Konya Mühendislik Bilimleri Dergisi, c. 11, s. 3, 715-729, 2023
Konya Journal of Engineering Sciences, v. 11, n. 3, 715-729, 2023
ISSN: 2667-8055 (Electronic)
DOI: 10.36306/konjes.1217013

MODELLING APPROACH FOR ENERGY DEMAND AND CONSUMPTION USING GENERALIZED LINEAR MODELS

*Fuad ALHAJ OMAR

Zonguldak Bülent Ecevit University, Electric and Energy Department, Zonguldak, TÜRKİYE
fuad.a@beun.edu.tr

Highlights

- Statistical analysis of energy consumption data using R-Studio.
- Energy consumption model for households in Syria.
- The most cost-effective energy sources in Syria.

Graphical Abstract

***Corresponding Author:** Fuad ALHAJ OMAR, fuad.a@beun.edu.tr

Konya Mühendislik Bilimleri Dergisi, c. 11, s. 3, 715-729, 2023
Konya Journal of Engineering Sciences, v. 11, n. 3, 715-729, 2023
ISSN: 2667-8055 (Electronic)
DOI: 10.36306/konjes.1217013

MODELLING APPROACH FOR ENERGY DEMAND AND CONSUMPTION USING GENERALIZED LINEAR MODELS

***Fuad ALHAJ OMAR**

Zonguldak Bülent Ecevit University, Electric and Energy Department, Zonguldak, TÜRKİYE
fuad.a@beun.edu.tr

(Geliş/Received: 09.12.2022; Kabul/Accepted in Revised Form: 25.05.2023)

ABSTRACT: Energy management is an important process for maintaining available energy resources and meeting basic household energy needs. Many studies seek to optimize the household energy consumption patterns to manage the load demand and minimize energy costs. Adopting such optimizations in conflict-affected countries is more beneficial due to limited energy sources. This study identifies the optimal energy consumption model for households in northern Syria. The objective is to identify the most cost-efficient energy sources while considering the prices, average monthly household income, the main source of electricity, battery storage capacity, and energy demands for space heating, water heating and cooking. One hundred and thirty-six (136) standardized surveys of residential households are collected and used as a test case. Statistical analysis of the data was carried out using the R-Studio software, where Poisson regression and negative binomial regression were employed. Findings revealed that the Negative Binomial (NB) model used has high explanatory power. In addition, the energy sources used for space heating and water heating have a direct impact on monthly expenditures. The produced model showed that the most cost-effective energy sources are coal for space heating and natural gas and kerosene for water heating.

Keywords: *Energy Efficient, Energy Management, Optimizing Energy, Poisson Regression, Negative Binomial Regression*

1. INTRODUCTION

Most of the electricity infrastructures during the Syrian civil war have been destroyed [1-3]. This has been having a devastating effect on more than 90% of Syrians. As the main source of power supplies dropped off, Syrians had to be creative in finding alternative sources for light and power. Across Syria, people survive with little or no energy. Energy poverty has a knock-on effect on other basic services like water supplies and health services [4, 5]. We have seen first-hand the damaging effects of power shortages in Syria. We witnessed the devastating effect of the energy shortage on millions of Syrians. Aleppo city was Syria's powerhouse and home to over four million people. It became a battleground and remained so for more than four years. Syrians were undergoing enormous suffering because of cuts to energy supplies, used by each of the opposing parties to pressure the other. Energy shortage became a matter of life or death. Our recent pilot research (submitted for publication) showed that energy shortages have a direct impact on displacement and immigration, create barriers to education, and reduce social and psychological well-being. Therefore, the most cost-effective energy sources must be urgently identified to help people secure their main energy needs at the lowest costs. Thus, improving the energy use efficiency according to the household's monthly income. This research focuses on creating societal and ecosystem changes and develops a novel solution to the energy efficiency problem in north Syria that is more effective, efficient, sustainable and is beneficial primarily to society as a whole rather than private individuals. The results will have an impact on improving the households' living conditions where perceived self-efficacy with energy-saving behaviors and knowledge about behaviors all relate to achieving sustainability.

***Corresponding Author:** Fuad ALHAJ OMAR, fuad.a@beun.edu.tr

Existing literature has highlighted the management and optimization of energy consumption. The challenge is that Syria has been ranked as one of the fragile states globally in the past decade [6]. In fact, the conflicts between two parties, which are the prevailing form in most countries in the Middle East, do not result in positive social change or military victory for any party [7, 8]. Moreover, the prolonged period of conflicts in the Middle East may generate successive waves of conflicts [9]. Armed conflict may be exacerbated by natural and environmental threats such as climate change and urbanization. As well, the rapid change in advanced technology, communication systems, and the media may produce new and unfamiliar forms of wars [10]. Wherefore, any conventional development plans may be condemned to failure [11], and hence energy planning models are also susceptible to failure [12]. Therefore, there is a need to understand the complex structure of conflicts in order to develop flexible and long-term plans and strategies to address the energy problem [13, 14].

Some studies have considered the context when seeking to solve energy problems facing fragile societies by attempting to achieve the best available options [15-18]. For instance, the cost of capital for investing in solar energy has been changed according to different political situations [19]. Pettersson and Wallensteen [20] examined how the energy and water infrastructure can be made more resilient in the face of armed conflicts between two parties.

Zerriffi et al. [21] explained how the system attributes that are useful under conflict might undervalue if the reliability assessment takes into account only normal operating situations. Brazilian et al. [22] explained that typical values of parameters in a fragile country are unrealistic, which makes the resulting recommendations also irrelevant. Instead of that, they considered the fragility when developing the least expensive planning models through lower available capital, higher interest rates, extended construction time, and damages over the entire planning period.

Patankar et al. [23] believed that the generating assets could be damaged as a result of the conflict in South Sudan, so they decided to use stochastic programming to test power system plans that are not affected by risks.

In any case, all the proposed models and strategies were simulating the context in the studied area, whose conditions differ from one country to another. In other words, the strategies and models proposed in a particular context cannot be fully adopted in another context because the conditions and the nature of the conflict are different.

After the 2011 Arab spring, the Middle East was affected by socioeconomic factors. That pushed states deeper into a vicious cycle of poverty and deterioration [24]. Socio-economic factors are associated with high unemployment, poverty, social inequality, and frustration, all of which lead to the loss of human capital that is exacerbated after adding the high cost of living and unemployment [25]. The depletion of available resources, inflation, dependence on imports and price fluctuations exacerbate the social-economic problems [26, 27]. All of these factors combined can lead to internal and international displacement, and these displacement movements can be exacerbated as a result of protracted armed conflicts [28]. The displacement movements and the concentration of population in a limited area can lead to the collapse of infrastructure and insufficient resources available to meet the growing need [13, 29, 30].

As a result of the ongoing conflict in Syria, there has been a significant exodus towards northern Syria. Our recent research showed that there are about four million people living in North and Northwest Syria. As a result of this population density, there is a huge gap between available energy and household requirements. The energy burden for limited-income households has increased due to fluctuating prices of fossil fuels, outdated appliances, and energy inefficient homes [4]. Accordingly, there is an urgent need to build an energy consumption model to identify the most cost-effective energy sources to help people in northern Syria optimize their energy use.

In this paper, we have developed a feasible and actionable model that identifies the most cost-effective energy sources according to the household's monthly income. Two models were built based on Poisson regression and negative binomial regression, and according to the regression model evaluation criteria, the optimal model was determined. Note that this model only fits the context in northern Syria. However, the modelling and analysis method can be applied to propose a model compatible with any context elsewhere.

2. METHODOLOGY

In the framework of this study, the methodology was based on the collection of statistical information by using structured questionnaires. These served to explore deeply the reality of energy in northern Syria. We focused on families (head of household) for determining the daily need and consumption of energy in the target region. The questionnaire was coded to be used by the program “KoBo Toolbox”; this is a free open-source tool to collect data using mobile devices without a network connection. Before starting the questionnaire, the team consisting of 22 specialized data collectors (12 males and 10 females) was well trained on the questionnaire, and several experimental questionnaires were conducted to ensure the quality of the method used, also to ensure that the questions are clear, understandable, and appropriate to the context of the target region.

Also, considering the Syrian context, we were keen to form teams for data collectors, each team consisting of a male and a female, and therefore we did not face any challenges related to conducting the interviews with the female adults when the head of household is not available at home, so the sample was representative of females as well. The sample size was 136 samples of both genders and of different ages. The collected samples were randomly selected from 35 communities located in northern Syria as shown in Fig. 1. After reviewing the collected data, the analysis and modelling process was performed. Parameters of the Negative Binomial and Poisson regression models are estimated using the maximum likelihood (ML) method with "Newton Raphson (NR)” and “Fisher-Scoring (FS)" iterative algorithms in R software. Figure 2 shows the flow chart of the methodology stages.

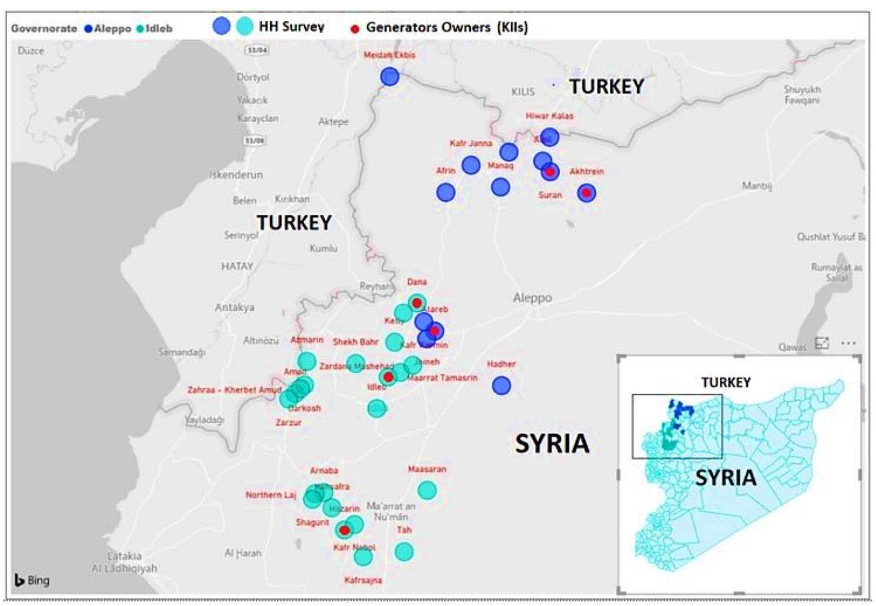

Figure 1. Mapping of targeted communities studied by household survey

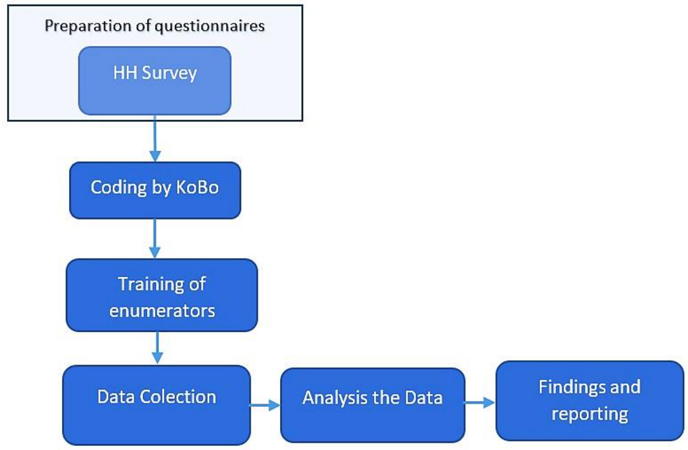

Figure 2. Flow chart of the methodology

3. REGRESSION MODELS

Regression analysis is one of the methods of inferential statistics that is concerned with estimating the relationships between a dependent variable (often called the 'outcome' or 'response' variable) and one or more independent variables [31-33].

The concept of regression dates to the 18th century when it was used to solve astronomical navigation problems. In 1805 Legendre developed the method of least squares to determine the averages of a regression model [34]. Then Gauss showed that this method is really the best solution when the distribution of random errors is normal [35]. This methodology was later used in the physical sciences until the 19th century. Galton was one of the first to use the term "meaning regression," and then regression, in his study of the height of fathers and their children in the late nineteenth century [36, 37].

If the regression model contains one independent variable, it is called a simple regression model, and if it contains more than one independent variable, it is a multiple regression model, and the regression model can be linear or non-linear [33]. The mentioned models are based on quantitative variables, but if the dependent variable is categorical (it has a finite number of categories and therefore has a non-normal distribution), then another model should be used, such as a logistic regression model. Which studies the relationship between a response variable of two classes and the independent variable which can be continuous (quantitative) or discontinuous [38]. When the dependent variable follows a distribution from the exponential family, such as the Poisson distribution, the gamma distribution, and the negative binomial distribution, then in this case generalized linear models are used [39].

3.1. Poisson Regression model

The Poisson regression model is also called the log-linear model and is often used to model count data. In addition, it allows the frequencies of events to depend on one or more variables, provided that the events are independent given these variables [40, 41]. If Y is the number of occurrences, its probability distribution can be written as [42, 43]:

$$P(y) = \frac{e^{-\mu}\mu^{y}}{y!}; \quad y = 0,1,2,\ldots \qquad (1)$$

where μ is the average number of occurrences and it requires careful definition. Often it needs to be described as a rate.

Let $Y_1, \ldots, Y_N$ be independent random variables with Y_i denoting the number of events observed from exposure n_i for the ith covariate pattern. The expected value of Y_i can be written as :

$$E(Y_i) = \mu_i = n_i.\theta_i \qquad (2)$$

The dependence of θ_i on the explanatory variables is usually modelled by [39]:

$$\theta_i = e^{x_i^T \beta} \qquad (3)$$

Therefore, the generalized linear model is:

$$E(Y_i) = \mu_i = n_i e^{x_i^T \beta} \quad ; \qquad Var(Y_i) = \mu_i \qquad (4)$$

According to the generalized linear model concept, three components are required to form a Poisson model. The first of these is the Poisson distribution, the second one is a linear predictor that describes the relationship between the independent variables and coeffıcıents as follows:

$$\eta_i = \beta_0 + \beta_1 y_{i1} + \beta_2 y_{i2} + \ldots\ldots + \beta_p y_{i(p)} \qquad (5)$$

The third component is a link function to obtain a functional relationship between the expectation of the response variable to linear predictors. There are many link functions for the studied model and the most significant link function is the log link function, which is given as follows:[38, 44, 45];

$$log\, \mu_i = log\, n_i + x_i^T \beta \qquad (6)$$

Equation (6) differs from the usual specification of the linear component due to the inclusion of the term $log\, n_i$. This term is called the offset. It is a known constant which is readily incorporated into the estimation procedure. As usual, the terms x_i and β describe the covariate pattern and parameters, respectively.

3.2. Negative Binomial (NB) Regression

The negative binomial regression model is derived from Poisson-gamma distribution [46]. The classic derivation of the negative binomial distribution is as the number of failures in Bernoulli trials until r successes. If π is the probability of success on each Bernoulli trial, then the number of failures y has the probability function [42]:

$$f(y) = \pi \times \binom{r+y-1}{r-1} \pi^{r-1}(1-\pi)^y \qquad (7)$$

$$f(y) = \binom{r+y-1}{r-1} \pi^{r}(1-\pi)^y \ , \quad y = 0, 1, 2, \ldots .... \qquad (8)$$

which depends on π and r.

The above formulation supposes r is a positive integer. However, the negative binomial distribution can be defined for any positive values of r, by using the gamma function in place of factorials:

$$f(y) = \frac{\Gamma(y+r)}{y!\,\Gamma(r)} \pi^r (1-\pi)^y , \qquad y = 0, 1, 2, ... \qquad (9)$$

In generalized linear modeling the following parametrization is convenient:

$$\mu = \frac{r(1-\pi)}{\pi} , \qquad \kappa = \frac{1}{r} \qquad (10)$$

Using this notation, the probability function of y is:

$$f(y) = \frac{\Gamma(y+\frac{1}{\kappa})}{y!\,\Gamma(\frac{1}{\kappa})} \left(\frac{1}{1+\kappa\mu}\right)^{\frac{1}{\kappa}} \left(\frac{\kappa\mu}{1+\kappa\mu}\right)^{y} , \quad y = 0, 1, 2, \ldots \qquad (11)$$

The mean and variance of response variable of NB distribution are given as follows:

$$E(y) = \mu , \qquad Var(y) = \mu(1 + \kappa\mu) \qquad (12)$$

The parameter κ is called the "overdispersion" or "shape" parameter.

The log link function is used for Negative Binomial Regression Model as link function and given as follows:

$$log(\mu_i) = X'\underline{\beta} \qquad (13)$$

Where:

$X_{i1}, \ldots, X_{iq}$ are the independent variables.

π is the probability of a claim,

μ is the mean claim size (the average number of occurrences).

3.3. Poisson and NB regression models assumptions

There are some assumptions related to both the Poisson regression and the Negative Binomial regression:

1. Assumption: the conditional probability function of the response variable follows the Poisson distribution with μ (the average number of occurrences) [47, 48].

2. Assumption: the Poisson distribution parameter μ for the i_{th} subject equals the exponential form of the linear predictor [39, 45]:

3. Assumption: there is independence between the explanatory and the response variables. By applying the properties of the Poisson distribution to the Poisson regression, the expected value and the variance of the response variable can be written as follows: [41];

$$E(Y_i/X) = Var(Y_i/X) = \mu_i = n_i e^{x_i^T \beta} \tag{15}$$

The conditions of negative binomial models are similar to those of the Poisson model, except that the conditional means do not equal the conditional variance. This variance is calculated by estimating the dispersion coefficient [46].

3.4. Overdispersion

The Poisson distribution is often suggested for count data but found to be inadequate because the data displays far greater variance than that predicted by the Poisson. This is termed overdispersion or extra-Poisson variation. Overdispersion occurs when $Var(Yi)$ is greater than $E(Yi)$, although $Var(Yi) = E(Yi)$ for the Poisson distribution. Solving the overdispersion problem in the Poisson model is by using an alternative model as the Negative Binomial model where the variance is greater than the mean [45]. The negative binomial distribution provides an alternative model with $Var(Yi) = \varphi\, E(Yi)$, where $\varphi > 1$ is a parameter that can be estimated [42, 43].

3.5. Goodness of fit of Statistical regression models

Information criteria (IC) are used as the method of choosing the best generalized linear models such as Akaike information criteria (AIC) and Bayesian information criteria (BIC) [45].

The two criteria require the sum of two main parts, the natural logarithm of a maximum likelihood and the number of parameters of the model as follows: [45];

$$AIC = -2(\log likelihood) + 2S \tag{16}$$

$$BIC = -2(\log likelihood) + S.\log(N) \tag{17}$$

Where:

N: The sample size studied.

S: Estimated number of parameters in the model.

Loglikelihood: The maximum likelihood function of the model.

The decision to choose the best generalized linear model by using the AIC and BIC depends on getting the lowest value for these two criteria.

4. RESULT

Statistical analysis of the data for 136 households in northern Syria was carried out using the R-Studio software. The descriptive statistical analysis was conducted for the dependent variable and the percentage and frequency of the independent variables were calculated. The data was modeled using Poisson and negative binomial regression models, and a comparison was made between the two models according to the regression model evaluation criteria.

The response and the independent variables used in this study are as follows:

Y: Average monthly household income in US dollars.

$\mathbf{X_1}$: The type of energy used for cooking (X_{10}: firewood, X_{11}: natural gas, X_{12}: kerosene oil).

$\mathbf{X_2}$: The type of energy used for heating (X_{20}: firewood, X_{21}: coal, X_{22}: Diesel, X_{23}:no energy is used, X_{24}: kerosene oil, X_{25}: natural gas).

$\mathbf{X_3}$: The capacity of the batteries used to store energy (X_{30}: not found, X_{31}: small, X_{32}: medium, X_{33}: big).

$\mathbf{X_4}$: The type of energy used to heat water (X_{40}: firewood, X_{41}: coal, X_{42}: Diesel, X_{43}: kerosene oil, X_{44}: natural gas, X_{45}: solar system, X_{46}: Burning household waste).

$\mathbf{X_5}$: Electricity source (X_{50}: No electricity at home, X_{51}: Subscribe amperes, X_{52}: Diesel generator, X_{53}: Gasoline generator, X_{54}: solar energy).

According to the data collected from the target area, the descriptive statistics of the response and independent parameters are summarized in Tables 1, 2, 3, 4, 5 and 6.

Table 1: Descriptive statistics for the average monthly household income in US dollars for 136 households.

Descriptive Statistics		
income		
N	Valid	136
	Missing	0
Mean		128.75
Median		100.00
Std. Deviation		93.178
Range		480
Minimum		20
Maximum		500

Table 2: The type of energy used for cooking: out of 136 households, 88.2% of them use natural gas, 11% use firewood and 0.7% use kerosene oil.

		Frequency	Percent	Valid Percent
Valid	firewood	15	11.0	11.0
	natural gas	120	88.2	88.2
	kerosen	1	0.7	0.7
	Total	136	100.0	100.0

Table 3: The type of energy used for space heating: Out of 136 households, 40.04% use firewood, 6.6% use coal, 33.1% use diesel, 2.9% have no energy source, 16.2% use kerosene oil, and 0.7% use natural gas.

		Frequency	Percent	Valid Percent
Valid	firewood	55	40.4	40.4
	coal	9	6.6	6.6
	Diesel	45	33.1	33.1
	No energy source	4	2.9	2.9
	kerosene oil	22	16.2	16.2
	natural gas	1	0.7	.7
	Total	136	100.0	100.0

Table 4: The capacity of the batteries used to store energy: 66.2% of 136 households do not use batteries, 0.7% use small capacity batteries, 10.3% use medium capacity batteries and 22.8 use big capacity batteries.

		Frequency	Percent	Valid Percent
Valid	no batteries	90	66.2	66.2
	small	1	0.7	0.7
	medium	14	10.3	10.3
	big	31	22.8	22.8
	Total	136	100.0	100.0

Table 5: The type of energy used to heat water: Out of 136 households, 44.9% of them use firewood, 5.9% use coal, 5.1% use kerosene oil, 21.3% use natural gas, 19.1% use solar energy and 3.7% burn household waste.

		Frequency	Percent	Valid Percent
Valid	firewood	61	44.9	44.9
	coal	8	5.9	5.9
	kerosene	7	5.1	5.1
	natural gas	29	21.3	21.3
	solar system	26	19.1	19.1
	Burning household waste	5	3.7	3.7
	Total	136	100.0	100.0

Table 6: Electricity source: 2.9% of households have no electricity, 66.9% depend on ampere subscriptions, 21.3% use small diesel generators, 0.7% use small gasoline generators, and 8.1% use solar energy systems.

		Frequency	Percent	Valid Percent
Valid	No electricity at home	4	2.9	2.9
	Subscribe amperes	91	66.9	66.9
	Diesel generator	29	21.3	21.3
	Gasoline generator	1	0.7	0.7
	solar energy	11	8.1	8.1
	Total	136	100.0	100.0

The data were analyzed according to the Poisson model and the Negative Binomial model. The results are presented in Table 7.

Table 7: Results of Poisson and NB regression models

Type of Model	cod	Poisson regression model			NB regression model		
Explanatory variables		**Estimate**	**Std. Error**	**Pr(Z>|z|)**	**Estimate**	**Std. Error**	**Pr(Z>|z|)**
Intercept	-	4.372	0.056	2e-16***	4.440	0.298	2e-16***
Energy used for cooking	X_{10}	-	-	-	-	-	-
	X_{11}	0.127	0.033	0.000	0.196	0.207	0.342
	X_{12}	-0.829	0.136	1.13e-09***	-0.702	0.621	0.258
Energy used for space heating	X_{20}	-	-	-	-	-	-
	X_{21}	0.273	0.035	1.83e-16***	0.217	0.232	0.350
	X_{22}	0.263	0.020	2e-16***	0.272	0.128	0.033*
	X_{23}	0.143	0.056	0.010*	0.094	0.312	0.763
	X_{24}	0.169	0.052	0.001**	0.114	0.330	0.730
	X_{25}	1.309	0.063	2e-16***	1.323	0.565	0.0192*
Capacity of the batteries	X_{30}	-	-	-	-	-	-
	X_{31}	0.290	0.083	0.000***	0.254	0.595	0.668
	X_{32}	-0.049	0.028	0.078	-0.040	0.175	0.815
	X_{33}	0.224	0.023	2e-16***	0.218	0.152	0.151
Energy used to heat water	X_{40}	-	-	-	-	-	-
	X_{41}	-0.176	0.035	4.36e-07***	-0.239	0.226	0.291
	X_{42}	-0.138	0.039	0.000***	-0.110	0.250	0.658
	X_{43}	-0.500	0.033	2e-16***	-0.483	0.193	0.012*
	X_{44}	-0.382	0.025	2e-16***	-0.396	0.155	0.010*
	X_{45}	-0.265	0.045	4.18e-09***	-0.267	0.294	0.363
	X_{46}	-0.435	0.101	1.97e-05***	-0.443	0.566	0.433
Electricity source	X_{50}	-	-	-	-	-	-
	X_{51}	0.277	0.063	1.13e-05***	0.138	0.345	0.687
	X_{52}	0.655	0.067	2e-16***	0.560	0.388	0.149
	X_{53}	-0.460	0.152	0.002**	-0.528	0.634	0.404
	X_{54}	0.426	0.069	7.42e-10***	0.365	0.398	0.359

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
P=0.05

As a starting point in modelling the count data, the data was modelled using the Poisson model. After estimating the regression parameters, we calculated the dispersion in the Poisson model by dividing the ratio of deviance by the degrees of freedom [49]. The result was (5889.2 / 115) = 51.21, which indicates an overdispersion of our data and that the Poisson regression model is inappropriate while the negative binomial model is more suitable for such a large rate of overdispersion.

From Table 7, there are differences between the coefficients as follows: The coefficients $X_{12}, X_{21}, X_{23}, X_{24}, X_{31}, X_{33}, X_{41}, X_{42}, X_{45}, X_{46}, X_{51}, X_{52}, X_{53}$ *and* X_{54} are highly statistically significant at the α = 0.05 level in the Poisson model. While in the Negative Binomial model the coefficients $X_{12}, X_{21}, X_{23}, X_{24}, X_{31}, X_{33}, X_{41}, X_{42}, X_{45}, X_{46}, X_{51}, X_{52}, X_{53}$ *and* X_{54} aren't statistically significant at the α = 0.05 level. The coefficients X_{11} *and* X_{32} aren't statistically significant at the α = 0.05 level in both regression models where the coefficients $X_{10}, X_{20}, X_{30}, X_{40}$, *and* X_{50} are reference variables.

The reason for the significant difference between the results of the two models is due to the over-dispersion in the data by 51.21, which causes some variables to appear statistically significant when in fact they are completely opposite.

Goodness-of-fit test according to the evaluation criteria (AIC, BIC, and log-likelihood) of the two regression models is shown in Table 8, where it is known that the smaller the evaluation criteria, the better the model. Accordingly, the negative binomial model fitted the data better than the Poisson model. It can be explained by taking into account the "overdispersion" problem when the variance of the response variable is greater than its mean. On the other hand, Poisson regression model is based on the "equidispersion" assumption, meaning that the variance of the response variable is equal to its mean.

Table 8: The evaluation criteria of Poisson and NB regression models

Type of model	AIC	BIC	Log -likelihood value
Poisson model	6815.56	6876.72	-3386.78
NB model	1549.03	1613.10	-752.51

To test the ability of the negative binomial model in fitting the data as shown in Fig. 3, the relationship between residuals and fitted values is a linear relationship because the sediment is spread randomly. This is a good indication of the commitment of the negative binomial model to the linearity condition, that is, there is no non-linearity relationship between the independent and dependent variables that the model could not explain.

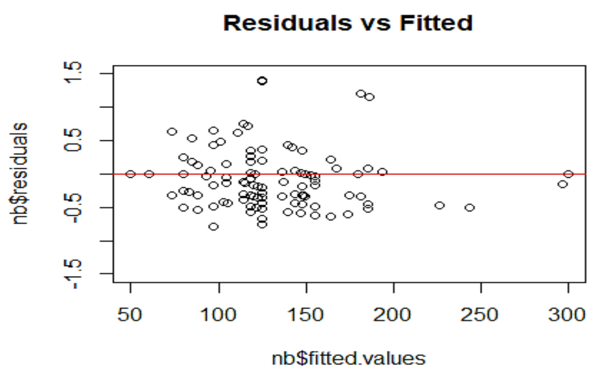

Figure 3. The relationship between Fitted values and residuals for NB regression model.

Before interpreting the coefficients in terms of incidence rate ratios (IRR), it is necessary to clarify how to move from interpreting the regression coefficients and the incidence rate ratios.

The outputs in Table (9) refer that the incidence rate ratio for cooking = 1 is 1.21 times compared with the incident rate for the reference group (cooking=0). Similarly, the incident rate ratio for cooking=2 is 0.49 times comparing the incident rate for the reference group holding the other variables constant. In a similar way, the results can be interpreted for the rest of the variables.

Table 9: Incidence rate ratios (IRR) for coefficients of the NB model.

Explanatory variables	cod	Estimate	2.5%	97.5%
Intercept	-	84.83	50.38	155.08
Energy used for cooking	X_{10}	-	-	-
	X_{11}	1.21	0.78	1.84
	X_{12}	0.49	0.16	2.01
Energy used for space heating	X_{20}	-	-	-
	X_{21}	1.24	0.79	2.01
	X_{22}	1.31	1.02	1.68
	X_{23}	1.09	0.30	2.13
	X_{24}	1.12	0.56	2.17
	X_{25}	3.75	1.42	14.06
Capacity of the batteries	X_{30}	-	-	-
	X_{31}	1.29	0.44	5.04
	X_{32}	0.96	0.68	1.36
	X_{33}	1.24	0.92	1.69
Energy used for water heating	X_{40}	-	-	-
	X_{41}	0.78	0.50	1.27
	X_{42}	0.89	0.54	1.53
	X_{43}	0.61	0.042	0.91
	X_{44}	0.67	0.50	0.90
	X_{45}	0.76	0.44	1.42
	X_{46}	0.64	0.24	2.39
Electricity source	X_{50}	-	-	-
	X_{51}	1.14	0.56	2.21
	X_{52}	1.75	0.81	3.63
	X_{53}	0.58	0.18	2.40
	X_{54}	1.44	0.64	3.13

By using the statistically significant parameter estimates of the NB regression model given in Table (7), and using the IRLS parameter estimates of the NB regression model belonging "an average monthly household income in US dollars" data, the expected value (mean) of the NB regression model is given as follows;

$$log(\hat{\lambda}) = 4.40 + 0.272X_{22} + 1.323X_{25} - 0.483X_{43} - 0.396X_{44} \quad (13)$$

Where $\hat{\lambda}$ indicates the estimation of average monthly household income in US dollars.

The equation can be written as:

$$\hat{\lambda} = e^{4.40 + 0.272X_{22} + 1.323X_{25} - 0.483X_{43} - 0.396X_{44}} \quad (14)$$

From the obtained model, the result can be summarized as follows:

- The fuel used for space heating and water heating has a direct impact on households' monthly income.
- The expected value of the monthly expenditure as a result of using diesel for heating will be 1.32% higher than the expected value of the monthly expenditure when using coal.
- The expected value of the monthly expenditure as a result of using natural gas for heating will be 3.754 % higher than the expected value of monthly expenditure when using coal.
- The expected value of the monthly expenditure from using natural gas to heat water will be 1.620 % less than the expected value of the monthly expenditure when using coal

- The expected value of the monthly expenditure from using kerosene to heat water will be 1.485 % less than the expected value of the monthly expenditure when using coal.

Based on these results, coal is the least expensive energy source for space heating compared to other energy sources, meanwhile, natural gas and kerosene are the least expensive energy sources for heating water.

5. CONCLUSION

Efficient use of energy sources is understood differently depending on the context and the country of reference. It may be challenging to provide a universal definition for energy efficiency but based on the results of the various studies conducted it is evident that there are still many people worldwide who use energy sources inefficiently. The environment targeted in this study is a crisis-affected region where fuel and energy are subject to restrictions and high prices. Thus, relieving the energy burden of low-income families requires a joint effort to ensure tangible benefits for these families. However, there is no published work on optimizing energy use or identifying cost-effective energy sources at local levels in Syria.

In this study, a mathematical model identifying the most cost-efficient energy sources for households in northern Syria is proposed. The proposed model considers average household monthly income, the main available energy sources, and prices. Poisson regression and Negative Binomial (NB) regression were used to apply the statistical analysis of the data and produce the required model. According to the evaluation criteria, evidence indicates that the NB model has a better fit than the Poisson regression model for a large rate of overdispersion. According to statistically significant parameters of the NB regression model given, only the fuel used for space heating and water heating has a direct impact on the monthly income. The model showed that the most cost-effective fuel for space heating is coal, and the most cost-effective fuel for water heating is natural gas and kerosene. The model can be used as a rapid indicative approach to reducing energy expenditure as It has functional flexibility integrating cost reduction, mitigation of energy problems and helping families to obtain energy services at affordable prices commensurate with their monthly income.

A future study will focus on whether energy issues can be addressed through investment in solar energy systems. In light of this, it will be concluded whether solar energy systems can protect families, especially those with low incomes, from energy poverty and price fluctuations in the energy market. Moreover, increasing citizens' investment in renewable energy can contribute to the desired transition to a low-carbon energy system, climate protection, and advancing sustainable development by encouraging individuals and policymakers to solve the energy problem, relying on sustainable and environmentally friendly energy sources.

Declaration of Ethical Standards

The author declares that the study complies with all applicable laws and regulations and meets ethical standards.

Credit Authorship Contribution Statement

The author contributed to the data collection and modelling, the results analysis and the manuscript writing.

Declaration of Competing Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding / Acknowledgements

The author declares that there is no financial support.

Data Availability

Data will be available upon reasonable request.

REFERENCES

[1] S. Ri, A. H. Blair, C. J. Kim, and R. J. Haar, "Attacks on healthcare facilities as an indicator of violence against civilians in Syria: an exploratory analysis of open-source data," *PLoS One,* vol. 14, no. 6, p. e0217905, 2019.
[2] L. Tichý, "Energy infrastructure as a target of terrorist attacks from the Islamic state in Iraq and Syria," *International Journal of Critical Infrastructure Protection,* vol. 25, pp. 1-13, 2019.
[3] S. Jabbour *et al.*, "10 years of the Syrian conflict: a time to act and not merely to remember," *The Lancet,* vol. 397, no. 10281, pp. 1245-1248, 2021.
[4] F. A. Omar, I. Mahmoud, A. Hussian, L. Mohr, H. O. Abdullah, and A. Farzat, "The effect of the Syrian crisis on electricity supply and the household life in North-West Syria: a university-based study," *Education and Conflict Review,* vol. 3, pp. 77-86, 2020.
[5] F. Alhaj Omar, I. Mahmoud, and K. G. Cedano, "Energy poverty in the face of armed conflict: The challenge of appropriate assessment in wartime Syria," *Energy Research & Social Science,* vol. 95, p. 102910, 2023/01/01/ 2023, doi: https://doi.org/10.1016/j.erss.2022.102910.
[6] J. Messner, "Fragile States Index 2017: Factionalization and Group Grievance Fuel Rise in Instability. The Fund for Peace," ed, 2017.
[7] C. International Committee of the Red, *Urban services during protracted armed conflict: a call for a better approach to assisting affected people*. International Committee of the Red Cross, 2015.
[8] S. Gates, H. Hegre, H. M. Nygard, and H. Strand, "Consequences of armed conflict in the Middle East and north Africa region," mimeo, 2010.
[9] I. P. Conflict and H. Action, "some recent ICRC experiences," *Geneva: ICRC,* 2016.
[10] C. Bennett, M. Foley, and S. Pantuliano, "Time to let go: Remaking humanitarian action for the modern era," *ODI, London,* 2016.
[11] P. Collier, "Fragile States and International Support," *Development,* vol. 175, 2016.
[12] E. Spyrou, B. F. Hobbs, M. D. Bazilian, and D. Chattopadhyay, "Planning power systems in fragile and conflict-affected states," *Nature energy,* vol. 4, no. 4, pp. 300-310, 2019.
[13] P. B. Anand, *Getting infrastructure priorities right in post-conflict reconstruction*. WIDER Research Paper, 2005.
[14] J. Goodhand, "Violent conflict, poverty and chronic poverty," *Chronic Poverty Research Centre Working Paper,* no. 6, 2001.
[15] K. Lyra, S. Mirasgedis, and C. Tourkolias, "From measuring fuel poverty to identification of fuel poor households: a case study in Greece," *Energy Efficiency,* vol. 15, no. 1, p. 6, 2022/01/07 2022, doi: 10.1007/s12053-021-10017-6.

[16] A. Acakpovi, G. Botwe-Ohenewaa, and D. M. Sackey, "Impact of energy efficiency and conservation programs on the national grid in some selected households in Ghana," *Energy Efficiency,* vol. 15, no. 1, p. 5, 2022/01/06 2022, doi: 10.1007/s12053-021-09998-1.
[17] S. Oguah and D. Chattopadhyay, "Planning in fragile and conflict states:: Case study for West Bank and Gaza," in *2018 IEEE Power & Energy Society General Meeting (PESGM)*, 2018: IEEE, pp. 1-5.
[18] E. L. Roach and M. Al-Saidi, "Rethinking infrastructure rehabilitation: Conflict resilience of urban water and energy supply in the Middle East and South Sudan," *Energy Research & Social Science,* vol. 76, p. 102052, 2021.
[19] M. Labordena, A. Patt, M. Bazilian, M. Howells, and J. Lilliestam, "Impact of political and economic barriers for concentrating solar power in Sub-Saharan Africa," *Energy Policy,* vol. 102, pp. 52-72, 2017.
[20] T. Pettersson and P. Wallensteen, "Armed conflicts, 1946–2014," *Journal of peace research,* vol. 52, no. 4, pp. 536-550, 2015.
[21] H. Zerriffi, H. Dowlatabadi, and N. Strachan, "Electricity and conflict: Advantages of a distributed system," *The Electricity Journal,* vol. 15, no. 1, pp. 55-65, 2002.
[22] M. Bazilian and D. Chattopadhyay, "Considering power system planning in fragile and conflict states," *Energy for Sustainable Development,* vol. 32, pp. 110-120, 2016.
[23] N. Patankar, A. R. de Queiroz, J. F. DeCarolis, M. D. Bazilian, and D. Chattopadhyay, "Building conflict uncertainty into electricity planning: A South Sudan case study," *Energy for Sustainable Development,* vol. 49, pp. 53-64, 2019.
[24] J. L. Sowers, E. Weinthal, and N. Zawahri, "Targeting environmental infrastructures, international law, and civilians in the new Middle Eastern wars," *Security Dialogue,* vol. 48, no. 5, pp. 410-430, 2017.
[25] P. Justino, "Violent conflict and human capital accumulation," *IDS Working Papers,* vol. 2011, no. 379, pp. 1-17, 2011.
[26] H. Sun, S. Lu, and S. Solaymani, "Impacts of oil price uncertainty on energy efficiency, economy, and environment of Malaysia: stochastic approach and CGE model," *Energy Efficiency,* vol. 14, no. 2, p. 21, 2021/02/04 2021, doi: 10.1007/s12053-020-09924-x.
[27] A. O'Sullivan, M.-E. Rey, and J. G. Mendez, "Opportunities and Challenges in the MENA Region," *Arab world competitiveness report,* vol. 2012, pp. 42-67, 2011.
[28] N. R. C. Idmc, "Global Report on Internal Displacement 2019," ed, 2019.
[29] T. A. A. People, "URBAN SERVICES DURING PROTRACTED ARMED CONFLICT," 2015.
[30] C. International Committee of the Red, "Bled Dry: How War in the Middle East is Bringing the Region's Water Supplies to Breaking Point," 2015: ICRC Geneva.
[31] N. R. Draper and H. Smith, *Applied regression analysis*. John Wiley & Sons, 1998.
[32] A. O. Sykes, "An introduction to regression analysis," 1993.
[33] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*. John Wiley & Sons, 2021.
[34] A. M. Legendre, *Nouvelles méthodes pour la détermination des orbites des comètes; par AM Legendre*. chez Firmin Didot, libraire pour lew mathematiques, la marine, l …, 1806.
[35] C. F. Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae*. H. Dieterich, 1823.
[36] F. Galton, "Regression towards mediocrity in hereditary stature," *The Journal of the Anthropological Institute of Great Britain and Ireland,* vol. 15, pp. 246-263, 1886.
[37] S. Weisberg, *Applied linear regression*. John Wiley & Sons, 2005.
[38] U. Olsson, "Generalized linear models," *An applied approach. Studentlitteratur, Lund,* vol. 18, 2002.
[39] J. Fox, *Applied regression analysis and generalized linear models*. Sage Publications, 2015.
[40] Ö. Deniz, "Poisson regresyon analizi," *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi,* vol. 4, no. 7,

pp. 59-72, 2005.
[41] R. Winkelmann, *Econometric analysis of count data*. Springer Science & Business Media, 2008.
[42] P. De Jong and G. Z. Heller, "Generalized linear models for insurance data," *Cambridge Books,* 2008.
[43] A. J. Dobson and A. G. Barnett, *An introduction to generalized linear models*. CRC press, 2018.
[44] A. Agresti, *Analysis of ordinal categorical data*. John Wiley & Sons, 2010.
[45] J. M. Hilbe, *Negative binomial regression*. Cambridge University Press, 2011.
[46] A. C. Cameron and P. K. Trivedi, *Regression analysis of count data*. Cambridge university press, 2013.
[47] C. Walck, "Hand-book on statistical distributions for experimentalists," *University of Stockholm,* vol. 10, pp. 96-01, 2007.
[48] K. J. Hastings, *Probability and statistics*. Addison-Wesley Reading, MA, 1997.
[49] E. H. Payne, J. W. Hardin, L. E. Egede, V. Ramakrishnan, A. Selassie, and M. Gebregziabher, "Approaches for dealing with various sources of overdispersion in modeling count data: Scale adjustment versus modeling," *Statistical methods in medical research,* vol. 26, no. 4, pp. 1802-1823, 2017.