Research Article

Multivariate statistics of biogas and national grid energy consumers for countries with limited and scarce data: Nepal as a case study

Jyoti U. Devkota[1]

Abstract

This paper analyzes the energy consumption pattern of rural households with respect to multiple variables. Some of these variables are distance travelled for the collection of firewood, firewood saved, sources of firewood collection, type of house, location and type of latrine used, no. of residents more than 15 years of age and composition of adult women more than 25% of the total family size. A survey of 400 households of biogas users and 300 households of national grid energy users of Nepal generated more than 350 multivariate data. Multivariate statistics is used here to quantify and explain the relationship between these variables. Firstly from the factor analysis, three factors are identified. These three factors can explain 78% of the variance of the energy consumption pattern of rural households. Then, multivariate analysis of variance is also conducted. After this, multinomial logistic regression is used for predicting benefits of biogas (time saved) on data 400 households of biogas users. In this, the probability of change in time saved is predicted for a unit change in distance travelled in the collection of firewood before and after the construction of plant, reduced pollution, reduced fuel expenses and amount of firewood saved. Thus, evidence-based analysis of dynamics of improvement of people in general and women in particular when biogas is used as an energy source for cooking is done. The results obtained here can be used for other least developed country in Asia and Africa having limited and scarce official records.

1 Introduction

Energy is needed for sustaining the rapid pace of economic development. Developed countries need energy to keep up with this accelerated pace of development. Least developed world depends on energy for meeting their basic needs. And here, a critical shortage of energy primarily for electricity exists and threatens to remain. This is detrimental to the populations living here [1]. For countries like Nepal without their own natural reserves of fossil fuel, there is an increased dependence on other countries even for meeting its basic energy needs. Thus, a switch over to a renewable energy source can answer all the problems.

Biogas is one of such source of renewable energy that uses animal and human litter for energy generation.

The fragile Himalayan environment here is suffering from rapid deforestation and soil erosion. Total energy consumption here in the years 2008 and 2009 was about 9.3 million tons of oil equivalent (401 million GJ), out of which 87% were derived from traditional resources, 12% from commercial sources and less than 1% from the alternative sources [2].

Non-commercial energy forms, largely fuel wood and other biomass, are still the main source of direct energy use in Nepal for a large proportion of the population. According to 2011 census, 64% of the households use wood/firewood for meeting their cooking needs [3]. The

✉ Jyoti U. Devkota, drjdevkota@ku.edu.np; devkotajb@gmail.com | [1]Department of Mathematics, Kathmandu University, Dhulikhel, Kavre, Nepal.

rural urban differential in the dependence on firewood as a source of cooking is reflected by the fact that 59.03% in the rural areas and 5% in the urban areas depend on firewood for meeting their cooking needs. Cow dung provides energy source for cooking in 10.3% of the households with rural urban differential of 10% and 0.29%, respectively. The dependence of households for cooking on commercial fuel source namely kerosene is 0.387% and 0.638% in rural area and urban areas, respectively. Only 1% households in Nepal use kerosene for meeting their cooking needs. According to WECS energy report 2010, fuel wood is the largest energy source in Nepal providing 77% of the total energy demand in the year 2008/2009. Here, 99.2% is consumed by the residential sectors, 0.2% by the industrial and 0.6% by the commercial sector. Other sources of energy are agriculture residue and animal dung contributing about 4% and 6%, respectively. Share of petroleum fuel in the energy consumption profile of the nation is about 8%. Other sources of commercial energy are coal and electricity with both contributing about 4% in the total energy supply. So, in nutshell, the overall energy consumption of Nepal is largely dominated by the use of traditional non-commercial forms of energy such as fuel wood, agriculture residue and animal waste. The share of traditional biomass resources, commercial energy and renewable energy resources is 87%, 12% and 1%, respectively. There has been a slight decline in the use of traditional fuel as an energy source for over the years with its share of 91%, 88% and 87% in 1995/1996, 2004/2005 and 2008/2009, respectively. The commercial source provided remaining 13% of the energy.

Generation, verification, rectification and prediction of data lead to evidence-based study. This can be applied to data from any discipline. It is thus an interdisciplinary approach to the solution of a research problem. This makes the research objective and undisputable. Halvorsen et al. [4] applied multinomial logistic regression to explore association between general practitioners preferences for time spent on preventive health care activities and age, gender and practice characteristics. Chambers and Whitehead [5] applied multinomial logistic regression for modeling the willingness to pay for wolf protection plan. Similarly, contraceptive use was modeled using multinomial logistic regression by Angeles et al. [6]. Crompton and Wu [7] have used evidence-based method of forecasting energy consumption in China by using Bayesian vector autoregressive methodology. Cai and Jiang [8] analyzed the rural urban differential in energy consumption in China using analysis of variance techniques for parametric and nonparametric data. Bhattacharyya [9] studied the access of energy to India's poor, whereas an overview of energy consumption pattern by available data and the analysis of some relevant aspects of energy policy in rural China are presented

by Zhang et al. [10]. Petrides and Furnham [11] analyzed several dimensions human's emotional intelligence with exploratory factor analysis.

In this paper, the socioeconomic benefits of energy are studied on the basis of collected data. It is shown that the benefit in developing countries far outweigh any directly attributable advantage. Here, the novelty lies in the fact that unlike many papers, the use of statistics is not superficial. Extensive statistical methods are developed and used in drawing in-depth conclusions and inferences. Categorical data are generated from the survey of 700 households. Nepal like many countries in the developing world suffers from limited and scarce data. Countries with limited and scarce data do not have a strong backbone of good quality official data. Remote geographical location, lack of awareness of stakeholders and lack of incentives are the main causes of limited and scarce data. Such data and results of such studies are needed by policy makers and planners for making and achieving developmental goals and initiatives [12]. In the absence of accurate measurement instruments for such countries, the use of methods of generation and prediction of categorical data was considered appropriate to fill this knowledge void. Section 2 explains Data Sources and Methods, and Sect. 3 presents Exploratory Data Analysis. Section 4 explains Models and Estimation, and Sect. 5 presents Results and Discussion. The Conclusion section gives the concluding remarks with possible question for future research.

2 Data sources and methods

2.1 Samples and data collection

This study is based on combined results of 700 rural households based on sample surveys of 300 households of national grid energy users and 400 households of biogas users. The survey areas are numbered in the map of Nepal shown in Fig. 1. Here, 2 which represent Kavre district is the survey area for 300 households and 1, 3 and 4 representing Bhaktapur, Simara and Sarlahi, respectively, is the survey area for 400 households. This detailed sample survey of 300 households was conducted with an aim to have a detailed energy consumption profile. Similarly, a sample survey of 400 households of biogas consumers was also conducted. In both these surveys, the base questionnaire is the same. It is a structured questionnaire with answers given as multiple choices that resulted in categorical data. This categorical data could be analyzed on ordinal scale. Because of large sample size this could be treated as a continuous data by the application of central limit theorem. The questionnaire was pretested. The survey of biogas consumers collected data on their energy consumption
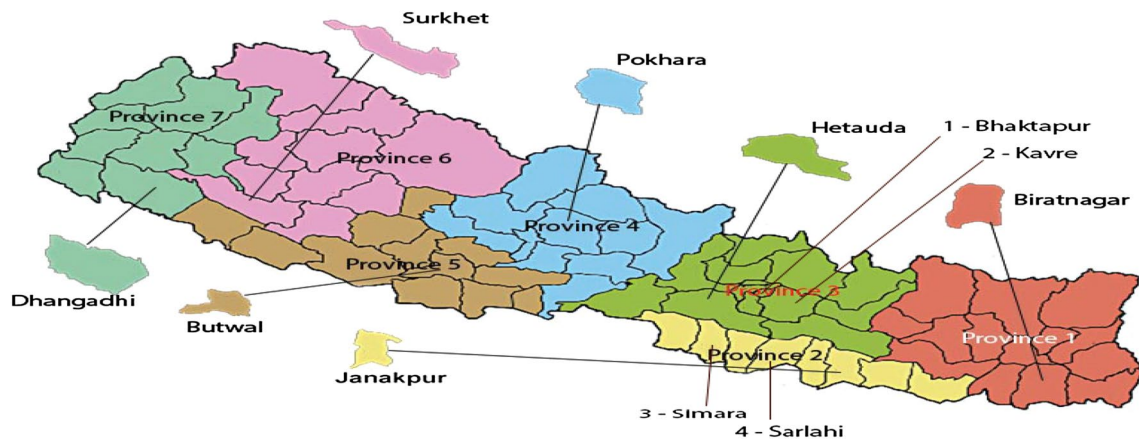
Fig. 1 Survey areas

Table 1 Respondent's profile

Construct and items	Frequency	Percent
Family size		
1–3	55	7.86
4–8	562	80.29
9–14	80	11.43
15–24	03	0.43
Location		
Hill	330	47.9
Plain	370	52.1
Biogas users		
Yes	400	57.1
No	300	42.9

pattern before and after the installation of biogas plant in their household. Their response on before questions helped assess their needs as normal users and potential biogas users. After construction of biogas, questions helped assess the change in their energy consumption pattern after they switched to biogas. The family structure and the location of 700 households are shown in Table 1.

2.2 Statistical methods

Factor analysis: It is a statistical technique studying the interrelationship among the variables in an effort to find a new set of factors that are fewer in number than the original variables. These factors are common among original variables and identify the hidden relationship between the variables. For categorical data, exploratory factor analysis based on principal components analysis method with varimax rotation produces stable results. For this method, normal distribution is not a prerequisite. Elliptical symmetry of the data is required for this method. Kaiser–Meyer–Olkin

measure of sampling adequacy (KMO) is an indicator of level of factorability of the data. KMO > 0.5 indicates good factorability of correlation matrix. The common factor analytic model assumes that a variable consists of common and unique parts. Therefore, the factor model can be considered as

$$X = \Lambda F + U \tag{1}$$

where $X = [X_1, X_2, \ldots, X_p]'$, $F = [f_1, f_2, \ldots, f_q]'$, $U = [e_1, e_2, \ldots, e_p]'$.

X is the data vector, F is the vector of common factor and U is the unique factor. The elements of F and U cannot be observed from investigation. The matrix of factor loadings is given in Eq. (1) where λ_{ij}'s ($i = 1, 2, \ldots, p; j = 1, 2, \ldots q$) are unknown.

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \ldots & \lambda_{1q} \\ \lambda_{21} & \lambda_{22} & \ldots & \lambda_{2q} \\ \ldots & \ldots & \ldots & \ldots \\ \lambda_{p1} & \lambda_{p2} & \ldots & \lambda_{pq} \end{bmatrix} \tag{2}$$

It is assumed that $e_1, e_2, \ldots, e_n$ are mutually independent and they are also independent of the elements in F. That is $\mathrm{cov}(U, F') = 0$. Also, it is assumed that

$$E[UU'] = \begin{bmatrix} \psi_1 & 0 & 0 & \ldots & 0 \\ 0 & \psi_2 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \psi_p \end{bmatrix} \tag{3}$$

$$E(F) = 0, \ V(F) = 1$$

Multinomial logistic regression: It is used to predict categorical placement in or the probability of category membership on a dependent variable based on multiple independent variables. It is a simple extension of binary logistic regression that allows for more than two categories of the dependent or outcome variable. Multinomial

logistic regression uses maximum likelihood estimation to evaluate the probability of categorical membership.

Let J denote the number of categories of Y. Here, Y is a multinomial response variable. Let $\{\pi_1, \pi_2 \ldots \ldots \pi_j\}$ denote the response probabilities, satisfying the condition that their sum is equal to 1. Logit models for multinomial response pair each category with a baseline category [13]. When the last category (J) is the baseline, the baseline category logits are

$$\log(\pi_i/\pi_j), \quad j = 1, \ldots J - 1 \qquad (4)$$

Given that the response falls in the category j or J, this is the log odds that the response is j.

The baseline category logit model with predictor x is

$$\log(\pi_j/\pi_j) = \alpha_j + \beta_j x, \quad j = 1, \ldots J - 1 \qquad (5)$$

The model has $J - 1$ equations with separate parameters for each. The effects vary according to the category paired with the baseline. When $J = 2$, this model simplifies to single linear equation for $\log(\pi_1/\pi_2) = \text{logit}(\pi_1)$, resulting in ordinary logistic regression for binary responses.

There is an odds ratio associated with each predictor. It is denoted by Exp(B). It is more than 1 in cases where predictors increase the logit, Exp(B) is equal to 1 in cases where predictor do not have any influence on the logit and Exp(B) is less than 1 in cases where predictors decrease the logit.

Multivariate analysis of variance for two-way classification: The multivariate generalization of the analysis of variance for testing the equality of mean vectors of several populations if the total covariance matrix is partitioned into two components, viz., between—populations and within populations covariance matrix. This analytical procedure is a special case of the multivariate general linear hypothesis to carry out test on the parameter in the model. Each component of response in the model is composed of the p values observed from p variables. So let us consider that Y is an observed matrix of order $n \times p$ of p response variables on each of n objects. These response variables depend on some other variables X_i's where there are ($q - 1$) independent variables observed on each of the n objects. Then, the multivariate regression model is written as

$$Y = XB + U \qquad (6)$$

where X is of order $n \times q$, B is a matrix of unknown regression parameters and is of order $q \times p$ and U is a matrix unobserved random disturbances. The matrix X in (2) may be a random matrix and the model is called multivariate regression model. This is multivariate regression analysis.

The elements in the X matrix may be 0 and 1 indicating the absence or presence, respectively, of some factors. Then, Eq. (6) is called a general linear model and we have a case of multivariate analysis of variance.

3 Exploratory data analysis

Primary data generated for this study result from the survey of 700 households. This survey generated more than 350 multivariate categorical data. The questionnaire used for data generation had answers given as a multiple choice option. These answers generated categorical data that could be classified on ordinal scale. The categorical data generated from variables later used in factor analysis are in Table 2. These variables measure the amount of energy consumed. For example as seen from Table 2, variables time spent in the collection of firewood and amount of firewood consumed reflect the quantity of energy consumed. Due to the large sample size, these categorical data, classified on ordinal scale can be treated as continuous. The results of statistical analysis can be physically interpreted. This is by the application of central limit theorem. This type of classification of these data reduces the ambiguity in response of the energy users. Here, data related to types of house and types of toilet which are proxy asset indicators are used to study the contribution of socioeconomic status on energy consumption. The impact of family size and family composition through adult female composition of more than 25% of the family size and number of residents more than 15 years is also analyzed.

The mean values of these variables along with the results of factor analysis are in Table 3. Due to the large sample size and through the application of central limit theorem, these values can be interpreted physically. For example, the mean time spent in the collection of firewood is 3.22 for all 700 households, implying that a household spends on average around 45 min per day for the collection of firewood. This time spent is directly proportional to the amount of energy spent in fuel wood collection. Distance travelled per day for the collection of firewood, amount of time spent per day in this process and the source of this firewood are the indicators of energy spent by a household for the consumption of energy gained from non-renewable energy source. The human energy invested in the collection of firewood could have been diverted to the conduction of income-generating activities. The benefits of biogas to women and utilization of time saved by biogas in specific income-generating activity are shown in Figs. 2 and 3. As seen from Fig. 2, out of 400 households, women in 396 say that a switch over to biogas saves their time and out of which 349 got involved in income-generating activities. The spare time is used in

Table 2 Details of the variables 700 households used in factor analysis

| S. no.	Variable name	Classification Likert Scale	Frequency
1	Type of house		
	Concrete	1	170
	Tile/asbestos	2	106
	Modern lite roof	3	262
	Mud house	4	162
2	Own latrine		
	No	0	66
	Yes	1	633
	Public latrine	2	01
3	Location of latrine		
	Inside the house	1	09
	Near the house	2	610
	Far from house	3	13
	No	4	68
4	Type of latrine		
	Well built	1	613
	Temporary	2	21
	Open defecation	3	61
	No	4	05
5	Time spent in the collection of firewood (per day)		
	No	0	52
	Less than 15 min	1	45
	15–30 min	2	194
	30–45 min	3	95
	45–60 min	4	57
	1–2 h	5	257
6	Source of firewood		
	From own farm	1	312
	From neighbor's farm	2	1
	From distant forest	3	249
	From own and neighbor's farm	4	7
	From own farm and distant forest	5	55
	From own farm and buying	6	16
	Buy firewood	7	44
	Distant farm and buying	8	15
	NA	9	1
7	Adult female population more than 25% of the family size		
	No	0	290
	Yes	1	410
8	No. of residents more than 15 years	Ratio data	

the involvement in children's education by 26 households. As shown in Fig. 3, the number of households engaged in farming is 331 and those rearing livestock is 243. These choices are not mutually exclusive as many families have chosen farming as their first choice and livestock rearing as their second choice. Similarly, many women have chosen time saved as their first choice and clean atmosphere as their second choice. Daily wages of farm workers in normal season amount to USD 8 from 11:00 A.M. to 4:00 P.M. This amounts to around 2 USD per hour. Distance covered for the collection is longer and sources of firewood are farther for the household without biogas. Further, the percentage of adult women and adult members constituting a rural family plays an important role in decision related to choice of a renewable energy source. The total variables explaining the social and economic landscape of biogas

Table 3 Result of factor analysis of 700 households

S. no.	Variable name	All users (sample size=700 HH)					Biogas user (sample size=400 HH)				
		Mean	KMO	SD	Factor loading	Factor group	Mean	KMO	SD	Factor loading	Factor group
			0.730					0.700			
1	Time spent in the collection of firewood	3.22		1.6269	0.740	Status	4.0575		1.60175	0.661	Resources
2	Type of house	2.59		1.092	−0.920	Infrastructure	2.48		1.336	−0.856	Infrastructure
3	Own latrine	0.91		0.303	0.941	Infrastructure	0.98		0.193	0.904	Infrastructure
4	Location of latrine	2.20		0.617	0.921	Infrastructure	2.07		0.407	0.870	Infrastructure
5	Type of latrine	1.23		0.626	0.791	Time for fuel	1.11		0.440	−.728	Resources
6	Source of firewood before	2.71		1.95	0.699	Time for fuel	4.12		1.683	0.865	Resources
7	Adult female population more than 25% of the family size	0.59		0.493	−0.646	Status	0.94		0.242		
8	No. of residents more than 15 years of age in the house	4.40		1.919			4.12		1.683	0.509	Size of household
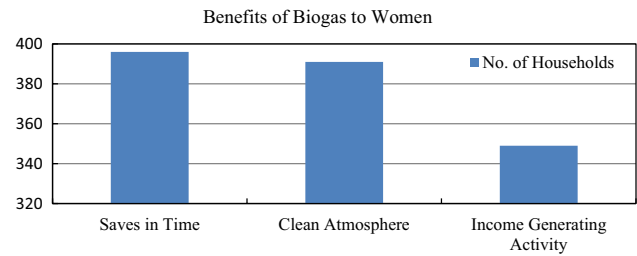
Fig. 2 Benefits of biogas to women
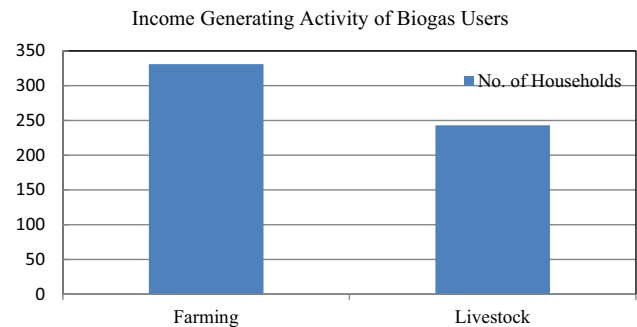
Fig. 3 Income-generating activity

users are in Table 4. These variables also form a part of biogas consumer profile database.

4 Models and estimation

4.1 Factor analysis of energy consumption profile of normal users and biogas users

Sources of firewood, time spent in the collection of firewood, type of house, location and type of latrine used, no. of residents more than 15 years of age and composition of adult women more than 25% of the total family size are used to assess and cross-validate the energy pattern of rural families. The last two variables namely no. of residents more than 15 years and composition of adult women more than 25% of the total family size will help validate that families with large adult composition of adult members in general and women in particular spend more time and cover longer distances in the collection of firewood. Further, such households are less likely to switch over to alternative sources of energy. These variables along with their values on ordinal scale are in Table 2. They are then analyzed using factor analysis. This analysis is conducted for 700 households of energy users in general and 300 households of biogas users. Exploratory factor analysis is done using principal component analysis method (PCA) and varimax rotation. The sampling adequacy factors

Table 4 Overview of biogas consumer profile database

Table nos.	Heading of the table	Variables	Types of data
1	Family background	10	Yes/no and nominal
2	Family description	60	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	Yes/no
9	Socioeconomic condition	54	Yes/no, nominal
10	Water needs	17	Yes/no
11	Biogas questions	75	Nominal, ordinal
12	Size of plant	18	Yes/no
13	Woman empowerment	44	Nominal
14	Source of firewood	14	Yes/no
15	Distance covered for firewood	8	Yes/no
16	Comparison of types of fuel used	40	Yes/no, continuous
17	Time spent on firewood collection	10	Yes/no
18	Frequency of feeding	15	Yes/no
19	Cultivation of crops	32	Yes/no
20	Application of fertilizers	20	Yes/no
21	Health related issues	8	Nominal
22 and 23	Positive impacts of biogas	9	Yes/no
	Total	467	

range over 0.60. This study is based on categorical data of 700 households, KMO > 0.5 indicates that sample size is adequate. The determinant of correlation matrix is 0.02. It is > 0.00001 indicating that there is no multi-collinearity. Similarly, Bartlett test of sphericity rejects the null hypothesis that the correlation matrix is identity at a very high level of significance. The factor loadings for normal users and biogas users are given in Table 3.

4.2 Multinomial regression of benefit of biogas users

A consumer profile data base of biogas consumers was constructed and information on 476 variables was stored.

The overview of this data base is in Table 4. The impact of benefits of use of renewable energy in particular and biogas in general is highlighted and quantified. Multinomial logistic regression is fitted to variable related to the benefits of biogas. The descriptive statistics of these variables and their categorical values are shown Table 5. These values of mean, mode and median can be physically interpreted by central limit theorem. For example, for Variable A representing distance travelled for the collection of firewood before the plant, takes values 1, 2, 3, 4 and 5 on the categorical scale. The mean value is 4.08 implying that a house on average covered 200–500 m for the collection of firewood. The variable saves time is regressed using logistic regression on distance travelled

Table 5 Descriptive statistics

Variables	Mean	Mode	Median	SD	Values and value taken on ordinal scale by the categorical data				
A	4.08	5.0	5.0	1.3248	Not travelled (1)	100 m (2)	100–200 m (3)	200–500 m (4)	> 500 m (5)
B	2.37	2.0	1.0	1.633	Not travelled (1)	100 m (2)	100–200 m (3)	200–500 m (4)	> 500 m (5)
C	2.85	3.0	3.0	0.931	No time (1)	Up to 60 min (2)	1–3 h (3)	3–5 h (4)	> 5 h (5)
D	1.94	2.0	2.0	0.233	Yes (1)	No (2)			
E	1.44	1.0	1.0	0.496	Yes (1)	No (2)			
F	1.99	2.0	2.0	0.100	Yes (1)	No (2)			
G	1.72	2.0	2.0	0.448	Yes (1)	No (2)			
H	3.51	4	4	0.67	No (1)	30 kg (2)	30–50 kg (3)	More than 50 kg (4)	

before, distance travelled after, female benefitting from reduced fuel expenses, male benefitting from reduced fuel expenses, male benefitting from reduced pollution, female benefitting from reduced pollution and amount of firewood saved per month. The result of this logistic regression is in Tables 6 and 7. As seen from the table, the results are highly significant.

4.3 Multivariate analysis of variance (MANOVA)

Multivariate analysis of variance is conducted with the three extracted factor scores as dependent variables and time taken for the collection of firewood (after the installation of plant) and source of firewood (after the installation of plant) as independent variables. After the installation of biogas plant, the sources of firewood collection and the interaction between the time taken for the collection of firewood and the source of firewood are significantly different with respect (with p value of 0.008 and 0.001) to the three factors identified after factor analysis. They are namely status, infrastructure and time for fuel. These three factors extracted from seven variables explore the role social and economic status classified in terms of status, time for fuel and infrastructure in the energy consumption pattern of a rural household. These factors differ significantly with respect to source of firewood and the interaction between time and firewood.

5 Results and discussion

Eight observed variables related to various aspects of energy consumption of rural household are reduced to three unobserved factors using factor analysis. Thus, the

hidden structure in data of 700 households (300 households of national grid energy users and 400 households of biogas users) for eight variables is revealed by three factors. These factors explain 78% of the total variability. As seen by the factor analysis results in Table 3, three factors are infrastructure (basic), status (type of house) and time (spent on collection of fuel), and explain 42%, 14% and 21% of the total variance. The exploratory factor analysis with varimax rotation is conducted here as it not based on the assumption of normal distribution. By using central limit theorem, sample size of 700 households for energy users of rural households (national grid energy and biogas energy) and 400 for biogas users is large enough for the assumption for treating this data as a continuous data. Similarly, 400 households of biogas consumers had three factors namely resources (time and money), infrastructure (basic) and size of the family. These factors explain 66% of the total variability of the data. The factor loadings are also given in Table 3. The factor analysis is done on the basis of before the installation of plant questions. The biogas consumers were asked questions related to energy consumption pattern before and after they installed the biogas plant.

The variables used in logistic regression are given in Table 5. We represent distance travelled for the collection of firewood before the construction of plant by A, distance travelled after by B, saves time by C, reduced fuel expenses by female by D, reduced fuel expenses by male by E, reduced pollution by male by F, reduced pollution by female by G and firewood saving by H. Although these are ordinal and nominal data, but the measures of central tendencies such as mean, mode and median give us an idea of the average response and most popular response. The standard deviation gives an idea of the consistency of the response. By the application of central limit theorem, these data can be treated as continuous data. Multinomial logistic regression is fitted to these variables. The variable saves time is regressed on distance travelled before, distance travelled after, female benefitting from reduced fuel expenses, male benefitting from reduced fuel expenses, male benefitting from reduced pollution,

Table 6 Model fitting information

Model	− 2 Log likelihood	Chi-square	df	Sig
Intercept only	642.490			
Final	364.621	277.869	56	0.000

Table 7 Likelihood ratio tests

Effect	− 2 Log likelihood	Chi-square	Degree of freedom	Significance
Distance travelled after	426.16	61.54	16	0.000
Male benefit from reduced fuel	374.286	9.666	4	0.046
Female benefit from reduced fuel	373.394	8.774	4	0.067
Male benefit from reduced pollution	365.784	1.164	4	0.884
Female benefit from reduced pollution	436.983	72.363	4	0.000
Firewood saved	390.633	26.012	8	0.001
Distance travelled before	403.043	38.422	16	0.001

female benefitting from reduced pollution and amount of firewood saved per month.

The likelihood ratio test of regression of saves time (dependent variables) on ordinal data distance travelled after, distance travelled before and firewood saved shows highly significant results. Similarly, the response of reduced fuel expenses by men and reduced pollution by women has also shown significant results. Through multinomial logistic models, the benefits have been quantified. The results can be summarized in the following manner.

The results of fitting a multinomial regression model are given in Tables 6 and 7. We see that this regression is highly significant with a Chi-square value of 56. The likelihood ratio tests in Table 7 also show significant results at 5% level of significance for most of the independent variables. The accuracy of this model is 62.8% with respect to the correspondence of these values with observed and expected. This is quiet high for a data based on response of human beings. These results are not based on closely controlled laboratory conditions. Although all the possible sources of error have been traced out and eliminated in this survey, experiments based on humans are more likely influenced by other psychological factors. Hence, an accuracy of 62.8% is a good result.

The improvement in lives of biogas users outweighs its direct advantage of merely getting to cook in gas stoves is displayed in Table 8. The improvement in the lives of biogas consumers is quantified in terms of odds ratio here. We see that those who travelled up to 100 for the collection of firewood before the construction of plant their likelihood of claiming the biogas saved their time increased by 19,730,000, 23,230,000 and 21,670,000 for up to 60 min, 1–3 h and 3–5 h, respectively. Similarly, those who covered 100–200 m for the collection of firewood, their likelihood of responding that it saved their time increased by 1.945 and 1.582 times for up to 60 min and 1–3 h, respectively. This is in comparison with those households who had to cover more than 500 m for the collection of firewood before the construction of plant. Similarly, those who do

not have to travel at all for the collection of firewood after the construction of the plant are 3.478 and 4.706 times more likely to respond that biogas saved 3–5 h and more than 5 h than those who still cover more than 500 m. These are all the Exp(B) values. They are also called the odds ratio. Odds ratios in this table are different from 1, showing significant impact of use of biogas. There were 229 households which covered more than 500 m before and now, after the construction of the plant, it is 88. This shows that there is a considerable reduction in the distance travelled. Less distance travelled implies more time saved which in turn implies more money saved. As this time saved is used in income-generating activities by these household. So a switch over to renewable energy is not only an adaptation strategy to climate change but it can be also increase the socioeconomic status of areas hit with scarcity of fossil fuel. Similarly, the odd of those women favoring biogas as source of energy that has reduced effects of pollution also supporting its time saving benefits is very highly significant. This feature is reflected more prominently in women as they have to suffer from the pollution of traditional cooking stoves.

Multivariate analysis of variance is done with the factor scores obtained from the three factors. The aim is to study the effect of time spent in the collection of firewood and source of firewood collection for the biogas users on these factor scores. The source of firewood (SourceAfter) and interaction between time spent in the collection of firewood and source of firewood (AFTIMES * SourceAfter) are significantly different with p values of 0.008 and 0.001 for Wilks Lambda. Sources of firewood collection and its interaction with the time taken for the collection are significantly different with respect to the three factors.

6 Conclusion

Three major multivariate analysis techniques are applied to the data collected from 400 households of biogas users and 300 households of national grid electricity users. They are namely factor analysis, multiple analysis of variance and logistic regression. The results of these techniques show the quantification of social and economic benefits that outweigh directly attributable advantages of energy use which is its use as a cooking gas. Categorical data generated from the primary data collected in the sample surveys have been shown to be suitable for the study of intangible factors and on impact on merit of energy use among rural households.

The result of factor analysis states that three hidden factors are namely status (type of house), infrastructure (basic) and time for fuel govern the energy consumption dynamics. These factors explain 78% of the energy

Table 8 Quantification of impact of biogas in terms of time saved-odds ratio

Distance travelled for firewood	Odds in favor	Time saved after plant construction
Before plant construction		
100–500 m	19,730,000	60 min
	23,230,000	1–3 h
	21,670,000	3–5 h
After plant construction		
0–500 m	3.478	3–5 h
	4.706	More than 5 h

consumption pattern of rural households. Multiple analysis of variance tests the significance of difference between the factor scores computed from these three factors with respect to independent variables time for the collection of firewood and source of collection of firewood for the biogas users. Multinomial regression has been used in predicting and quantifying the benefit of biogas in terms of time saved. Time saved is money saved. This time saved has been used in income-generating activities. Large values of odds ratio thus obtained from the multinomial regression have quantified great benefits obtained from biogas plants. These benefits are more for women than men.

Similar survey with exact quantification is planned for the further extension of this work. In future, required instruments for exact measurement along with the questionnaire will be carried by the data collection team. This will be one of many strategies employed for the accuracy of the collected continuous ratio data.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Goldsmith K (1993) Economic and financial analysis of hydropower projects, vol 6. Norwegian Institute of Technology, Division of Hydraulic Engineering, Trondheim, pp 165–169
2. WECS (2010) Energy sector synopsis report. Water and Energy Commission Secretariat, Kathmandu, Nepal
3. CBS (2011) National report on national population and housing census 2011, vol 1. Central Bureau of Statistics Nepal, pp 25–29
4. Halvorsen PA et al (2013) What professional activities do general practitioners find most meaningful? Cross sectional survey of Norwegian general practitioners: family practice. BMC Ser 14:41
5. Chambers CM, Whitehead JC (2003) A contingent valuation estimate of the benefits of wolves in minnesota. Environ Resource Econ 26:249–267
6. Angeles G, Dietrich J et al (2001) A meta-analysis of the impact of family planning programs on fertility preferences, contraceptive method choice and fertility. Carolina Population Center, University of North Carolina. https://www.measureevaluation.org/resources/publications/wp-01-30. Accessed 19 Mar 2019
7. Crompton P, Wu Y (2005) Energy consumption in China: past trends and future directions. Energy Econ 27(1):195–208
8. Cai J, Jiang Z (2008) Changing of energy consumption patterns from rural households to urban households in China: an example from Shaanxi Province. Renew Sustain Energy Rev 12:1667–1680
9. Bhattacharya SC (2006) Energy access problem of the poor in India: Is rural electrification a remedy? Energy Policy 34(3387–33):97
10. Zhang L, Yang L et al (2009) Rural energy in China: pattern and policy. Renew Energy 34:2813–2823
11. Petrides KV, Furnham A et al (2000) On the dimensional structure of emotional intelligence. Personal Individ Differ 29(2):313–320
12. Devkota JU (2012) Mortality and fertility models for countries with limited data. Lambert Academic Publishing, Saarbruecken, pp 1–4
13. Lindsey JK (2006) Introduction to applied statistics a modeling approach. Oxford University Press, New Delhi, pp 84–105

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.