



Binary Logistic Model to Identify the Factors Associated with Households with Bank Accounts in Nepal

Santosh Kumar Shah

Lecturer, Central Department of
Statistics, Tribhuvan University,
Kathmandu, Nepal
(santoshshah1234@yahoo.com)

Received: 12 August, 2020

Revised: 02 November, 2020

Accepted: 18 November, 2020

Published: 15 December, 2020

How to cite this paper:

Shah, S.K. (2020). Binary logistic model to identify the factors associated with households with bank accounts in Nepal. *Quest Journal of Management and Social Sciences*, 2(2), 323-336.

Copyright © 2020 by authors and
Quest Journal of Management
and Social Sciences.

This work is licensed under a
Creative Commons Attribution-Non
Commercial-No Derivatives 4.0
International License.

<https://creativecommons.org/licenses/by-nc-nd/4.0/>



Abstract

Introduction: Banks play an important role in ensuring the economic and social stability, and the sustainable growth of the economy. The savings and other accounts in financial institutions, including banks, finances, microfinances and cooperatives, enable people to execute important financial functions. Thus, households that have accounts in any of financial institutions can have access to various banking services.

Objective: The objective of the study is to identify the factors associated with households having bank accounts in Nepal.

Methods: The analysis is based on household data extracted from the dataset of Nepal Demographic and Health Survey, 2016. The dependent variable is dichotomous, as the households with bank accounts and without bank accounts in any formal financial channels. In order to identify the factors associated with households receiving financial services in Nepal, multiple logistic regression models were developed by examining the model adequacy test.

Results: The study finds that a total of 66.9% of the households had bank accounts. Several variables were found to be significant at 1% level. The predictive power of the model is found to be 31.2% and multicollinearity among the independent variables was absent. The Hosmer-Lemeshow goodness of fit test revealed that the data were poorly (p -value=0.056) fitted by the model. However, Osius-Rojek goodness of fit test ($z=0.11$; p -value=0.911), Stukel test ($Z=0.683$, p -value=0.494), likelihood ratio test ($\chi^2=2770$; p -value<0.0001) and area under receiver operating curve (79.8%) revealed that fitted model was good.

Conclusion: Multiple logistic regression model revealed that in mountainous and hilly regions, women headed households have less chances of not having bank accounts compared to the Terai region and men-headed households. The chances of having a bank account in province-2 is even worse than in Karnali and other provinces. The odds of not having bank accounts gradually decreased with the increase in size of agricultural land, wealth index, family size and the number of family members who have completed secondary education.

Keywords: Multiple Logistic Regression, Goodness of Fit, Osius and Rojek Test, Stukel's score test, Receiving Operating Curve.

JEL Classification: D10, G21, G51

Introduction

Banks are financial institutions that are known to keep deposits and flow loans. They play a very important role for the economic development of a country because they provide vital services for both consumers and businesses. Access to financial services can have a strong positive impact on economic growth and alleviating poverty. It becomes an important key to development and growth for developing countries and the countries where the economy is just emerging (Rabobank, 2005). According to Dey and Majumder (2017), access to banking services in low-income countries is viewed as a key determinant of economic well-being for households. The financial sector can help households to better manage financial decisions. They do not only make it easier to manage financial emergencies such as job loss, crop failure but also help people to establish enterprises and businesses. Increasing access to financial services for the poor can have a positive impact on economic growth, while also alleviating poverty.

Financial services in a region is typically measured by a number of people who have access to bank accounts (Beck & de la Torre, 2006; Littlefield et al. 2006). Many poor people around the world lack access to these services and thus do not have bank accounts. Despite the increase in the number of adults having accounts in a formal financial institution, only 69% of people have bank accounts in a formal financial institution worldwide, in the year 2017. The percentage in developing countries was even lesser, only 63% (Demirgüç-Kunt et. al., 2018) and only 45% in Nepal (World Bank, 2018). These “unbanked” adults lack a safe place to save their monetary capital and are likely to have only limited access to credit. Altogether, Nepal has five categories of formal banking institutions. They are Class ‘A’, Class ‘B’, Class ‘C’ and Class ‘D’ institutions, and cooperatives. Class ‘A’ refers to commercial banks, Class ‘B’ to development banks, Class ‘C’ to finance companies and Class ‘D’ to microfinance companies (Acharya, 2019). The first four categories of banks and financial institutions in Nepal are regulated by Nepal Rastra Bank under Nepal Rastra Bank Act, 2002 and Banks and Financial Institutions Act, 2017 (Nepal Rastra Bank [NRB], n.d.). And the cooperatives are under the Cooperative Department under the Cooperative Act, 2017.

Total of 241 banks and other financial institutions were registered under Nepal Rastra Bank provided services all over Nepal through 3,838 branches to 7,206 people per branch (NRB, 2015). Out of the 241 banks and financial institutions, 30 were “A” class commercial banks, 76 “B” class development banks, 47 “C” class finance companies, 38 “D” class micro-credit development banks, 15 saving and credit co-operatives and 27 NGOs and 8 other institutions (ibid). Similarly, 34,512 cooperatives were registered under the Department of Cooperative and had 6,305,581 general members with 247,827 board members in Nepal (Department of Cooperative, 2017). Bank accounts enable people to receive important financial functions like access to savings schemes, access to credit, taking loan, insurance, money transfer etc. It invariably determines access to many other financial services (Mohan, 2006). The status of households having formal bank accounts in Nepal remains limited, and lags far behind even other parts of the developing world. That could potentially have important repercussions on people’s lives. People might have difficulty in saving up large sums. They might not obtain credit from formal financial institutions, be forced to take loans from local moneylenders, and pay high interest. Thus, a study is needed to identify the factors associated with households having bank accounts in Nepal. The present study attempts to investigate the socioeconomic and demographic factors associated with the households with bank accounts using multiple binary logistic models.

Research Methods

Data and Variables

The data of Nepal Demographic Health Survey (NDHS, 2016) was used for this study. It is the fifth national representative survey conducted in Nepal, as part of the worldwide Demographic and Health

Surveys Program (Ministry of Health [MoH] et. al., 2016). Altogether, 11,040 samples were collected at household levels during the survey. Out of them, 10,893 samples were only analyzed for this study. The remaining 147 cases were omitted because these households responded with “I did not know” for the questions “How many bigha/ropani of agricultural land do members of this household own?” and “what is your level of education?”. The study is focused on the households’ access to banking services; thus the unit of analysis is the household. Details of the survey procedure and sampling design are available in the NDHS report (MoH et. al., 2017).

The dependent variable is “households with bank accounts in any formal financial institution”. This means that the households with at least one member(s) has a savings or other account in any bank, finance and microfinance, and cooperative. It is a dichotomous variable. Based on the literature review, a number of independent variables including categorical as well as continuous were taken into account. These variables were family size, ecological zone (Mountain/Hill/Terai), sex of the household head (men/women), age of the household head, education level of household head (no education/below secondary level/above secondary level), marital status of the household head (unmarried/currently, formerly and ever married), wealth index (poorest/poorer/ middle/richer/richest), size of own agriculture land (no land/less than 0.3 Ha/more than 0.3 Ha), HH rearing cattle including cows, bulls, buffalo (yes/no) in a household, whether any person had gone for employment in last 5 years in countries other than India (yes/no), place of residence (rural/urban) and the province of country (Province 1/Province 2/ Bagmati Province/Gandaki Province/Lumbani Province/Karnali Province/Sudur Paschim Province).

Data Analysis

Descriptive analysis was carried out to assess the nature and characteristics of dependent variables and independent variables. The statistical tools such as mean and standard deviation for continuous variables, and percentage for the categorical variables, were used. The independents of two attributes were compared by using Chi-square test. In NDHS, 2016, samples were weighted in order to get the actual estimates of the country. But sample weights were not considered in the entire analysis because the estimates of parameters, with and without using sample weights, were found to be similar.

As the outcome variable “household having bank account” is dichotomous, simple and multiple logistic regression models were used. The covariates for the logistic regression were screened based on the simple logistic regression model. The covariates that were found significant were further analyzed through multiple logistic model. At last, the final multiple logistic regression model was obtained after applying different measures of models adequacy test.

Logistic regression is a statistical technique for predicting the probability of an event, given a set of predictor variables. Consider that a random variable y_i denotes household’s access to banking service that can be taken as value 1, if the household has no bank account and the value is 0, if the household has a bank account:

$$\left\{ \begin{array}{l} 1, \text{ if the household has no bank account : } \\ y_i = 0, \text{ if the household has a bank account} \end{array} \right\}$$

Let $\pi_i = P(y_i=1)$ denote the probability that the i-th household has no bank account and $1-\pi_i = P(y_i=0)$ indicates the probability that the i-th household has a bank account. Consider that the multivariate logistic regression model is the model of log-odds for m covariates associated with i-th and the individual is given by:

$$\ln \frac{\pi_i}{1-\pi_i} = \beta_0 + \sum_{j=1}^m \beta_j X_j$$

Therefore,

$$\frac{\pi_i}{1 - \pi_i} = e^{\beta_0 + \sum_{k=1}^m \beta_k X_k}$$

The above model can also be written in terms of π_i as:

$$\pi_i = \frac{e^{\beta_0 + \sum_{k=1}^m \beta_k X_k}}{1 + e^{\beta_0 + \sum_{k=1}^m \beta_k X_k}}$$

$$1 - \pi_i = \frac{1}{1 + e^{\beta_0 + \sum_{k=1}^m \beta_k X_k}}$$

Where $0 \leq \pi_i \leq 1$, $i = 1, 2, \dots, n$ and x_k ($k = 1, 2, \dots, m$) are the explanatory variables. The parameters of the models β_0 and β_j were estimated on the basis of the maximum likelihood estimation, which follows Newton-Raphson iterative estimation method (McCullagh & Nelder, 1989).

Model Adequacy Test

Model adequacy test helps to check, as far as possible, whether the fitted model is correctly specified or not. Failure to address model adequacy may lead to misleading or incorrect inferences. Several measures of model adequacy tests were considered in order to inspect how well the fitted multiple logistic model matched with the observed data.

Evaluating the goodness of fit of the logistic regression models is crucial to ensure the accuracy of the estimated probabilities. The Hosmer and Lemeshow, the Osius and Rojek normal approximation test, and Stukel’s score test are considered as better methods for overall assessment of goodness-of-fit tests, when continuous covariates are present in the model (Hosmer & Lemeshow, 2000 cited in LIU, 2007). Thus, overall goodness-of-fit was assessed through Hosmer-Lemeshow goodness-of-fit test, likelihood ratio test, Osius-Rojek large sample approximation test, Stukel test and area under Receiver Operating Characteristic curve.

As the correlation increases among the independent variables, the variances of the logit parameters become inflated. Consequently we might incorrectly infer about the relationships between explanatory and response variables. Thus, multicollinearity was assessed by computation of variance inflation factor (VIF) having the threshold value of 5.

Wald statistic was used to test the significance of the estimated parameter where the null hypothesis states that particular coefficients do not differ from zero. The rejection of null hypothesis (i.e. p-value < 0.05) means that independent variable has significant contribution on the particular categories of outcome variable, relative to reference category. The predictive power of the fitted model was measured through the Pseudo R² that is Nagelkerke R². Higher Nagelkerke R² indicates that more of the variation is explained by the fitted model. The whole analysis was carried out using SPSS 20.0 and R 4.0.2 software.

Data Analysis and Results

Descriptive Analysis

The descriptive analysis is carried out to the dependent variable and all independent variables, by considering equal sample weight. Out of the total, 66.9% of the households had bank accounts in formal financial institutions and received services either from commercial and development banks, finances and microfinances, and cooperatives. Remaining of the 33.1% of the households had no bank accounts (Figure 1).

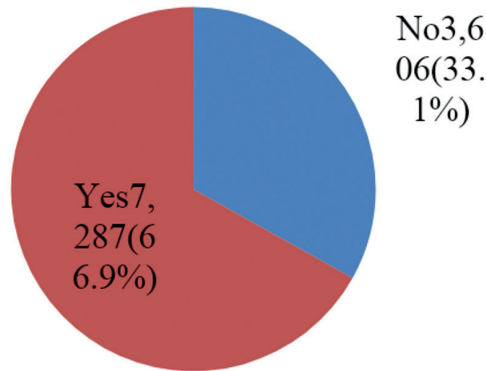
Figure 1: Distribution of household with bank account

Table 1 shows the distribution of socioeconomic and demographic characteristics of the respondents. Since sample weight is not considered during analysis, it represents only the distribution of the sample respondents but not the national scenario. Total of 47.8% of the samples were selected from the hilly regions followed by the Terai (43.9%) and mountainous regions (8.3%). However, Terai constitutes of 50.27% of the total population while the hilly and mountainous regions constitute only 43% and 6.73%, respectively (CBS, 2012). This indicates that the hilly and mountainous regions had been oversampled. Similarly, the urban area constitutes with 17% of the total population (CBS, 2012). However, the sample contains two third of the total. The spreading of financial institutions across the country was the highest in urban areas and Terai regions. This might be due to high population density. Total of 32.4% sample households were women-headed. Most of the household decisions change from men-headed to women-headed, just because the husbands were departing to work elsewhere (CBS 2005). The poor people are less likely to use banking facilities. Out of the 47.4% of the population, 25.2% were poorest and 22.2% were poorer and the rest belonged to the middle or richest categories. Literature reveals that the use of banking services increases with income, wealth and education (Beck and Brown, 2011). Total of 40.4% were uneducated. Further, about 17.2% reported that at least one of the family members had gone for foreign employment other than India in the past 5 years. Approximately 20.2% households had not owned agricultural farms and 58.2% did not have cattle including cows, buffaloes and oxen.

Table 1: Distribution of characteristics of the households

Variable	Description	Frequency	Percent
Ecological region	Mountain	906	8.3
	Hill	5,210	47.8
	Terai	4,777	43.9
Province	Province 1	1,642	15.1
	Province 2	1,619	14.9
	Bagmati	1,587	14.6
	Gandaki	1,470	13.5
	Lumbani	1,612	14.8
	Karnali	1,483	13.6
	SudurPaschim	1,480	13.6

Variable	Description	Frequency	Percent
Place of residence	Rural	4,037	37.1
	Urban	6,856	62.9
Wealth index	Poorest	2,747	25.2
	Poorer	2,416	22.2
	Middle	2,183	20.0
	Richer	1,892	17.4
	Richest	1,655	15.2
Sex of the HHH	Women	3,529	32.4
	Men	7,364	67.6
Education status of HHH	No education	4,396	40.4
	Below secondary	4,519	41.5
	Secondary & above	1,978	18.2
Marital status of HHH	Unmarried	261	2.4
	Currently, formerly, ever married	10,632	97.6
Foreign employment	No	9,019	82.8
	Yes	1,874	17.2
Size of own agriculture land	No land	2,202	20.2
	<0.3 Ha	3,988	36.6
	>0.3 Ha	4,703	43.2
Rearing cattle	No	4,558	41.8
	Yes	6,335	58.2
Total		10,893	100.0

Table 2 illustrates the summary of statistics of the age of the household head, family size and the number of members who have completed secondary education. Mean and standard deviation (SD) of the age of the household's head were 46.04 and 14.77, respectively, ranging from 15 to 95 years. This clearly indicates that approximately 68% of the respondents had ages in between 31 to 61 and 95% had ages in between 17 to 76 years. Similarly, average family size was 4.45 and the mean number of the members who had completed secondary education in the household was 0.66. Access to bank credit was positively and significantly influenced by age, household size, and education level. (Okurut 2006, Diagne and Zeller 2001).

Table 2: Summary statistics of household characteristics of continuous variables

Variable	Mean	Standard Deviation	Minimum	Maximum
Age of HHH	46.04	14.77	15.00	95.00
Number of members completed secondary education	.66	1.01	.00	13.00
Family size.	4.45	2.28	1.00	38.00

Inferential analysis

Table 3 depicts that households with bank accounts were the highest in Terai (69.5%) compared to mountainous (55.6%) and hilly (66.4%) regions. Similarly, the percentage in seven provinces ranged

from 61.9% in Province-2 to 74.6% in Gandaki. The percentage of households having bank accounts was the highest in urban area and men-headed households. The number of bank accounts was the highest (86.4%) among the household heads that had completed secondary education and above. That was followed by below secondary level (69.7%) and no education (55.2%). This clearly reveals that the percentage of households with bank accounts increases if the education level of the head of the household increases.

The result was quite interesting that the number of bank accounts increased with those households which did not have their own land, compared to those having less than 0.3 hectares of land. Total of 73.2% of households who received remittance from other than from India had bank accounts.

Table 3: Household with bank account

Variable	Description	Access to bank account		Chi-square (P-value)
		No	Yes	
Ecological region	Mountain	402 (44.4)	504 (55.6)	67.3 (<0.0001)
	Hill	1,748 (33.6)	3,462 (66.4)	
	Terai	1,456 (30.5)	3,321 (69.5)	
Province	Province-1	472 (28.7)	1,170 (71.3)	256.0 (<0.0001)
	Province-2	617 (38.1)	1,002 (61.9)	
	Bagmati	417 (26.3)	1,170 (73.7)	
	Gandaki	374 (25.4)	1,096 (74.6)	
	Lumbani	469 (29.1)	1,143 (70.9)	
	Karnali	691 (46.6)	792 (53.4)	
	Sudur Paschim	566 (38.2)	914 (61.8)	
Place of residence	Rural	1,662 (41.2)	2,375 (58.8)	188.4 (<0.0001)
	Urban	1,944 (28.4)	4,912 (71.6)	
Wealth index	Poorest	1,720 (62.6)	1,027 (37.4)	1795.0 (<0.0001)
	Poorer	839 (34.7)	1,577 (65.3)	
	Middle	572 (26.2)	1,611 (73.8)	
	Richer	345 (18.2)	1,547 (81.8)	
	Richest	130 (7.9)	1,525 (92.1)	
Sex of the HHH	Women	1,245 (35.3)	2,284 (64.7)	11.1 (<0.01)
	Men	2,361 (32.1)	5,003 (67.9)	
Education status of HHH	No education	1,970 (44.8)	2,426 (55.2)	628.6 (<0.0001)
	Below secondary	1,367 (30.3)	3,152 (69.7)	
	Secondary & above	269 (13.6)	1,709 (86.4)	
Marital status of HHH	Unmarried	91 (34.9)	170 (65.1)	0.375 (>0.05)
	Currently, formerly, ever married	3,515 (33.1)	7,117 (66.9)	
Foreign employment	No	3,060 (33.9)	5,959 (66.1)	16.1 (<0.0001)
	Yes	546 (29.1)	1,328 (70.9)	

Size of own agriculture land	No land	790 (35.9)	1,412 (64.1)	153.0 (<0.0001)
	<0.3 Ha	1,554 (39.0)	2,434 (61.0)	
	>0.3 Ha	1,262 (26.8)	3,441 (73.2)	
Rearing cattle	No	1,301 (28.5)	3,257 (71.5)	73.6
	Yes	2,305 (36.4)	4,030 (63.6)	(<0.0001)

Note: figure in parentheses shows the raw percentage

The chi-square statistics and corresponding p-values clearly revealed that the variables ecological region, provinces of Nepal, place of residence, wealth index, sex and education of household head, foreign employment, size of own agriculture land, household rearing cattle were significantly associated with the dependent variable, households having bank accounts even at 1% of level of significance. Moreover, marital status is found to be insignificant at 5%.

Statistical modeling

Simple Logistic Regression Model

Table 4 depicts estimated coefficients, odds ratios with 95% confidence intervals (CI) and their p-value obtained from the simple binary logistic model for explaining the household having a bank account. The unadjusted or crude odds ratio of the poorest households was 19.65 (CI: 16.17-23.88) with reference to the richest household, which were followed by poorer (OR-6.24, CI: 5.12-7.61), middle (OR-4.16, CI: 3.4-5.10) and richer households (OR-2.62, CI: 2.11-3.24). This clearly indicates that the risk of not having bank accounts were 19.65 times more for the poorest households, 6.24 times more for the poorer households, 4.17 times more for the middle households and 2.62 times more for the richer households compared to the richest households. It further reveals that the households with higher wealth indexed have a greater chance of having bank accounts in formal financial institutions.

Table 4: Simple binary logistic model for explaining the household access to bank account

Variable	Description	Coefficient (B)	SE (B)	P-value	Odds Ratio Exp (B)	95% Confidence Interval	
						Lower	Upper
Ecological region	Mountain	.598	.074	.000	1.819	1.574	2.103
	Hill	.141	.043	.001	1.152	1.059	1.253
	Terai (R)						
Province	Province 1	-.429	.076	.000	.651	.561	.757
	Province 2	-.006	.074	.939	.994	.860	1.150
	Bagmati	-.552	.078	.000	.576	.494	.671
	Gandaki	-.596	.080	.000	.551	.471	.645
	Lumbani	-.412	.077	.000	.663	.570	.770
	Karnali	.343	.075	.000	1.409	1.217	1.631
	Sudur Paschim (R)						
Place of residence	Rural	.570	.042	0.000	1.768	1.629	1.919
	Urban (R)						

Wealth index	Poorest	2.978	.100	.000	19.646	16.165	23.878
	Poorer	1.831	.101	.000	6.241	5.122	7.605
	Middle	1.427	.104	.000	4.165	3.400	5.102
	Richer	.962	.109	.000	2.616	2.113	3.240
	Richest (R)						
Sex of the HHH	Women	.144	.043	.001	1.155	1.061	1.257
	Men (R)						
Age of the household head		.005	.001	.001	1.005	1.002	1.007
Educational status of HHH	No education	1.641	.072	.000	5.159	4.478	5.944
	Below secondary	1.014	.073	.000	2.755	2.387	3.180
	Secondary & above (R)						
Marital status of HHH	Unmarried	.081	.132	.540	1.084	.838	1.403
	Currently, formerly, ever married (R)						
Family size		-.088	.010	.000	.916	.899	.933
Number of members completed secondary education		-.828	.032	.000	.437	.411	.465
Foreign employment	No	.222	.055	.000	1.249	1.120	1.392
	Yes (R)						
Size of own agriculture land	No land	.422	.055	.000	1.526	1.369	1.700
	<0.3 Ha	.554	.046	.000	1.741	1.590	1.906
	>0.3 Ha (R)						
Rearing cattle	No	-.359	.042	.000	.698	.643	.758
	Yes (R)						

R= Reference category

Multiple Logistic Regression Model

Simple logistic regression model was used for screening the covariates. The covariates ecological region; provinces of Nepal; place of residence; wealth index; sex, age and education of household head; foreign employment; size of own agriculture land; and household rearing cattle were found significant in the simple logistic regression (Table 4). Hence, these variables were finally run in the multiple logistic regression model. The covariate marital status was found insignificant and was dropped from the multiple logistic regression model. Both forward and backward likelihood ratio methods were run in SPSS 24.0 and showed the same result as in the final model. The only one covariate age of the household head was found insignificant at 5% level of significance, and hence it was excluded from the final model. The covariates ecological region, provinces of Nepal, place of residence, wealth index, sex and education of household head, family size, any member(s) went to foreign employment in last 5 years other than India, size of own agriculture land, household member(s) rearing cattle were found to

be significant with p -value <0.01 and included in the final model.

Table 5 illustrates the estimated coefficients, odds ratios with 95% confidence intervals (CI) and their significances of each covariate obtained from the multiple binary logistic model for explaining the household with a bank account. It provides the adjusted coefficient and odds ratio of each covariate.

Table 5: Multiple binary logistic model for explaining the household access to bank account

Variable	Description	Coefficient (B)	SE (B)	P-value	Odds ratio Exp (B)	95% Confidence Interval	
						Lower	Upper
Constant		-2.625	0.178	0.000			
Ecological region	Mountain	-0.365	0.104	0.000	0.694	0.566	0.851
	Hill	-0.353	0.072	0.000	0.702	0.61	0.809
	Terai (R)						
Province	Province-1	-0.118	0.089	0.183	0.888	0.746	1.058
	Province-2	0.392	0.094	0.000	1.48	1.231	1.779
	Bagmati	-0.143	0.096	0.135	0.867	0.718	1.046
	Gandaki	-0.278	0.098	0.005	0.757	0.625	0.918
	Lumbani	-0.086	0.09	0.338	0.917	0.769	1.094
	Karnali	0.19	0.091	0.037	1.209	1.011	1.445
	Sudur Paschim (R)						
Place of residence	Rural	0.769	0.052	0.000	2.157	1.946	2.39
	Urban (R)						
Wealth index	Poorest	3.039	0.119	0.000	20.886	16.531	26.388
	Poorer	1.822	0.113	0.000	6.181	4.956	7.709
	Middle	1.324	0.111	0.000	3.759	3.024	4.672
	Richer	0.839	0.114	0.000	2.313	1.85	2.893
	Richest (R)						
Sex of the HHH	Women	-0.185	0.054	0.001	0.831	0.747	0.924
	Men (R)						
Educational status of HHH	No education	0.487	0.098	0.000	1.627	1.344	1.971
	Below secondary	0.182	0.096	0.058	1.2	0.994	1.449
	Secondary & above (R)						
Family size		-0.107	0.013	0.000	0.899	0.877	0.921
Number of members completed secondary education		-0.259	0.041	0.000	0.772	0.712	0.837
Foreign employment	No	0.142	0.065	0.028	1.152	1.015	1.308
	Yes (R)						

Size of own agriculture land	No land	0.692	0.071	0.000	1.997	1.738	2.295
	<0.3 Ha	0.22	0.054	0.000	1.246	1.12	1.385
	>0.3 Ha (R)						
Rearing cattle	No	0.246	0.06	0.000	1.279	1.138	1.438
	Yes (R)						

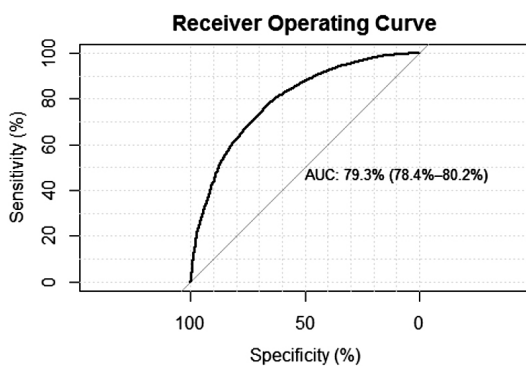
R= Reference category

Concerning the overall fit of the multiple logistic regression model, Hosmer-Lemeshow test is widely used. Hosmer-Lemeshow test statistic gives χ^2 statistics value 15.16 with p-value>0.05 and revealed that the fitted model is good. However, the p-value of test 0.056 is just greater than the threshold value of 5%. Literature shows that as the sample size gets large, the rejection rate of Hosmer and Lemeshow test statistic for acceptable models becomes high, thus a significant or poor Hosmer-Lemeshow test does not mean that a predictive model is not useful or suspicious (Kramer & Zimmerman, 2007). Therefore, other confirmatory tests including likelihood ratio test, Osius-Rojek large sample approximation test, Stukel test and area under Receiver Operating Characteristic curve were carried out to assess the goodness of fit of the model.

The likelihood ratio test (LRT) which is based on the -2 Log Likelihood (LL) produces a chi-square statistic to test the null hypothesis that there is no significant difference between the models without explanatory variables and the model with explanatory variables. The test revealed that models with explanatory variables and without explanatory variables is significantly different ($\chi^2=2770$; p-value<0.0001). This means explanatory variables had a significant effect on the outcome variable. Hence the fitted model is good.

Osius and Rojek (1992) proposed normal goodness of fit tests for multinomial models with large degrees of freedom, which can be applied to the binary cases also. It is the normal approximation to the Pearson Chi-square statistic. Osius-Rojek goodness of fit tests was found insignificant ($z=0.11$; p-value=0.911), which means that the fitted model is good.

Stukel (1988) proposed a test derived from score test, in which two new variables are created based on linear predictor from the fitted model and then extended logistic regression is run by keeping two additional variables in the fitted model. In the Stukel test, the null hypothesis states that coefficients of new variables are zero. For the fitted model, there was no evidence to reject the null hypothesis ($Z=0.683$, p-value=0.494). This also shows that fitted the model is good.



Receiver Operating curve (ROC) is a graphical plot to test the goodness of fit. ROC curve is simply a plot of the values of sensitivity against specificity where sensitivity and specificity is calculated at different cut-points from 0 to 1. The area under ROC curve (AUC) provides an overall measure of fit of the model as 0.7-0.8 considered “good”, 0.8-0.9 excellent and >0.9 outstanding (Dardis, 2016). Figure 2 revealed that AUC was found to be 79.3% (CI: 78.4%-80.2%) which lies in the range of almost excellent. Hence, the fitted model is good.

The VIF of all covariates were found to be less than 3, which indicates that there is no multicollinearity between the covariates. The predictive power of the fitted model is 31.2% assessed through Nagelkerke R^2 (pseudo R^2). This shows that 31.2% of the variation in the dependent variable is explained by the model.

Discussion

The multiple logistic regression model explored the factors associated with households having bank accounts. Total of 11 factors, out of 13, were significant predictors that changed the probability of households without bank accounts to households with bank accounts. The people who were living in the mountainous regions had 1.819 (CI: 1.571-2.10, p -value<0.0001) times and the hilly regions had 1.152 (CI: 1.06-1.25, p -value<0.01) times more chances of not having bank accounts compared to the people living in the Terai regions. However, controlling the effect of other covariates, the relation became reversed. Thus, the people living in the mountainous and the hilly regions had respectively 30.6% and 29.8% less chances of not having bank accounts compared to the people living in the Terai regions.

Odds of not having bank accounts were higher among women-headed households and lower among the household who did not rear cattle. However, the scenario became reversed after adjusting the covariates. Women-headed households had less and households who rear cattle had more chances of not having accounts in the banks or cooperatives. Dey and Majumder (2017) also justify the results by stating that women-headed households have greater access to banking services as compared to men-headed households. As the financial institutions mostly existed in urban areas, it is obvious that the access to bank accounts is higher for the urban people. The study also verifies it. The likelihood of not having account was 1.768 (CI: 1.63-1.92; p -value<0.0001) times higher among the rural people compared to the urban and the figure was increased (OR=2.157; CI: 1.95-2.39; p -value<0.0001) even more after controlling the other covariates used in the model.

Dey and Majumder (2017) analyzed the data of Demographic and Health Survey (DHS) 2014 in Bangladesh and found that age, sex, marital status of household's head and education level of household's head, wealth index, access to information, land ownership of household, own homestead, division of household, and place of residence are the significant predictors of access to banking services. Orkut (2006) uncovered that age, sex, education level of household head, household size and household per capita expenditure significantly influenced the access to bank credit. Beck and Brown (2011) found that in most countries the use of banking services increased with income, wealth and education. Wealth index is one of the highly significant predictors of households having bank accounts. The probability of not having bank accounts for the poorest households were 19.65 times (CI: 16.17-23.88; p -value<0.0001) higher than the richest households. Similarly, the poorer households were 6.24 times (CI: 5.12-7.61; p -value<0.0001), middle households 4.17 times (CI: 3.40-5.10; p -value<0.0001) times and richer households 2.62 times (CI: 2.11-3.24; p -value<0.0001) more likely to not have bank accounts compared to the richest households, adjusted for other covariates. The result was also justified by Dey and Majumder (2017) and Beck and Brown (2011). Wealth index, a composite measure of a household's cumulative living standard, is the most influential factor of households having bank accounts. Higher access to the physical capitals such as televisions and bicycles; materials used for housing construction; and drinking water and sanitation facilities will decrease the risk of not having a bank account.

The result of multiple logistic regression models revealed that the people living in Province-2 and Province Karnali were significantly more likely to not get a bank account compared to the Sudur Paschim Province. Similarly, the risk of not having bank accounts was 1.152 times more for the households whose at least one member(s) had not gone for foreign employment. The result further revealed that increase in the educational status of the household head, the numbers of member(s) completed secondary education and size of agriculture land will decrease the risk of not having a bank account. Similarly, increase in the size of the family decreases the risk of not having bank accounts. Hosmer-Lemeshow test statistic revealed that the fitted model is good. However, p -value was just greater than the threshold value of 5%. Literature shows that this might be due to a large sample size.

The goodness of fit of the model was confirmed through likelihood ratio test, Osius-Rojek large sample approximation test, Stukel test and area under Receiver Operating Characteristic curve test also and found it to be a good fit.

Conclusion

Access to banking services is one of the key determinants of economic well-being for households and it helps people to manage financial emergencies better. The study is based on secondary data. Multiple logistic regression models revealed that mountain and hill people have less chances of not having bank accounts compared to Terai. Similarly, women headed households have higher chances of having bank accounts compared to men-headed. The chance of having bank accounts in Province-2 is even worse than Karnali and other provinces. People living in rural areas had less likelihood of having bank accounts. If any member of the household has gone out for foreign employment, and those households who rear cattle, have higher chances of having bank accounts. The odds of not having bank accounts gradually decreased with increase in size of agriculture land, wealth index, family size, and number of members completed secondary education. Concerning the goodness of fit, the Hosmer-Lemeshow poorly fits the model. However, Osius and Rojek normal approximation test, Stukel's score test, likelihood ratio test and Receiver Operating Characteristic fits well. Further research is needed to identify why the Hosmer-Lemeshow fits the model poorly for large samples whereas the other tests fit well.

Conflict of Interest

The authors declared that there is no conflict of interest.

Acknowledgement

Authors would like to express sincere gratitude to Dr. Bijay Kumar Singh for insightful comments and encouragement to complete this paper.

References

- Acharya, S. (2019). Legal measures of Nepalese financial system: An overview. Retrieved from <http://dx.doi.org/10.2139/ssrn.3389607>.
- Beck, T. & de la Torre, A. (2006). The basic analytics of access to financial services. World Bank Policy Research Working Paper 4026. Retrieved from http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2006/10/03/000016406_20061003154257/Rendered/PDF/wps4026.pdf.
- Beck, T. & Brown, M. (2011). "Which households use banks? Evidence from the transition economies," Working Paper Series 1295, European Central Bank. Retrieved from <https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1295.pdf>
- CBS. (2005). Poverty trends in Nepal (1995-96 and 2003-04). Nepal Planning Commission Secretariat, Government of Nepal.
- CBS. (2012). National Population and Housing Census 2011 (National report-Volume 01), National Planning Commission Secretariat, Government of Nepal.
- Dardis, C. (2016). Diagnostic Tests for Models with a Binomial Response, Package 'LogisticDx'. Retrieved from <https://cran.r-project.org/web/packages/LogisticDx/LogisticDx.pdf>.
- Demirgüç-Kunt, A., Klapper, L., Singer, D., Ansar, S., & Hess, J. (2018). The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution. Washington, DC: World Bank. Retrieved from: https://globalfindex.worldbank.org/sites/globalfindex/files/2018-04/2017%20Findex%20full%20report_0.pdf.

- Department of Cooperative, (2077). Co-operative Statistics in Details, 2074 . Ministry of Land Management, Co-operative & Poverty Alleviation, Government of Nepal. Retrieved from <http://www.deoc.gov.np/english/downloadsdetail/3/2017/76236913/>.
- Dey, S. & Majumder, S. (2017). Identifying factors that influence access to banking services in Bangladesh: A household level analysis. *The Bangladesh Journal of Scientific Research*, 30, 1-2. DOI: 10.3329/bjsr.v30i1-2.36122.
- Diagne, A. and M. Zeller. (2001). Access to credit and its impact on welfare in Malawi. International Food Policy Research Institute, Washington, D.C. Retrieved from <http://ebrary.ifpri.org/utils/getfile/collection/p15738coll2/id/86048/filename/86049.pdf>.
- Kramer, A. A. & Zimmerman J. E. (2007). Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Crit Care Med*, 35(9), 2052-2056. doi:10.1097/01.CCM.0000275267.64078.B0.
- Littlefield, E., Brigit. H. & Porteous, D. (2006). Financial Inclusion 2015: Four Scenarios for the Future of Microfinance. CGAP Focus Note No. 39. Retrieved from <https://www.cgap.org/sites/default/files/CGAP-Focus-Note-Financial-Inclusion-2015-Four-Scenarios-for-the-Future-of-Microfinance-Oct-2006.pdf>.
- LIU, Y. (2007). On Goodness of Fit of Logistic Regression Model. An abstract of a dissertation, Ph.D., Department of Statistics, College of Arts and Sciences. Retrieved from <https://krex.k-state.edu/dspace/bitstream/handle/2097/530/YingLiu2007.pdf;sequence=1>.
- MoH [Nepal], New ERA & ICF. (2017). Nepal Demographic and Health Survey 2016. Kathmandu, Nepal: Ministry of Health, Nepal. Retrieved from <https://www.dhsprogram.com/pubs/pdf/fr336/fr336.pdf>.
- McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models (2nd ed.)*. London, Chapman and Hall.
- Mohan, R. (2006). Agricultural credit in India - Status, issues and future agenda. *Economic and Political Weekly*, 61(1), 18-24. Retrieved from <http://www.rakeshmohan.com/docs/AgricreditBulletin.pdf>.
- NRB, (2015). Banking and Financial Statistics. Bank and Financial Institution Regulation Department, Statistics Division, no. 61. Retrieved from https://archive.nrb.org.np/bfr/statistics/bank_fina_statistics/Banking_and_Financial_Statistics--No_61%20July%202015.pdf.
- NRB.(n.d.). Laws & Legislation. Retrieved from https://archive.nrb.org.np/lgd/acts_ordinances.php
- Okurut, F. N. (2006). Access to credit by the poor in South Africa: Evidence from household survey data 1995 and 2000. Department of Economics, University of Botswana, Stellenbosch Economic Working Papers: 13/2006.
- Osius, G., & Rojek, D. (1992). Normal goodness-of-fit tests for multinomial models with large degrees of freedom. *Journal of the American Statistical Association*, 87(420), 1145-1152. DOI: 10.1080/01621459.1992.10476271.
- Rabobank Nederland (2005), Access to financial services in developing countries. Retrieved from https://economic.rabobank.com/contentassets/b693fc22988f44d68a29a87c7bc19d4d/access_tcm64-75165.pdf
- Stukel, T. A. (1988). Generalized logistic models. *Journal of the American Statistical Association*, 83, 426–431.
- World Bank, (2018). The Global Index Database, 2017. Retrieved from <https://globalindex.worldbank.org>.