Principal Components Analysis of the Socioeconomic Conditions of Biogas Users - with Example from Nepal

Jyoti U. Devkota*‡, Swechhya Singh**, Chanda Prajapati**, Binu Hada**

*Department of Natural Sciences, Mathematical Sciences Group, Kathmandu University, G. P. O. 9303, Kathmandu, Nepal.

**Department of Environmental Science and Engineering, Kathmandu University, G. P. O. 6250, Kathmandu, Nepal.

(drjdevkota@ku.edu.np, swechhya.singh@gmail.com, chanda_prajapati@hotmail.com, binuhada@hotmail.com)

‡ Corresponding Author; Jyoti U. Devkota, G. P. O. Box 9303, Kathmandu, Nepal, Tel: +9779841539845,

drjdevkota@ku.edu.np

Received: 29.06.2014 Accepted: 29.08.2014

Abstract- Severe energy crisis in developing countries such as Nepal is reflected by power outages of 16 hours daily from the national grid in the lean season. Currently in the monsoon season it is 8 hours daily. A detailed study of impact of energy crisis and a switch over to a source of renewable energy is very crucial here, where even officially recorded data are faulty and erroneous. The agrarian nature of Nepalese economy is ideally suited for the use of Biogas as a source of renewable energy. A survey of 400 households using biogas as a source of renewable energy was conducted in three different rural settings of Nepal during September to November 2010. Out of 467 variables studying various socio-economic and performance parameters in the consumer profile database 47 proxy asset variables are identified. They indirectly estimate the socioeconomic status of the family. In an agrarian economy where many economic transactions take place outside the market, these methods provide more accurate data. The socio-economic status is thus objectively quantified. The data on assets ownership (television, refrigerator, motorcycle, bicycle etc) and type of house (type of house, type of toilet, type of water source etc) owned called assets indicators are used in constructing asset index by using Principal Components Analysis (PCA). The dimension is reduced to ten orthogonal variables explaining 60 percent of the variability. Here interest is stimulated in the interdisciplinary applications of statistical methodologies to problems from renewable energy in general and the application of PCA in particular.

Keywords: Principal components analysis, socioeconomic survey, biogas, consumer profile database, statistical analysis

1. Introduction

The development status of any country is also governed by the availability of the energy from fossil fuel. But the demand of this energy is ever increasing and will never meet the supply. A switch over to clean energy source is a cost effective and sustainable alternative for developing world instead of fossil fuels, which pollute the lower layer of the environment. Adapting to renewable energy provides answer to both the problems of energy scarcity and climate change. The developing countries are mostly agrarian economies. They depend also on forest for firewood which has resulted in deforestation of the forest. Most of the rural population has the tradition of raising cattle as an integral part of their farm. In addition to draft power and milk, cattle produce them necessary manure in the form of dung. This dung is the most important component of bio fuel. Nepal has a good potential of renewable energy mainly solar and energy from biomass. According to Nepal labor force survey 2010 while 68.4% of the households use fuel wood for cooking, 78.4% have agricultural land holdings. There are still more than 200,000 households which are potential biogas consumers.

Data based research requires sound statistical analysis and its interdisciplinary application. Many statistical concepts and theories specially tailored to address a particular are being developed. Principal components analysis retains the variability of the original data by transforming the correlated variables into fewer orthogonal variables. It is especially useful in the case of multivariate analysis where several variables might be interrelated and the true information of the data might be disturbed due to this multicollinearity. Hebert et. al. applied principal components analysis (PCA) to reduce the number of predictor variables with minimal loss of information [1]. The seven variables related to socioeconomic conditions included in computing the socioeconomic status principal components (SESPCs) were as follows: gross national product, life expectancy at birth, infant mortality rate, proportion of the population without safe water or excreta disposal facilities, and the number of physicians or hospital beds per 10, 000 people. A survey using a structured questionnaire to collect data on socio-economic characteristics and malaria beliefs and practices among more than 400 net-owning and non-net-owning households was conducted by Howard et al. [2]. A composite socio-economic index was created using principal components analysis, and survey households were divided into socio-economic quartiles. Vyas et al. have applied PCA to Brazil and Ethiopia demographic health survey data and also reviewed various issues related to choice of variables and data preparation [3]. Filmer and Pritchett studied wealth by constructing a linear index from asset ownership indicators, using principal components analysis to derive weights [4]. Different methods to be adopted to improve the running of PCA on discrete data were suggested by Kolenikov and Angeles [5]. Polonsky et al. not only discussed about members of a random sample of 506 households in villages operating insurance schemes in rural Armenia but also elaborated on household wealth scores based on ownership of assets generated using principal components analysis [6]. Logistic and Poisson regression analyses were performed to identify the determinants of health facility utilization and equity of access across socio-economic strata. Mwageni et al. used asset-based wealth index based on PCA to determine the relationship between household socio-economic characteristics and inequalities of access to health interventions, and to health outcomes in rural Tanzania [7]. Similarly the mortality and fertility data of developing countries especially from Nepal are analyzed by using deterministic and mixed effect models by Devkota [8][9]. Several problems associated with the continual collection of vital statistics especially in countries with limited and defective data are discussed by Devkota [10]. The importance of digitization of survey data is elucidated in details by Devkota [11].

This paper primarily aims to indentify households in various socio-economic groups on the basis of principal components analysis. This is done on the basis of identification of asset variables from the consumer profile database, comprising of 467 variables. These variables resulted from a questionnaire comprising of 59 questions with answers classified into several categories. A background of the statistical methodology including the design and implementation of this survey, data collection and its digitization are also elucidated. This gives a theoretical background of the foundation on which the entire principal components analysis rests. With the help of PCA the dimension of 47 socio economic asset proxy variables is reduced to 10 variables. Then these components were used to classify the household into socio-economic groups.

The remainder of this paper is arranged as follows. Section 2 covers the theoretical background In Section 3 called materials and methods, the steps followed for design as well as implementation of the survey and the resulting dataset that motivates this study are described. Steps taken to minimize the error in the collected data at different stages of this study, its characteristics, diagnostics and patterns are also explained in detail. The concluding Section 4 gives the overall conclusions.

2. Theory

As mentioned by Hair et al. [12], multivariate analysis methods are not only related to analytic aspect of research but also to the design and approach to data collection for decision making and problem solving. The use of multiple variables and the reliance on their combination also requires attention on a very important complementary issue that is measurement error. Data entry error, imprecision of measurement, inability of the respondent to accurately provide information are some of the major sources. The impact of the measurement error is to add "noise" to the observed or measured variables. The observed value thus comprises of the true value and the "noise". Efforts should be made to minimise the error as much as possible by identifying all the possible sources of error in the study from the design of experiments to the final inference. Principal component analysis is a statistical approach that can be used to analyse the interrelationship among a large number of variables such that the information contained in a number of original variables is condensed into a smaller set of variates (factors) with minimum loss of information. This data summarisation helps identify the underlying dimension or factor, estimates of factors and contribution of each variable to the factors (termed loadings). Unrotated factor matrix comprising of factor loadings is used when the main objective of research is in best linear combination of variables where the a particular combination of original variables account for more of variance in the data as a whole than any other linear combination.

Suppose we have a set of N variables, a^*_{1j} to a^*_{Nj}, representing the ownership of N assets by each household j. Further, let us standardize each variable by its mean and standard deviation: for example $a_{ij} = \frac{a^*_{1j}-a^*_1}{s^*_1}$, where a^*_1 is the mean of a^*_{1j} across households and s^*_1 is its standard deviation. These selected variables are expressed as linear combination of a set of underlying components for each household j: (where, $j = 1, \ldots, J$)

$$a^*_{1j} = v_{11} * A_{1j} + v_{12} * A_{2j} + \cdots + v_{1N} * A_{Nj}$$

$$\vdots$$

$$a^*_{Nj} = v_{N1} * A_{1j} + v_{N2} * A_{2j} + \cdots + v_{NN} * A_{Nj} \quad (1)$$

A s are the components and v s are the coefficient on each component for each variable. The "scoring factors" from the model are recovered by inverting the system implied by Eq. (1), and yield a set of estimates for each of the N principal components: ($j = 1 \ldots J$)

$$A_{1j} = f_{11} * a_{1j} + f_{12} * a_{2j} + \cdots + f_{1N} * a_{Nj}$$

.
.

$$A_{Nj} = f_{N1} * a_{1j} + f_{N2} * a_{2j} + \cdots + f_{NN} * a_{Nj} \quad (2)$$

The first principal component, expressed in terms of the original (unnormalized) variables, is therefore an index for each household based on the expression

$$A_{1j} = f_{11}\left(\frac{a_{1j}^{*}-a_{1}^{*}}{s_{1}^{*}}\right)+\ldots+f_{1N}\left(\frac{a_{Nj}^{*}-a_{N}^{*}}{s_{N}^{*}}\right) \quad (3)$$

With respect to this study there are 47 asset variables and 400 households. So, N = 47 and J = 400.

3. Materials and Methods

The role of data and its digitization has a crucial role to play in the modern world. To ensure data with minimum statistical and background noise, steps from design of experiment, conduction of data collection process to its analysis and interpretation have to be carefully scrutinized. Error free data and the data with minimum white noise is the foundation of good statistical analysis. A detailed survey results in multivariate data that are nominal, ordinal, interval or ratio in nature. The pattern and behaviour of the variables have to be analysed in details. Some variables are correlated with each other and emit same kind of information. A well planned household survey and a well tested questionnaire are very important pre requisites for a statistically sound research. A judiciously designed questionnaire tested and refined before the major survey, with all the possible answers classified and mentioned in several categories, collects information with minimum possible error.

A survey was conducted on 400 households of biogas consumers inhabiting in rural areas of Nepal in three different regions during Oct-Nov 2010. With a primary objective of making the entire process error free, all the possible sources of error from the stage of questionnaire design to the data entry into the database were clearly worked out.

3.1. Questionnaire and Survey

The draft questionnaire comprised of 62 questions before it was tested on 30 households in Sudal VDC, Bhaktapur, Nepal. The response of the consumers was noted and the answer options were accordingly refined and updated to remove errors, ambiguity of answers and smoothness in the flow of answers. It was polished according to the responses of the interviewee to 59 questions, after the pre test.

The questionnaire collected detailed information about the degree of change in their life style after biogas was used by them as a source of renewable energy. It also enquired about the role of gender in different activities of the household and their empowerment after the use of renewable energy. Thus the questionnaire was designed with an objective of keeping biogas use in the core and getting all the possible information about a typical middleclass Nepalese family inhabiting in rural areas, its economic and social background and change after biogas was used in their household. In these 59 questions information was collected on very relevant topics such as the age distribution of 400 households comprising of 2272 individuals of different age groups, amount of landholdings, livestock, their fuel wood expenses before and after the installation of plants, performance of the plant in summer and winter etc. So with the structure of the questions information can be obtained about the households that haven't installed biogas as a source of renewable energy. Most of the questions were yes/no in nature. The data was collected on 400 households, where 370 households were the consumers of Rapti Renewables and Energy Services, our industrial partner in that project. The entire data collection process of 400 households comprising of 2272 individuals was completed in 15 days. The details of the consumer profile database, variable names and their types are explained in detail in Table 1. For example, the table number 11 called biogas questions, enquires about the performance of their biogas plant from the consumers. Different yes/no questions pertaining to various variables related with the performance monitoring are asked. They are 75 such categories and are either nominal or ordinal in nature. After the data was collected in filled up printed questionnaire then it was digitized into an electronic form in a database. Different variables in the questionnaire were carefully analyzed and the design of the database was carefully worked out. Use of input masks and default values ensured the quality and correctness of the input data for data digitization.

In any survey direct questions assessing the socio-economic status of the respondents are subject to maximum bias. Validity of the answers to important and sensitive questions was cross checked by reframing it and asking it again at a later part of the questionnaire. Normally correct answers relating to the income and expenditure are difficult to get. Human beings are very sensitive to such questions and don't provide correct answers to the enumerators. Further in a developing country like Nepal if the respondents are farmers from rural area such information is more difficult to get as large fraction of economic activity is carried out outside the market. Among biogas consumers of Nepal inhabiting in the rural area who are primarily using cow dung for their plant, income and expenditure transactions are carried out outside the market. The respondents of this study were 400 households of farmers inhabiting in rural area, raising cattle and using biogas plants as a source of renewable energy. Among them 30 lived in one area in relatively close proximity to each other in Bhaktapur. 300 lived in Simara and 70 in Sarlahi. In such a situation other proxy's of household wealth was used to assess their socio-economic status [4], [5]. Such proxies were questions related to the possession of goods and living conditions which indirectly assessed their socio-economic status. They were pretested and their response was noted. Unlike direct questions on income such proxy questions have less probability of being false as they are less offending and the response can be cross checked by observation. Indirect

Table 1. An overview of the consumer profile database

Table Number	Heading	Variables	Types of Data
1	Family Background	10	Nominal and Binary
2	Family Description	62	Discrete
3	Occupation	6	Nominal
4	Livestock	12	Discrete and Continuous
5	Business	3	Nominal and Continuous
6	Other Occupation	8	Discrete
7 and 8	Land	6	Binary
9	Socio-economic condition	54	Binary and Categorical-Nominal
10	Water Needs	17	Binary
11	Biogas Questions	75	Categorical-Nominal, Categorical-Ordinal,
12	Size of Plant	18	Binary
13	Woman Empowerment	44	Categorical-Nominal
14	Source of Firewood Collection	14	Binary
15	Distance Travelled for Firewood Collection	8	Binary
16	Comparison of Firewood	40	Binary and Continuous
17	Comparison of time spent on Firewood collection	10	Binary
18	Frequency of bio feed in biogas plant	15	Binary
19	Cultivation of crops	32	Binary
20	Application of Fertilizers	16	Binary
21	Health Related Issues	8	Categorical-Nominal
22 and 23	Positive Impact after Installation of Plant	9	Binary
	Total	**467**	

questions with few number of clear response categories suffer fewer reporting errors than direct questions on income.

3.2. Data Analysis

Detailed consumer household surveys result in data structures which are multivariate and hence multidimensional. Although exploratory data analysis gives an idea on the patterns existing within and between the different variables in the specific subset of the survey data, in depth idea is obtained with the help of multivariate analysis. In multivariate analysis several key variables are studied simultaneously. A pattern between several variables is minutely analyzed. The inter relationship and interdependence between several variables can be minutely analyzed and their dimensions can be reduced on the basis of commonality or independence.

3.3. Principal Components Analysis

Here a task of objective quantification of socio economic classification is done, although in the developing world they are subjective in nature. Being an agrarian economy many socio-economic transactions are outside the market. So many benefits cannot be evaluated monetarily. Additionally this data is of biogas consumers who are mostly farmers and keep cattle. The data on asset ownership (e.g. owning a bicycle or radio) and housing characteristic (e.g. type of house, type of toilet, and type of water source etc) called asset indicators or asset variables were used to construct asset index by using PCA. Socio-economic impacts cannot be directly identified and quantified in monetary terms. In our case the interviewee were farmers using biogas and living in close proximity.

So the problem was to quantify the socio-economic status of the households which were fairly similar with respect to one another with respect to the following characteristic namely a) biogas consumers b) farmers c) reared cattle d) lived in close proximity with each other.

Various features of biogas consumer were characterized into 467 variables and classified into 23 tables comprising primarily of binary, ordinal and nominal data. Out of 467 variables 47 variables were related to the socio-economic aspect of the consumer. These variables related to the ownership of assets like land holding, private ownership of water sources for drinking, bathing, irrigating, possession of cars, tractor, bicycle, radio, location of toilet, amount of loan incurred were used to assess the economic prosperity. As direct sensitive questions on amount of income earned by the household was subject to white noise due to untrue answers, the socio-economic status of the household was assessed by asking questions on possession of assets with all the possible answers as multiple choices. Further the rural setting of the households fails to give an idea of the economic prosperity of the household if questions related to monetary income earned were only asked. Most of the survey data were binary and categorical in nature. Some of the variables were continuous and discrete in nature. The continuous and discrete data were bundled together for analysis and the binary and categorical data were considered together for the principal components analysis. The discrete and continuous data were sometimes fitted into some classifications and converted into categorical data. The most important issue in this study was how to convert 47 variables to produce a range of critical points differentiating socio economic levels among consumers of biogas living in a rural setting of Nepal.

Following steps were followed in the PCA. First step was the identification of assets variables. The identified asset variables were converted into indicator variables or binary data. This survey included 47 indicator asset variables that can be classified into four groups. In the first group there were 8 questions related to asset ownership which comprised of land, house, electronic equipments such as computer, TV, radio, mobiles, telephones and refrigerator, means of transportation. There were 10 indicator variables in this category. Characteristics of house dwellings and toilet were the second group and comprised of 6 questions and 10 indicator variables. Here questions on materials used in the construction of house were asked. Open latrine, latrine far away from the house or close to the house etc were queries related to the type of latrine. Water source and needs was the third group with 6 questions and 20 indicator variables. Detailed question on the use of water such as for bathing or cooking etc were asked. Different sources of water such as private well, open well, community water supply etc were asked. Fourth and the last group was the amount of land owned and comprised of 5 indicator variables. Then the descriptive statistics of the variables including mean, standard deviation and correlation were calculated. This gave an idea on the average value of the variable per household and its spread across the household. Since most of the asset variables data are 0 or 1 as they were in response to yes/no questions, the mean of most common asset variable is close to 1 and the standard deviation is very low. Referring to Table 2 for example in response to the yes/no question of Do you have a latrine, the mean is 0.97 and the standard deviation is 0.17. Whereas in response to the yes/no question Do you have a well-built latrine, the mean is 0.93 and standard deviation is 0.26. This suggests that although latrine is almost universal among the consumers well-built latrine is somewhat less and not so common. The data were mainly ordinal and ratio in nature. Table 1 gives the detail of the consumer profile database and types of data.

Principal components analysis was applied in the second step. Table 2 gives the detail of the application of PCA to the data. R software has been used for the principal components analysis of the data. Unrotated factor solution was done for data reduction as the objective was to identify the first best summary of linear combination of variables existing in the data. Here this particular linear combination accounts for more variance in the data as a whole than any other linear combination of variables. Other types of rotations were also tested on the data but there was not much difference in the factor loadings and in the final inference

The interpretation of results is the third step. The first factor shows high loading. Own source of water for bathing, drinking and other day to day activities is an indicator of economic well being of the biogas community as indicated by high loading of the first principal component on these variables. The first principal component is an indicator of socio economic status. So, the first principal component is an indicator of economic affluence of the community which is reflected by the use of own water source for daily activities. PCA II reflects the closeness to the community for their daily day to day water needs. It is reflected by high loading of PCA II on water source available for common use such as Kuwa (community owned water spouts) and common wells. It is a measure of extent of closeness to the community. PCA III loads highly on good well built latrine and is an indicator of belief and hygiene. These results are plausible. The biogas consumers surveyed for this study were mainly from three areas. They lived in close vicinity and were fairly homogenous with respect to the socio economic status. As we see from our analysis in such cases the ownership of own source of water is the most important factor which identifies a person from high socio economic group. This phenomenon is true in developing world were the water supplies cannot reach to all part of the country. People have a search for their own water source. This is privately owned in cases to households of higher economic groups. Similarly the type toilet is the second important criterion which identifies a household of higher socio economic group. Thus among the biogas consumers living in a community water source for daily needs plays an important role in the determination of the socio economic status of a community. It is followed by the type of toilet used, which is also an indicator of economic affluence and belief of the household. The highest eigen value is 5.436 and it explains 11.56% of the total variance. Ten principal components account for 60% variability of 47 socio economic asset indicators.

Then in the fourth step the asset index was calculated on the basis of the first principal component. On the basis of the values of the asset index the consumers were classified into rich that were the top 20 % which was followed by middle 40% and bottom 40%. It can also be classified as rich, middle income and poor respectively. The asset indices are elaborated in detail in Table 3. The distribution of income based on asset index is also shown in Fig. 1.

4. Conclusion

Adequate electricity supply is as important in the developing country as in the developed country. But due to limited resources there exists a critical shortage and it is manifested by long hours of load shedding in Nepal. This has hindered the social and economic advancement of the population as all the economic and social activities come to a halt after dark. Some affluent sections of the community switch over to inverters and generators, but they are also dependent on diesel. So a switch over to a source of renewable energy such as biogas reduces the amount of dependence on electricity to fulfill our day to day needs. An attempt is made to quantify socio economic variables on the basis of data of 400 households of biogas consumers. The respondents are basically farmers and

Table 2. Results from the Principal Components Analysis

Sr. No.	Socio Economic Variables	Factor Score (FS)	Mean	Standard Deviation (SD)	FS/SD
Material Assets					
1	Motorcycle	-0.123000	0.340000	0.14	-0.36176
2	Cycle	-0.035710	0.330000	0.89	-0.10821
3	Car	-0.034900	0.050000	0.0025	-0.698
4	Computer	0.065820	0.184000	0.035	0.357717
5	TV	-0.034580	0.390800	0.8125	-0.08849
6	Telephone	-0.008405	0.170800	0.03	-0.04921
7	Fridge	-0.131000	0.196200	0.04	-0.66769
8	Radio	0.101000	0.382700	0.8375	0.263914
9	Cell Phone	-0.038580	0.346800	0.8675	-0.11125
Type of House					
10	Concrete House	0.177000	0.42	0.490000	0.361224
11	Mud House	-0.096070	0.35	0.480000	-0.20015
12	Modern Light Roof	-0.005543	0.19	0.390000	-0.01421
13	Tile and Asbestos House	-0.195000	0.0425	0.200000	-0.975
Type of Latrine if any					
14	Latrine	-0.196000	0.97	0.170000	-1.15294
15	Well-built Latrine	-0.135000	0.93	0.260000	-0.51923
16	Temporary Latrine	0.033690	0.0525	0.220000	0.153136
17	Open Latrine	0.189000	0.0225	0.150000	1.26
18	Inside the House	0.049190	0.02	0.140000	0.351357
19	Near the House	-0.166000	0.92	0.270000	-0.61481
20	Far from the House	-0.008764	0.03	0.170000	-0.05155
Loan					
21	Any Loan	0.043720	0.36	0.480000	0.091083
22	Amount of loan	0.085520	1.02	1.600000	0.05345
Land Possession(in sq. m)					
23	Less Than 127.185	-0.061060	0.005	0.070600	-0.86487
24	158.9813 to 254.37	-0.033690	0.0125	0.110000	-0.30627
25	286.1663 to 476.9438	-0.005983	0.0725	0.260000	-0.02301
26	508.74 to 1526.22	0.041580	0.12	0.330000	0.126
27	More than 1526.22	-0.025730	0.78	0.410000	-0.06276
Source of Water					
28	Own Well/Springs/Ponds Drinking	-0.973000	0.66	0.480000	-2.02708
29	Own Well/Springs/Ponds Bathing and Washing	-0.975000	0.66	0.470000	-2.07447
30	Own Well/Springs/Ponds Cooking	-0.973000	0.66	0.480000	-2.02708
31	Own Well/Springs/Ponds Irrigating Crops	-0.357000	0.19	0.390000	-0.91538
32	Commonly used private well/spring/pond Drinking	0.492000	0.11	0.310000	1.587097
33	Commonly used private well/spring/pond Bathing and Washing	0.495000	0.11	0.310000	1.596774
34	Commonly used private well/spring/pond Cooking	0.501000	0.11	0.320000	1.565625
35	Commonly used private well/spring/pond Irrigating Crops	-0.037260	0.16	0.360000	-0.1035
36	Common Well/Spring/Pond Drinking	0.475000	0.12	0.320000	1.484375
37	Common Well/Spring/Pond Bathing and Washing	0.424000	0.12	0.320000	1.325
38	Common Well/Spring/Pond Cooking	0.483000	0.12	0.320000	1.509375
39	Common Well/Spring/Pond Irrigating Crops	0.130000	0.22	0.410000	0.317073
40	River/Stream Drinking	0.113000	0.0125	0.110000	1.027273
41	River/Stream Bathing and Washing	0.113000	0.0125	0.110000	1.027273
42	River/Stream Cooking	0.113000	0.0125	0.110000	1.027273
43	River/Stream Irrigating Crops	0.023030	0.22	0.420000	0.054833
44	Kuwa/Common Tap Drinking	0.471000	0.1	0.300000	1.57
45	Kuwa/Common Tap Bathing and Washing	0.483000	0.0925	0.290000	1.665517
46	Kuwa/Common Tap Cooking	0.483000	0.0925	0.290000	1.665517
47	Kuwa/Common Tap Irrigating Crops	0.264000	0.06	0.240000	1.1

Table 3. Comparison between the three socio economic groups by Asset Index

Economic Group	Variable	N	Minimum	Maximum	Average Asset Index	SD
Total	Asset Index	400	-6.0585	12.1740	0.0000008	5.435213
Rich-Top 20%	Asset Index	80	-6.0585	-4.3003	-4.88347	0.4138973
Middle-40%	Asset Index	160	-4.2132	-2.6325	-3.57765	0.353197
Poor-lowest 40%	Asset Index	160	-2.6323	12.174	6.0194	3.5472
Material Assets						
Rich-Top 20%	Motorcycle	80	0	1	0.33	0.47
Middle-40%	Motorcycle	160	0	1	0.0813	0.27
Poor-lowest 40%	Motorcycle	160	0	1	0.0938	0.29
Rich-Top 20%	Cycle	80	0	1	0.81	0.39
Middle-40%	Cycle	160	0	1	0.93	0.25
Poor-lowest 40%	Cycle	160	0	1	0.87	0.37
Rich-Top 20%	Car	80	0	1	0	0
Middle-40%	Car	160	0	1	0.00625	0.07
Poor-lowest 40%	Car	160	0	1	0	0
Rich-Top 20%	Telephone	80	0	1	0.0625	0.3426
Middle-40%	Telephone	160	0	1	0.0125	0.1115
Poor-lowest 40%	Telephone	160	0	1	0.03125	0.1745
Rich-Top 20%	Fridge	80	0	1	0.125	0.3328
Middle-40%	Fridge	160	0	1	0.01875	0.1361
Poor-lowest 40%	Fridge	160	0	1	0.01875	0.1361
Rich-Top 20%	Radio	80	0	1	0.7	0.4611
Middle-40%	Radio	160	0	1	0.8562	0.3519
Poor-lowest 40%	Radio	160	0	1	0.8875	0.3544
Rich-Top 20%	Cell Phone	80	0	1	0.90	0.3019
Middle-40%	Cell Phone	160	0	1	0.8562	0.3694
Rich-Top 20%	TV	80	0	1	0.8635	0.3465
Middle-40%	TV	160	0	1	0.8063	0.3965
Poor-lowest 40%	TV	160	0	1	0.7938	0.4059
Rich-Top 20%	Computer	80	0	1	0.025	0.1571
Middle-40%	Computer	160	0	1	0.0125	0.1115
Poor-lowest 40%	Computer	160	0	1	0.0625	0.2528
Rich-Top 20%	Any Loan	80	0	1	0.19	0.39
Middle-40%	Any Loan	160	0	1	0.42	0.49
Poor-lowest 40%	Any Loan	160	0	1	0.38	0.49
Type of House and Type of Latrine						
Rich-Top 20%	Concrete House	80	0	1	0.18	0.38
Middle-40%	Concrete House	160	0	1	0.44	0.50
Poor-lowest 40%	Concrete House	160	0	1	0.51	0.50
Rich-Top 20%	Mud House	80	0	1	0.34	0.48
Middle-40%	Mud House	160	0	1	0.43	0.50
Poor-lowest 40%	Mud House	160	0	1	0.28	0.45
Rich-Top 20%	Modern Light Roof	80	0	1	0.29	0.46
Middle-40%	Modern Light Roof	160	0	1	0.13	0.33
Poor-lowest 40%	Modern Light Roof	160	0	1	0.21	0.41
Rich-Top 20%	Tile Asbestos House	80	0	1	0.20	0.40
Middle-40%	Tile Asbestos house	160	0	1	0.00625	0.079
Poor-lowest 40%	Tile Asbestos house	160	0	1	0	0
Rich-Top 20%	Latrine(Yes/No)	80	0	1	0.99	0.11
Middle-40%	Latrine(Yes/No)	160	0	1	1.00	0.00
Poor-lowest 40%	Latrine(Yes/No)	160	0	1	0.93	0.25
Rich-Top 20%	Well Built Latrine	80	0	1	0.99	0.11
Middle-40%	Well Built Latrine	160	0	1	0.96	0.21
Poor-lowest 40%	Well Built Latrine	160	0	1	0.86	0.35
Rich-Top 20%	Temporary Latrine	80	0	1	0.0125	0.11
Middle-40%	Temporary Latrine	160	0	1	0.0437	0.21
Poor-lowest 40%	Temporary Latrine	160	0	1	0.0813	0.27
Rich-Top 20%	Open Latrine	80	0	1	0.0	0.0
Middle-40%	Open Latrine	160	0	1	0.0	0.0
Poor-lowest 40%	Open Latrine	160	0	1	0.0563	0.23
Rich-Top 20%	Commonly used private well/ spring/pond Irrigating Crops	80	0	1	0.19	0.39
Middle-40%	Commonly used private well/ spring/pond Irrigating Crops	160	0	1	0.16	0.37
Poor-lowest 40%	Commonly used private well/ spring/pond Irrigating Crops	160	0	1	0.13	0.34

Table 3. Comparison between the three socio economic groups by Asset Index (Cont.)						
Rich-Top 20%	Common Well/Spring/ Pond Bathing and Washing	80	0	1	0.00	0.00
Middle-40%	Common Well/Spring/ Pond Bathing and Washing	160	0	1	0.00	0.00
Poor-lowest 40%	Common Well/Spring/ Pond Bathing and Washing	160	0	1	0.29	0.46
Rich-Top 20%	Common Well/Spring/ Pond Cooking	80	0	1	0.00	0.00
Middle-40%	Common Well/Spring/Pond Cooking	160	0	1	0.00	0.00
Poor-lowest 40%	Common Well/Spring/Pond Cooking	160	0	1	0.28	0.45
Rich-Top 20%	Common Well/Spring/Pond Irrigating Crops	80	0	1	0.0	0.0
Middle-40%	Common Well/Spring/Pond Irrigating Crops	160	0	1	0.16	0.37
Poor-lowest 40%	Common Well/Spring/Pond Irrigating Crops	160	0	1	0.13	0.34
Rich-Top 20%	River/Stream Drinking	80	0	1	0.00	0.00
Middle-40%	River/Stream Drinking	160	0	1	0.00	0.00
Poor-lowest 40%	River/Stream Drinking	160	0	1	0.031	0.17
Rich-Top 20%	River/Stream Bathing and Washing	80	0	1	0.00	0.00
Middle-40%	River/Stream Bathing and Washing	160	0	1	0.00	0.00
Poor-lowest 40%	River/Stream Bathing and Washing	160	0	1	0.0313	0.17
Rich-Top 20%	River/Stream Cooking	80	0	1	0.00	0.00
Middle-40%	River/Stream Cooking	160	0	1	0.00	0.00
Poor-lowest 40%	River/Stream Cooking	160	0	1	0.0313	0.17
Rich-Top 20%	River/Stream Irrigating Crops	80	0	1	.0625	0.24
Middle-40%	River/Stream Irrigating Crops	160	0	1	0.28	0.45
Poor-lowest 40%	River/Stream Irrigating Crops	160	0	1	0.24	0.43
Rich-Top 20%	Kuwa/Common Tap Drinking	80	0	1	0.00	0.00
Middle-40%	Kuwa/Common Tap Drinking	160	0	1	0.00	0.00
Poor-lowest 40%	Kuwa/Common Tap Drinking	160	0	1	0.25	0.43
Rich-Top 20%	Kuwa/Common Tap Bathing and Washing	80	0	1	0.0	0.0
Middle-40%	Kuwa/Common Tap Bathing and Washing	160	0	1	0.0	0.0
Poor-lowest 40%	Kuwa/Common Tap Bathing and Washing	160	0	1	0.23	0.42
Rich-Top 20%	Kuwa/Common Tap Cooking	80	0	1	0.0	0.0
Middle-40%	Kuwa/Common Tap Cooking	160	0	1	0.0	0.0
Poor-lowest 40%	Kuwa/Common Tap Cooking	160	0	1	0.23	0.42
Rich-Top 20%	Kuwa/Common Tap Irrigating Crops	80	0	1	.00625	.079
Middle-40%	Kuwa/Common Tap Irrigating Crops	160	0	1	.00625	.079
Poor-lowest 40%	Kuwa/Common Tap Irrigating Crops	160	0	1	0.14	0.35
Rich-Top 20%	Less Than 127.185	80	0	1	0.0125	0.11
Middle-40%	Less Than 127.185	160	0	1	0.00625	0.0791
Poor-lowest 40%	Less Than 127.185	160	0	1	0.0	0.0
Rich-Top 20%	158.9813 to 254.37	80	0	1	0.0125	0.11
Middle-40%	158.9813 to 254.37	160	0	1	0.0188	0.14
Poor-lowest 40%	158.9813 to 254.37	160	0	1	0.00625	0.0791
Rich-Top 20%	286.1663 to 476.9438	80	0	1	0.0	0.0
Middle-40%	286.1663 to 476.9438	160	0	1	0.12	0.33
Rich-Top 20%	508.74 to 1526.22	80	0	1	0.11	0.32
Middle-40%	508.74 to 1526.22	160	0	1	0.11	0.32
Poor-lowest 40%	508.74 to 1526.22	160	0	1	0.14	0.35
Rich-Top 20%	More than 1526.22	80	0	1	0.85	0.36
Middle-40%	More than 1526.22	160	0	1	0.74	0.44
Poor-lowest 40%	More than 1526.22	160	0	1	0.79	0.41

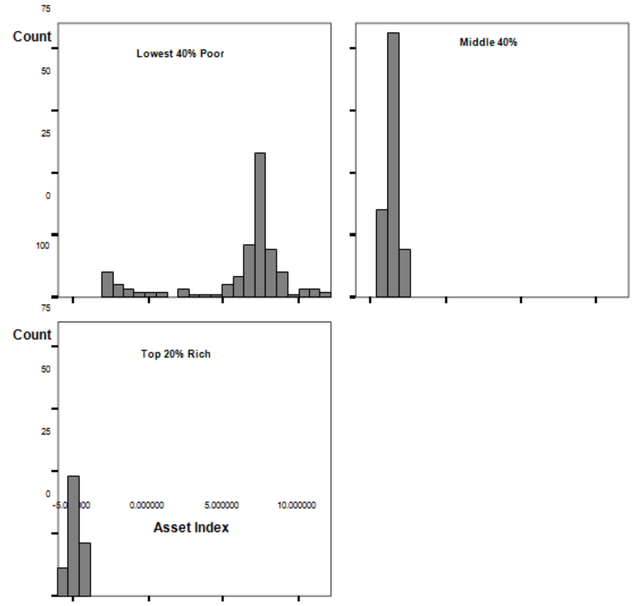

Fig. 1. Distribution of assets among three income groups

many of the socio economic obligations and benefits cannot be calculated in monetary terms.

This paper describes the process of development and implementation of socio-economic survey of 400 households of biogas consumers of Nepal. It also describes the process to derive the socio- economic index on the basis of PCA on durable asset ownership, access to utilities and infrastructure and housing characteristics data.

Various steps taken to minimize errors from the stage of questionnaire design to PCA are also elaborated. PCA of 47 asset variables from 467 variables of entire socio-economic survey extracted 10 components explaining 60% of the total variance. Then the socio-economic categorization is done classifying the households into three socio-economic groups, where rich comprised of top 20% of the socio-economic group of the society, the middle income group comprised of middle 40% and the lowest 40% are economically most deprived. This was on the basis of asset index which is linear combination of factor loadings and normalized asset ownership variables. The robustness of this classification is tested by the data on those assets which are conventionally owned by people who are more economically well off. Thus the dimensionality of the data comprising of 47 interrelated asset variables is reduced to 10 orthogonal variables. It is also seen that ownership of own water source is the main factor differentiating economically well off households with poor households. The respondents of this survey are 400 households of biogas consumers living in three different rural settings of Nepal. They are fairly similar to each other as they all have cattle, live in rural areas, use a source of renewable energy primarily for cooking needs, are farmers by profession and live in close proximity to each other. The results of this paper help understand and quantify socio-economic status which is very subjective and sensitive topic. Biogas users are mainly farmers and are a part of an agrarian economy where many transactions are carried out outside the market. Hence these cannot be quantified terms of money. So here PCA has been used in classifying the families in different income groups. This will be useful to policy makers and planners in gaining a better understanding of their consumers and will help them make an optimum market strategy for biogas in particular and any other source of renewable energy in general.

Acknowledgements

This work is funded by NORAD, SINTEF under Renewable Nepal Project grant number RENP-10-06-PID-379. The authors thank the partner industry in this project, Rapti Renewable energy services and all the students of the data collection team.

References

[1] Hebert, J. R., Hurley, T. G. et. al., 1998. Nutritional and Socioeconomic Factors in Relation to Prostate Cancer Mortality: a Cross-National Study. Journal of National Cancer Institute, 90(21), 1637-1647.

[2] Howard, N., Chandramohan, D., et al., 2003. Socio-economic factors associated with the purchasing of insecticide-treated nets in Afganisthan and their implications for social marketing..Tropical Medicine and International Health, 8(12), 1043-1050.

[3] Vyas, S., Kumaranayake, L., 2006. Constructing Socio-Economic status indices: how to use principal component analysis. Health Policy and Planning, 21, 459-468.

[4] Filmer, D., and Pritchett, L. 2001. Estimating wealth effect without expenditure data - or tears: An application to educational enrollments in states of India. Demography, 38, 115-132.

[5] Kolenikov, S., and Angeles, G., 2009. Socioeconomic status measurement with discrete proxy variables: is principal components analysis a reliable answer? Review of Income and Wealth, 55(1), 128-165.

[6] Polonsky, J., Balabonova, D., et al., 2009. Equity in community health insurance schemes: evidence and lessons from Armenia. Health and Policy Planning, 24, 209-216.

[7] Mwageni, E., Masanja, H., et al., 2002. Initial studies on health inequalities in the Rufiji River Basin, Coastal Tanzania: Evidence from a Demographic Surveillance System. Tanzania Ministry of Health. TEHIP Discussion Paper No. 3. http://idl-bnc.idrc.ca/dspace/bitstream/10625/40988/1/129039 accessed on 28 June 2014.

[8] Devkota J. U., Singh R. S., 2011. Mathematical modeling of mortality for countries with limited and defective data. Journal of Applied Statistical Sciences, 19 (1), 24-36.

[9] Devkota J. U., Singh R. S., 2010. Deterministic and probabilistic models with applications to modeling fertility data, Journal of Applied Statistical Sciences, 18 (2), 161-176.

[10] Devkota J. U., 2012a. Mortality and Fertility Models for Countries with Limited Data - Results Based on Demographic Data of Nepal, India and Germany. Lambert Academic Publishing, Saarbruecken, Germany, 2012, 2 – 5.

[11] Devkota J. U., Hada B., Prajapati C., Singh S., 2012b. The importance of research data digitization and its statistical analysis-with examples of biogas consumers of Nepal. International Journal for Environmental Science and Development, 3 (2), 103-108, April 2012.

[12] Hair J. F. and Anderson R. E. 1998. Mutivariate Data Analysis. Singapore Pearson 1998.