

COMPARISON OF POVERTY STATUS AMONG PADDY GROWERS IN MID-HILLS, NEPAL USING PRINCIPAL COMPONENT ANALYSIS

¹*Sabin Dawadi, ²Ajit Khanal, ³Thaneshwar Bhandari,
⁴Prayash Pathak Chalise, ⁵Sudeep Poudel

Department of Social Sciences, Tribhuvan University, Lamjung, Nepal

*Corresponding author

ABSTRACT

Paddy contributes 20.75% to the Agricultural Gross Domestic Product in Nepal, where agriculture solely contributes 27.6 % to the National GDP. Socio-economic conditions including poverty status is directly related to paddy production, especially in the rural agricultural community. A detailed study of socio-economic status of paddy growing household is crucial to determine relevant determinants of poverty. This study compares the status of rural household poverty level of paddy growing farmers in 2 mid-hill districts of Gandaki Province: Lamjung and Tanahun. The survey research was conducted with randomly selected 696 households from 9 Municipalities across two districts from November to December in 2018. The study identified and analyzed data on 19 proxy variables which directly indicate socio-economic status. Principal Component Analysis was carried out to determine relevant variables explaining poverty. The result showed that household characteristics (foundation, outer -wall, number of sleeping rooms) and ownership status of assets like Motorcycle and Refrigerator are the indicators of poverty. An asset index was developed using a first principal component. Based on score obtained in the asset index, households were categorized into three different socio-economic groups: The Rich, The Average and The Poor. The result showed that, in both Tanahun and Lamjung district, all households categorized into The Poor, had mud bonded bricks and stones as their outer wall. None of the households categorized into The Poor, had Cement bonded foundation of the house. This study sheds light upon economic status of The Poor targeting their reduction of poverty via implementation of rice-based program.

Keywords: Asset index, Asset ownership, Principal component analysis, Poverty, Socio-economic status

1. INTRODUCTION

The development status of any country hinges upon major economic activity of its citizens. If citizens make hatful of money from their major occupation, then the country is, socio-economically, a prosperous one. However, if economic condition of people involved in the major occupation of the country is below par, then the country is plagued with poverty. The situation is analogous to rural setting of Nepal, where agriculture, or to be more precise, Paddy cultivation is primary occupation of most of the people. However, these peoples are the one, who are among the poorest class of the society. Apparently, people make a generalization that farmers are epitome of poor socio-economic status. According to CBS, 2011, 65.7% of Nepalese undertake agriculture as primary occupation. Meanwhile, 25% of the Nepalese population still lives under absolute poverty line. The situation is even worse in rural areas, where 27.4% of the population live under absolute poverty line. Despite poor economic status of farmers, agriculture still makes substantial contribution in the National GDP, with a considerable 27.6% contribution. The fact that sole share of paddy on Agricultural GDP is 20.75% corroborates that paddy cultivation is patently one of the major occupation of people in Nepal.

It is crucial to study about the relevant determinants of socio-economic condition. Such determinants help policy makers to map out the target group before implementation of any program. Principal Component Analysis (PCA), the tool used in this study, is a multivariate statistical tool which manipulates large set of correlated but meaningless data to elicit meaningful information. The most beautiful aspect of PCA is that it provides meaningful information from the data set, which was previously difficult to interpret because of its coiling. The coiling, here, refers to multi-collinearity. In simple words, when large set of data are so interrelated that true information is cloaked, PCA serves as an ideal tool to disclose the information. (Gwatkin et al. 2000; Filmer and Pritchett 2001; McKenzie 2003) have applied PCA to such data to create a socio-economic status (SES) index and then grouped households into quintiles, representing different SES levels.

This study aims to assess the socio-economic status of the paddy growing farmers. Not only the study aims to shower light upon the assets entertained by those farmers, it also focus on determining which of those assets can serve as reliable indicators of poverty. Secondly, based on the variance in asset ownership, the study aims to classify paddy growers into different socio-economic classes. It aims to compare socio-economic differences between rich and poor paddy growers. The study finally compares SES of the poorest farmers of one district with another to find out potential variations or similarities.

Measurement and mapping out of socio-economic status is crucial to understand the present scenario of the community, especially when reliable statistical tool like PCA is used. Poverty in Nepal is generally measured in terms of monetary indicator. Using data from National Living

Standard Surveys, monetary value of expenditure on multiple aspects of life, including food, education, housing and assets are calculated and simply compared against the standard poverty line. However, the collection of accurate monetary information is an exacting task (Montgomery et al. 2000), requiring extensive resources for household surveys. The new statistical method used in this study is expected to divulge true information than ever without using monetary information. Moreover, outcome of the study is expected to serve as a reliable source of basic information to the stake holders before designing poverty alleviation program in that area.

2. THEORY

Principal Components Analysis, or PCA, is a data analysis tool that is usually used to reduce the dimensionality of a large number of interrelated variables, while retaining as much of the information (variation) as possible. PCA calculates an uncorrelated set of variables, known as factors or principal components. These factors are ordered so that the first few retain most of the variation present in all of the original variables. Filmer and Pritchett (2001) validated the use of PCA for estimating SES using asset indicators to overcome the difficulty faced in collecting income or consumption data. Based on their analysis of household assets for India, Pakistan, and Nepal, they concluded that PCA serves as tenable tool to assess SES. “One of the main objectives of PCA is to try and organize the data to reduce its dimensionality with minimal loss of information as possible in the total variation” (Giri 2004).

Let ‘N’ be the number of households. Suppose we have a set of k variables. Here, each household ‘i’ represents ownership of k assets, x_{1i} to x_{ki} . A principal component is a linear combination of variables. Mathematically, a linear equation of a single component can be represented as:

$$Y_{ij} = w_{1i} x_{1j} + w_{2i} x_{2j} + w_{3i} x_{3j} + \dots + w_{ki} x_{kj}$$

The output from a PCA is a table of factor scores or weights (w) for each variable. Basically, a variable with a positive factor score contributes to the higher SES, and contrastingly, a variable with a negative factor score is associated with lower SES. The estimation of relative wealth using PCA is based on the first principal component. The asset index for household ‘i’ is the linear combination,

$$Y_i = \frac{w_1 (x_1 - \bar{x}_1)}{S_1} + \frac{w_2 (x_2 - \bar{x}_2)}{S_2} + \frac{w_3 (x_3 - \bar{x}_3)}{S_3} + \dots + \frac{w_k (x_k - \bar{x}_k)}{S_k}$$

Where, w_k is the factor score or weight carried by each variable. x_k and S_k are the mean and standard deviation of an asset x_k .

In some studies, ownership of durable assets such as a bicycle have been ascribed a negative weight by PCA (Gwatkin et al. 2000; Houweling et al. 2003; McKenzie 2003). This implies that a household with a bicycle will contribute negatively in the asset index score, subsequently lowering the SES, than a household that does not own a bicycle. The main reason for this might be due to ownership of bicycle being more strongly correlated with variables that are associated with lower SES, for instance lower housing quality.

3. MATERIALS AND METHODS

3.1 Research design and study area:

The study is based upon survey type of research design. Primary data were collected via pre-tested questionnaire. Initially, the questionnaire was tested in Harrabot Village of Lamjung district. Problems faced while asking the question by the enumerators, ambiguity of answers given by respondents were carefully contemplated, discussed, and eventually, changes were made in the questionnaire. The Survey collected data from 696 households explicitly involved in paddy production in the monsoon season of 2018. Here, data collection is completely focused do elicit reliable information from farmers strictly involved in paddy production. The study selected two districts of western mid-hills of Nepal: Lamjung and Tanahun as sampling frame.

3.2 Data Collection:

The study used structured and semi-structured questionnaire to collect data from three districts of mid-hills from November to December 2018. While selecting villages for data collection, simple random sampling was prioritized. In Lamjung, data was collected from 337 households from randomly selected 28 villages. From each of these villages, 12 households were randomly selected. Those 28 villages represent 6 municipalities and rural municipalities (Besisahar, Sundarbazar, Dordi, Dudhpokhari , Madhyanepal and Rainas). In Tanahun, data was collected from 359 households from randomly selected 30 villages representing 3 different municipalities (Bhanu, Byas and Bandipur). All in all, primary data was collected from total 696 households from the mid-hills, indicating that the collected source of information is representative of the entire western mid-hills. The study collected Binary and categorical, data as shown in the Table 1.

Table 1: Types of ownership assets

Variables	Data type
Housing Characteristics	
I. Foundation of house	
1. Mud –bonded foundation	Yes/No
2. Wooden pillar	Yes/No
3. Cement –bonded (without pillar)	Yes/No
4. Cement –bonded (with pillar)	Yes/No
II. Outer wall of house	
5. Mud –bonded bricks or stones	Yes/No
6. Cement –bonded bricks or stones	Yes/No
III. Roof of house	
7. Straw or tiles	Yes/No
8. Galvanized Iron or Tin	Yes/No
9. Concrete cement	Yes/No
IV. Number of sleeping rooms	
10. One or two	Yes/No
11. Three or four	Yes/No
12. Five or more	Yes/No
V. Window type	
13. Glass window	Yes/No
Farmer’s Household Assets	
14. Television	Yes/No
15. Refrigerator	Yes/No
16. Motorcycle	Yes/No
17. More than two Cows/Ox	Yes/No
18. More than two Buffaloes	Yes/No
19. Adoption of Stress Tolerant Rice Varieties	Yes/No

3.3 Data Analysis:

To measure SES, studies have used variables such as ownership of land (Filmer and Pritchett 2001), farm animals and type of housing, rented or self-owned (Schellenberg et al. 2003), level of education of head of household, and other economic proxies such as primary occupation of household head (Cortinovis et al. 1993). Montgomery et al. (2000) identified the absence of a 'best practice' approach of selecting variables to proxy living standards, as, in many studies, variables were chosen on an 'ad-hoc' basis. This study identifies 19 proxy variables which are directly associated with SES of paddy growers of mid-hills. Information about ownership of durable assets like Television, Computer, Refrigerator etc are binary data (Yes/No). Regarding some variables like number of Cows, number of Buffaloes and number of sleeping rooms, data were initially collected numerically, however, they were categorized into continuous set of data prior to analysis. Since, most of the data are binary or categorical in nature, the study has managed to parry potential source of misleading information, caused due to dishonest answers given by respondents.

The study used SPSS 23.0 version for data analysis. After identification of variables directly associated with SES, they were converted into binary and categorical data to facilitate the analysis and the interpretation. Descriptive statistics of each variable was carried out to determine their mean and standard deviation. Mean value, which generally remain between 0 and 1 for binary data, gives the average ownership of the asset. The study assigns value '0' to 'No ownership' and value '1' for 'Yes ownership' of any asset. In this regard, for instance, the mean value of the asset Television is 0.82 means that 82% of respondents entertain the asset.

The number of principal components extracted can be manually demarcated, and a common method used is to extract components having Eigen value greater than one. McKenzie (2003) considered the use of more than one principal component in characterizing household SES, scrutinizing whether they are related to non-durable consumption, and asserted that only the first principal component (one with the maximum Eigen value) is enough for measuring wealth. It is presumed that the first principal component is the best measure of economic status (Houweling et al. 2003).

The study considered Kaiser criterion as a basis for factor extraction and extracted Components having Eigen value greater than 1. Moreover, comprehensive study through observation of scree-plot and through Parallel Analysis using Monte Carlo PCA software was also carried out to confirm the number of factors to be extracted. Bartlett's test of sphericity at 5% level of significance, was used to check level of significance of data set. Similarly, Kaiser Meyer Olkin measure of sampling adequacy was tested which must be at least greater than 0.6. Positive result in both of the relevance -assessment tool provides green signal to proceed to further analysis.

First round of Principal Component Analysis was carried out and component correlation matrix was studied to check the correlation among the components. Since, the components have low correlation, orthogonal rotation of type: Varimax was selected. Selection of orthogonal factor rotation also leads to simple and easier interpretation. Total variance explained by Principal Component 1 was studied and few more rounds of PCA were carried out such that first principal component (PC_1) explains maximum variance. Factor score of all the variables on the first component were studied and based on the first component, an asset index was developed. Based on score obtained by a household in the asset index, they were categorized into five quintiles having equal number of households. First quintile, representing 20% of the households having highest asset index score is termed as The Rich. Middle two quintiles, second and third, representing 40% of the households, is termed as The Average. Finally, remaining 2 quintiles, representing 40% of households having lowest asset index score is termed as The Poor. The study emphasizes on the poorest quintiles targeting their reduction of poverty. Cross tabulation of three poverty groups (The Rich, The Average and The Poor) is carried out to study the distribution of assets among the groups. The comprehensive method, explained above, is separately carried out for both districts. The study finally compares SES of The Poor from one district with another to find out potential variations or similarities.

4. RESULTS AND DISCUSSIONS

Before proceeding into further analysis, it is crucial to assess the suitability of data set. Significant result was found in Bartlett's test of Sphericity in data set of both districts. The value of Kaiser Meyer-Olkin Measure of Sampling Adequacy was also found to be greater than the threshold level of 0.6 in both districts. Positive result in KMO and Bartlett's test assured relevance of data set to interpret meaningful result.

Using Kaiser criterion—extraction of components having Eigen Value greater than one, eight and ten components were extracted Lamjung and Tanahun respectively. In Tanahun district, ten components explaining 81.79 % of variance were extracted where first principal component alone explained 16.95% of variance with the Eigen Value of 3.22. In Lamjung district, eight components explaining 73.14% variance were extracted. PC_1 alone explained 19.2% of variance with the Eigen value of 3.64.

Result of PCA is a factor score table, where each variable is assigned with a score, known as factor score or loadings as shown in Table 2. Higher and positive factor score is associated with higher asset index score and eventually higher SES. The result showed that variables like ownership of housing characteristics like foundation of house (cement –bonded pillar system), outer wall of house (cement –bonded bricks or stones), ownership of Refrigerators and Television have relatively higher factor scores. These variables serve as relevant determinants to

map out poverty. Similarly, negative factor score is associated with lower SES. The result displayed that asset variables like foundation of house (mud-bonded foundation), mud bonded outer wall, roof of house (tiles or straw) carry negative factor scores. Ownership of such assets contributes to lower SES of households. Socio-economic classification (into three groups) was cross tabulated with all variables and the result was displayed showing distribution of assets among the groups (Table 3). The table shows how ownership of any asset helps to predict SES of a household. For instance, according to the result, as displayed in Table 3, none of the households having cement bonded foundation of the house (pillar system) were categorized into poor in both districts. Similarly, none of the households having cement –bonded bricks or stones as their outer wall, were categorized into The Poor category. This indicates that, based on this study, households having cement –bonded foundation and cement bonded outer wall cannot be considered poor at all.

Ownership of assets like Refrigerator and Motorcycle also showed significant result, particularly in Lamjung district. Out of all the households possessing Refrigerator in their house, only 3.3% of them were categorized in the Poor, meanwhile, 1.2 % of households owning Motorcycle, were categorized under the Poor. In this way, inter validity of all the variables or their distribution among socio-economic groups are explicitly explained in Table 3.

Since, one of the objectives of the study is to portray the socio-economic condition of poorest paddy growing farmers, the result focused on proportion of these farmers enjoying these assets. Table 4 and Table 5 depict some significant results regarding percentage of The Poor possessing the assets. Interpretation of result from these tables proves that ownership of assets like cement – bonded foundation, cement –bonded outerwall, Refrigerator, Motorcycle is associated with economic prosperity while ownership of assets like mud –bonded foundation of house, mud – bonded outer wall are associated with poor economic condition . in other words, ownership of such assets is epitome of poverty.

Table 2: Results from Principal Component Analysis

Asset Variables	LAMJUNG			TANAHUN		
	Factor score	Mean	Standard deviation	Factor score	Mean	Standard deviation
Foundation of house						
Cement –bonded pillar system	0.846	0.34	0.47	0.541	0.22	0.41
Cement –bonded without pillar	-0.024	0.17	0.37	0.450	0.25	0.43
Wooden –pillar foundation	-0.390	0.24	0.42	-0.130	0.19	0.39

Mud –bonded foundation	-0.512	0.26	0.44	-0.777	0.34	0.47
Outer wall of house						
Cement –bonded bricks or stones	0.834	0.54	0.5	0.949	0.43	0.5
Mud –bonded bricks or stones	-0.834	0.46	0.5	-0.949	0.57	0.5
Roof of house						
Concrete cement	0.19	0.36	0.48	0.032	0.15	0.36
Galvanized Iron or Tin	-0.009	0.47	0.50	0.006	0.52	0.50
Tiles or Straw	-0.13	0.17	0.37	-0.037	0.32	0.46
Sleeping rooms						
One or two rooms	-0.147	0.09	0.28	-0.137	0.41	0.49
Three or four rooms	-0.104	0.51	0.50	0.048	0.50	0.50
Five or more rooms	0.191	0.40	0.49	0.152	0.09	0.28
Other assets						
Glass window	0.078	0.41	0.49	0.021	0.45	0.49
Refrigerator	0.632	0.82	0.42	0.247	0.28	0.45
Television	0.002	0.26	0.44	-0.026	0.88	0.45
Motorcycle	0.580	0.23	0.42	0.097	0.17	0.37
More than 2 cows	0.035	0.07	0.25	-0.005	0.08	0.27
More than 2 buffaloes	-0.232	0.13	0.34	0.068	0.09	0.28
Stress tolerant rice varieties adoption	0.2	0.61	0.48	0.065	0.60	0.49

Table 3: Internal Validity of Asset Index: Results based on the First Principal Component (Lamjung and Tanahun districts)

Variables	District	The Rich (top 20%)	The Average (middle 40%)	The Poor (lowest 40%)	Total
Foundation of house					
Cement –bonded pillar system	Lamjung	59.3%	40.7%	0%	100%
	Tanahun	64.1%	35.9%	0%	100%
Cement –bonded without pillar	Lamjung	0%	87.5%	12.5%	100%
	Tanahun	25.3%	74.7%	0%	100%

Wooden –pillar foundation	Lamjung	0%	27.5%	72.5%	100%
	Tanahun	0%	60.9%	39.1%	100%
Mud –bonded foundation	Lamjung	0%	20.5%	79.5%	100%
	Tanahun	0%	1.7%	98.3%	100%
Outer wall of house					
Cement –bonded bricks or stones	Lamjung	37%	63%	0%	100%
	Tanahun	47.4%	52.6%	0%	100%
Mud –bonded bricks or stones	Lamjung	0%	13.5%	86.5%	100%
	Tanahun	0%	28.8%	71.2%	100%
Roof of house					
Concrete cement	Lamjung	20.7%	39.7%	39.7%	100%
	Tanahun	16.4%	49.1%	34.5%	100%
Galvanized Iron or Tin	Lamjung	21.3%	38.1%	40.6%	100%
	Tanahun	20.7%	38.3%	41%	100%
Tiles or Straw	Lamjung	14.3%	46.4%	39.3%	100%
	Tanahun	21.7%	34.8%	43.5%	100%
Sleeping rooms					
One or two rooms	Lamjung	3.3%	46.7%	50%	100%
	Tanahun	17.7%	39.5%	42.9%	100%
Three or four rooms	Lamjung	16.4%	36.8%	46.8%	100%
	Tanahun	18.3%	39.4%	42.3%	100%
Five or more rooms	Lamjung	27.9%	42.6%	29.4%	100%
	Tanahun	43.8%	34.4%	21.9%	100%
Other assets					
Glass window	Lamjung	25.9%	39.6%	34.5%	100%
	Tanahun	21.9%	37.5%	40.6%	100%
Refrigerator	Lamjung	60.7%	36%	3.3%	100%
	Tanahun	41.2%	42.4%	16.5%	100%
Television	Lamjung	19.9%	39.3%	40.8%	100%
	Tanahun	18.6%	40.4%	41%	100%
Motorcycle	Lamjung	60.8%	38%	1.2%	100%
	Tanahun	30%	40%	30%	100%
More than 2 cows	Lamjung	29.2%	37.5%	33.3%	100%
	Tanahun	20.7%	34.5%	44.8%	100%
More than 2 buffaloes	Lamjung	0%	57.8%	42.2%	100%
	Tanahun	21.2%	54.5%	24.2%	100%
Stress tolerant rice varieties adoption	Lamjung	21.8%	45.6%	32.5%	100%
	Tanahun	21.4%	40.5%	38.5%	100%

Table 4: Percentage of households within The Poor category having ownership of the assets (Lamjung)

Assets	Percentage of The Poor
Foundation of house (cement –bonded pillar system)	0%
Foundation of house (cement –bonded without pillar)	5.2%
Outer wall of house (mud –bonded bricks or stones)	100%
Outer wall of house (cement –bonded bricks or stones)	0%
Motorcycle	0.7%
Refrigerator	2.2%
More than two cows/Ox	5.9%
More than two Buffaloes	14.1%

Table 5: Percentage of household with in The Poor category having ownership of the assets (Tanahun)

Assets	Percentage of The Poor
Foundation of house (cement –bonded pillar system)	0%
Foundation of house (cement –bonded without pillar)	0%
Foundation of house (mud –bonded foundation)	98.3%
Outer wall of house (mud –bonded bricks or stones)	100%
Outer wall of house (cement –bonded bricks or stones)	0%
Number of sleeping rooms(Five or more)	4.8%
More than two cows/Ox	8.9%
More than two Buffaloes	5.5%

5. CONCLUSION

Despite attempts are being made by institutions and organizations to reduce poverty and to improve SES of paddy growers in mid-hills Nepal, the economic condition of the farmers is vulnerable. Institutions like International Rice Research Institute (IRRI) in co-operation with Institute of Agriculture and Animal Science (IAAS) have fostered farmers to adopt stress tolerant rice varieties (STRV) in these regions. Since, farming activities in mid –hills are entirely rainfed, farmers are happy to harvest comparatively larger share of production even in those seasons inflicted by drought; however, their economic condition is not any less aggravated. Hence, it is crucial to implement rice based programs to eliminate their poverty. Due to scarcity of human

power, high labor cost has been scourge in their main economic activity. So, policies must include mechanization and provision of subsidies in rice farming to tackle against the extant problem. The programs should also focus on distribution of superior rice varieties in minimal cost. This study makes complete assessment of socio-economic status of paddy growers in mid-hills, Lamjung and Tanahun districts, and more precisely focus on condition of the poorest members of rice farming community. Using PCA, the socio-economic variables are objectively quantified, moreover, poverty is properly assessed without using income and expenditure information. The result of the study is expected to provide basic source of information to correctly map out the target groups, prior to implementation of poverty alleviation programs.

ACKNOWLEDGEMENTS

This study is funded by International Fund for Agricultural Development (IFAD) for the assessment of Consortium for Unfavorable Rice Environments (CURE) to improve farmers' livelihoods. The author thanks partnership organization iDE-Nepal and all the students of IAAS, who actively participated in data collection.

REFERENCES

1. Devkota, J., Singh, S., Prajapati, C., & Hada, B. (2014). Principal components analysis of the socio –economic conditions of biogas users -with example from Nepal. *International Journal of Renewable Energy Research*. Retrieved from : https://www.researchgate.net/publication/281445887_Principal_Components_Analysis_of_the_Socioeconomic_Conditions_of_Biogas_Users_-with_Example_from_Nepal
2. Filmer, D., & Pritchett, L. (2001). Estimating wealth effects without expenditure data-or tears: An application to educational enrollments in states of India. *Demography*, 38(1), 115-132. Retrieved from: www.jstor.org/stable/3088292
3. Gwatkin, DR., Rutstein, S., Johnson, K., Suliman, E., Wagstaff, A., & Amouzou, A. (2000). Socio-economic differences in health, nutrition, and population in Ethiopia (English). *Country reports on HNP and poverty*. Washington, DC : World Bank Group. Retrieved from: <http://documents.worldbank.org/curated/en/435741468330985969/Socio-economic-differences-in-health-nutrition-and-population-in-Ethiopia>.
4. Kolenikov, S., & Angeles, G. (2009). Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer? *Review of Income and Wealth*. 55. 128-165. 10.1111/j.1475-4991.2008.00309.x.
5. McKenzie, DJ. (2005). Measuring inequality with asset indicators. *Journal of population economics*. 18. 229–260. 10.1007/s00148-005-0224-7.

6. Polonsky, J., & Balabanova, D. et. al. (2009). Equity in community health insurance schemes: Evidence and lessons from Armenia. *Health policy and planning*. 24. 209-16. 10.1093/heapol/czp001.
7. Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: How to use principal components analysis. *Health policy and planning*. 21. 459-68. 10.1093/heapol/czl029.