z. H., & Asif, A. (2011). Female households and poverty: A case study of Faisalabad District. *International Journal of peace and development studies*, **2**(2): 37-44.
- Jennings, D.E. (1986). Judging inference adequacy in logistic regression. *Journal of American Statistical Association*; **81**: 471-476.
- Katz, K.A. (2006). The (relative) risks of using odds ratios. *Archives of Dermatology*, **142**(6): 761–764.

- Khan, R. E. A., Rehman, H., & Abrar ul Haq, M. (2015). Determinants of rural household poverty: the role of household socioeconomic empowerment. *American-Eurasian J. Agric. & Environ. Sci*, **15**(1): 93-98. <https://doi.org/10.5829/idosi.aejaes.2015.15.1.1050>
- Khanal, S.P., Sreenivas, V., & Acharya, S. K. (2019). Comparison of Cox Proportional Hazards Model and Lognormal Accelerated Failure Time Model: Application in Time to Event Analysis of Acute Liver Failure Patients in India. *Nepalese Journal of Statistics*, **3**: 21–40. DOI: <https://doi.org/10.3126/njs.v3i0.25576>
- Khudri, M. M., & Chowdhury, F. (2013). Evaluation of socio-economic status of households and identifying key determinants of poverty in Bangladesh. *European Journal of Social Sciences*, **37**(3): 377-387.
- Kleinbaum D.G. & Klein M.(2010). *Logistic Regression: A Self Learning Text*. New York, Springer Publications.
- Kona, M. P., Khatun, T., Islam, N., Mijan, A., & Noman, A. (2018). Assessing the impact of socio-economic determinants of rural and urban poverty in Bangladesh. *International Journal of Science & Engineering Research*, **9**(8): 178-184. <https://www.researchgate.net/publication/329252093>
- Kousar, R., Makhdum, M. S. A., & Ashfaq, M. (2015). Impact of land ownership on the household welfare in rural Pakistan. In: *2015 World Bank Conference on Land and Poverty*. The World Bank-Washington DC, March 23-27.
- Kvalseth, T.O. (1985). Cautionary note about  $R^2$ . *The American Statistician*: **39**: 279-285.
- Lamichhane, K., Paudel, D. B., & Kartika, D. (2014). Analysis of Poverty between People with and without Disabilities in Nepal. *JICA-RI Working Paper*, 77.
- Landwehr, J.M., Pregibon, d., & Shoemaker, A.C. (1984). Graphical methods for assessing logistic regression models. *Journal of the American Statistical Association*, **79**: 61-71.
- Lee, J. (1994). Odds ratio or relative risk for cross-sectional data? *International Journal of Epidemiology*, **23**(1): 201-203.

- Leekoi, P., Jalil, A. Z. A., & Harun, M. (2014). An empirical on risk assessment and household characteristics in Thailand. *Middle-East Journal of Scientific Research*, **21** (6): 962-967. <https://doi.org/10.5829/idosi.mejsr.2014.21.06.21539>
- Lemeshow, S. A., & D. W. Hosmer (1982). A review of goodness of fit statistics for the use in logistic regression models. *Journal of the American Statistical Association*, **79**: 61-71.
- Lumley, T., Kronmal, R., & Ma, S. (2006). Relative risk regression in medical research: models, contrasts, estimators, and algorithms. <http://biostats.bepress.com/uwbiostat/paper293>
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological methods*, *7*(1): 19. DOI: 10.1037//1082-989X.7.1.19
- Majeed, M. T., & Malik, M. N. (2015). Determinants of household poverty: Empirical evidence from Pakistan. *The Pakistan Development Review*, 701-717. <https://ideas.repec.org/a/pid/journal/v54y2015i4p701-718.html>
- Makame, I. H., & Mzee, S. S. (2014). Determinants of Poverty on Household Characteristics in Zanzibar: A logistic Regression Model. *Developing Country Studies*, *4*(20): 188-195.
- Maloma, I. (2016). The socioeconomic determinants of household poverty status in a low-income settlement in South Africa. *International Journal of Social Sciences and Humanity Studies*, *8*(2): 122-131.
- Mamo, B. G., & Abiso, M. (2018). Statistical analysis of factors affecting poverty status of rural residence. *American Journal of Theoretical and Applied Statistics*, *7*(5): 188-192. <https://doi: 10.11648/j.ajtas.20180705.14>
- Margwa, R. S., Onu, J. I., Jalo, J. N., & Dire, B. (2015). Analysis of poverty level among rural households in Mubi region of Adamawa State, Nigeria. *J. Sci. Res. Stud*, *2*(1): 29-35.

- McCullagh, P., & Nelder, J.A. (1989). *Generalized Linear Models*. Second Edition. Chapman Hall , London.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. P. Zarembka (ed.), *Frontiers in Econometrics*. Academic Press, 105-142.
- McNutt, L.-A., Xiaonan Xue C. W., & Hafner J.P. (2003). Estimating the relative risk in cohort studies and clinical trials of common outcomes. *American Journal of Epidemiology*, **157**: 940–943. DOI: 10.1093/aje/kwg074
- Menard, S. (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statistician* **54**: 17-24.
- Miller, S., & Roby, P. (1967) ‘Poverty: Changing Social Stratification’, in P. Townsend (ed.), *the Concept of Poverty*, London: Heinemann.
- Ministry of Finance. (2005). *Economic Survey*. Government of Nepal, Kathmandu.
- Ministry of Finance. (2012). *Economic Survey*. Government of Nepal, Kathmandu. [https://mof.gov.np/uploads/document/file/Economic%20 Survey%202011-12\\_20141224054554.pdf](https://mof.gov.np/uploads/document/file/Economic%20Survey%202011-12_20141224054554.pdf)
- Ministry of Finance. (2020). *Economic Survey 2019/20*. Government of Nepal, Kathmandu. [https://www.mof.gov.np/uploads/document/file/Economic%20Survey%202019\\_20201125024153.pdf](https://www.mof.gov.np/uploads/document/file/Economic%20Survey%202019_20201125024153.pdf)
- Ministry of Health. (2011). *Nepal Demographic Health Survey 2011*. Ministry of Health and Population, New ERA, and ICF International. [https://dhsprogram.com/pubs/pdf/fr257/ fr257 \[13april2012\].pdf](https://dhsprogram.com/pubs/pdf/fr257/fr257 [13april2012].pdf)
- Ministry of Health; New ERA; and ICF. (2017). *Nepal Demographic and Health Survey 2016*. Kathmandu, Nepal. <https://dhsprogram.com/pubs/pdf/FR336/FR336.pdf>
- Mittlbock, M. & M. Schemper (1996). Explained variation in logistic regression. *Statistics in Medicine* **15**: 1987-1997.

- Mohammed, M. B. (2017). Measurement and determinants of urban poverty in case of Southern Nations, Nationalities, and Peoples' Region (SNNPR), Ethiopia. *International Journal of Scientific and Research Publications*, 7(3): 181-189.
- Myftaraj, E., Zyka, E., & Bici, R. (2014). Identifying household level determinants of poverty in Albania using logistic regression model. *International Journal of Sustainable Development*, 7(3): 35-42. <http://www.ssrn.com/link/OIDA-Intl-Journal-Sustainable-Dev.htm>
- Nagelkerke, N.J.D. (1991). A note on a general definition of the coefficient of determination." *Biometrika* 78: 691-692.
- Nepal, U. N. F. P. A. (2017). Population situation analysis of Nepal. *Kathmandu: UNFPA Nepal*.
- Neter, J., Kutner, M.H., Nachtsheim, C.J. & Wasserman, W. (1996). *Applied Linear Statistical Models*. Fourth Edition. MCB MCGraw-Hill.
- Niemietz, K. P. (2011). A new understanding of poverty. *Institute of Economic Affairs Monographs, Forthcoming*.
- NPC. 1978. Employment, Income Distribution and Consumption Patterns in Nepal. National Planning Commission.
- NRB. (1988). Multipurpose Household Budget Survey. Kathmandu: Nepal Rastra Bank.
- Okojie, C. E. (2002). *Gender and education as determinants of household poverty in Nigeria* (No. 2002/37). Wider discussion paper.
- Olkin, I. (1998). Letter to the editor. *Evidence-Based Medicine* 3(71).
- Omoregbee, F. E., Ighoro, A., & Ejembi, S. A. (2013). Analysis of the effects of farmers characteristics on poverty status in Delta State. *International Journal of Humanities and Social Science Invention*, 2(5): 11-16.

- Onu, J. I., & Abayomi, Z. (2009). An analysis of poverty among households in Yola Metropolis of Adamawa State, Nigeria. *Journal of Social Sciences*, **20**(1): 43-48.
- Osowole, O. I., Ugbechie, R., & Uba, E. (2012). On the Identification of Core Determinants of Poverty: A Logistic Regression Approach. *Mathematical Theory and Modeling*, **2**(10): 45-53.
- Osowole, O. I., Ugbechie, R., & Uba, E. (2012). On the Identification of Core Determinants of Poverty: A Logistic Regression Approach. *Mathematical Theory and Modeling*, **2**(10): 45-53.
- Pant, D. (2017). *An analysis of the determinants of remittances and effect of remittance on expenditure behavior and child welfare in the households of Nepal* (Doctoral dissertation, University of Reading).
- Patel, S. P. (2012). Poverty incidence in Nepal by caste/ethnicity: Recent levels and trends. *Academic Voices: A Multidisciplinary Journal*, **2**: 59-62.
- Pregibon D. (1981). Logistic regression diagnostics. *The Annals of Statistics*, **9**(4): 705 – 724.
- Rahman Mustafa A. (2013). Household characteristics and Poverty: A Logistic Regression Analysis. *The Journal of developing Areas*, **47**.
- Ranganathan, P. Aggarwal, R., & Pramesh, C.S., (2015). Common pitfalls in statistical analysis: Odds versus risk. *Perspectives in Clinical Research*, **6**(4): 222-224. DOI: <https://doi.org/10.4103%2F2229-3485.167092>
- Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd ed. New York: Wiley.
- regression model when the estimated probabilities are small. *Biometrical Journal*, **30**: 911–924.
- Ravallion, M., & Chen, S. (1997). What can new survey data tell us about recent changes in distribution and poverty? *The World Bank Economic Review*, **11**(2): 357-382. <https://doi.org/10.1093/wber/11.2.357>
- Robbins, A. S., Chao, S.Y., & Fonseca, V.P. (2002). What's the relative risk? A method to directly estimate risk ratios in cohort studies of common outcomes. *Annals of Epidemiology*, **12**: 452–454. DOI: 10.1016/s1047-2797(01)00278-2

- Rowntree, Benjamin S., (1901). *Poverty: A Study of Town Life*, London: Macmillan.
- Rusnak, Z. (2012). Logistic regression model in poverty analysis. *Econometrics/Ekonometria*, **35**.
- Sackett, D. L., Deeks J.J, & Altman, DG (1996). Down with odds ratios! Evidence Based Medicine, **1**: 164–166. <https://ebm.bmj.com/content/ebmed/1/6/164.full.pdf> Accessed on 17<sup>th</sup> Dec., 2022.
- Sakuhuni, R. C., Chidoko, C., Dhoru, N. L., & Gwaindepi, C. (2011). Economic determinants of poverty in Zimbabwe. *International Journal of Economic Research*, **2**(6): 1-12.
- Salami, L. A., & Atiman, K. (2013). An Analytical Study of Determinants of Poverty Level among Households in Adamawa North District, Nigeria. *Mediterranean Journal of Social Sciences*, **4**(16): 73-73.
- Sanusi, R. A., Owagbemi, T. S., & Suleman, M. (2013). Determinants of poverty among farm households in Ikorodu local government area of Lagos state, Nigeria. *International Journal of Accounting and Finance Studies*, **4**: 538-552.
- Saurbrei, W., & Schumacher, M. (1992). A bootstrap resampling procedure for model building application to the Cox regression model. *Statistics in Medicine*, **11**: 2093-2109. DOI: <https://doi.org/10.1002/sim.4780111607>
- Schechtman, E. (2002). Odds ratio, relative risk, absolute risk reduction, and the number needed to treat—which of these should we use? *Value in health*, **5**(5): 431-436.
- Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**: 461–464.
- Schwendinger, F., Wagner, J., Infanger, D., Schmidt-Trucksäss, A., & Knaier, R. (2021). Methodological aspects for accelerometer-based assessment of physical activity in heart failure and health. *BMC Medical Research Methodology*, **21**, 1-12.



- Sekhampu, T. J. (2013). Determinants of poverty in a South African township. *Journal of Social Sciences*, **34**(2): 145-153.
- Sekwati, L., Narayana, N., & Raboloko, M. (2012). Understanding the nature of household poverty in Botswana. *PULA: Botswana Journal of African Studies*, **26**(1): 71-81.
- Sen, A. (1976). Poverty: an ordinal approach to measurement. *Econometrica* **44** (2): 219 – 231. URL: <http://www.jstor.org/stable/1912718>
- Shaga, H. H., Mega, T. L., & Senapathy, M. (2021). Determinants of rural household poverty: the case of Sodo Zuria Woreda, Wolaita Zone, Southern Ethiopia. *European Journal of Sustainable Development Research*, **5**(2). <https://doi.org/10.21601/ejosdr/10844>
- Shively, G., & Thapa, G. (2017). Markets, transportation infrastructure and food prices in Nepal. *American Journal of Agricultural Economics*, **99**: 660–682. <https://doi.org/10.1093/ajae/aaw086>
- Sikander, M. U., & Ahmed, M. (2008). Household determinants of poverty in punjab: a logistic regression analysis of MICS (2003-04) data set. In *8th Global Conference on Business & Economics*, 1-41.
- Skove, T., Deddens, J., Petersen, M. R., & Endahl, L. (1998). Prevalence proportion ratios: estimation and hypothesis testing. *International journal of epidemiology*, **27**(1): 91-95.
- Slesnick, D. T. 1993. Gaining Ground: Poverty in the Postwar United States. *The Journal of Political Economy*, **101**(8): 1 – 38.
- Spaho, A. (2014). Determinants of poverty in Albania. *Journal of Educational and Social Research*, **4**(2): 157-163. <https://doi.org/10.5901/jesr.2014.v4n2p157>
- Spicker, P. (1990). Charles Booth: the examination of poverty. *Social Policy & Administration*, **24**(1), 21-38. <https://doi.org/10.1111/j.1467-9515.1990.tb00322.x>

- Stewart, F., & Samman, E. (2013). Inequality and development: an overview. <https://www.researchgate.net/publication/299898763>
- Taylor, J. E., Gurkan, A. A., & Zezza, A. (2009). *Rural poverty and markets*. Agricultural development economics division, the Food and Agriculture Organization of the United Nations. <http://www.fao.org/3/a-ak424e.pdf>
- Teka, A. M., Woldu, G. T., & Fre, Z. (2019). Status and determinants of poverty and income inequality in pastoral and agro-pastoral communities: Household-based evidence from Afar Regional State, Ethiopia. *World Development Perspectives*, **15**: 100123 <https://doi.org/10.1016/j.wdp.2019.100123>
- Thapa, A. K., Dhungana, A. R., Tripathi, Y. R., & Aryal, B. (2013). Determinants of poverty in rural parts of Nepal: A study of Western Development Region. *Pinnacle Economics & Finance*, 1-6. [https://www.pjpub.org/pef/pef\\_105.pdf](https://www.pjpub.org/pef/pef_105.pdf)
- Thapa, S., & Acharya, S. (2017). Remittances and household expenditure in Nepal: Evidence from cross-section data. *Economies*, **5**(2). DOI: 10.3390/economies5020016
- Tuyen, T. Q. (2015). Socio-economic determinants of household income among ethnic minorities in the North-West Mountains, Vietnam. *Croatian Economic Survey*, **17**(1): 139-159.
- Uematsu, H., Shidiq, A. R., & Tiwari, S. (2016). Trends and drivers of poverty reduction in Nepal: A historical perspective. *World Bank Policy Research Working Paper* (No. 7830). <https://doi.org/10.1596/1813-9450-7830>
- United Nations, (2017). Basic Consumption and Income based Indicators of Economic Inequalities in Bosnia and Herzegovina: evidence from Household Budget Surveys. United Nations Economic Commission for Europe Conference of European Statistician. Working paper 13, 11 August 2017.
- UNDP. (2014). Nepal human development report 2014. *United Nations Development Programme: In Kathmandu*.

- Viera, A.J. (2008). Odds Ratios and Risk Ratios: What's the Difference and Why Does It Matter? *Southern Medical Journal*, **101**(7): 730-734. DOI: <https://doi.org/10.1097/smj.0b013e31817a7ee4>
- Wacholder, S. (1986). Binomial regression in GLIM: Estimating risk ratios and risk differences. *American Journal of Epidemiology*, **123**: 174–184. DOI: 10.1093/oxfordjournals.aje.a114212
- Wagle, B. K. (2014). An analysis of changing inter-group economic inequalities among different caste/ethnic groups in Nepal.
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*: **54**: 426-482.
- Walter, S. (1998). Letter to the editor. *Evidence-Based Medicine* **3**(71).
- Williamson, T., Eliasziw, M., & Fick, G. H. (2013). Log-binomial models: exploring failed convergence. *Emerging Themes in Epidemiology*, **10**(14). DOI: <https://doi.org/10.1186/1742-7622-10-14>
- World Bank Institute. (2005). *Introduction to poverty analysis*. Poverty Manual, All, J. H. Revision of August 8. World Bank.
- Khafaj, E., & Nurja, I. (2014). Determination of the key factors that influence poverty through econometric models. *European Scientific Journal*, **10**(24).
- Yelland, L. N., Salter, A. B., & Ryan, P. (2011). Relative risk estimation in randomized controlled trials: a comparison of methods for independent observations. *The international journal of biostatistics*, **7**(1). <https://doi.org/10.2202/1557-4679.1278>
- Yusuf, H. M., Daninga, P. D., & Xiaoyun, L. (2015). Determinants of Rural Poverty in Tanzania: Evidence from Mkinga District, Tanga Region. *Developing Country Studies*, **5**(6): 40-48.
- Yusuf, S. A., Adesanoye, A. O., & Awotide, D. O. (2008). Assessment of poverty among urban farmers in Ibadan Metropolis, Nigeria. *Journal of Human Ecology*, **24**(3): 201-207.

## APPENDIX

### APPENDIX- A

**Appendix A1:** Sample size of three NLSSs

1995/96	2003/04	2010/11
3388	5240	5988

**Appendix A2:** Estimated minimum calorie requirement/person/day in three NLSSs

1995/96	2003/04	2010/11
2124	2124	2220

**Appendix A3:** Estimated food and non-food poverty lines in NRS/person/year in three NLSSs

	1995/96 <sup>1</sup>	2003/04 <sup>1</sup>	2010/11 <sup>2</sup>
Food Poverty line	3114.1	3143.7	11929
Non-food poverty line	1540.5	1624.3	7332
Poverty line	4654.6	4768	19261

Source: <sup>1</sup>(CBS, 2005: 55) and <sup>2</sup>(CBS, 2010/11; 16)

**Appendix A4:** Growth in urban areas and population in three population censuses

	1991	2001	2011
Number of urban areas <sup>1</sup>	33	58	58
Population size <sup>1</sup>	1,695,719	3,227,879	4,523,820
Population per urban area	51385	55653	77997
Percentage change of population per urban area		8.3	40.1

Source: Population Monograph, 2014 Vol. III, p 109-110

## APPENDIX– B

### Appendix B1: Computer Program for Constructing CQGs

```
Sort cases by PCCE.  
CREATE CHS = csum(HS).  
Variable Labels CHS 'cumulated household size'.  
Execute.  
Compute PCHS = 100*CHS/28670.  
Variable labels PCHS 'Percentage of CHS'.  
Execute.  
if (PCHS <= 20) CQG =1.  
if (PCHS > 20 and PCHS <=40) CQG =2.  
if (PCHS > 40 and PCHS <=60) CQG =3.  
if (PCHS > 60 and PCHS <=80) CQG =4.  
if (PCHS > 80 ) CQG =5.  
Variable labels CQG 'Consumption Quintile Groups'.  
Value labels CQG 1 'Poorest' 2 'Second' 3 'Third' 4 'Fourth' 5 'Richest'.  
Execute.
```

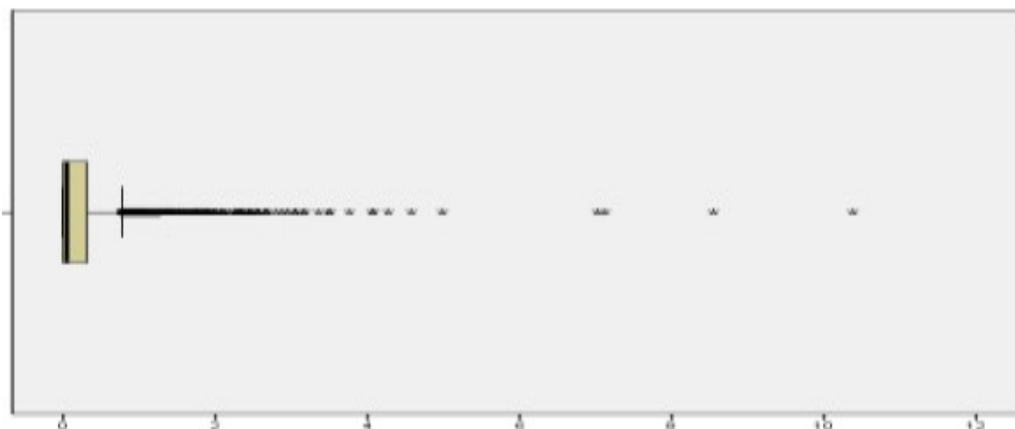
### Appendix B2: Number of individuals and household within each quintile

	Poorest	Second	Third	Fourth	Richest	Total
Number of Individuals	5733	5734	5732	5737	5734	28670
Number of Households	939	1051	1159	1301	1538	5988

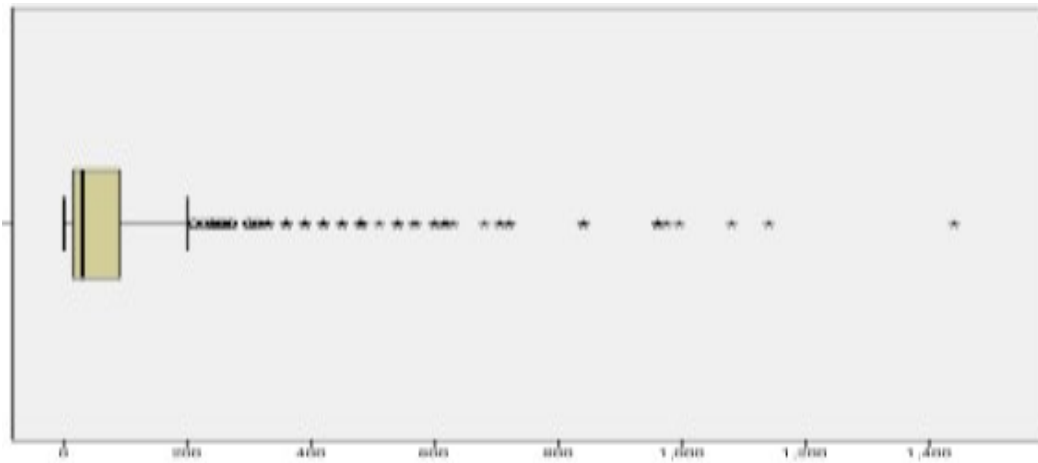
**Appendix B3:** Summary measures four quantitative variables and the detection of outliers

Quantitative variable	Mean	Standard deviation	3-sigma rule bounds		Min	Max	<u>Conclusion:</u> Outliers are detected in each variable
			Lower	Upper			
Area of land holding	0.26	0.49	-1.22	1.73	0	10.38	
Distance to the Nearest market center	80.63	124.58	-293.10	454.36	0	1440	
Number of children under 15	1.68	1.52	-2.87	6.23	0	11	
Number of literate WAP	1.8	1.45	-2.55	6.15	0	14	

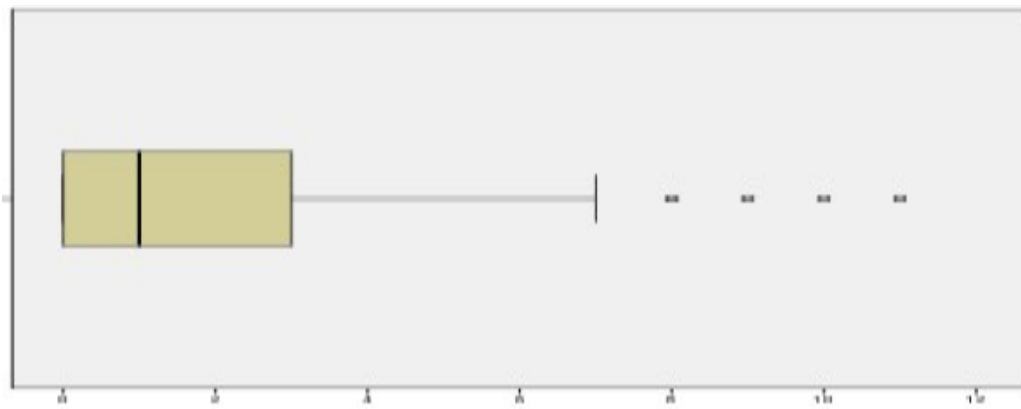
**Appendix B4:** Box-plots of four quantitative variables for the detection of outliers  
Area of land holding



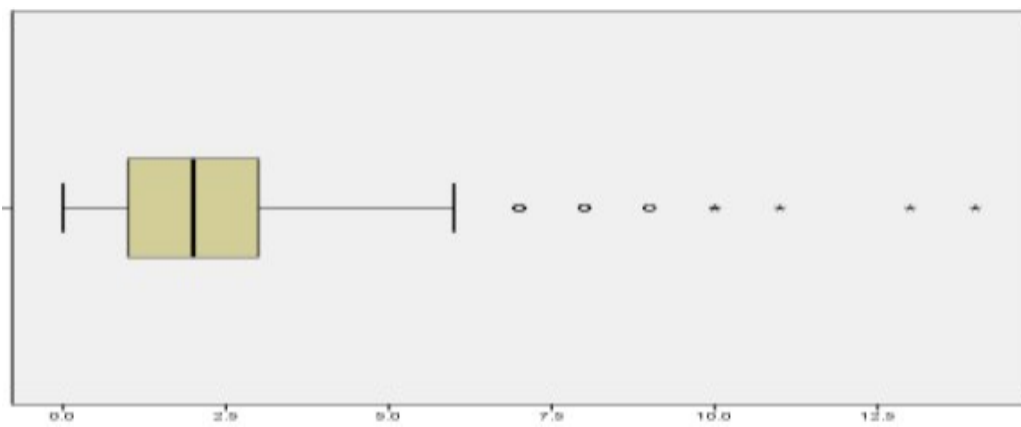
Distance to market center



Number of children



Number of literate WAP



**Conclusion:** Outliers are detected in each variable

**Appendix B5:** Number of outliers in each of four quantitative variables by Rules

Rule	Number of children under 15	Number of literate member of WAP	Area of land owned	Distance to market center
3-sigma rule	40	37	112	186
Box plot	21	37	586	583

**Conclusion:** A substantial number of outliers in each variable are detected which may raise questions when these variables are entered as covariates into the two proposed models.



## APPENDIX– C

### Published Papers

Acharya, K.P., Khanal, S.P., & Chhetry, D. (2022a). Factors Affecting Poverty in Nepal - A Binary Logistic Regression Model Study. *Pertanika Journal of Social Science and Humanities*, **30**(2). DOI: <https://doi.org/10.47836/pjssh.30.2.12>

Acharya, K. P., Khanal, S. P., & Chhetry, D. (2022b). Dichotomization of quantitative variables in poverty analysis. *BIBECHANA*, **19**(1-2): 142-149. DOI: <https://doi.org/10.3126/bibechana.v19i1-2.46407>

Acharya, K. P., Khanal, S. P., & Chhetry, D. (2022c). On the use of logistic regression model and its comparison with log-binomial regression model in the analysis of poverty data of Nepal. *Nep. J. Stat*, **6**, 63-79. DOI: [10.3126/njs.v6i01.50806](https://doi.org/10.3126/njs.v6i01.50806)

## **Factors Affecting Poverty in Nepal - A Binary Logistic Regression Model Study**

**Krishna Prasad Acharya, Shankar Prasad Khanal\* and Devendra Chhetry**

*Central Department of Statistics, Institute of Science and Technology, Tribhuvan University, 44618, Kirtipur, Nepal*

### **ABSTRACT**

One of the key factors in reducing monetary poverty is the identification of its determinants. Using a logistic regression model and considering household poverty status (poor/non-poor) as the response variable, this paper attempts to identify the most promising factors associated with monetary poverty based on nationally representative data of 5,988 households from the Nepal Living Standard Survey (2010/11). The goodness of fit, classification, discrimination, and diagnostics of the fitted model is performed. Six factors, namely illiteracy of household head (OR: 2.20; 95% CI: 1.86–2.61), households receiving no remittance (OR: 1.90; 95% CI: 1.64–2.20), households with no landholdings (OR: 1.53; 95% CI: 1.31–1.78), households with poor access to market centers (OR: 1.77; 95% CI: 1.52–2.07), households having more than two children under the age of 15 (OR: 4.69; 95% CI: 4.06–5.42) and households having no literate persons of working age (OR: 1.29; 95% CI: 1.07–1.56) are significantly associated with the likelihood of poverty. Male-headed households are not better positioned than female-headed households concerning poverty level. The developed regression model has satisfied the test of goodness of fit of the model

and reasonably satisfied the regression diagnostics through visual assessment. As several risk factors associated with poverty increase, the likelihood of a household being poor increases substantially. This analysis is expected to be helpful for the concerned authority to reframe the policy.

### ARTICLE INFO

#### *Article history:*

Received: 10 September 2021

Accepted: 13 April 2022

Published: 15 June 2022

DOI: <https://doi.org/10.47836/pjssh.30.2.12>

#### *E-mail addresses:*

[acharyakrishna20@gmail.com](mailto:acharyakrishna20@gmail.com) (Krishna Prasad Acharya)

[drshankarcds@gmail.com](mailto:drshankarcds@gmail.com) (Shankar Prasad Khanal)

[chhetrydevendra@gmail.com](mailto:chhetrydevendra@gmail.com) (Devendra Chhetry)

\*Corresponding author

*Keywords:* Covariate pattern, diagnostics, goodness of fit, logistic regression, Nepal, poverty

## INTRODUCTION

Poverty reduction in developing countries like Nepal is a central issue. One of the key factors in reducing monetary poverty, poverty conceptualized and measured in economic dimensions (in terms of income or consumption), is the identification of its determinants. Based on empirical studies in several countries, it can be inferred that poverty is partially determined by internal household characteristics and partially by external factors. Internal household characteristics include gender and education level of the household head, number of dependents, household size, place of residence, human capital, remittance, and area of landholdings. The effects of these characteristics on poverty have been researched by many scholars, including Abrar ul Haq et al. (2019), Teka et al. (2019), Imam et al. (2018), R. E. A. Khan et al. (2015), Spaho (2014), Leekoi et al. (2014), Thapa et al. (2013), Omoregbee et al. (2013), Osowole et al. (2012), Achia et al. (2010). In addition, external factors such as access to health care facilities (M. M. Khan et al., 2006; Peters et al., 2008), access to market centers (Obi et al., 2012), access to micro-credit (Chowdhury et al., 2005), access to infrastructure (John & Scott, 2002), economic growth (Adams, 2003), are also reported to be associated with poverty.

Nepal made remarkable progress in the reduction of monetary poverty in very unfavorable situations from 1996 to 2011, a period characterized by a decade long (1996–2006) violent, armed conflict

between the State and the Maoist. The conflict was formally ended by signing the Comprehensive Peace Agreement in November 2006 between the State and the then Communist Party of Nepal- Maoist. The prolonged political instability manifested by frequent changes in government, which lasted till a single political party came into power through a general election that took place under the Constitution of Nepal 2015. There was a sluggish economic growth of around 4.0% per annum (Ministry of Finance [MoF], 2013). However, the percentage of the population below the poverty line at the national level declined from 41.8 in 1996 to 30.9 in 2004 and further declined to 25.2 in 2011 (Central Bureau of Statistics [CBS], 2011a), which is still 4 in 1 person remained as poor due to several factors which are not yet known.

The main objective of this paper is to identify the most promising factors influencing household-level poverty using binary logistic regression on the nationally representative sample survey data of the Nepal Living Standard Survey III (NLSS-III) conducted by the Central Bureau of Statistics (CBS) in the fiscal year 2010/11, since to the best of our knowledge, no rigorous work on said data has yet to be done. Since 2010/11, the NLSS has not yet been conducted again for several reasons, such as the devastating twin earthquakes of 2015 and COVID-19. As a result, the NLSS-III conducted in 2010/11 is the latest estimate of poverty based on nationally representative survey data.

In order to identify the potential factors affecting poverty, a review of relevant literature is essential, and it is done in the next section.

## LITERATURE REVIEW

A brief review of the literature is made below to identify policy-driven factors affecting household-level poverty in Nepal.

Several drivers were responsible for the amazing progress in the reduction of poverty. The three main drivers identified by the World Bank are a drastic increase in personal remittances received from abroad, a rise in labor incomes, and an improvement in household demographics. These factors contributed to a 27, 52, and 15% reduction in poverty from 1996 to 2011 (Uematsu et al., 2016).

A large volume of Nepalese laborers migrated abroad for employment during the 1996–2011 period. As a result, the absent population reported in 2011 was 1,921,494, a big jump from the number of 762,181 reported in the census of 2001 (CBS, 2014). The outmigration brought many changes in Nepal's socio-economic and demographic sectors.

The two visible economic impacts of remittances are as follows. First, at the micro-level, the nominal average amount of remittance per recipient household in Nepali currency increased from 15,160 in 1996 to 80,436 in 2011 (CBS, 2011b). At the macro level, the percentage share of remittances in GDP increased from 1.8 in 1996 (MoF, 2005) to 18.5 in 2011 (MoF, 2012).

The average annual population growth rate had sharply declined from 2.25% during the census period of 1991–2001 to 1.35% during the census period of 2001 to 2011 (CBS, 2014); the total fertility rate had decreased from 4.6 births per woman in 1996 to 2.6 births per woman in 2011, (Ministry of Health, 2011); the percentage of female-headed households had increased from 13.6 in 1996 to 26.6 in 2011; the percentage of children under 15 had declined from 42.4 in 1996 to 36.7 in 2011 (CBS, 2011c).

Such demographic changes and many more others had several intertwined implications on the socio-economic life of millions of Nepali peoples. First, the outmigration of millions of literate youths had created a shortage of productive labor (or loss of human capital) within Nepal. The other positive and negative impacts of the outmigration of labor are discussed elsewhere (International Organization of Migration, 2019; Kunwar, 2015; Uematsu et al., 2016).

In addition to households directly benefitting from remittances sent by migrant members, non-migrant households also benefitted from the spillover effects of migration (Uematsu et al., 2016). As a result, household income increased by almost fivefold over a decade and a half: the nominal average household income in Nepali currency increased from 43,732 in 1996 to 202,374 in 2011 (CBS, 2011c).

Correlates of poverty are also reported in CBS (2005, 2011a). For example, the poverty rate increases with an increase

in household size, such as increasing the number of children. Conversely, the poverty rate decreases with an increase in the level of education of the household head. Households headed by someone working in the agricultural sector, self-employed persons, or wage workers are poorer than those headed by people in other sectors or professions.

The land has multidimensional roles: key factors in production, collateral in credit markets, security against natural disasters or shocks, and symbol of social, economic, and political prestige (Kousar et al., 2015). This statement also holds in the context of Nepal. Further, the computation based on the NLSS-III data showed that 28.8% of households have no land. The problems of the landless are discussed elsewhere (Wickeri, 2011).

Without good access to markets, a poor household cannot market its products, obtain inputs, sell labor, obtain credit, learn about, or adopt new technologies, insure against risks, obtain consumption goods at low prices, or use its scarce resources like land and labor efficiently (Taylor et al., 2009). For example, CBS (2011a) shows the link between poverty and access to facilities, including a market center in Nepal, where the percentage of poor living within 30 minutes of the market center is 16.3, while the remaining 83.7% live beyond 30 minutes of a market center.

Using multinomial logit regression on 962 household-level panel data between NLSS-I and NLSS-II, Bhatta and Sharma (2006) identified factors affecting chronic

and transient poor households under three scenarios. The relative risk ratio (base category non-poor [= 0]) of each of the two factors—household size and % of individuals under 15 or over 59 years of age—was significantly greater than 1 for the chronic poor. On the other hand, the relative risk ratio for a percentage of the household adults who can read and write and the value of livestock owned each was significantly less than 1 for the chronic poor.

Thapa et al. (2013), using a binary logistic regression model on data obtained from 279 households from six districts of western Nepal, reported that the literacy of the household head, family size, family occupation, size of landholding, females' involvement in service, occupation of household head and social involvement was significantly associated with the rural poverty.

R. E. A. Khan et al. (2015) studied the factors affecting rural household poverty in one district of Pakistan based on 600 households' data. The probability of poverty decreases considerably in households with members having only an agricultural occupation, households with higher socio-economic empowerment indexes, and remittance-receiving households. In contrast, the probability of poverty increases significantly with an increase in the female to male ratio and the number of household members.

Abrar ul Haq et al. (2018a) assessed the role of household empowerment (developed by Abar ul Haq in his Ph. D. dissertation) in alleviating participatory poverty of

600 rural households in Pakistan. Their assessment suggested that participatory poverty can be reduced by improving household empowerment in the studied area. Abrar ul Haq et al. (2018b) provided a detailed framework for measuring the household empowerment index (HEMI) and measured the index using the data of 42 variables collected from 600 rural households in Pakistan. Abrar ul Haq et al. (2019) found that household empowerment has a significant positive impact on monetary poverty in the studied area. This series of studies open a new window in poverty analysis in a developing country like Nepal and the monitoring and spatial comparison of household empowerment. In the present study, the 42 variables selected in constructing HEMI are useful in justifying the reason for the selected covariates in our study.

After an extensive literature review, seven factors were tentatively identified, and the rationale for their selection in the context of Nepal is elaborated in the next section.

### **Selection of Factors**

The factors selected in this paper are directly or indirectly related to some of the items Abrar ul Haq et al. (2019) used to develop the household empowerment index (HEMI). For example, the two items 'status of landholding' and 'sex of household head' selected in this study correspond to the variables 'land owned' and 'gender of household head' selected in the development of HEMI. The other three factors 'literacy status of household head,' 'number of

literate members of working age' and 'number of children under 15' selected in this study are modified versions of the items 'education of household head,' 'average education of the household' and 'size of the house' selected in the development of HEMI. These modifications are necessary due to the unavailability of data and need in the context of Nepal, as described below.

The NLSS-III data showed that the average number of children under 15 among poor households is almost two times higher than among non-poor households (2.81 versus 1.43). Likewise, the average working age population (15–64) among poor households is slightly higher than among non-poor households (2.95 versus 2.79). On the contrary, the average number of elders (65+) among poor and non-poor households in the same (0.24). These results indicate that instead of investigating the effect of household size on household poverty, it is more realistic from a policy perspective to investigate the effect of 'number of children' and 'number of literate working age members (or human capital)' separately. Investigating the effect of human capital on poverty is essential since a huge number of skilled or semiskilled individuals have out-migrated. Likewise, investigating the effect of children on poverty is essential since it is a perennial problem in Nepal.

Considering the contribution of remittance to Nepal's GDP and the source of income of most households in Nepal, the factor, 'status of remittance recipient,' has been included in this paper. Moreover, many scholars in contemporary studies have

included it as a covariate; for example, see Abrar ul Haq et al. (2018b), R. E. A. Khan et al. (2015).

Considering over 50% of Nepal's population were reported to dwell beyond a 30 minutes reach of the nearest market centers (CBS, 2011a), and realizing the direct/indirect role of market centers (Joshi & Joshi, 2016; Shively & Thapa, 2017; Taylor et al., 2009) in reducing poverty, the factor 'access to nearest market' has been included in this paper.

In summary, based on the extensive review of the literature and empirical evidence, the present study identified seven factors, each of which is related to two pillars—economic empowerment and social empowerment—of household empowerment, formulating the hypothesis that each of these factors will have a significant effect in reducing poverty in Nepal.

The source of data, the process of dichotomization of four tentatively identified quantitative factors, the appropriate statistical model with its goodness-of-fit test, the diagnostic criteria of the fitted model, and the risk assessments of the identified factors are discussed in the next section.

## METHODS

The main data source for this study is NLSS-III which provides household-level data on several variables of 5,988 households and individual-level data on several variables of 28,670 individuals. The available data on the variable "household poverty status" (poor/non-poor) was taken as the response variable

by assigning code values 1 for poor and 0 for non-poor. In this study, a household is defined as *poor (non-poor) if the per capita expenditure of the household members falls below (above) the poverty line of Nepali currency, 19,261*. The unweighted and weighted proportions of poor households were correspondingly 18.5% and 20.0%.

The available data on three household level dichotomous variables—sex (male/female) and literacy status (literate/illiterate) of household head and the remittance-receiving status (yes/no)—were used as one set of covariates in this study. Also, the available household level numeric data on two variables—area of landholding measured in hectares and access to the nearest market center measured in walking distance time in minutes to reach the nearest market was also used as covariates after converting them into dichotomous variables.

The available data on the variables "age" and "literacy status" of individuals were used to construct the two household-level numeric variables—the number of children under 15 and the number of literate members of working age (15–64 years) within each household. These two numeric variables were also used as covariates after converting them into dichotomous variables. The main reason for dichotomizing each of the four numeric variables is to make a meaningful comparison between the two mutually exclusive and exhaustive households, namely the *disadvantaged* and *advantaged groups*. The process of dichotomizing, particularly choosing the demarcating value for each of the four quantitative variables, is described below.



**Households Dichotomized by Area of Land Holding**

Considering the importance of land possession in households in Nepal, the demarcating value for the area of landholdings (numeric variable) was chosen to be 0, which demarcates households into two groups—one group of households in which each had no land (disadvantaged group) and the other group of households in which each had land (advantaged group).

**Households Dichotomized by Access to Nearest Market**

Realizing the importance of access to markets in poverty reduction, the demarcating value of this numeric variable was chosen to be 30 minutes of walking distance, which demarcates the households into two groups—one group of households in which each was beyond 30 minutes reach of the nearest market (disadvantaged group) and the other group of households in which each was within 30 minutes reach of the nearest market (advantaged).

**Households Dichotomized by the Number of Children Under 15**

Children under 15 are considered dependents. Therefore, even if families desire to have multiple children, many children in a household create an economic burden that aggravates household poverty. In order to determine the demarcating value for dichotomizing households by the number of children under 15, a little exercise was carried out. The results (Table 1) show that for each group of households with less than or equal to two children, the poverty incidence falls below the national level of 25.2%. On the contrary, the poverty incidence exceeds the national level for each group of households with more than two children. Therefore, the demarcating value was chosen as two, which demarcates the households into two groups—one group of households in which each had more than two children (disadvantaged group) and the other group of households in which each had less than or equal to two children (advantaged group). The poverty incidence of the former group is estimated at 41.4%, and for the latter group is estimated at 13.5%.

Table 1  
*The rationale for choosing two children as demarcating value*

Group of households with several children	0	1	2	3	4	5+
Within-group incidence of poverty (%)	5.9	11.6	19.6	33.5	42.3	55.7

Source: Computed from data of NLSS III



### Households Dichotomized by the Number of Literate Members of Working-Age

In order to investigate the impact of the loss of human capital due to outmigration on household poverty, the household level numeric variable “number of literate members of working age” was selected. They are converted into a dichotomous variable by grouping the households into two groups: no literate members of working age (disadvantaged group) and at least one literate member of working age (advantaged group). The rationale behind choosing a demarcating value of 0 is as follows: *a household with no literate member of working age is in a more difficult position than a household with at least one literate member of working age in fighting against poverty.*

### The Statistical Model and its Goodness of Fit

The seven household level dichotomous variables, namely sex of household head (female vs. male), literacy status of household head (illiterate vs. literate), remittance-receiving status (no vs. yes), market access (poor vs. better), landholding status (no vs. yes), number of children (more than two vs. at most two), number of literate members of working age (none vs. at least one) were identified as potential covariates in this study. The Chi-square test of independence assessed the association of each potential covariate with the response variable. The binary logistic regression analysis included only the covariates significantly associated (at a 5% significance level) with the response variable. The usual binary logistic regression model with a p-number of covariates (yet to be determined) is expressed below. The model is estimated with the aid of a statistical software package.

$$\ln(\text{odds}(\pi(x))) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p \tag{1}$$

The model adequacy was assessed by Pseudo R<sup>2</sup> proposed by McFadden (1974), Omnibus test, and Wald  $\chi^2$  test. The goodness-of-fit test was carried out by the Hosmer and Lemeshow (H-L)  $\chi^2$  test.

### Classification, Discrimination, and Diagnostics of the Model

The classification of the fitted binary

logistic regression model was examined by sensitivity, specificity, and accuracy. Furthermore, the ability of the fitted model to discriminate between the poor and non-poor was assessed through the area under the Receiver Operating Characteristics (ROC) curve.

Among the different diagnostics approaches reported in the statistics

literature, mainly two scatter plots were used for the fitted logistic regression model. Firstly, as an influential statistic suggested by Pregibon (1981), the delta beta statistic ( $\Delta\hat{\beta}$ ) was computed, which measures the changes in estimated regression coefficients for each covariate pattern if we were to exclude that pattern, where ( $\Delta$ ) stands for the difference. A scatter plot was prepared, keeping the values of ( $\Delta\hat{\beta}$ ) in the vertical axis and predicted probabilities based on the fitted logistic regression model on the horizontal axis to identify the large influence on the estimated coefficients. Secondly, the delta Chi-square ( $\Delta\chi^2$ ) based on Pearson's residuals was computed, which measures the effects of patterns on the model's fit in general. A scatter plot keeping delta Chi-square in the vertical axis and predicted probability in the horizontal axis to examine the influence of pattern on overall fit with symbol size proportional to delta beta was also prepared. Besides these two, the model specification test was attempted to examine whether the fitted model needs independent covariates or not by regressing the original response variable on the model predicted variable ( $\hat{y}$ ) and ( $\hat{y}^2$ ) with the null hypothesis that there is no specification of error at a 5% level of significance.

### **Risk Assessment based on Presence of Factors**

Finally, after fitting the model and assessing the model diagnostics, the risk assessment of the factors by quantifying their effects presented in the model was attempted by regressing the same response variable used

in the finally developed model with the newly generated indicator variable, ( $x_i$ , for  $i = 0, 1, 2, 3, \dots$ ), where 0 stands for no factors present, and 1,2,3,...,p stand for the presence of any one or two factors, and finally all factors in the final model respectively. Finally, statistical analysis was performed by using statistical software IBM SPSS version 20 and STATA 13 Stata Corp LP, College Station, Texas, USA.

The empirical results regarding the screening of the tentatively identified factors, the estimated binary logistic regression model with discussion, the classification and discriminating power of the fitted model, the diagnostic outcomes of the fitted model, and the risk assessments of the finally selected factors are discussed in the next section.

## **RESULTS AND DISCUSSION**

The following sub-sections deal with the results and discussions of the association of covariates with the response variable, fitted binary logistic regression model, classification and discrimination, diagnostics of the fitted model, and risk assessment.

### **Association of Covariates with the Response Variable**

Descriptions of the seven covariates, such as their categories, coding schemes, distributions of households over two categories of each proposed covariate, an association of each proposed covariate with the response variable assessed by the Chi-square test, and the effect size of each Chi-square test measured by the phi-coefficient are presented in Table 2.

Table 2

*Association of covariates with the response variable*

Description of household-level dichotomous covariates	Percentage distribution of households	Association of covariates with poverty			Phi-coefficient
		% of poor households within a category	Chi-square value	p-value	
Sex of household head:					
Male (0)	73.3	18.9	1.7	.193	-0.02
Female (1)	26.7	17.4			
Literacy status of household head:					
Literate (0)	60.2	12.2	240.7	<.001	0.20
Illiterate (1)	39.8	28.1			
Status of remittance recipient:					
Yes (0)	53.1	15.7	35.7	<.001	0.08
No (1)	46.9	21.7			
Status of land holdings:					
Yes (0)	71.2	15.1	114.9	<.001	0.14
No (1)	28.8	27.0			
Access to nearest market:					
Better (0)	52.0	11.6	206.7	<.001	0.19
Poor (1)	48.0	26.0			
Number of children under 15:					
At most two (0)	73.8	10.9	653.0	<.001	0.33
More than two (1)	26.2	40.1			
Number of literate members of working age:					
At least one (0)	80.7	15.6	142.0	<.001	0.15
None (1)	19.3	30.8			

*Note.* Figures within parentheses are binary codes; Sample size (n) = 5,988. *Source:* Computed from data of NLSS-III

All covariates except the sex of the household head are significantly associated with poverty. Male-headed households were not better positioned than female-headed households concerning poverty level. This finding contradicts the findings of other studies (Kona et al., 2018; Omoregbee et al., 2013). Nonetheless, our finding is analogous to the findings reported by some studies (Bhatta & Sharma, 2006; Edoumiekumo et al., 2014; Spaho, 2014). In order to explore this issue, a chi-square test of independence was also performed to determine whether there is an association between the sex of the household head and the status of the remittance receiver. A significant association was found ( $\chi^2(1) = 491.5, p < .001$ ). Among the female-headed households, 76.8% were remittance receivers, while only 44.4% were remittance receivers among the male-headed households. This result partially explains why male-headed households were not in a better position than female-headed households regarding the poverty measurement.

Among the significantly associated covariates, the effect size of remittance is the smallest, and the number of children is the highest. Therefore, the smallest effect size of remittance indicates that remittance alone is not responsible for reducing poverty, which is consistent with the result of the World Bank (Uematsu et al., 2016).

The effect size of the number of children being the highest is due to several socio-demographic factors, including the varying fertility levels among different social groups of women educationally

disadvantaged groups of women, since the adult literacy rate of women is 44.5% (CBS, 2011c). In the context of Nepal, the level of fertility is inversely related to women's educational attainment, decreasing rapidly from 3.7 births among women with no education to 1.7 births among women with a School Leaving Certificate (SLC) or above (Ministry of Health, 2011). As a result, it will take more years to see the benefits of improvement in household demographics.

### Results of Binary Logistic Regression

The six significant covariates obtained from the previous analysis are candidates for the binary logistic regression model. The estimated binary logistic regression model results are presented in Table 3. The estimated model is statistically significant, as shown by the omnibus Chi-square test ( $\chi^2(6) = 938.97, p < .001$ ). In addition, each beta coefficient is significant at a level  $< 0.001$ .

The regression model is fitted well as assessed by Hosmer-Lemeshow Chi-square test ( $\chi^2(8) = 7.24, p = .51$ ). A little exercise shows no severe problem of multicollinearity assessed through Variance Inflation Factor (VIF) as it varies from 1.01 to 1.47. Sixteen percent of the variation of the outcome variable (McFadden pseudo  $R^2 = 0.16$ ) has been explained by the variations of independent covariates in terms of log-likelihood.

Table 3  
*Results of estimated binary logistic regression model*

Characteristics	Beta	OR	S.E.	P-value	95% C.I. for OR
Literacy status of household head:					
Literate	0.79	1.00	0.09	<.001	(1.86, 2.61)
Illiterate		2.20			
Status of remittance recipient:					
Yes	0.64	1.00	0.08	<.001	(1.64, 2.20)
No		1.90			
Status of land holdings:					
Yes		1.00			(1.31, 1.78)
No	0.43	1.53	0.08	<.001	
Access to nearest market:					
Better		1.00			(1.52, 2.07)
Poor	0.57	1.77	0.08	<.001	
Number of children under 15:					
At most two	1.55	1.00	0.07	<.001	(4.06, 5.42)
More than two		4.69			
Number of literate members of working age:					
At least one	0.25	1.00	0.10	<.001	(1.07, 1.56)
None		1.29			
Constant	-3.27	0.04	0.09	<.001	

Source: Computed from data of NLSS III

The sign of each regression coefficient is positive, which indicates that each disadvantaged group identified in this study is more likely to be poorer than the corresponding advantaged group. This fact is elaborated on below.

The head of the household in Nepal is considered the household leader and is responsible for the entire household resource management. If the household head is illiterate, he/she is likely to get a low-paying job, have less bargaining power, and

not be engaged in other economic activities. Consequently, the household income will be less, and the households' poverty level will be increased. In our study, the households headed by illiterate heads are 2.2 times more likely to be poorer than those headed by literate heads (OR: 2.20; 95% CI: 1.86 – 2.61), keeping the effects of all other covariates fixed. Our finding is supported by the findings of Teka et al. (2019), Imam et al. (2018), and Botha (2010).

The households not receiving remittance are 1.9 times more likely to be poorer than those receiving remittance (OR: 1.90; 95% CI: 1.64–2.20), keeping the effects of all other covariates fixed. Similar findings were found in the study carried out in Pakistan. Majeed and Malik (2015) reported that the risk of poor households was 43% less (OR = 0.57) among remittance-receiving households compared to households receiving no remittance. The findings of our study also aligned with the findings of Abrar ul Haq et al. (2019) and R. E. A. Khan et al. (2015). In this study, the remittance

association with each remaining covariate is examined using the Chi-square test, and the results are presented in Table 4. Interestingly, the percentage of households receiving remittance is significantly higher among the five disadvantaged groups than their corresponding counterparts, except for the group of households having more than two children. Despite this fact, the odds ratio for the likelihood of households being poor among the disadvantaged groups continues to be greater than one compared to their counterparts.

Table 4  
*Role of remittance*

		% of households receiving remittance	Chi-square value	p-value
Literacy status of household head	Literate	49.0	59.8	< .001
	Illiterate	59.2		
Status of land holdings	Yes	49.8	62.8	< .001
	No	61.1		
Access to the nearest market	Better	49.3	37.5	< .001
	Poor	57.2		
Number of children under 15	At most 2	53.2	0.1	.739
	More than 2	52.7		
Number of literate members of working age	At least one	51.2	34.5	< .001
	None	60.8		

Source: Computed from data of NLSS III

In order to escape from rural poverty, in the context of Nepal, the availability and access to different resources such as job opportunities, availability of land, and access to loans are very important. A person

having (not having) land is directly related to social prestige. A household not having a single piece of land generally has very limited access to getting loans, starting businesses, and getting land on rent, which

brings constraints on the economic activities of such households, and ultimately the household poverty level increases. Our study has indicated that households with no land are 1.5 times more likely to be poorer than those with land (OR: 1.53; 95% CI: 1.31–1.78), keeping the effects of all other covariates fixed. Other studies corroborate this finding (Farah, 2015; Imam et al., 2018; Kousar et al., 2015).

In rural parts of Nepal, if the market center is far away and roads and feeder roads are not developed, it is very difficult for farmers and smallholders to sell their products and have access to credit. Postharvest food loss due to lack of cold storage centers and inadequate infrastructure significantly affects household poverty (Shively & Thapa, 2017). Our estimates have shown that the households with poor access to the nearest market are 1.8 times more likely to be poorer than households with better market access (OR: 1.77; 95% CI: 1.52 – 2.07), keeping the effects of all other covariates fixed. This finding is similar to the finding reported by Mamo and Abiso (2018).

Children are dependents, and households with more children require more income for education, health, food, and clothing. Because of this, the household poverty level will increase. Regarding this issue, our study has identified that households with more than two children are 4.7 times more likely to be poorer than households with less than or equal to two children (OR: 4.69; 95% CI: 4.06–5.42), keeping the effects of all other covariates constant. This finding is similar

to the findings of Myftaraj et al. (2014), who indicated that households that had two children decreased the possibility of being poor by 20% (OR = 0.8) but increasing one more dependent child increased the risk of becoming poor (OR = 1.03) for three children.

Supposed all members of working age in a household are illiterate. In that case, they are likely to get fewer opportunities for good jobs, be less aware of the opportunities provided by the government and market demand and be less familiar with the latest information and technology; consequently, they lag in social and economic activities. In this context, our study has found that households having no literate members of working age are 1.3 times more likely to be poorer than those with at least one literate member of working age (OR: 1.29; 95% CI: 1.07–1.56) keeping the effects of all other covariates constant. A comparable result was reported by Mamo and Abiso (2018) in rural residencies of Ethiopia (OR =1.4). Omoregbee et al. (2013) also found that the odds of less-educated farmers were 1.3 times more likely to be poorer than more educated farmers in Nigeria. Another study conducted in Pakistan concluded that an increase of one educated earner of any level in the household significantly reduces the risk of the household being poor by 11% (OR = 0.89) compared to the households having uneducated earners (Majeed & Malik, 2015).

### **Results of Classification and Discrimination of the Model**

The sensitivity, specificity, and correct

model classification values are presented in Table 5 for two cutoff points, 0.5 and 0.16. The later cutoff point, 0.16, was identified by plotting the sensitivity/specificity in the

vertical axis against various probability cutoffs in the horizontal axis, as presented in Figure 1.

Table 5

*Sensitivity, specificity, and correct classification value*

Cutoff	Sensitivity	Specificity	Correct classification
0.50	20.80%	97.00%	82.50%
0.16	74.12%	65.57%	67.15%

Source: Computed from data of NLSS III

The percentage of poor cases correctly predicted by the model is 20.80 when the cutoff point is 0.50, whereas it is 74.12 when the cutoff point is 0.16. The overall correct

classification of the model considering a cutoff value of 0.50 is 82.50%, and it reduces to 67.15% when considering a cutoff value of 0.16.

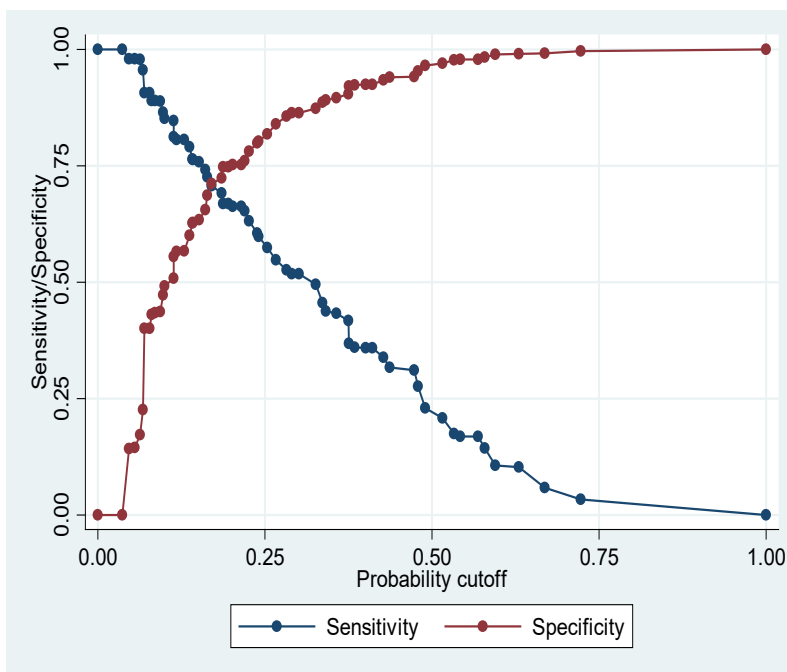


Figure 1. Plot of the sensitivity/specificity against the predicted probability



The ROC curve in Figure 2 shows discrimination of the developed model that the area under the curve (AUC) is 0.78, which can be considered acceptable (Hosmer & Lemeshow, 2000).

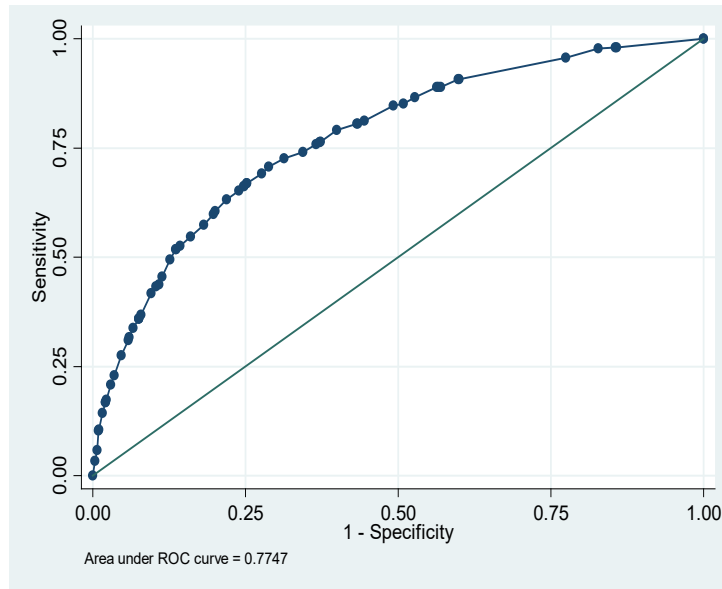


Figure 2. Plot of sensitivity versus 1- specificity

### Diagnostics of the Fitted Model

In order to assess the diagnostics of the model, two plots are used. The plot of delta beta ( $\Delta\beta$ ) versus estimated probability and the plot of delta chi-square ( $\Delta\chi^2$ ) versus estimated probability with a symbol size proportional to delta beta ( $\Delta\beta$ ) and the model specification test results are presented below.

### Plot of Delta Beta ( $\Delta\beta$ ) versus Estimated Probability

The influential statistic ( $\Delta\beta$ ) was plotted with estimated probability based on the fitted logistic regression model with 60 covariate patterns, as shown in Figure 3.

It can be seen clearly that only two data points are falling somewhat far away from the rest of the data. In the scatter plot of delta beta and the estimated probability, if the values of delta beta are greater than 1, there is an indication for an individual covariate pattern to influence the estimated regression coefficients (Hosmer & Lemeshow, 2000). Hence, this curve has indicated that overall, there is not much influence of the individual covariate pattern on the estimated regression coefficients except for two covariate patterns based on visual assessment.

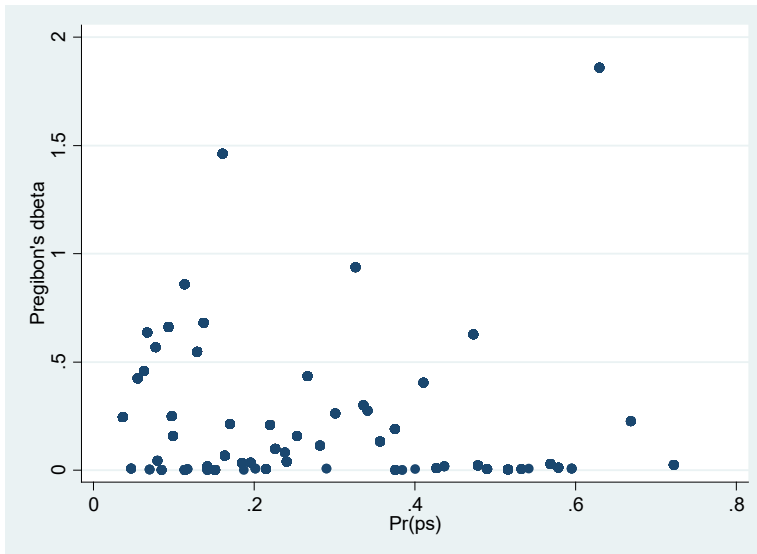


Figure 3. Plot of Pregibon's dbeta ( $\Delta\beta$ ) versus estimated probability

**Plot of Delta Chi-square ( $\Delta x^2$ ) versus Estimated Probability with Symbol Size Proportional to Delta Beta ( $\Delta\beta$ )**

A scatter diagram of ( $\Delta x^2$ ) versus estimated probability based on the fitted logistic

regression model with the size of the symbol proportional to ( $\Delta\beta$ ) is presented in Figure 4. This measure is used to assess the influence of pattern on the overall fit with symbol size proportional to delta beta.

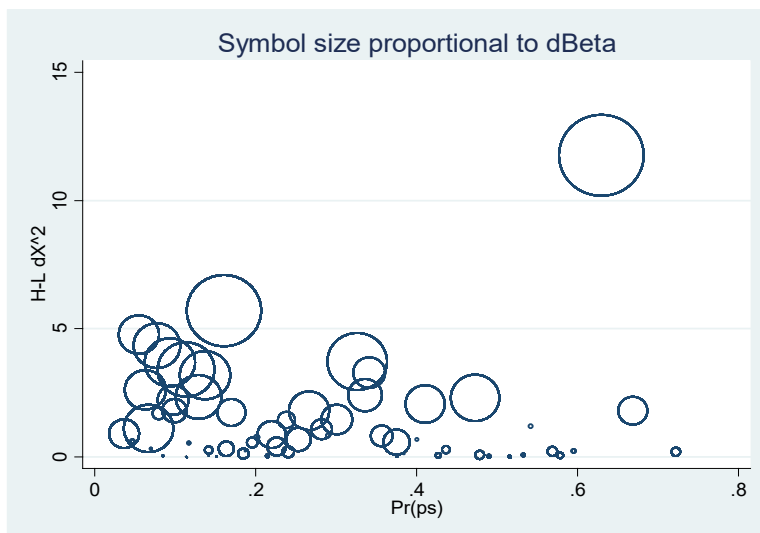


Figure 4. Plot of H-Ldx<sup>2</sup> ( $\Delta x^2$ ) versus estimated probability with symbol size proportional to ( $\Delta\beta$ )

It can be observed clearly in Figure 4 that a few extremely large circles differently appearing are noted in the plot, and for all these circles except one, the value of ( $\Delta\chi^2$ ) is small. It indicates an influence of the individual covariate pattern on the delta chi-square and the regression coefficients but only for one covariate pattern.

Both figures (3 and 4) show very few (one or two) covariates outlying patterns. Further, the value of ( $\Delta\chi^2$ ) is not much higher, and only two covariate patterns have a ( $\Delta\beta$ ) value of more than 1. So, it can be concluded that the overall fit of the

developed model based on the considerable data size is not violated in diagnostic prospects.

**Model Specification**

In order to assess whether the final fitted model may need other independent covariates or not, a new regression model was run considering the model predicted value ( $\hat{y}$ ) and the square of the predicted value ( $\hat{y}^2$ ) as the independent variable with the original outcome variable. The results are presented in Table 6.

Table 6  
*Model predicted value and the square of the predicted value*

	Coefficient	S. E.	Z	p-value	95% C. I.
$\hat{y}$	0.97	0.09	11.26	<.001	(0.80, 1.14)
$\hat{y}^2$	-0.01	0.03	-0.39	.696	(-0.08, 0.05)
Constant	-0.01	0.06	-0.10	.923	(-0.12, 0.11)

Source: Computed from data of NLSS III

The non-significant result of the regression coefficient of  $\hat{y}^2$  indicates that the model is correctly specified.

**Risk Assessment based on Factors Present in the Model**

The risk of a household being poor (in terms of odds ratio) was computed based on several factors identified in the model, shown in Figure 5.

The risk of poor households increases continuously as the number of factors increases. The risk of poor households is

six times more for households even only presenting any two factors than households not presenting any factor (reference category). This risk is likely to increase by ten times for households presenting any three factors. The conclusions and recommendations based on the empirical results obtained are presented in the next section.

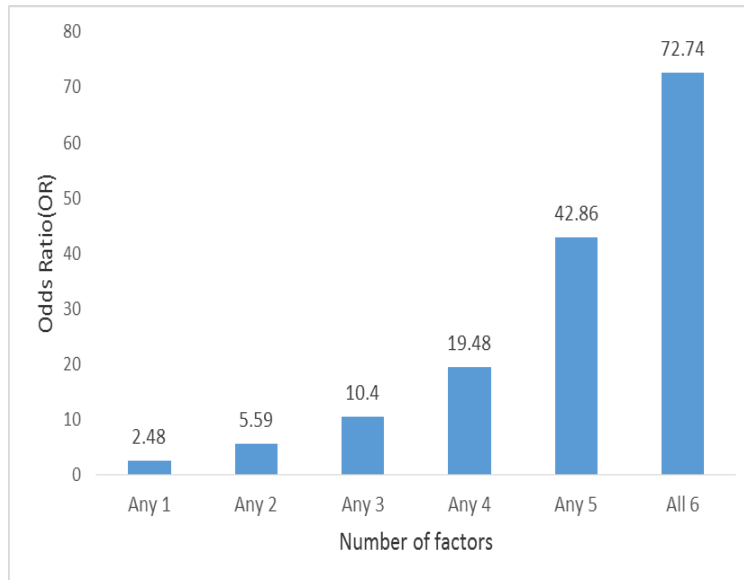


Figure 5. Risk of the household being poor in the presence of several factors

## CONCLUSIONS AND RECOMMENDATIONS

This study identified six factors affecting household level poverty by developing a binary logistic regression model on nationally representative sample data of Nepal. The developed logistic regression model with these six covariates has satisfied the test of goodness of fit of the model and reasonably satisfied the regression diagnostics.

The identified factors are related to a broader construct of socio-economic empowerment of households. Moreover, the selected factors being household-level and policy-driven, the concerned authorities can easily implement poverty alleviation programs. Therefore, it can be considered a practical contribution of this study.

The study concludes that even a single literate member of working age in

household assists in reducing poverty as much as having a literate household head. It is an indication that many households are suffering from the problem of human capital shortage. Therefore, policies and poverty alleviation programs are to be directed toward building human capital, particularly in those households with inadequate human capital.

It can also be deduced that remittance is an important factor in reducing poverty. The household income increases as the number of remittance recipients increases and reduces poverty. Therefore, the government of Nepal must create a conducive environment where remittance recipients can utilize their money, and foreign-employment returnees can employ their skills in productive areas.

The results further infer that more than two children in a household aggravates household poverty. If the children of poor

households are not given a proper education, then those households may get into the vicious cycle of poverty characterized by an intergeneration poverty cycle. Therefore, the government of Nepal must invest in providing proper education to children of poor households, particularly focusing on those households having more than two children.

In addition, the study identifies a household being landless as a factor that increases household poverty. Therefore, the government of Nepal must address the problems of landless households, either through official government documents or other reliable sources, formulate policies and prepare programs for reducing their problems. We anticipate that these measures will reduce the poverty of landless households.

The results also indicate that poor access to the nearest market center increases the likelihood of household poverty. Therefore, the government of Nepal needs to take the initiative to improve access to markets by developing infrastructure such as road networks, transport networks, cold storage, and electricity, particularly in the rural areas of the country. These measures will increase the connectivity between rural and urban areas and eventually reduce poverty.

This study might have missed incorporating some internal household characteristics (such as the occupation of the household head) and external factors (such as distance to health center) associated with poverty. Future research can be planned with the upcoming NLSS IV data, incorporating

other relevant variables. Different composite indices such as the household empowerment index may also be incorporated. The subgroup analysis for different provinces may also be attempted within the same statistical analysis framework based on these indices. Moreover, new studies can also be recommended to capture other community variables associated with poverty and the variables identified in this study in a wider domain using advanced statistical modeling such as multilevel modeling.

## ACKNOWLEDGMENT

The authors want to thank the Central Bureau of Statistics (CBS) for providing NLSS III 2010/11 data and the department research committee of the Central Department of Statistics, Tribhuvan University (TU), Nepal, for their comments and suggestions for this study. In addition, the University Grant Commission (UGC) of Nepal is also acknowledged for providing Ph.D. fellowship as this is part of Ph.D. research work.

The authors would also like to acknowledge anonymous reviewers' critical comments and suggestions. Finally, we would like to thank Dr. Anirudra Thapa, Professor of Central Department of English, TU, Nepal, for editing the English language version of the manuscript.

## REFERENCES

- Abrar ul haq, M., Jali, M. R. M., & Islam, G. M. N. (2018a). Assessment of the role of household empowerment in alleviating participatory poverty among rural household of Pakistan. *Quality*

- & *Quantity*, 52(6), 2795-2814. <https://doi.org/10.1007/s11135-018-0710-0>
- Abrar ul haq, M., Jali, M. R. M., & Islam, G. M. N. (2018b). The development of household empowerment index among rural household of Pakistan. *Pertanika Journal of Social Sciences & Humanities*, 26(2), 787-809.
- Abrar ul haq, M., Jali, M. R. M., & Islam, G. M. N. (2019). Household empowerment as the key to eradicate poverty incidence. *Asian Social Work and Policy Review*, 13(1), 4-24. <https://doi.org/10.1111/aswp.12152>
- Achia, T. N., Wangombe, A., & Khadioli, N. (2010). A logistic regression model to identify key determinants of poverty using demographic and health survey data. *European Journal of Social Sciences*, 13(1), 38-45.
- Adams, R. H. (2003). *Economic growth, inequality and poverty: Findings from a new data set* (Vol. 2972). World Bank Publications. <https://doi.org/10.1002/sim.1956>
- Bhatta, S. D., & Sharma, S. K. (2006). The determinants and consequences of chronic and transient poverty in Nepal. *Chronic Poverty Research Centre* (Working paper No. 66). <http://doi.org/10.2139/ssrn.1753615>
- Botha, F. (2010). The impact of educational attainment on household poverty in South Africa. *Acta Academica*, 42(4), 122-147. <https://www.researchgate.net/publication/287486539>
- Central Bureau of Statistics. (2005). Poverty Trends in Nepal (1995-96 and 2003-04). Central Bureau of Statistics, National Planning Commission.
- Central Bureau of Statistics. (2011a). *Nepal Living Standard Survey* (2010/11). Poverty in Nepal (A brief report based on NLSS-III), Central Bureau of Statistics, National Planning Commission Secretariat, Government of Nepal. <https://cbs.gov.np/poverty-in-nepal-2010-11/>
- Central Bureau of Statistics. (2011b). *Nepal Living Standard Survey* (2010/11). Statistical Report, Volume Two, Central Bureau of Statistics, National Planning Commission Secretariat, Government of Nepal. [https://time.com/wp-content/uploads/2015/05/statistical\\_report\\_vol2.pdf](https://time.com/wp-content/uploads/2015/05/statistical_report_vol2.pdf)
- Central Bureau of Statistics. (2011c). *Nepal Living Standard Survey* (2010/11). Statistical Report, Volume One, Central Bureau of Statistics, National Planning Commission Secretariat, Government of Nepal. [https://cbs.gov.np/wp-content/uploads/2018/12/Statistical\\_Report\\_Vol1.pdf](https://cbs.gov.np/wp-content/uploads/2018/12/Statistical_Report_Vol1.pdf)
- Central Bureau of Statistics. (2014). *Population Monograph of Nepal 2014: , Population Dynamics* (Vol. 1). <https://mohp.gov.np/downloads/Population%20Monograph%20V01.pdf>
- Chowdhury, M. J. A., Ghosh, D., & Wright, R. E. (2005). The impact of micro-credit on poverty: Evidence from Bangladesh. *Progress in Development Studies*, 5(4), 298-309. <https://doi.org/10.1191/1464993405ps1160a>
- Edoumiekumo, S. G., Karimo, T. M., & Tombofa, S. S. (2014). Determinants of households' income poverty in the South-South Geopolitical Zone of Nigeria. *Journal of Studies in Social Sciences*, 9(1), 101-115.
- Farah, N. (2015). Impact of household and demographic characteristics on poverty in Bangladesh: A logistic regression analysis. *Eastern Illinois University*. [https://thekeep.eiu.edu/lib\\_awards\\_2015\\_docs/3/](https://thekeep.eiu.edu/lib_awards_2015_docs/3/)
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). John Wiley & Sons.
- Imam, M. F., Islam, M. A., & Hossain, M. J. (2018). Factors affecting poverty in rural Bangladesh: An analysis using multilevel modelling. *Journal of the Bangladesh Agricultural University*, 16(1), 123-130. <https://doi.org/10.3329/jbau.v16i1.36493>

- International Organization of Migration. (2019). *Migration in Nepal: A country profile 2019*. <https://publications.iom.int/books/migration-nepal-country-profile-2019>
- John, G., & Scott, R. (2002). Poverty and access to infrastructure in Papua New Guinea. *UC Davis Working Paper* (Working paper No. 02-008). <https://doi.org/10.2139/ssrn.334140>
- Joshi, G. R., & Joshi, N. B. (2016). Determinants of household food security in the eastern region of Nepal. *SAARC Journal of Agriculture*, 14(2), 174-188. <https://doi.org/10.3329/sja.v14i2.31257>
- Khan, M. M., Hotchkiss, R., Berruti, A. A., & Hutchinson, P. L. (2006). Geographic aspects of poverty and health in Tanzania: Does living in a poor area matter? *Health Policy and Planning*, 21(2), 110-122. <https://doi.org/10.1093/heapol/czj008>
- Khan, R. E. A., Rehman, H., & Abrar ul Haq, M. (2015). Determinants of rural household poverty: The role of household socioeconomic empowerment. *American-Eurasian Journal of Agricultural and Environmental Science*, 15(1), 93-98.
- Kona, M. P., Khatun, T., Islam, N., Mijan, A., & Noman, A. (2018). Assessing the impact of socio-economic determinants of rural and urban poverty in Bangladesh. *International Journal of Science & Engineering Research*, 9(8), 178-184. <https://www.researchgate.net/publication/329252093>
- Kousar, R., Makhdum, M. S. A., & Ashfaq, M. (2015, March 23-27). *Impact of land ownership on the household welfare in rural Pakistan* [Conference session]. 2015 World Bank Conference on Land and Poverty, The World Bank-Washington DC, United States. [https://www.researchgate.net/publication/274713955\\_IMPACT\\_OF\\_LAND\\_OWNERSHIP\\_ON\\_THE\\_HOUSEHOLD\\_WELFARE\\_IN\\_RURAL\\_PAKISTAN](https://www.researchgate.net/publication/274713955_IMPACT_OF_LAND_OWNERSHIP_ON_THE_HOUSEHOLD_WELFARE_IN_RURAL_PAKISTAN)
- Kunwar, L. S. (2015). Emigration of Nepalese people and its impact. *Economic Journal of Development Issues*, 19-20(1-2), 77-82. <https://doi.org/10.3126/ejdi.v19i1-2.17705>
- Leekoi, P., Jalil, A. Z. A., & Harun, M. (2014). An empirical on risk assessment and household characteristics in Thailand. *Middle-East Journal of Scientific Research*, 21(6), 962-967.
- Majeed, M. T., & Malik, M. N. (2015). Determinants of household poverty: Empirical evidence from Pakistan. *The Pakistan Development Review*, 701-717. <https://doi.org/10.30541/v54i4I-Ipp.701-718>
- Mamo, B. G., & Abiso, M. (2018). Statistical analysis of factors affecting poverty status of rural residence. *American Journal of Theoretical and Applied Statistics*, 7(5), 188-192. <https://doi.org/10.11648/j.ajtas.20180705.14>
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 104-142). Academic Press.
- Ministry of Finance. (2005). *Economic Survey*. Government of Nepal, Kathmandu.
- Ministry of Finance. (2012). *Economic Survey*. Government of Nepal, Kathmandu. [https://mof.gov.np/uploads/document/file/Economic%20Survey%202011-12\\_20141224054554.pdf](https://mof.gov.np/uploads/document/file/Economic%20Survey%202011-12_20141224054554.pdf)
- Ministry of Finance. (2013). *Economic Survey*. Government of Nepal, Kathmandu. [https://mof.gov.np/uploads/document/file/ES%202069-70\\_20140720094744.pdf](https://mof.gov.np/uploads/document/file/ES%202069-70_20140720094744.pdf)
- Ministry of Health. (2011). *Nepal Demographic Health Survey 2011*. Ministry of Health and Population, New ERA, and ICF International. [https://dhsprogram.com/pubs/pdf/fr257/fr257\[13april2012\].pdf](https://dhsprogram.com/pubs/pdf/fr257/fr257[13april2012].pdf)
- Myftaraj, E., Zyka, E., & Bici, R. (2014). Identifying household level determinants of poverty in Albania using logistic regression

- model. *International Journal of Sustainable Development*, 7(3), 35-42. <https://ssrn.com/abstract=2457441>
- Obi, A., van Schalkwyk, H. D., & van Tilburg, A. (2012). Market access, poverty alleviation and socio-economic sustainability in South Africa. In H. D. van Schalkwyk, J. A. Groenewald, G.C.G. Fraser, A. Obi, & A. van Tilburg (Eds.), *Unlocking markets to smallholders. Mansholt Publication Series* (Vol. 10). Wageningen Academic Publishers. [https://doi.org/10.3920/978-90-8686-168-2\\_1](https://doi.org/10.3920/978-90-8686-168-2_1)
- Omoregbee, F. E., Ighoro, A., & Ejembi, S. A. (2013). Analysis of the effects of farmers characteristics on poverty status in Delta State. *International Journal of Humanities and Social Science Invention*, 2(5), 11-16.
- Osowole, O. I., Ugbechie, R., & Uba, E. (2012). On the identification of core determinants of poverty: A logistic regression approach. *Mathematical Theory and Modeling*, 2(10), 45-53.
- Peters, D. H., Garg, A., Bloom, G., Walker, D. G., Brieger, W. R., & Rahman, M. H. (2008). Poverty and access to health care in developing countries. *Annals of the New York Academy of Sciences*, 1136(1), 161-171. <https://doi.org/10.1196/annals.1425.011>
- Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics*, 9(4) 705-724.
- Shively, G., & Thapa, G. (2017). Markets, transportation infrastructure and food prices in Nepal. *American Journal of Agricultural Economics*, 99, 660-682. <https://doi.org/10.1093/ajae/aaw086>
- Spaho, A. (2014). Determinants of poverty in Albania. *Journal of Educational and Social Research*, 4(2), 157-163. <https://doi.org/10.5901/jesr.2014.v4n2p157>
- Taylor, J. E., Gurkan, A. A., & Zezza, A. (2009). *Rural poverty and markets* (ESA Working Paper No. 09-05). Agricultural Development Economics Division, The Food and Agriculture Organization of the United Nations. <http://www.fao.org/3/a-ak424e.pdf>
- Teka, A. M., Woldu, G. T., & Fre, Z. (2019). Status and determinants of poverty and income inequality in pastoral and agro-pastoral communities: Household-based evidence from Afar Regional State, Ethiopia. *World Development Perspectives*, 15, 100123. <https://doi.org/10.1016/j.wdp.2019.100123>
- Thapa, A. K., Dhungana, A. R., Tripathi, Y. R., & Aryal, B. (2013). Determinants of poverty in rural parts of Nepal: A study of Western Development Region. *Pinnacle Economics & Finance*, 1-6. [https://www.pjpub.org/pef/pef\\_105.pdf](https://www.pjpub.org/pef/pef_105.pdf)
- Uematsu, H., Shidiq, A. R., & Tiwari, S. (2016). Trends and drivers of poverty reduction in Nepal: A historical perspective. *World Bank Policy Research Working Paper* (No. 7830). <https://doi.org/10.1596/1813-9450-7830>
- Wickeri, E. (2011). "Land is life, land is power": Landlessness, exclusion, and deprivation in Nepal. *Fordham International Law Journal*, 34(4), Article 6. <https://ir.lawnet.fordham.edu/ilj/vol34/iss4/6>





# BIBECHANA

ISSN 2091-0762 (Print), 2382-5340 (Online)

Journal homepage: <http://nepjol.info/index.php/BIBECHANA>

Publisher: Department of Physics, Mahendra Morang A.M. Campus, TU, Biratnagar, Nepal

## Dichotomization of quantitative variables in poverty analysis

Krishna Prasad Acharya<sup>1</sup>, Shankar Prasad Khanal<sup>1\*</sup>, and Devendra Chhetry<sup>1</sup>

<sup>1</sup>Central Department of Statistics, Institute of Science and Technology, Tribhuvan University, Kirtipur, Nepal

Email: drshankarcds@gmail.com

### ABSTRACT

It has been proposed four schemes of dichotomization for the four household level quantitative variables – area of land holding, geographic accessibility to the nearest market centre, number of children under 15 and number of literate members of working-age – with justification in the selection of threshold value for each variable to dichotomize into disadvantaged and advantaged group of households using the Nepal Living Standard Survey 2010/11 data with 5988 households and 28,670 of their household members. Association of each dichotomized variable with household level poverty status (poor/non-poor) was found highly significant. Finally, the proposed schemes of dichotomization have tested empirically for their ability to differentiate the poor people into two categories - ‘more vulnerable’ and ‘less vulnerable’ - by first estimating the three measures of poverty – head count index, poverty gap index and squared poverty gap index - of each group of population and comparing the estimated measures between the disadvantaged and advantaged group of populations. Statistical analysis has been performed by using IBM SPSS version 20. To a large extent the proposed schemes of dichotomization have found to differentiate the poor people into two groups; for example, the head count index of the disadvantaged group of the number of children under 15 is 3.1 times higher than that of the advantaged group. The results of this paper are expected to be useful to the policy makers and development planners of Nepal for focusing their poverty reduction program towards the more vulnerable group of population as well as academicians.

### Article Information:

Received: Dec 15, 2021

Accepted: Jan 30, 2022

### Keywords:

Dichotomization

Headcount index

Poverty gap index

Square poverty gap index

Vulnerable

DOI: <https://doi.org/10.3126/bibechana.v19i1-2.46407>

This work is licensed under the Creative Commons CC BY-NC License.

<https://creativecommons.org/licenses/by-nc/4.0/>

## 1. Introduction

Several statistical methods are available for assessing the association of a dichotomous variable with a set of quantitative variables, not necessarily all continuous. When dichotomous variable is treated as dependent variable and the set of quantitative variables is treated as

independent variables, the association can be assessed through *logistic regression* which assesses the effect of independent variables on dependent variable simultaneously. When the dichotomous variable is treated as grouping variable and the set of quantitative variables are

treated as test variables, the association can be assessed through *independent samples t-test* which assess the association by comparing the two group means of each quantitative variable. The former method is more rigorous from both theoretical and practical point of view than the later method. However, the use of later method faces the problem of normality assumption which means each test variable has to follow normal distribution within each of the two categories of the grouping variables since the t-test was developed under this normality assumption. Conceptually the normality assumption is hard to justify when a test variable is discrete. When normality assumption fails, instead of independent samples t-test it is a common practice to use Mann-Whitney test. Nonetheless, the Mann-Whitney non parametric test transforms the quantitative variables into their rank orders and the test works on rank ordered data, and the test results are not easily understandable to wider users. The association between a binary variable with a set of quantitative variables having only two levels, the association can also be assessed through biserial correlation. The less frequently used method is to first dichotomize each independent variable using a rationally defined threshold value and then use either logistic regression or use Chi-square test for the dichotomous variable and each dichotomized quantitative variable. Several scholars have discussed the advantages and disadvantages on this less frequently used method [1-3]. Sometimes, dichotomization of quantitative variable is absolutely necessary. For example, dichotomization of per capita consumption expenditure using poverty line as threshold value is absolutely necessary in measuring monetary poverty.

This paper has two-fold objectives. First objective is to dichotomize the four household level quantitative variables by justifying in the selection of threshold value for each variable, and assess the association of the four dichotomized variables with the dichotomous variable - poverty status (poor/non-poor).

Second objective is to estimate the three measures of poverty – *head count index*, *poverty gap index*, and *squared poverty gap index* – for all the four dichotomized variables in order to investigate the ability to differentiate poor peoples into ‘more vulnerable’ and ‘less vulnerable’ through the estimated measures of poverty.

## 2. Materials and Methods

The main source of data of this study is NLSS III which provides household level data on socio-economic and demographic variables of 5,988 households and 28,670 household members. The available data on the variable ‘household poverty status’ was taken as binary variable by assigning code 1 for poor and 0 for non-poor. In this study a household is defined as *poor (non-poor) depending upon the per capita expenditure of the household members falls below (above) the poverty line of NRs 19,261*. The un-weighted and weighted proportions of poor households were correspondingly 18.5% and 20.0%. Similarly, un-weighted and weighted proportions of poor household members were correspondingly 23.4% and 25.2%.

Out of many household level variables that influence the monetary poverty, only the following four quantitative variables are considered in the present study.

1. Area of land holding
2. Geographic accessibility to market center (defined in the present study by time taken in minutes to reach the nearest market irrespective of transport mode)
3. Number of children under 15
4. Human capital (defined by number of literate members of working-age (15 – 64 years)).

The available data on the above four variables were dichotomized. The main reason for dichotomization of each of these variables is to divide the households into two groups:

*advantaged and disadvantaged group* with respect to each variable. The rationale behind such demarcation of households is that the disadvantaged group of households would be in a more difficult position to escape out of poverty than the advantaged group of households. The process of dichotomization, particularly choosing the threshold value for each of the four quantitative variables is rationalized below.

### **2.1 Dichotomizing households by area of land holding**

The available data on land holding is highly skewed (skewness = 5.55) with extremely high measure of kurtosis (excess of kurtosis = 65.15), and considerable number of households had no lands. As a result, analysis based on original data suffers from various problems. Moreover, farm size is not a good determinant of poverty [4]. In this context, analysis based on dichotomizing the quantitative variable is more sensible than analyzing the data as it is. In view of this fact, the threshold value for this quantitative variable was chosen to be 0 which demarcates households into two groups - one group of households each had no land (disadvantaged group) and the other group of households each had land (advantaged group).

### **2.2 Dichotomizing households by the number of children under 15**

Children under 15 are considered as dependent population in the sense that their basic needs have to be fulfilled by their parents. In this context, large number children would be burden to parents. As a result large number of children aggravates poverty [5]. However, small number of children is desirable. The ideal number of children responded by women respondents on an average was 2.1 and by men respondents was 2.3 in 2011 [6]. Based on these results, the threshold value of 2 is used for dichotomizing the quantitative variable. This threshold value demarcates the households into two groups - one group of households each had more than two children (disadvantaged group) and the other group of households each had less than or equal to two children (advantaged group).

### **2.3 Dichotomizing households by the number of literate working-age members**

In the context of Nepal, number of illiterate persons in a household is major disadvantage of the poor households [5]. In view of this fact, the household level quantitative variable “the number of literate members of working-age” was selected and converted it into dichotomous variable by grouping the households into two groups: one group of households each had no literate member of working-age (disadvantaged group) and the other group of households each had at least one literate member of working-age (advantaged group).

### **2.4 Dichotomizing households by access to nearest market**

*The available* data on access to nearest market center is highly skewed (skewness = 3.46) with high measure of kurtosis (excess of kurtosis = 16.74). The mean and median of the time taken to reach market center in minutes are correspondingly 80.63 and 30.00. The analysis based on dichotomizing the quantitative variable is more sensible rather than analyzing the data as it is. In view of this fact, the threshold value for this quantitative variable was chosen 30 minutes which demarcates households into two groups - one group of households each is required more than 30 minutes to reach market center (disadvantaged group) and the other group of households each is required less than or equal to 30 minutes (advantaged group). The threshold value of 30 minutes is taken because it is a common in Nepal [7].

### **2.5 Test of association**

The association of poverty status with each dichotomized variable is assessed using Chi-square test and effect size of each test is measured by Phi-coefficient whose values range from -1 to 1. Just like the correlation coefficient, a negative value of Phi-coefficient indicates that

when one variable increases, the other decreases and a positive value indicates that when one variable increases, so does the other.

### 2.6 The measures of poverty

In contemporary studies three measures of poverty are used. They were originated from a class of poverty measures  $P(\alpha)$  introduced by Foster-Greer-Thorbecke [8], and expressed as

$$P(\alpha) = \frac{1}{N} \sum_{i=1}^q \left( 1 - \frac{y_i}{z} \right)^\alpha$$

where  $\alpha$  is index  $\geq 0$ ,  $q$  is the number of poor peoples,  $N$  is the total number of individuals,  $z$  is the poverty line,  $y_i$  is the per capita consumption expenditure and sum of the expression within parentheses is total poverty gap expressed as a proportion of the poverty line. In particular,  $P(0)$ ,  $P(1)$  and  $P(2)$  correspondingly yield the three measures of poverty – such as head count index, poverty gap index and squared poverty gap index. The measure  $P(0)$  is also known as head count ratio, incidence of poverty or poverty rate. It is a simple concept to understand and, therefore, widely used in political debate. However, it does not take into account of how poor the poor are, and this issue is addressed by the measure  $P(1)$  which in simple term measures how far away the poor peoples are from the poverty line, consequently  $P(1)$  satisfies the Monotonicity Axiom of Amartya Sen [9] and larger the value of  $P(1)$  larger the investment and effort would require to alleviate poverty. However, the measure  $P(1)$  does not take into account of the inequality in distribution of per capita expenditure among poor, and this issue takes into account by the measure  $P(2)$  and consequently  $P(2)$  satisfies the Transfer Axiom of Amartya Sen [9]. The two measures –  $P(1)$  and  $P(2)$  – are difficult concept to understand and, therefore, they are not widely used in political debate but they are useful for policy makers as well as for academicians.

Several developing countries, including Nepal, have been estimating and using the three measures of poverty for monitoring, evaluation

and planning program of poverty reduction. Academicians are also using three measures in their academic work [10 - 21].

All the statistical analysis has been performed by using IBM SPSS version 20.

### 3. Results and Discussions

The results of this study are summarized in three tables where the first table displays the descriptive statistics of quantitative variables for the disadvantaged and advantaged group of households, the second table displays the association between the dichotomized variables with poverty status, and finally the third table provides weighted estimates of the three measures of poverty for the disadvantaged and advantaged group of population where weights are the population weights provided by CBS in the data file.

Table 1 shows that among the total 5,988 households, around 29% of have no land, 48% have at least three children, 26% have no literate persons of working age and 19% have poor access to market. The mean difference between advantaged and disadvantaged is highest in the variable ‘access to market’ and least is in the variable ‘area of land holding’. Skewness is positive in all variables.

Table 2 shows that the percentage of poor households is larger among the disadvantaged group of households than among the advantaged group. Within group difference in percentage of poor is highest in the variable ‘number of children under 15 and least in the variable ‘area of land holding’. All four dichotomized variables were found statistically significant with poverty status. The effect-size for each Chi-square test is positive and it is minimum for the test of association between the dichotomized variable of access to market center and poverty status, and it is maximum for the test of association between the dichotomized variable of the number of children under 15 and poverty status.

**Table 1:** Descriptive statistics of quantitative variables by group

	% of households	Mean	Std. Dev.	Skewness
Status of land holding:				
With land (0) - AG	71.2	0.36	0.55	5.00
Without land (1) – DG	28.8	0.00	0.000	NA
Number of children under 15:				
At most 2 (0) - AG	52.0	0.96	0.83	0.07
At least 3 (1) - DG	48.0	3.73	1.10	2.50
Number of literate working-age members:				
At least one (0) - AG	73.8	2.23	1.28	1.61
None (1) - DG	26.2	0.00	0.00	NA
Access to nearest market center:				
Having better access (0) - AG	80.7	13.91	11.42	0.12
Having poor access (1) - DG	19.3	152.96	148.86	2.77

Note: Figures within parentheses are coding scheme (0 for advantaged group and 1 for disadvantaged group). The dichotomous variable - poverty status – is coded as follows: 0 for non-poor and 1 for poor. AG = advantaged group and DG = Disadvantaged group NA: Not Applicable.

Computed from NLSS-III (2010/11)

**Table 2:** Association of dichotomized variables with poverty status

	% of poor households Within group	Chi-square & p-value	Phi-coefficient
Status of land holding:			
With land	15.0	114.9	0.14
Without land	27.0	(p<0.001)	
Number of children under 15:			
At most 2	11.0	653.0	0.33
At least 3	40.0	(p<0.001)	
Number of literate working-age members:			
At least one	16.0	142.0	0.15
None	31.0	(p<0.001)	
Access to nearest market center:			

Having better access	11.0	206.7	0.08
Having poor access	26.0	(p<0.001)	

Computed from NLSS-III (2010/11)

Table 3 shows that the scheme of dichotomization for each quantitative variable is able to differentiate poor peoples as ‘more vulnerable’ and ‘less vulnerable’ very distinctly according to each of the three measures of poverty. Comparison of the estimated three measures of poverty with the corresponding estimate of the national level, each estimate of each advantaged group of population is below the estimate of the national level, on the contrary each estimate of each disadvantaged group of population is above the estimate of the national level.

**Table 3:** Three measures of poverty for eight groups of household population

Variables	Head Count Index (P(0))×100	Poverty Gap Index P(1)×100	Square Poverty Gap Index P(2)×100
Status of land holding:			
With land (LV)	21.4	4.5	1.4
Without land (MV)	32.9	7.5	2.6
Number of children under 15:			
At most 2 (LV)	13.5	2.4	0.7
At least 3 (MV)	41.4	9.6	3.3
Number of literate persons of working age:			
At least one (LV)	21.5	4.3	1.4
None (MV)	41.7	10.4	3.9
Access to nearest market center:			
Having better access (LV)	16.3	3.3	0.9
Having poor access (MV)	32.1	7.1	2.5
National level of estimates	25.2	5.4	1.8

Note: LV = Less Vulnerable and MV = More Vulnerable.

Computed from NLSS-III (2010/11)

The extent of differentiations of poor people by the scheme of dichotomization varies across the dichotomized variables. For instance, the ratio of the head count index of the more vulnerable group to that of the less vulnerable group of population is highest for the dichotomized variable of ‘the number of children under 15’ and the ratio is 3.1 showing that the head count

index of the more vulnerable group is 3.1 times higher than that of the less vulnerable group of population. Such ratios for the poverty gap index and the squared poverty gap index of the same dichotomized variable are correspondingly 4.0 and 4.7. Whereas the ratio of the head count index of the more vulnerable group to that of the less vulnerable group of

population is lowest for the dichotomized variable of ‘the status of land holding’ and the ratio is 1.5 and such ratios for the poverty gap index and the squared poverty gap index are correspondingly 1.7 and 1.9.

#### 4. Conclusions

Dichotomization of a quantitative variable with the aid of a reasonable threshold value in poverty analysis is a useful strategy because this study to a large extent succeeded to show that such dichotomization scheme differentiates the poor people into two groups ‘more vulnerable’ and ‘less vulnerable’, so that the policy makers and development planners focus their poverty reduction program towards the more vulnerable people. Among the more vulnerable household populations of those households having three or more children under 15 is the most vulnerable. A baffling issue ‘why the poor households tend to have large number children?’ has to be resolved because for the reduction of the number of children from the most vulnerable group of households. The currently available data fail to resolve this issue. The impact of outmigration of literate and young population for employment as well as settlement in abroad was seen in the dichotomized variable of the quantitative variable ‘the number of literate persons of working-age group’.

#### Conflict of Interest

Authors declare that there is not any conflict of interest.

#### Acknowledgments

Authors would like to thank Central Bureau of Statistics (CBS) for providing NLSS III 2010/11 data, and the department research committee of Central Department of Statistics, Tribhuvan University, Nepal for their critical comments and suggestions for this study. University Grant Commission (UGC) of Nepal is also acknowledged for providing Ph.D. fellowship as this is the part of Ph.D. research work.

#### References

- [1]J. E. Hunter & F. L. Schmidt. Dichotomization of continuous variables: The implications for meta-analysis. *J. Appl. Psychol.* 75(3) (1990) 334–349.
- [2]R. C. MacCallum, S. Zhang, K. J. Preache, & D. D. Rucker. On the practice of dichotomization of quantitative variables. *Psychol. Methods* 7(1) (2002).
- [3]S. P. Nelson, V. Ramakrishnan, P. J. Nietert, D. L. Kaman, P. S. Ramos, & B. J. Wolf. An evaluation of common methods for dichotomization of continuous variables to discriminate disease status. *Commun. Stat. Theory and Methods* 46(21) (2017) 10823-10834.
- [4]D. Chhetry. Understanding rural poverty in Nepal. In *Asia and Pacific forum on poverty: Reforming policies and institutions for poverty reduction 1* (2001) 293-314.
- [5]D. Chhetry. Comparative analysis of socio-demographic and economic characteristics of the poor and non-poor group of households. *Nepal Population J.* (2000) 101-112.
- [6]Ministry of Health. *Nepal Demographic Health Survey*. Ministry of Health, New ERA Kathmandu, Nepal, the DHS Program, ICF, Rockville, Maryland, USA. (2011).
- [7]Central Bureau of Statistics. *Nepal Living Standard Survey (2010/11)*. Summary Report, Central Bureau of Statistics. National Planning Commission Secretariat, Government of Nepal (2011).
- [8]J. Foster, J. Greer, & E. Thorebecke. A Class of decomposable measure of poverty, *Econometrica* 52(1984).
- [9]A. Sen. Poverty: an ordinal approach to measurement. *Econometrica* 44 (2) (1976) 219 – 231.
- [10] S. A. Yusuf, A. O. Adesanoye, and D. O. Awotide. Assessment of poverty among urban farmers in Ibadan Metropolis, Nigeria. *J. Hum. Ecol.* 24(3) (2008) 201-207.
- [11]J. I. Onu, and Z. Abayomi. An analysis of poverty among households in Yola Metropolis



- of Adamawa State, Nigeria. *J. Soc. Sciences* 20(1) (2009) 43-48.
- [12]I. S. Chaudhry and S. Malik. The impact of socioeconomic and demographic variables on poverty: A village Study. *Lahore J. Economics* 14(1) (2009) 39-68.
- [13]D. Akerele, and S. A. Adewuyi. Analysis of poverty profiles and socioeconomic determinants of welfare among urban households of Ekiti State, Nigeria. *Curr. Res. J. Soc. Sci.* 3(1) (2011) 1-7.
- [14]R. A. Sanusi, T. S. Owagbemi, and M. Suleiman. Determinants of poverty among farm households in Ikorodu local government area of Lagos State, Nigeria. *IJAFS* 4 (2013) 538-552.
- [15]S. O.Akinbode, A. O. Salami, and O. T. Ojo. Impact of micro finance on poverty status of small scale crop farming households in Southwest Nigeria. *Am. J. Econ.* 3(6) (2013) 322-329.
- [16]L. A. Salami, and K. Atiman. An Analytical study of determinants of poverty level among households in Adamawa North District, Nigeria. *Mediterranean J. Social Sci.* 4(16) (2013) 73-73.
- [17]A. Adetayo. Analysis of farm households poverty status in Ogun states, Nigeria. *Asian Econo. Financ. Rev.* 4(3) (2014) 325-340.
- [18]R. S.Margwa, J. I. Onu, J. N. Jalo, and B. Dire. Analysis of poverty level among rural households in Mubi region of Adamawa State, Nigeria. *J. Sci. Res. Stud.* 2(1) (2015) 29-35.
- [19]R.Saramia. Analysis of poverty profile and determinants of welfare among rural households: A Case study of Udalguri District, Assam. *Int. J. Humanities & Social Studies (IJHSS)* 1(4) (2015) 138-144.
- [20] K. P. Acharya. Analysis of poverty profile by type of house of households in Nepal. *Management Dynamics* 22(1) (2019) 7-10. DOI: <https://doi.org/10.3126/md.v22i1.30231>
- [21]P.Uprety. Measures, distribution and decomposition of poverty: An empirical analysis in Nepal. *Nepalese J. Stat.* 4 (2020) 1-16.

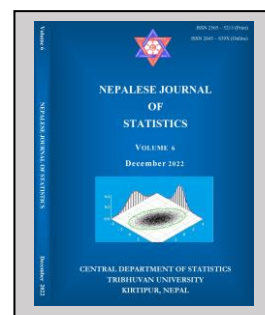
## On the Use of Logistic Regression Model and its Comparison with Log-binomial Regression Model in the Analysis of Poverty Data of Nepal

Krishna Prasad Acharya<sup>1</sup>, Shankar Prasad Khanal<sup>2\*</sup>  
and Devendra Chhetry<sup>3</sup>

Submitted: 17 December 2022; Accepted: 25 December 2022

Published online: 27 December 2022

DOI: 10.3126/njs.v6i01.50806



### ABSTRACT

**Background:** Previous literatures have indicated that log-binomial regression model is an alternative for the logistic regression model for frequent occurrence of event of outcome. The comparison of the performance of these two models has been found with reference to clinical/epidemiological data. Nonetheless, the application of log-binomial model and its comparison with the logistic model for poverty data has not been described.

**Objective:** To compare logistic and log-binomial regression model in terms of variable selection, effect size, precision of effect size, goodness of fit, diagnostics, stability of the model, and the issue of failure convergence.

**Materials and Methods:** Cross sectional data of 5988 households of Nepal Living Standard Survey 2010/11 has been used for the analysis. The performance of logistic and log-binomial model has been compared in terms of variable selection, effect size, and its precision for each covariate, goodness of fit using Hosmer - Lemeshow (H-L) test, diagnostics of the model, stability of the model using bootstrapping method, and the issue of failure convergence.

**Results:** Logistic model overestimates the effect size, yields wider 95% confidence interval than that of log - binomial model for each covariate. The greater elevation in risk for covariates varies from 13% to 173%. Logistic model satisfies goodness of fit of the model ( $p = 0.534$ ), diagnostics tests, and stability of the model. However, log-binomial model grossly violates the goodness of fit of the model ( $p = 0.0004$ ) but satisfies the model diagnostics and stability criteria.

**Conclusion:** Log-binomial model satisfies all criteria for model development and diagnostics except gross violation in goodness of fit of the model. However, logistic regression model satisfies all the criteria including goodness of fit of the model. On the basis of the entire comparison of model performance, logistic regression model is better fitted than the log-binomial model in fitting the poverty data set of Nepal.

**Keywords:** Diagnostics, elevation in risk, goodness of fit, log-binomial, logistic, poverty, stability, variable selection.

**Address correspondence to the authors:** Central Department of Statistics, Institute of Science and Technology, Tribhuvan University, Kirtipur, Kathmandu, Nepal.

Email: acharyakrishna20@gmail.com<sup>1</sup>; drshankarcds@gmail.com<sup>2\*</sup> (corresponding author email); chhetrydevendra@gmail.com<sup>3</sup>

## INTRODUCTION

The logistic regression model is being used as a common method to study the associations of independent variables with the categorical dichotomous outcomes. Its use is frequent in case control, cohort studies and clinical trials. It has also been used in cross-sectional studies (Barros & Hirakata, 2003). It does not only measure the association of outcome variable with the independent variables, but also help to quantify the effect of these variables on the response variable. Logistic regression yields both regression coefficient for each independent covariate and the odds ratio (OR) based on the regression coefficient itself. Odds ratios are commonly reported in the analysis of different studies under such scenario (Davies, Crombie & Tavakoli, 1998) and seem to be relatively more appealing and effective in interpretations compared to regression coefficients. A considerable number of previous studies also indicated the use of relative risk, risk ratio (RR), prevalence ratio (PR), or rate ratio under such scenario (Davies et al., 1998; Holcomb, Chaiworapongsa, Luke & Burgdorf, 2001; Martinez, Leotti, Silva, Nunes, Machado & Corbellini, 2017; Gallis & Turner, 2019). There is still an academic debate regarding the issue of reporting which one either 'OR' or 'RR' is better. Some authors favor to report OR, some favor RR. Walter (1998), Olkin (1998), Newman (2001), and Cook (2002) favor to report odds ratio (OR) as they claimed that it is symmetric with the outcome. On the other hand, Sackett, Deeks and Altman (1996), De Andrade and Carabin (2011), and Gallis and Turner (2019) favor to use relative risk (RR) claiming that it is easily understandable. Lee (1994) has also remarked that odds ratio has been described as incomprehensible. Williamson, Eliasziw, & Fick (2013) has encouraged to use relative risk in epidemiological studies wherever possible, and to advocate its use. The odds ratios and the risk ratios are closer if the outcome of interest is very rare i.e. generally considered as less than 10 % (Greenland & Thomas, 1982; Greenland, Thomas & Morgenstern, 1986; Viera, 2008). If the outcome of interest is common (i.e.  $\geq 10\%$ ), odds ratio will not be able to approximate risk ratio (Greenland, 1987; McNutt, Xiaonan Xue & Hafner, 2003; Katz, 2006; Viera, 2008; Ranganathan, Aggarwal & Pramesh, 2015; Gallis & Turner, 2019).

Initially, Wacholder (1986) recommended a simple approach of estimating risk ratios (RR) directly for studying the association of number of independent variables with the dichotomous response variable. Later, Barros and Vânia (2003) declared that generally the OR overestimates the RR in cross-sectional studies having frequent occurrences of event of interest. The basis of the log-binomial model is a generalized linear model with log link and binomial probability distribution, which results in risk ratio (RR). Robbins, Chao and Fonseca (2002), and McNutt et al. (2003) also highlighted their descriptions and applications. After that, Blizzard and Hosmer (2006) proposed the goodness of fit test and some diagnostics of the log- binomial regression model. There are

established methods to convert odds ratios into risk ratios. However, Robbins et al. (2002) had clearly indicated that these converted methods yielded inaccurate confidence intervals of estimates. It is also reported that there is failure convergence of log - binomial regression model for some applications (Williamson et al., 2013). The odds ratio and the relative risk can be computed in a different approach. The established technique for computing OR and RR in bivariate analysis is summarized in Table I.

**Table I.** Layout of computation of OR and RR.

Independent variable	Outcome variable		Total
	Present	Absent	
Group I	a	b	$n_{10}$
Group II (Reference category)	c	d	$n_{20}$
Total	$n_{01}$	$n_{02}$	$n$

With reference to table I, probability of occurrence of ‘a’ is  $p_1 = \frac{a}{n_{10}}$ , and its complementary probability is  $(1 - p_1)$  in Group I. Similarly the probability of occurrence of ‘c’ in Group II is denoted by  $p_2 = \frac{c}{n_{20}}$ , and its complementary probability is  $(1 - p_2)$ .

The odds ratio for the presence of outcome is defined as:

$$OR = \frac{p_1 / (1 - p_1)}{p_2 / (1 - p_2)}$$

The relative risk for the presence of defined outcome is simply defined as:  $RR = p_1 / p_2$

The odds ratio is the ratio of two odds whereas the risk ratio is the ratio of two probabilities. The value of OR suppose is 3, is interpreted as the odds of having the outcome is 3 times higher in Group I than the reference group. If the probability of occurrence of outcome in Group I is 0.9 and 0.3 in reference group respectively, then risk ratio is interpreted as the group I is thrice as likely to have the outcome as the reference group. There is still some confusion while interpreting the odds ratios and relative risks which had been well indicated by Schwartz, Woloshin and Welch (1999); Zocchetti, Consonni and Bertazzi (1995). Further, Holcomb et al. (2001), and Baicus (2003) also clearly indicated the misinterpretation of odds ratio as risk ratio in considerable number of published articles in medical journals. However, the interpretation of RR and OR is not the focus of this paper; these issues have been discussed for the sake of completeness of the paper. Poverty is a complex issue and it possesses broadly two dimensions such as monetary poverty and non-monetary poverty. The analysis of this paper is exclusively focused on monetary poverty of Nepal. There are still 18.5% of households under poverty based on data of Nepal Living Standard Survey (2010/11) (Acharya, Khanal & Chhetry, 2022).

Identification of important factors associated with poverty using appropriate statistical model plays very important role for policy point of view. This is an attempt to recommend a more suitable model by comparing logistic and log-binomial regression model based on different criteria.

On extensive review of literature, it is found that the use of log-binomial regression is almost rare in social science related data problems especially on poverty data. The comparison of logistic regression model and the log-binomial regression model is also found quite rarely in social science researches. The objective of this paper is to apply the log-binomial regression model to the variables of the poverty data set of Nepal Living Standard Survey (2010/11), and to compare the results of logistic regression model and log-binomial regression model in terms of selection of variables in the final model, estimates, precision of the estimates, goodness of the fit of the model, diagnostics of the model, issue of convergence of the model, and stability of the model in the context of household poverty of Nepal. The issues of variable selection and model comparison are based on the paper by Acharya et al. (2022).

## METHODOLOGY

### Data

The study is based on cross sectional data of 5988 households of Nepal extracted from Nepal Living Standard Survey 2010/11. The survey was conducted by Central Bureau of Statistics (CBS) Government of Nepal. The un-weighted data was used for both log-binomial and logistic regression model. The detail survey methodology of Nepal Living Standard Survey is explained in survey report (CBS, 2011). The response variable for the model is the poverty status of household (coded 0 for non-poor and 1 for poor). Based on extensive review of relevant literature, altogether seven independent variables namely sex of household head (male vs. female), literacy status of household head (literate vs. illiterate), remittance receiving status(yes vs. no), land holding status(yes vs. no), access to market(better vs. poor), number of children in the household under 15 years of age ( $\leq 2$  vs.  $> 2$ ), and number of literate members of working age( $\geq 1$  vs. none) are considered initially for log-binomial regression model as used in logistic regression model. All details about selection of variables and need of dichotomization of independent variables, etc. had been described in Acharya et al. (2022).

### Statistical model

The log-binomial regression model is a special type of generalized linear model for which the link function is log link. Logistic regression model is also a type of generalized linear model with logit link function. The response variable for both the models is dichotomous type.

The log-binomial regression model for  $p$  number of covariates  $(X_1, X_2, \dots, X_p)$  in association with binary response variable is given by:

$$\log \pi = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \tag{1}$$

where  $\pi = \text{Pr } ob[Y = 1 | X] = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)$  for binary outcome  $Y$ ,  $\beta_1, \beta_2, \dots, \beta_p$  are the regression coefficients for covariate

$(X_1, X_2, \dots, X_p)$ ,  $\beta_0$  is the constant term in the model. The link function for this model is log link. In this model RR can be computed as  $e^{\beta_j}$  for each considered covariate as done in the case of computing OR in logistic regression model. Regression coefficients are estimated by using the maximizing the likelihood function (for detail, please see McCullagh and Nelder, 1989)

$$l(\beta) = \sum_{i=1}^n [y_i \log \pi_i + (1 - y_i) \log (1 - \pi_i)] \tag{2}$$

where, 
$$\pi_i = e^{\left( \sum_{j=0}^p \beta_j x_{ij} \right)}$$

**Goodness of fit test and diagnostics of the fitted model**

The goodness of fit of the log - binomial regression model can also be assessed by using Hosmer and Lemeshow (H-L)  $\chi^2$  test with  $(10 - 2 = 8)$  degrees of freedom using the formula

$$\hat{c} = \sum_{j=0}^1 \sum_{k=1}^{10} \frac{(o_{jk} - \hat{e}_{jk})^2}{e_{jk}} \tag{3}$$

The observed and expected value in H-L  $\chi^2$  test in case of log-binomial regression model appear approximately equal but not exact. However, this test can also be applied for assessing the goodness of fit of the log-binomial regression model (Blizzard & Hosmer, 2006). The diagnostics of the fitted log-binomial regression model has been assessed graphically through the plot of (i) leverage in the vertical axis and the fitted model in the horizontal axis, and (ii) Chi-square displacement generated by the log- binomial model in the vertical axis and model fitted values in the horizontal axis (Blizzard & Hosmer, 2006).

The stability of the developed model has been evaluated by using bootstrapping method (Chen & George, 1985; Altman & Anderson, 1989; Saurbrei & Schumacher, 1992) as used for assessing the stability of Cox regression model. Same approach as applied by Khanal, Sreenivas & Acharya (2019) for comparing the stability of Cox proportional hazards model and two accelerated failure time models has been used to assess the stability of the logistic and log- binomial regression model. After fitting the final log-binomial regression model, the risk assessment of factors has been performed in terms of RR running the log-binomial model with same response variable on newly generated variable ( $X_i$  for  $i = 1, 2, 3, \dots, p$ ), where 1 represents the presence of any one factor, 2 for presence of any 2 factors, and finally the presence of all factors in the final model respectively. The values of RR obtained from log-binomial model and the values of OR computed in similar manner from logistic regression model are compared.

Finally, the logistic regression model developed in the same data set by Acharya et al. (2022) and the log-binomial regression model developed in this attempt are compared in terms of variable selection, estimates, precision of estimates, goodness of fit of the model, regression diagnostics, stability of the model, and model convergence issue. Bootstrapping procedure has been done by

using R software, and remaining all statistical analysis has been performed by using STATA version 13.0.

## RESULTS

There are altogether seven independent covariates associated with outcome variable used in the model. In order to identify the candidate variables for the final log-binomial regression model, simple log-binomial regression model considering one independent variable at a time separately is performed. Among these seven variables, only six variables; literacy status of household head - literate vs. illiterate (RR: 2.31,  $p < 0.001$ ), remittance receiving status - yes vs. no (RR: 1.38,  $p < 0.001$ ), land holding status - yes vs. no (RR: 1.79,  $p < 0.001$ ), access to market - better vs. poor (RR: 2.25,  $p < 0.001$ ), number of children under 15 years of age-less than or equal to 2 vs. more than 2 (RR: 3.68,  $p < 0.001$ , and number of literate members of working age- at least one vs. none (RR: 1.97,  $p < 0.001$ ), each has come out significantly associated with response variable at 5% level of significance except sex of household head - male vs. female (RR: 0.92,  $p = 0.195$ ). Though sex of household head has not come out statistically significant even in the simple log-binomial regression model, it is also considered as one of the candidate variables for developing multiple log-binomial regression model treating it as a known confounder. Hence, the seven variables including sex of household head are considered as potential candidate variables for the final multiple log-binomial regression model.

### Results of multiple log-binomial regression model

In order to select the significant variables in the final model both stepwise forward selection and backward selection procedure are adopted considering seven candidate variables. Both selection procedures have yielded six common set of significant variables at 5% level of significance except sex of household head. The values of RR, standard error (S.E.) of RR, p-values and 95% CIE for each independent factor associated with poverty obtained through multiple log- binomial regression model are presented in Table 2.

Among finally selected six independent predictors associated with poverty, the risk of household having more than two children under 15 is found to be the highest (RR: 2.96, 95% CIE: 2.66, 3.28) followed by household having illiterate household head, household with poor access to the market center, household not receiving remittance, household with no land respectively. The risk of household being poor among household not having single literate members of the working age in comparison with households having at least one literate member is found to be the least (RR: 1.16, 95% CIE: 1.05, 1.29). This can be interpreted as the household not having single literate members of the working age is 1.16 times as likely to have poorer than those household having at least one literate members of the working age. The goodness of fit of the model assessed by H-L ( $\chi^2$ ) test with 8 degrees of freedom is highly violated ( $p = 0.0004$ ) (Table 2).

**Table2.** Results of multiple log - binomial regression model.

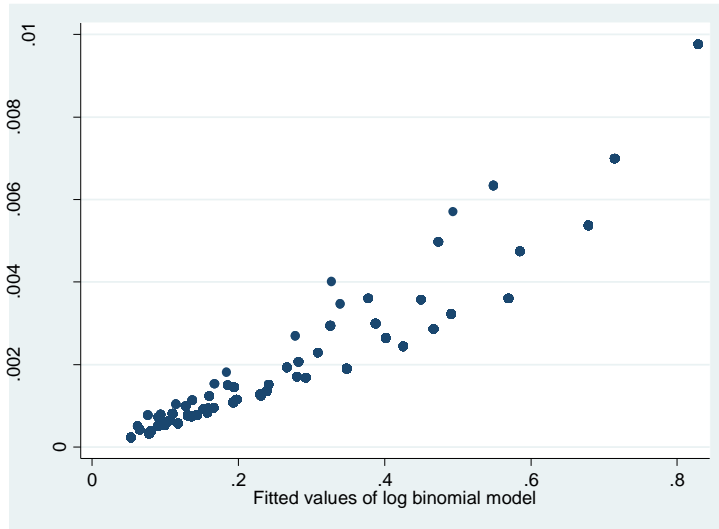
Independent variables	RR	S.E.	p-value	95% CIE
Literacy status of household head:				
Literate	1.00			
Illiterate	1.68	0.1006	< 0.001	(1.49 1.89)
Remittance receiving status:				
Yes	1.00			
No	1.45	0.0685	< 0.001	(1.33 1.59)
Land holding status:				
Yes	1.00			
No	1.22	0.0594	< 0.001	(1.11 1.34)
Access to market:				
Better	1.00			
Poor	1.51	0.0888	< 0.001	(1.34 1.69)
Number of children under 15:				
≤ 2	1.00			
> 2	2.96	0.1590	< 0.001	(2.66 3.28)
No. of literate members of working-age:				
≥ 1	1.00			
0	1.16	0.0606	< 0.001	(1.05 1.29)
Constant	0.05	0.0034	< 0.001	(0.05 0.06)
Log likelihood (only with intercept) = - 4068.888; Log likelihood (full model) = - 2412.336				
AIC = 0.808; BIC = - 47195.150; H-L ( $\chi^2$ ) with 8 d. f. = 28.602, p = 0.0004				

Source: computed based on NLSS 2010/11 data

### Results of diagnostics for log-binomial regression model

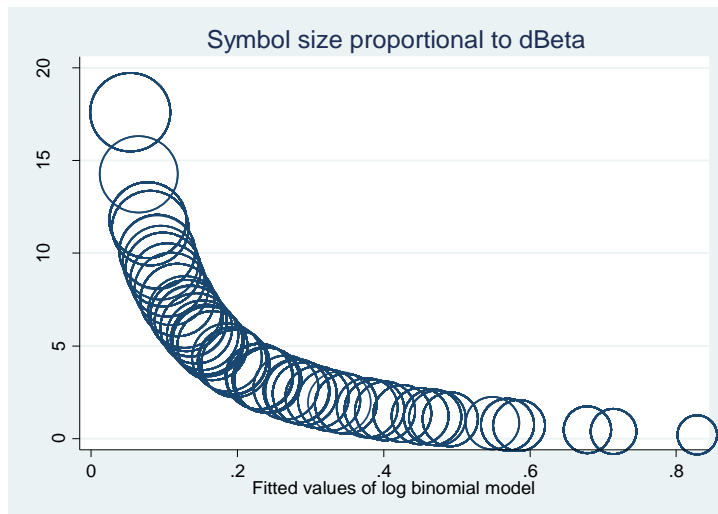
The plot of the leverage values in y-axis and the model fitted values in x-axis is presented in Figure1 (a). One data value seems to be in the top right corner having relatively greater leverage than others. Majority of the leverage values are found less than 0.008 and the extreme one leverage in this dataset is also less than 0.01. Hence, all the leverage values are found to be less than 0.08 which indicates that there is not violation of the diagnostics of the model assessed based on leverage (Blizzard & Hosmer, 2006).





**Fig. 1(a).** Leverage and fitted values of log - binomial model.

The diagnostic of the fitted model has also been assessed through the plot keeping  $\Delta\chi^2$  in vertical axis and the values of fitted log-binomial regression model in horizontal axis with plotting symbol proportional to Cook's distance (Figure 1(b)). There are four poorly fit data points with  $\Delta\chi^2 > 10$  (Blizzard & Hosmer, 2006). The circles of these four data points are observed to be larger than others. Two data points are lying in a bit far away and another one is farther away from others in the right lower corner. There is not much serious violation of the diagnostics of the fitted model evaluating on the basis of the plot of  $\Delta\chi^2$  vs. fitted log - binomial model.



**Fig. 1(b).** Graph of  $\Delta\chi^2$  vs. values of fitted log - binomial model with plotting symbol proportional to Cook's distance.

### Comparison of logistic and log-binomial regression model

Model building of both logistic and log-binomial regression models are started taking with same set of seven covariates. Out of these seven covariates both models have come up with six significant covariates except variable 'sex of household head'. While comparing the effect size (OR for logistic regression and RR for log - binomial regression model) for each covariate, logistic regression model overestimates the effect size (Table3) and wider width of 95% confidence interval estimation than that of log-binomial regression model (Table 3 ). Wider confidence interval estimation of effect size for each covariate in logistic regression model clearly indicates the lesser precision of the estimate than that of log-binomial regression model. The value of OR for each independent variable obtained from logistic regression model overestimates the value of RR obtained from log-binomial regression model. The reference value for each OR and RR is 1.

The elevation of risk percentage for a covariate is computed as  $\text{Elevation risk}(\%) = [(OR-1) - (RR-1)] \times 100$ . For example; let us consider the elevation risk (%) for the variable literacy status of the household head is computed with reference to Table3 as  $[(2.20 - 1) - (1.68 - 1)] \times 100 = (1.20 - 0.68) \times 100 = 52.0\%$ . It is computed in similar fashion for other covariates (Table3). The elevation of risk for the covariate estimates generated by logistic regression varies from 13% to 173% than those generated by log-binomial regression model. The highest elevation of risk (173%) is noted for the variable 'number of children less than 15 years of age' and the least elevation of risk (13%) is observed for the variable 'number of literate members of working age'.

Logistic regression model has satisfied the goodness of fit of the test as assessed by H-L ( $\chi^2$ ) test (8 d.f.) with non-significant result ( $p = 0.534$ ) whereas the goodness of fit test of the log-binomial regression model as assessed by H-L ( $\chi^2$ ) test (8 d.f.) is grossly violated ( $p = 0.0004$ ). The value of AIC is less, and the value of BIC is greater with negative sign in log-binomial model compared to logistic regression model. Neither logistic regression nor log-binomial regression model has faced the model failure convergence i.e. both the models do not show the misbehavior.

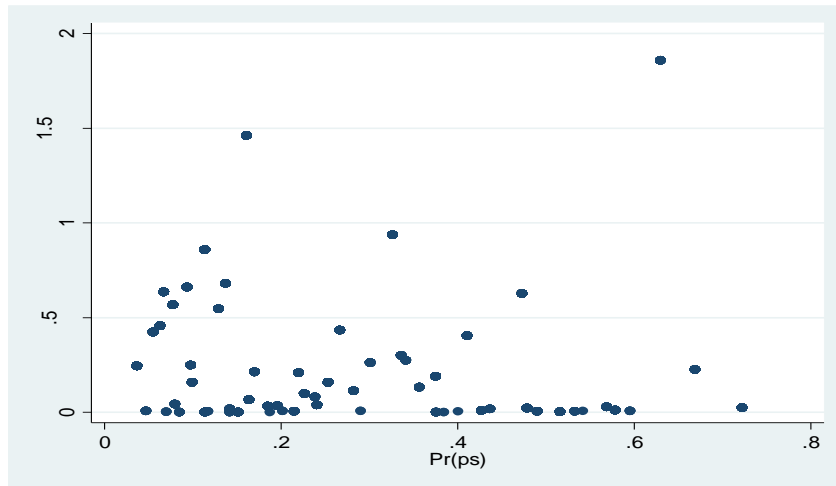
**Table 3.** Comparison of logistic and log - binomial regression model in terms of variable selection, estimates, precision of the estimates, goodness of fit of the model, AIC and BIC.

Independent variables	Logistic regression model		Log - binomial regression model		Elevation in risk (%)
	OR(95% CIE)	Width of interval	RR(95% CIE)	Width of interval	
Literacy status of household head:					
Literate	1.00		1.00		
Illiterate	2.20 (1.86 2.61)	0.75	1.68 (1.49 1.89)	0.4	52
Remittance receiving status:					
Yes	1.00		1.00		
No	1.90 (1.64 2.20)	0.56	1.45 (1.33 1.59)	0.26	45
Land holding status:					
Yes	1.00		1.00		
No	1.53 (1.31 1.78)	0.47	1.22 (1.11 1.34)	0.23	31
Access to market:					
Better	1.00		1.00		
Poor	1.77 (1.52 2.07)	0.55	1.51(1.34 1.69)	0.35	26
Number of children under 15:					
≤ 2	1.00		1.00		
> 2	4.69(4.06 5.42)	1.36	2.96(2.66 3.28)	0.62	173
No. of literate members of working-age:					
≥ 1	1.00		1.00		
0	1.29(1.07 1.56)	0.49	1.16(1.05 1.29)	0.24	13
H-L ( $\chi^2$ ) with 8 d.f	6.05, p = 0.534		28.602, p = 0.0004		
AIC	4813.844		0.808		
BIC	4860.727		- 47195.150		

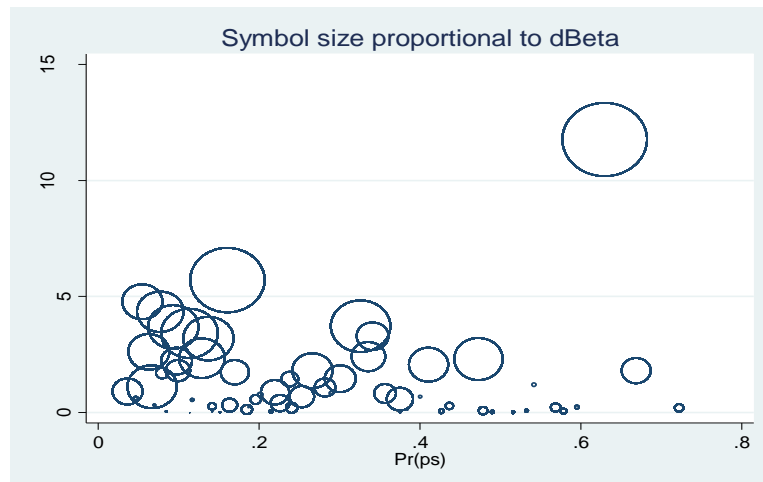
Source: Results of logistic regression are adopted from Acharya et al. (2022); Results of log-binomial regression are computed based on NLSS 2010/11 data.

### Comparison based on diagnostics of the model

The diagnostics of the fitted logistic regression model was assessed graphically through the (i) plots of  $\Delta\beta$  vs. model estimated probability, and (ii)  $\Delta\chi^2$  vs. model estimated probability with symbol size proportional to  $\Delta\beta$  (Figure2(a) and 2(b)). Both the figures have reasonably satisfied the diagnostics of the model through visual assessment except 2 data points greater than 1 in Figure2 (a), and the value of  $\Delta\chi^2$  and  $\Delta\beta$  are not influenced by covariate patterns except for one covariate (Figure2 (b)).



**Fig. 2(a).** Plot of  $\Delta\beta$  versus logistic regression model estimated probability.  
(Source: Acharya et al. (2022))



**Fig. 2(b).** Plot of  $\Delta\chi^2$  versus logistic model estimated probability with  
symbol size proportional to  $(\Delta\beta)$ .  
(Source: Acharya et al. (2022))

The diagnostics of the fitted log - binomial regression model is assessed graphically by (i) leverage versus predicted value of log - binomial regression model (Figure I (a)), and by (ii) graph of  $\Delta\chi^2$  versus values of fitted log - binomial model with plotting symbol proportional to Cook's distance (Figure I (b)). Based on the visual assessment of the plots, the fitted log-binomial model (Figure I (a) & (b)) reasonably satisfies the diagnostics of the model.

### Comparison based on stability of the model

High repetition of each variable in each final model is assessed through bootstrapping resampling technique running each model 1000 times with final set of independent variables. The major objective of this method was to identify the importance of each variable in each final model through the maximum number of occurrences of each variable in each model. Naturally, the higher the repetition of occurrence of variable indicates the more importance in the model and consequently indicates the stability of the fitted model. In each model same five variables are found repeated 100% of times, and only one variable ' number of literate members of working age group' is repeated 97.4% of times. Hence, both models have satisfied the stability criteria indicating that the selected variables are almost equally important in each model.

### Comparison based on risk assessment

The risk assessment has been performed for each model based on the presence of any one, any two risk factors, etc. by running logistic and log - binomial model separately. The risk of households being poor is found increasing continuously as the number of factors increases in each model (Figure3). However, logistic regression model overestimates the risk for each factor than that of log - binomial regression model analogous to the results of the original logistic and log - binomial model we used in the analysis.

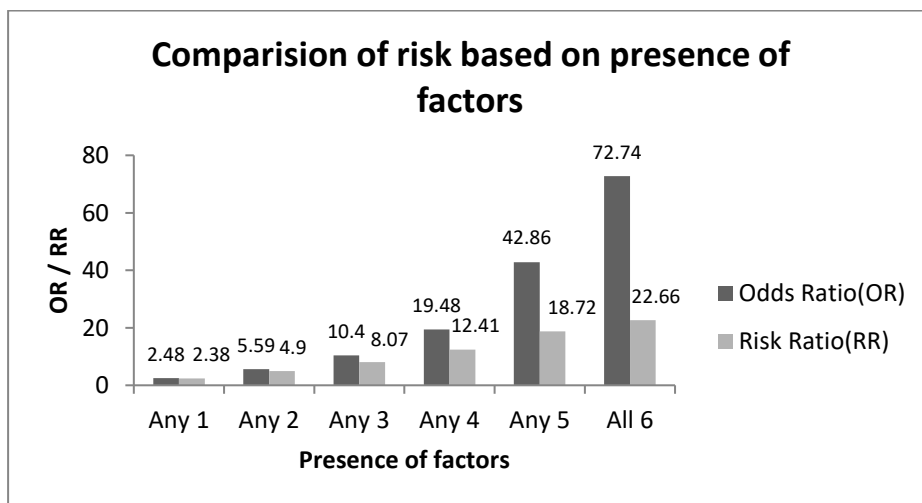


Fig. 3. Risk asseemsnt based on presence of factors for logistic and log-binomial regression model.

## DISCUSSION

The findings of the study have clearly indicated that each model has picked up the same set of six independent predictors in the final model from the same pool of variables. Both stepwise forward and backward selection procedures have been applied to select the variables in each of the final model to akin whether different selection methods serve differently in each model. However, both selection procedures have behaved in a similar manner in each model by selecting the same

set of variables. The effect size of each independent variable is overestimated in logistic regression model as compared to that of log-binomial regression model. Similar findings have been reported by other studies (Barros and Hirakata, 2003; Espelt et al., 2017; Diaz-Quijano, 2012). In logistic regression model OR varies from 1.29 to 4.69, and in log - binomial regression model RR varies from 1.16 to 2.96. There is clear greater elevation of risk in logistic regression model as compared to log-binomial regression model for each independent variable, and it varies from 13% to 173%. While comparing effect size for each variable within the model and between the models, if it is smaller or larger in a variable in logistic model, it is also smaller or larger in log - binomial model for the same variable. Just for example; the highest OR value of 4.69 for a variable 'number of children under 15 years' in logistic regression model, and the highest RR value of 2.96 for the same variable in log - binomial regression model. The precision of effect size of each variable in log-binomial model is better than that of logistic regression model as assessed by 95% CIE.

There is remarkable wider interval width of effect size in each variable in logistic regression model than that of log-binomial regression model. This finding is similar to the findings of Deddens and Petersen (2008); Barr, et al. (2016). While comparing the goodness of fit of two models, logistic regression model has satisfied the goodness of fit criteria but log - binomial regression model has grossly violated as assessed by H-L ( $\chi^2$ ) test. The violation in this regard in log-binomial model might be because of considering only categorical independent variables, but the exact reason is not known. There is not any problem of failure convergence in both models. Some studies have reported the issue of failure convergence specially while running log-binomial regression model (Williamson, et al., 2013; Barros & Hirakata, 2003; De Andrade & Carabin, 2011). The value of AIC of log - binomial model is smaller than that of logistic model; the value of BIC is larger in magnitude in log - binomial model than that of logistic model but with negative sign. The diagnostics of the logistic model assessed through (i) the graph of  $\Delta\beta$  versus model estimated probability, and (ii) the graph of  $\Delta\chi^2$  versus model estimated probability with symbol size proportional to  $(\Delta\beta)$  are reasonably satisfied. The diagnostics of the fitted log - binomial model assessed through graph of (i) leverage versus model fitted value, and (ii) the graph of  $\Delta\chi^2$  versus model fitted value with symbol proportional to Cook's distance is also reasonably satisfied. Similar findings are reported by (Blizzard & Hosmer, 2006) regarding the regression diagnostics based on the comparison of these two models using follow up study of infants. However, our finding regarding the goodness of fit of the model for log - binomial model does not support the study of Blizzard & Hosmer (2006) but supports for the logistic regression model. While comparing the stability of the fitted model evaluated using bootstrapping resampling method of running each model 1000 times, all five variables are repeated 100% times except one variable's repetition of 97% of times. This signifies that both finally fitted models can be considered as stable.

### Limitation

All independent variables used in both logistic and log - binomial model are of categorical type. The reasons behind the consideration of categorized variables are ease of interpretations of effect size and effective implementation in policy implications for comparing the groups such as advantaged vs. non advantaged groups, etc.

## CONCLUSION

Both logistic and log- binomial model possess same behavior in terms of selection of variables in the final model, diagnostics of the fitted model, stability of the model and issue of failure convergence. However, logistic regression model overestimates the effect size, wider CIE of effect size than that of log-binomial model. The value of AIC is smaller in log- binomial model than that of logistic model. Comparison based on the estimates, precision of estimates, and AIC, log-binomial model is better than logistic regression model in this cross sectional poverty data of Nepal. Logistic regression model satisfies the goodness of fit but log - binomial model grossly violates. Logistic regression model is better than log - binomial regression model for this poverty data comparatively based on the entire comparison including goodness of fit of the model. Nonetheless, log - binomial model is a good alternative for logistic regression model, especially for not overestimating effect size and its better precision.

## CONFLICT OF INTEREST

The authors declared absence of conflict of interest.

## ACKNOWLEDGEMENTS

We would like to acknowledge Research Committee, Central Department of Statistics, TU for comments and suggestions, and to University Grants Commission Nepal for Ph.D. fellowship of this work as it is a part of Ph.D. research work. We also like to thank Prof. Leigh Blizzard, Menzies Institute for Medical Research, University of Tasmania for providing us STATA codes for computing H-L Chi-Square in log -binomial model, and would like to acknowledge unknown reviewers whose comments and suggestions have greatly helped to improve the manuscript.

## REFERENCES

- Acharya, K. P., Khanal, S. P., & Chhetry, D. (2022). Factors Affecting Poverty in Nepal - A Binary Logistic Regression Model Study. *Pertanika Journal of Social Science and Humanities*, 30(2). doi: <https://doi.org/10.47836/pjssh.30.2.12>
- Altman, D. G., & Anderson, P. K. (1989). Bootstrap investigation of the stability of a Cox regression model. *Statistics in Medicine*, 8(7), 771-783. doi: 10.1002/sim.4780080702
- Baicus, C. (2003). Relative risks or odds ratios? *Canadian Medical Association Journal*, 168(12), 1529.
- Barr, Margo L., Clark, Robert, & Steel, D. G. (2016). *Examining associations in cross-sectional studies*. National Institute for Applied Statistics Research Australia, University of Wollongong. Retrieved from <https://ro.uow.edu.au/niasrawp/35>
- Barros, A. J. D., & Hirakata, V. N. (2003). Alternatives for logistic regression in cross-sectional studies: An empirical comparison of models that directly estimate the prevalence ratio. *BioMed Central Medical Research Methodology*, 3(21). doi: <https://doi.org/10.1186/1471-2288-3-21>
- Blizzard, L., & Hosmer, D. W. (2006). Parameter estimation and goodness-of-fit in log binomial regression. *Biometrical Journal*, 48, 5–22. doi: 10.1002/bimj.200410165

- Central Bureau of Statistics. (2011). *Nepal Living Standard Survey (2010/11)*. Statistical Report, Volume One. Central Bureau of Statistics, National Planning Commission Secretariat, Government of Nepal.
- Chen, C. H., & George, S. L. (1985). The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. *Statistics in Medicine*, 4(1), 39-46. doi: 10.1002/sim.4780040107
- Cook, T. D. (2002). Advanced statistics: Up with odds ratios! A case for odds ratios when outcomes are common. *Academic Emergency Medicine*, 9, 1430-1434. doi: 10.1111/j.1553-2712.2002.tb01616.x
- Davies, H. T. O., Crombie, I. K., & Tavakoli, M. (1998). When can odds ratios mislead?. *British Medical Journal*, 316(7136), 989-991. doi: 10.1136/bmj.316.7136.989
- De Andrade, B. B., & Carabin, H. (2011). On the estimation of relative risks via log binomial regression. *Revista Brasileira de Biometria*, 29(1), 15.
- Deddens, J. A., & Petersen, M. R. (2008). Approaches for estimating prevalence ratios. *Occupational and environmental medicine*, 65(7), 501-506. doi: <https://doi.org/10.1136/oem.2007.034777>
- Diaz-Quijano, F. A. (2012). A simple method for estimating relative risk using logistic regression. *BMC Medical Research Methodology*, 12(1), 1-6. doi: 10.1186/1471-2288-12-14.
- Espelt, A., Mari-Dell'Olmo, M., Penelo, E., & Bosque-Prous, M. (2017). Applied prevalence ratio estimation with different Regression models: An example from a cross-national study on substance use research. *Adicciones*, 29(2), 105-112. doi: 10.20882/adicciones.823
- Gallis, J. A., & Turner, E.L. (2019). Relative measures of association for binary outcomes: Challenges and recommendations for the global health researcher. *Annals of Global Health*, 85(1): 137, 1-12. doi: <https://doi.org/10.5334/aogh.2581>
- Greenland, S., & Thomas, D. C. (1982). On the need for the rare disease assumption in case-control studies. *American Journal of Epidemiology*, 116(3), 547-553.
- Greenland, S. (1987). Interpretation and choice of effect measures in epidemiologic analyses. *American Journal of Epidemiology*, 125(5), 761-768.
- Greenland, S., Thomas, D. C., & Morgenstern, H. (1986). The rare-disease assumption revisited: A critique of "estimators of relative risk for case-control studies. *American Journal of Epidemiology*, 124(6), 869-883.
- Holcomb, W. L., Chaiworapongsa, T., Luke, D. A., & Burgdorf, K. D. (2001). An odd measure of risk: use and misuse of the odds ratio. *Obstetrics & Gynecology*, 98(4), 685-688. doi: [https://doi.org/10.1016/s0029-7844\(01\)01488-0](https://doi.org/10.1016/s0029-7844(01)01488-0)
- Katz, K. A. (2006). The (relative) risks of using odds ratios. *Archives of Dermatology*, 142(6), 761-764.
- Khanal, S.P., Sreenivas, V., & Acharya, S. K. (2019). Comparison of Cox proportional hazards model and lognormal accelerated failure time model: Application in time to event analysis of acute liver failure patients in India. *Nepalese Journal of Statistics*, 3, 21-40. doi: <https://doi.org/10.3126/njs.v3i0.25576>



- Lee, J. (1994). Odds ratio or relative risk for cross-sectional data. *Int J Epidemiol*, 23(1), 201–203. doi: <https://doi.org/10.1093/ije/23.1.201>
- Martinez, B. A. F., Leotti, V. B., Silva, G. D. S. E., Nunes, L. N., Machado, G., & Corbellini, L. G. (2017). Odds ratio or prevalence ratio? An overview of reported statistical methods and appropriateness of interpretations in cross-sectional studies with dichotomous outcomes in veterinary medicine. *Frontiers in Veterinary Science*, 4, 193. doi: <https://doi.org/10.3389/fvets.2017.00193>
- McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall
- McNutt, L.-A., Xiaonan Xue C. W., and Hafner J. P. (2003). Estimating the relative risk in cohort studies and clinical trials of common outcomes. *American Journal of Epidemiology*, 157, 940–943. doi: [10.1093/aje/kwg074](https://doi.org/10.1093/aje/kwg074)
- Newman, S. C. (2001). *Biostatistical Methods in Epidemiology* (pp 35-40). New York: Wiley
- Olkin, I. (1998). Letter to the editor. *Evidence-Based Medicine* 3, 71.
- Ranganathan, P., Aggarwal, R., & Pramesh, C. S., (2015). Common pitfalls in statistical analysis: Odds versus risk. *Perspectives in Clinical Research*, 6(4), 222-224. doi: <https://doi.org/10.4103%2F2229-3485.167092>
- Robbins, A. S., Chao, S. Y., & Fonseca, V. P. (2002). What's the relative risk? A method to directly estimate risk ratios in cohort studies of common outcomes. *Annals of Epidemiology*, 12, 452–454. doi: [10.1016/s1047-2797\(01\)00278-2](https://doi.org/10.1016/s1047-2797(01)00278-2)
- Sackett, D. L., Deeks J. J., & Altman, D. G. (1996). Down with odds ratios!. *Evidence Based Medicine*, 1, 164–166. Retrieved from <https://ebm.bmj.com/content/ebmed/1/6/164.full.pdf>
- Saurbrei, W., & Schumacher, M. (1992). A bootstrap resampling procedure for model building application to the Cox regression model. *Statistics in Medicine*, 11, 2093-2109. doi: <https://doi.org/10.1002/sim.4780111607>
- Schwartz, L. M., Woloshin, S., & Welch, H. G. (1999). Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. *New England Journal of Medicine*, 341(4), 279–283.
- Viera, A. J. (2008). Odds ratios and risk ratios: What's the difference and why does it matter?. *Southern Medical Journal*, 101(7), 730-734. doi: <https://doi.org/10.1097/smj.0b013e31817a7ee4>
- Wacholder, S. (1986). Binomial regression in GLIM: Estimating risk ratios and risk differences. *American Journal of Epidemiology*, 123, 174–184. doi: [10.1093/oxfordjournals.aje.a114212](https://doi.org/10.1093/oxfordjournals.aje.a114212)
- Walter, S. (1998). Letter to the editor. *Evidence-Based Medicine*, 3(71).
- Williamson, T., Eliasziw, M., & Fick, G. H. (2013). Log-binomial models: exploring failed convergence. *Emerging Themes in Epidemiology*, 10(14). doi: <https://doi.org/10.1186/1742-7622-10-14>
- Zocchetti C., Consonni D., and Bertazzi P. A. (1995). Estimation of prevalence rate ratios from cross-sectional data. *International Journal of Epidemiology*, 24(5), 1064–1067.

---

**Reference** to this paper should be made as follows:

Acharya, K. P., Khanal, S. P., & Chhetry, D. (2022). On the use of logistic regression model and its comparison with log-binomial regression model in the analysis of poverty data of Nepal. *Nep. J. Stat*, 6, 63-79.

---

## **APPENDIX– D**

### **Paper Presented in the Conference**

Acharya K.P., Khanal S.P., & Chhetry D. (2023). Comparison of Logistic and Log-binomial Regression Model with Reference to Household Poverty of Nepal. A paper presented on the the 3<sup>rd</sup> International Conference on Application of Mathematics to Nonlinear Sciences ( May, 25-28), Pokhara Nepal.

Acharya K.P., Khanal S.P., & Chhetry D. (2022). Dichotomization of Quantitative Variables in Poverty Analysis. A paper presented on the the 9th National Conference on Science and Technology (June 26-28), Kathmandu Nepal.

Acharya K.P., Khanal S.P., & Chhetry D. (2021). Factors Affecting Poverty in Nepal: A Binary Logistic Regression Model. A paper presented on the the 12th ORSN National Conference (February, 1-2), Kathmandu Nepal.

Acharya K.P., Khanal S.P., & Chhetry D. (2016). Cross-sectional Analysis of Poverty Profiles and Socioeconomic Factors of Welfare among Nepalese households. A paper presented on the 7th National Conference on Science and Technology (March, 29-31), Kathmandu Nepal.

### **Poster presented in the Ph. D. Festival**

Acharya K.P., Khanal S.P., & Chhetry D. (2023). Identification of Risk Factors of Household Poverty in Nepal Using Statistical Models in the Ph. D. Festival 2023 IOST (October, 9-10), Tribhuvan University.

**Third International Conference on  
Applications of Mathematics to Nonlinear Sciences  
(AMNS-2023)**

May 25-28, 2023, Pokhara, Nepal

<http://anmaweb.org/AMNS-2023>



This certificate is awarded to

**Krishna Acharya**

for presenting a talk on

*Comparison of Logistic and Log-binomial Regression Model  
with Reference to Household Poverty of Nepal*

in a contributed session of the

**Third International Conference on Applications of  
Mathematics to Nonlinear Sciences  
(AMNS-2023)**

held in Pokhara, Nepal from May 25-28, 2023.

Dr. Gokarna Aryal

Dr. Shree Ram Khadka

Date: May 28, 2023



**Conveners**

Dr. Gokarna Raj Aryal  
Department of Mathematics and Statistics  
Purdue University Northwest, USA  
[aryalg@pnw.edu](mailto:aryalg@pnw.edu), <http://www.anmaweb.org>

Dr. Shree Ram Khadka  
Central Department of Mathematics  
Tribhuvan University, Nepal  
[shree.khadka@cdmath.tu.edu.np](mailto:shree.khadka@cdmath.tu.edu.np), <https://www.cdmath.tu.edu.np>



**Nepal Academy of Science and Technology (NAST)**

**CERTIFICATE OF PARTICIPATION**

Awarded to

**Krishna Prasad Acharya**

for Presentation in Oral / ~~Poster~~ / ~~Participation~~ in the  
**9th National Conference on Science and Technology**

June 26-28, 2022 (Asar 12-14, 2079)

Khumaltar, Lalitpur, Nepal

Ms. Luna Vajra  
Chief, Promotion Division

Prof. Dr. Mahesh K. Adhikari  
Secretary

Dr. Sunil Babu Shrestha  
Vice Chancellor





**12<sup>th</sup> ORSN National Conference**  
**Operations Research and Development**

**February 1-2, 2021**

**certificate**

**presented to**

**Mr. Krishna Prasad Acharya**

**Central Department of Statistics, Tribhuvan University**

**For presenting paper titled**  
**"Factors Affecting Poverty in Nepal: A Binary Logistic**  
**Regression Model"**

**On the occasion of**  
**14<sup>th</sup> Annual Day of**  
**Operational Research Society of Nepal**

**Prof. Dr. Sunity Shrestha**  
**Conference Chair**

**Dr. Govinda Tamang**  
**President, ORSN**



# Nepal Academy of Science and Technology

## CERTIFICATE

Awarded to

*Krishna Prasad Acharya*

for active participation/paper presentation/poster presentation

in

THE 7<sup>th</sup> NATIONAL CONFERENCE ON SCIENCE AND TECHNOLOGY

SCIENCE, TECHNOLOGY AND INNOVATION FOR NEPAL'S GRADUATION TO DEVELOPING COUNTRY STATUS

March 29-31, 2016

Kathmandu, Nepal

*Ramila Raut*

Mrs. Ramila Shrestha Raut  
Chief, Promotion Division

*Buddhi*

Dr. Buddhi Ratna Khadge  
Secretary

*Jr*

Prof. Dr. Jitendra Raj Pokharel  
Vice-Chancellor