Research Article

Open Access

Neslihan Iyit*

Modelling world energy security data from multinomial distribution by generalized linear model under different cumulative link functions

<https://doi.org/10.1515/chem-2018-0053>

received January 24, 2018; accepted March 19, 2018.

Abstract: *Energy security* is one of the major components of energy sustainability in the world's energy performance. In this study, *energy security* is taken as an ordinal response variable coming from the multinomial distribution with the energy grade levels *A*, *B*, *C*, and *D*. Thereafter, the world *energy security* data is tried to be statistically modelled by using *generalized linear model (GLM)* approach for the ordinal response variable under different cumulative link functions. The cumulative link functions comparatively used in this study are cumulative logit, cumulative probit, cumulative complementary log-log, cumulative Cauchit, and cumulative negative log-log. In order to avoid a multicollinearity problem in the data structure, principal component analysis (*PCA*) technique is integrated with the *GLM* approach for the ordinal response variable. In this study, statistically, the importance of determining the best cumulative link function on the accuracy of parameter estimates, confidence intervals, and hypothesis tests in the *GLM* for the multinomially distributed response variable is highlighted. In terms of energy evaluation, by using *cumulative logit* as the best cumulative link function, *energy sources consumptions*, *electricity productions from nuclear energy*, *natural gas*, *oil*, *coal*, and *hydroelectric*, *energy use per capita* and *energy imports* are found to have statistically significant effects on *energy security* in the world's energy performance.

Keywords: Energy security; multinomial distribution; principal component analysis; generalized linear model; cumulative link function.

1 Introduction

In the World Energy Trilemma Index 2016 report, the World Energy Council's definition of energy sustainability is based on three core dimensions as energy security, energy equity, and environmental sustainability (https://www.worldenergy.org/wp-content/uploads/2016/10/Full-report_Energy-Trilemma-Index-2016.pdf). In this study, the focus is on *energy security*. The World Energy Council defines "energy security" as the "effective management of primary energy supply from domestic and external sources, reliability of energy infrastructure, and ability of energy providers to meet current and future demand" in the World Energy Trilemma Index, 2016. The World Energy Council used indicators given in Table 1 to calculate energy security's grade levels.

After the data for each indicator given in Table 1 was collected and verified, scores have been calculated for each indicator, the indicator-level results were standardized by using the z-score and then rescaled to the range between 0–100. The balance score grade for energy security was assigned from *A* to *D* based on the mean and standard deviation of energy security in the World Energy Trilemma Index, 2016.

In this study, following The World Energy Council's above-stated classification, *energy security* is taken as the ordinal response variable coming from the multinomial distribution with four ordinal categories as *A*, *B*, *C*, and *D* grade levels.

In order to investigate statistically significant effects on the world *energy security* by using *generalized linear model (GLM)* approach, 15 countries from Asia, 26 countries from Europe, 8 countries from Latin America and Caribbean, 8 countries from Middle East and North Africa, 3 countries from North America, and 1 country from Sub-Saharan Africa regions are taken as subjects into the study.

Some studies on *energy security* in the literature can be given as follows;

Jacobson [1] reviewed and ranked major proposed energy-related solutions to global warming, air pollution

*Corresponding author: Neslihan Iyit, Selcuk University, Faculty of Sciences, Statistics Department, Alaeddin Keykubat Campus, Postal code:42130, Konya, Turkey, E-mail: niyit@selcuk.edu.tr

Table 1: Indicators used to calculate energy security's grade levels in the World Energy Trilemma Index, 2016.

Indicator category	Indicator
Security of supply and energy delivery	Diversity of primary energy supply
	Energy consumption in relation to GDP growth
	Import dependence
Resilience	Diversity of electric generation
	Energy storage
	Preparedness(human factor)

mortality, and energy security. He also estimated CO₂ emissions due to the leakage for different residence times of carbon dioxide stored in a geological formation by using exponential equations over time. Krut et al. [2] provided an overview of available indicators for the long-term energy security of supply by using computation methods as dual concept diversity indices, mean variance portfolio theory, and Shannon index. Winzer [3] reviewed the multitude of definitions of energy security characterized according to some basic concepts. He calculated the system average interruption duration index of electricity supplies, and heat supply. He also estimated the total gross domestic product loss of Austria, Italy, and U.K. due to electricity interruptions by the additive mathematical model with multiplicative terms. Chester [4] examined energy security and its complex nature in full detail without giving any mathematical model. Sovacool and Mukherjee [5] provided a synthesized and workable framework for analyzing national energy security policies and performance in the aspect of energy security indicators and metrics. They also investigated demographic details of the energy security survey by using pie charts and bar graphs. Bahgat [6] reviewed the concept of energy security by examining Europe's energy mix and then analyzed European efforts to establish and strengthen energy partnerships in all aspects without giving any mathematical model. Löschel et al. [7] suggested an additional dimension along which indicators of energy security were classified by calculating energy security market concentration measure, ex-post and ex-ante indicators of energy security. Skodras et al. [8] suggested that production of coal-based synthetic natural gas might be a good alternative to carbon intensity in the energy generation and energy security by using nonlinear regression analysis. Yergin [9] investigated the concept of energy security very broadly in terms of the return of energy security, the dimensions of the energy security, the limits

of energy independence and the strategic significance of the energy security. Awerbuch [10] investigated the effect of mitigating fossil price volatility on the energy security by using portfolio equation and calculating portfolio risks for the EU, the US and Mexico. Sovacool and Brown [11] analyzed the energy security performances of the US and 21 OECD countries based on availability, affordability, efficiency and environmental stewardship between 1970 and 2007. Cherp and Jewell [12] examined energy security from three different perspectives as sovereignty, robustness and resilience. Belkin [13] examined energy security challenges of the EU member states and the efforts of EU to determine the energy strategy policy. Ang et al. [14] surveyed 104 studies in the literature about 83 different energy security definitions with energy security indicators and indexes. Further literature studies on energy can be found in [15-20]. During the detailed examination of energy security literature, it can be stated that a contribution of the findings of *GLM* on energy security with respect to the different energy components as investigated in this study have not been made before. Therefore this study on energy security is expected to make significant contribution to the literature.

Based on the importance of *energy security* in the energy sustainability of the world's energy performance, the main motivation of this study is to bring a new statistical evaluation of *energy security* in terms of energy imports, energy use per capita, energy sources consumptions, electricity productions from energy sources, and some energy reserves by *GLM* for ordinal response variable under different cumulative link functions. From this point of view, the statistical assessment of energy security presented in this study has not been included in any of the previous studies in the literature.

The statistical significance of this study is to highlight the importance of the most appropriate cumulative link function selection between the systematic component and the cumulative probabilities of the ordinal response variable in *GLM* by using major information criteria as *AIC*, *AICC*, *BIC*, and *CAIC*. And then, to emphasize the importance of the best cumulative link function selection on the accuracy of parameter estimates, confidence intervals, and hypothesis tests in the systematic component of *GLM* for the multinomially distributed response variable. The necessity to model the energy security data by *GLM* is the flexibility of modelling the ordinal response variable coming from the exponential family including multinomial distribution as given in this study.

2 Generalized linear model for ordinal response variable coming from multinomial distribution

In this study, attention is paid on regression models for the analysis of ordinal response variable with more than two response categories, especially which are ordered. Multinomial distribution is an extension of the binomial distribution to more than two response categories.

Consider a response variable $Y_i; i = 1, 2, \dots, N$ taking one of several discrete values. Let $P(Y_i = j) = \pi_{ij}; i = 1, 2, \dots, N, j = 1, 2, \dots, c$ denotes the probability that the response belonging to the i^{th} subject falls in the j^{th} ordinal response category. In this study, the response variable is “energy security” and it takes the ordered values A, B, C , and D as categories indexed by 1, 2, 3, and 4 for 61 different countries from 6 regions in the world.

Let $P(Y_i = j) = \pi_{ij}; i = 1, 2, \dots, 61, j = 1, 2, 3, 4$ denotes the probability that the i^{th} country's response ($i = 1, 2, \dots, 61$) falls in the j^{th} energy security grade level ($j = 1, 2, 3, 4$). In this study, π_{i1} is the probability that the i^{th} response is in A energy security grade level and so on. Assume that the response categories are mutually exclusive and exhaustive, $\sum_{j=1}^c \pi_{ij} = 1$. Additionally there are N independent trials for 61 different countries and each trial results in 1 of 4 mutually exclusive and exhaustive outcomes as energy security grade levels. Then the “energy security” ordinal response variable comes from the multinomial probability distribution. In *generalized linear model (GLM)* approach for the energy security ordinal response variable, the main interest is in the cumulative probability of the i^{th} response falling into or below the j^{th} energy security grade level as follows;

$$P(Y_i \leq j) = \pi_{i1} + \dots + \pi_{ij}, \quad j = 1, 2, 3, 4 \quad (1)$$

GLM for the multinomial distribution has three components; the random component, the systematic component, and the cumulative link function between the random and systematic components [15, 21, 22].

The random component identifies the ordinal response variable as energy security and assumes that the response variable comes from the exponential family including the multinomial distribution. The systematic component specifies the covariates $\mathbf{x}_i' = (x_{i1}, \dots, x_{ip})$ belonging to the i^{th} country in the *GLM*. Parameters in the systematic component of the *GLM* for the multinomial distribution are estimated by using maximum likelihood (*ML*) method with one of the accompanying iterative methods such

as Newton-Raphson (*NR*), Fisher scoring (*FS*) or hybrid method [21-24].

Let G^{-1} denotes a cumulative link function given in Table 2 as the inverse of the continuous cumulative distribution function G . Then the general form of a cumulative link model for the i^{th} response relates the cumulative probabilities to the linear predictor depended on the covariates as follows [22];

$$G^{-1}[P(Y_i \leq j)] = \alpha_j + \mathbf{x}_i' \boldsymbol{\beta}; \quad j = 1, 2, \dots, c-1 \quad (2)$$

Let the value of $Y_i = j$ be denoted by y_{ij} with the corresponding probability value π_{ij} . Then the multinomial likelihood function for the cumulative link model given in Eq.(2) is as follows [22];

$$\prod_{i=1}^{N=61} \left(\prod_{j=1}^{c=4} \pi_{ij}^{y_{ij}} \right) = \prod_{i=1}^{N=61} \left\{ \prod_{j=1}^{c=4} [P(Y_i \leq j) - P(Y_i \leq j-1)]^{y_{ij}} \right\} \quad (3)$$

In Eq.(3), $\alpha_j; j = 1, 2, \dots, c-1$ are the intercept parameters and $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$ are the parameters belonging to the covariates in the systematic component of the *GLM* for the multinomial distribution. Then the multinomial log-likelihood function for the cumulative link model given in Eq.(2) is as follows [22];

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{N=61} \sum_{j=1}^{c=4} y_{ij} \log \left[G(\alpha_j + \mathbf{x}_i' \boldsymbol{\beta}) - G(\alpha_{j-1} + \mathbf{x}_i' \boldsymbol{\beta}) \right] \quad (4)$$

Let g be the probability density function belonging to the derivative of the cumulative distribution function G . δ_{jk} denotes the Kronecker delta as $\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = 0$ otherwise. Then the likelihood equations belonging to the $\alpha_j; j = 1, 2, \dots, c-1$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$ parameters in the *GLM* for the multinomial distribution are as follows [22];

$$\frac{\partial L}{\partial \beta_k} = \sum_{i=1}^{N=61} \sum_{j=1}^{c=4} y_{ij} x_{ik} \frac{g(\alpha_j + \mathbf{x}_i' \boldsymbol{\beta}) - g(\alpha_{j-1} + \mathbf{x}_i' \boldsymbol{\beta})}{G(\alpha_j + \mathbf{x}_i' \boldsymbol{\beta}) - G(\alpha_{j-1} + \mathbf{x}_i' \boldsymbol{\beta})} = 0 \quad (5)$$

$$\frac{\partial L}{\partial \alpha_k} = \sum_{i=1}^{N=61} \sum_{j=1}^{c=4} y_{ij} \frac{\delta_{jk} g(\alpha_j + \mathbf{x}_i' \boldsymbol{\beta}) - \delta_{j-1,k} g(\alpha_{j-1} + \mathbf{x}_i' \boldsymbol{\beta})}{G(\alpha_j + \mathbf{x}_i' \boldsymbol{\beta}) - G(\alpha_{j-1} + \mathbf{x}_i' \boldsymbol{\beta})} = 0 \quad (6)$$

The likelihood equations given in Eq.(5) and Eq.(6) can be solved by using *NR*, *FS* or hybrid method. In this study, hybrid method, in which iterations with the *FS* method are performed before continuing iterations with the *NR* method, is used. If convergence is achieved before the

maximum number of Fisher iterations is reached, the hybrid algorithm continues with the NR method [23].

The scale parameter related to the variance of the response variable is estimated by using *Pearson chi-square* statistic as follows [22,23];

$$\chi^2 = \sum_{i=1}^{N=61} \sum_{j=1}^{c=4} \frac{(y_{ij} - \hat{\pi}_{ij})^2}{\hat{\pi}_{ij}} \quad (7)$$

The estimate of the scale parameter is computed as the ratio of the model Pearson chi-square statistic to the model degrees of freedom defined as $N(c-1)-p$ in the *GLM* for the multinomial distribution [22,25,26].

In this study, the cumulative link function $G^{-1}[P(Y_i \leq j)]$ describes the functional relationship between the systematic component of the *GLM* and the cumulative probabilities of the i^{th} response falling into or below the j^{th} energy security grade level.

Cumulative link functions given in Table 2 are used to permit the cumulative probabilities of the ordinal response variable to be linearly related to the covariates as in Eq.(2) (<http://share.uoa.gr/public/Software/SPSS/SPSS22/Manuals/IBM%20SPSS%20Advanced%20Statistics.pdf>).

Table 2: Cumulative link functions belonging to the ordinal response variable in the *GLM* from the multinomial distribution.

Cumulative link functions	Formula $G^{-1}[P(Y_i \leq j)]$
Cumulative logit	$\ln[P(Y_i \leq j)/(1 - P(Y_i \leq j))]$
Cumulative probit	$\Phi^{-1}[P(Y_i \leq j)]$
Cumulative complementary log-log	$\ln[-\ln(1 - P(Y_i \leq j))]$
Cumulative Cauchit	$\tan[\pi(P(Y_i \leq j) - 0.5)]$
Cumulative negative log-log	$-\ln[-\ln(P(Y_i \leq j))]$

where Φ^{-1} is the inverse of the standard normal cumulative distribution function.

Information criteria (IC) are used as goodness-of-fit test statistics for determining the best cumulative link function between the systematic component and the cumulative probabilities in the *GLM* for the ordinal response variable are given in Table 3.

(https://www.ibm.com/support/knowledgecenter/SSLVMB_22.0.0/com.ibm.spss.statistics.algorithms/alg_glnlin_gzlm_modeltest_goof.htm)

Table 3: Information criteria (IC) used as goodness-of-fit-test statistics for comparing different cumulative link functions in the *GLM* for the ordinal response variable.

Goodness-of-fit-test statistics	Formula
Akaike Information Criteria (AIC)	$-2l + 2d$
Finite Sample Corrected Akaike Information Criteria (AICC)	$-2l + \frac{2dN}{N - d - 1}$
Bayesian Information Criteria (BIC)	$-2l + d \ln(N)$
Consistent Akaike Information Criteria (CAIC)	$-2l + d(\ln(N) + 1)$

where l is the maximum value of the multinomial log-likelihood function given in Eq.(4) evaluated at the parameter estimates, $d = c - 1 + p$ is the number of parameters in the model, N is the total number of subjects. The smallest values of IC determine the best cumulative link function in the *GLM* for the ordinal response variable.

3 Generalized linear model for the world energy security data from multinomial distribution

In this study, “energy security” is taken as the ordinal response variable. Ordinal response variable categories are taken as A, B, C, and D energy grade levels as indicated in the World Energy Trilemma Index, 2016. The response variable probability distribution is taken as *multinomial*. Cumulative link functions associated with the ordinal response variable are taken as *cumulative logit*, *cumulative probit*, *cumulative complementary log-log*, *cumulative Cauchit* and *cumulative negative log-log*. The parameter estimation method for the systematic component or the mean structure of the *GLM* for multinomial distribution is taken as the *ML* method with the *hybrid* iterative algorithm. Also, the scale parameter is estimated by using the *Pearson chi-square* statistic.

Energy imports (% of net energy use), energy use per capita (kg of oil equivalent per capita), oil (million tonnes), natural gas, coal, nuclear energy, hydroelectricity, solar, wind, geothermal, biomass and other (GBO) renewable energy sources consumptions (million tonnes oil equivalent), electricity productions (% of total) from coal, hydroelectric, natural gas, nuclear energy, and oil energy sources, total proved oil (thousand million tonnes), and natural gas (trillion cubic metres) reserves are taken

Table 4: Rotated component matrix between the covariates and the components in *PCA* for the world energy security data.

Covariates/Components	1	2	3	4	5	6
Wind energy consumption	.943					
Oil consumption	.937					
GBO renewable energy sources consumption	.907					
Solar energy consumption	.852					
Natural gas consumption	.804					
Nuclear energy consumption	.764					
Coal consumption	.645					
Hydroelectricity consumption	.643					
Natural gas reserve		.804				
Total proved oil reserve		.700				
Electricity production from nuclear energy			.875			
Electricity production from natural gas			-.550			
Electricity production from oil			-.523			
Energy use per capita				-.826		
Energy imports				.766		
Electricity production from coal					.753	
Electricity production from hydroelectric						.938

as covariates for modelling the world energy security by *GLM* under different cumulative link functions. All computations and statistical data analysis are performed by using Microsoft Excel 2010, and IBM SPSS 22.0 (IBM Corp, IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY, 2013).

Before starting the data analysis, covariates are centered by using their arithmetic means called “mean centering” method to get more stabilized, and consistent parameter estimates of the *GLM* and to reduce multicollinearity in the data structure.

Before constructing *GLM* for the world energy security data, pairwise correlations between these covariates are examined to measure the strength of the linear relationships between them. By using Pearson correlation coefficient, highly correlated pairs of the covariates are only determined for the energy sources consumptions between oil-natural gas 0.865, oil-wind 0.905, oil-GBO 0.834, natural gas-nuclear energy 0.824, coal-hydroelectricity 0.891, solar-wind 0.819, solar-GBO 0.816, and wind-GBO 0.854 with statistically significant *p*-values 0.000 at $\alpha = 0.05$ significance level.

When the covariates are moderately or highly correlated, parameter estimates, hypothesis tests, and confidence intervals may yield inaccurate statistical

inferences and conclusions in the *GLM* for the world energy security data. In order to overcome this data-based multicollinearity problem, principal component analysis (*PCA*) technique [27] is used in this study. Ridge regression [28] and lasso regression [29], based on constrained minimization problem with penalty term and Bayesian inference, are alternative but also complex approaches to *PCA* in the case of multicollinearity.

As the measure of sampling adequacy, Kaiser-Meyer-Olkin (KMO) value is found to be 0.640. Especially for the values of $KMO > 0.5$, it can be concluded that the sample is adequate, and *PCA* is moderately suitable for the world energy security data. Approximate chi-square test statistic value for Bartlett’s test of sphericity is found to be 845.979 with statistically significant *p*-value 0.000 at $\alpha = 0.05$ significance level. This indicates that the original correlation matrix is not an identity matrix and therefore the covariates are related and *PCA* is suitable for detecting this correlated covariate structure.

PCA which is also known as a popular dimension reduction technique is performed to form uncorrelated linear combinations of the covariates. In the *PCA* technique, “varimax” is used as the orthogonal rotation method that minimizes the number of covariates that have high loadings on each component. Six different principal

Table 5: Results for goodness-of-fit test statistics comparing different cumulative link functions in the *GLM* for the world energy security data.

Cumulative link function	Deviance	Pearson chi-square	-2Log likelihood	AIC	AICC	BIC	CAIC
Cumulative logit*	104.982*	172.995*	-55.779*	122.982*	126.511*	141.980*	150.980*
Cumulative Cauchit	105.007	188.630	-53.285	123.007	126.536	142.005	151.005
Cumulative negative log-log	105.171	204.820	-52.585	123.171	126.700	142.169	151.169
Cumulative probit	106.570	207.409	-52.503	124.570	128.099	143.568	152.568
Cumulative complementary log-log	111.558	217.055	-52.491	129.558	133.088	148.556	157.556

components out of 17 original covariates are constituted based on eigenvalue criterion higher than 1, and the scores for each principal component are created as the values of the new covariates.

As a key output of *PCA*, rotated component matrix containing estimates of the correlations between each of the covariates and the estimated components is given in Table 4. As seen from Table 4, six different components consist of highly correlated energy sources consumptions; non-renewable energy (fossil fuels) reserves of natural gas and oil; electricity productions from nuclear energy, natural gas, and oil; energy use per capita and energy imports; electricity production from coal; and electricity production from hydroelectric, respectively.

Ethical approval: The conducted research is not related to either human or animals use.

4 Results and discussion

In this section, results for modelling the world energy security data by *GLM* under different cumulative link functions are given comparatively.

Firstly, *GLM* for the world energy security data is constituted with 5 different cumulative link functions in terms of 6 different principal components as the new covariates. The values of goodness-of-fit test statistics comparing these cumulative link functions in the *GLM* for the world energy security data are given in Table 5. The best cumulative link function between the cumulative probabilities of the ordinal response variable and the new candidate components to the mean structure of the *GLM* is determined as “cumulative logit” with the smallest values of *IC* as *AIC*, *AICC*, *BIC*, and *CAIC*, goodness-of-fit measures as the deviance, Pearson chi-square and the multinomial negative log-likelihood as seen from Table

5. Additionally, the worst cumulative link function is determined as “cumulative complementary log-log” with the largest values of the same *IC* and the goodness-of-fit measures from Table 5.

The estimates of the scale parameter are computed based on the value of the Pearson chi-square test statistic as 0.994, 1.084, 1.177, 1.192, and 1.247 for cumulative logit, cumulative Cauchit, cumulative negative log-log, cumulative probit, and cumulative complementary log-log link functions in the *GLM* for the world energy security data, respectively.

The results belonging to the parameter estimates, hypothesis tests of the parameters, odds ratios estimates and confidence intervals for the odds ratios in the *GLM* for the world energy security data under the best cumulative link function are given in Table 6.

As seen from Table 6, there are two statistically significant intercept parameter estimates, with $\hat{\alpha}_2 = 1.683$ for *B* energy security grade level, and $\hat{\alpha}_3 = 4.962$ for *C* energy security grade level with *p*-values 0.000 at $\alpha = 0.05$ significance level. The last grade level of “energy security” ordinal response variable as *D* is taken as the reference category. Then by using the parameter estimates given in Table 6, cumulative logit model for *A* energy security grade level, indexed by 1, can be given as follows;

$$\begin{aligned}
 \logit[P(Y_i \leq 1)] &= \ln[P(Y_i \leq 1)/(1 - P(Y_i \leq 1))] \\
 &= 1.176(\text{energy sources cons.}) \\
 &\quad + 1.566(\text{elec. prod. from nuclear energy, nat. gas and oil}) \\
 &\quad + .661(\text{energy use per capita and energy imports}) \\
 &\quad - .872(\text{elec. prod. from coal}) + .849(\text{elec. prod. from hydroelectric}) \quad i = 1, 2, \dots, 61
 \end{aligned} \tag{8}$$

Cumulative logit models for *B* and *C* energy security grade levels as $\logit[P(Y_i \leq 2)]$ and $\logit[P(Y_i \leq 3)]$ can be constituted in a similar manner by including the statistically significant intercept parameter estimates

Table 6: Results for GLM by using ML method with hybrid iterative algorithm under cumulative logit link function.

Variables		$\hat{\beta}$	$s.e(\hat{\beta})$	Hypothesis Test			$Exp(\hat{\beta})$	95% Wald Confidence Interval for $Exp(\hat{\beta})$	
				Wald Chi-Square	df	Sig.		Lower	Upper
Intercept	[Energysecurity=A]	-.375	.4001	.879	1	.349	.687	.314	1.505
	[Energysecurity=B]	1.683	.4782	12.391	1	.000*	5.384	2.109	13.745
	[Energysecurity=C]	4.962	.9907	25.081	1	.000*	142.841	20.490	995.789
Energy sources consumptions		1.176	.5637	4.355	1	.037*	3.241	1.073	9.785
Natural gas and oil reserves		.323	.3503	.851	1	.356	1.381	.695	2.744
Electricity productions from nuclear energy, natural gas and oil		1.566	.4591	11.635	1	.001*	4.787	1.946	11.773
Energy use per capita and energy imports		.661	.3188	4.306	1	.038*	1.936	1.036	3.617
Electricity production from coal		-.872	.3571	5.958	1	.015*	.418	.207	.841
Electricity production from hydroelectric		.849	.3595	5.574	1	.018*	2.337	1.155	4.728

*statistically significant at $\alpha = 0.05$ significance level

Table 7: Results for GLM by using ML method with hybrid iterative algorithm under cumulative complementary log-log link function.

Variables		$\hat{\beta}$	$s.e(\hat{\beta})$	Hypothesis Test			$Exp(\hat{\beta})$	95% Wald Confidence Interval for $Exp(\hat{\beta})$	
				Wald Chi-Square	df	Sig.		Lower	Upper
Intercept	[Energysecurity=A]	.320	.2522	1.608	1	.205	1.377	.840	2.257
	[Energysecurity=B]	1.666	.3382	24.268	1	.000*	5.291	2.727	10.265
	[Energysecurity=C]	4.199	.7549	30.934	1	.000*	66.612	15.168	292.525
Energy sources consumptions		.897	.3602	6.206	1	.013*	2.452	1.210	4.967
Natural gas and oil reserves		.093	.2047	.207	1	.649	1.097	.734	1.639
Electricity productions from nuclear energy, natural gas and oil		1.177	.3149	13.971	1	.000*	3.244	1.750	6.014
Energy use per capita and energy imports		.186	.1861	1.001	1	.317	1.204	.836	1.734
Electricity production from coal		-.727	.2191	11.024	1	.001*	.483	.314	.742
Electricity production from hydroelectric		.449	.2337	3.688	1	.055	1.566	.990	2.477

*statistically significant at $\alpha = 0.05$ significance level

as $\hat{\alpha}_2 = 1.683$ for B energy security grade level, and $\hat{\alpha}_3 = 4.962$ for C energy security grade level into the model.

To make a comparison between the best and the worst cumulative link functions, results belonging to the GLM for the world energy security data under the worst cumulative link function are given in Table 7.

As seen from Table 7, there are also two statistically significant intercept parameter estimates, with $\hat{\alpha}_2 = 1.666$ for B energy security grade level, and $\hat{\alpha}_3 = 4.199$ for C energy security grade level with p-values 0.000 at $\alpha = 0.05$ significance level. The last grade level of “energy security” ordinal response variable as D is again taken as the reference category. Then by using the parameter estimates given in Table 7, cumulative

complementary log-log model for A energy security grade level, indexed by 1, can be given as follows;

$$\ln[-\ln[P(Y_i \leq 1)]] = .897(\text{energy sources cons.}) + 1.177(\text{elec.prod.from nuclear energy, nat. gas and oil}) - .727(\text{elec.prod.from coal}) \quad i = 1, 2, \dots, 61 \quad (9)$$

Cumulative complementary log-log models for B and C energy security grade levels as $\ln[-\ln[P(Y_i \leq 2)]]$ and $\ln[-\ln[P(Y_i \leq 3)]]$ can be again constituted in a similar manner by including the statistically significant intercept parameter estimates as $\hat{\alpha}_2 = 1.666$ for B energy security grade level, and $\hat{\alpha}_3 = 4.199$ for C energy security grade level into the model.

As the final results of this study, by using the cumulative logit model, 1 million tons of oil equivalent increase in *energy sources consumptions*; 1% increase in *electricity productions from nuclear energy, natural gas, and oil*; 1 kg of oil equivalent per capita increase in *energy use* and 1% increase in *energy imports*; 1% increase in *electricity production from hydroelectric* increase odds of falling into or below any *energy security grade level* by $e^{1.176} = 3.241$, $e^{1.566} = 4.787$, $e^{0.661} = 1.936$ and $e^{0.849} = 2.337$ units, respectively. On the other hand, 1% increase in *electricity production from coal* decreases the odds of falling into or below any *energy security grade level* by $e^{-0.872} = .418$ units.

By using the cumulative complementary log-log model, 1 million tonnes oil equivalent increase in *energy sources consumptions*, and 1% increase in *electricity productions from nuclear energy, natural gas, and oil* increase the odds of falling into or below any *energy security grade level* by $e^{0.897} = 2.452$, and $e^{1.177} = 3.244$ units, respectively. On the other hand, 1% increase in *electricity production from coal* decreases the odds of falling into or below any *energy security grade level* by $e^{-0.727} = .483$ units.

5 Conclusions

In the light of this study, it can be concluded that overcoming the multicollinearity problem by integrating the mean centering transformation method with the PCA technique as an alternative to the ridge regression and LASSO regression is vital when the covariates in the GLM for the ordinal response variable are moderately or highly correlated. Additionally, the most appropriate cumulative link function selection between the systematic component and the cumulative probabilities of the ordinal response variable is one of the most important stages on the accuracy of parameter estimates, hypothesis tests, and confidence intervals in the GLM. Thus, statistically significant covariates such as *energy use per capita* and *energy imports*, and also *electricity production from hydroelectric* in the cumulative logit model become statistically insignificant in the cumulative complementary log-log model.

Furthermore, the inaccurate estimates of the scale parameter greater than 1, indicating over-dispersion in the data structure, are obtained from the misspecification of the cumulative link functions in the GLM for the world energy security data. These inaccurate estimates of the scale parameter may also cause misspecification of the systematic component in the GLM for the world energy security data. Therefore, larger inaccurate estimates of the scale parameter

lead the statistically significant covariates in the cumulative logit model to become statistically insignificant in the cumulative complementary log-log model.

In the light of this study, as a future work, it is intended to extend this study for the statistical evaluation of three core dimensions of energy sustainability as energy security, energy equity, and environmental sustainability altogether in the world's energy performance by using GLM for the ordinal response variable.

Acknowledgments: The author is grateful to Professor Iskender Akkurt, the editors and three anonymous referees for their valuable comments and contributions to the improvement of this paper.

Conflict of interest: The author states no conflict of interest.

References

- [1] Jacobson M.Z., Review of solutions to global warming, air pollution, and energy security, *Energy&Environmental Science*, 2009, 2, 148-173. doi: 10.1039/b809990c
- [2] Krut B., van Vuuren D.P., de Vries H.J.M., Groenenberg H., Indicators for energy security, *Energy Policy*, 2009, 37(6), 2166-2181. doi: 10.1016/j.enpol.2009.02.006
- [3] Winzer C., Conceptualizing energy security, *Energy Policy*, 2012, 46, 36-48. doi: 10.1016/j.enpol.2012.02.067
- [4] Chester L., Conceptualising energy security and making explicit its polysemic nature, *Energy Policy*, 2010, 38(2), 887-895. doi: 10.1016/j.enpol.2009.10.039
- [5] Sovacool B.K., Mukherjee I., Conceptualizing and measuring energy security: a synthesized approach, *Energy*, 2011, 36(8), 5343-5355. doi: 10.1016/j.energy.2011.06.043
- [6] Bahgat G., Europe's energy security: challenges and opportunities, *International Affairs*, 2006, 82(5), 961-975. doi: 10.1111/j.1468-2346.2006.00580.x
- [7] Lösche A., Moslener U., Rübbelke D.T.G., Indicators of energy security in industrialised countries, *Energy Policy*, 2010, 38(4), 1665-1671. doi: 10.1016/j.enpol.2009.03.061
- [8] Skodras G., Panagiotidou S., Kokorotsikos P., Serafidou M., Potassium catalyzed hydrogasification of low-rank coal for synthetic natural gas production, *Open Chemistry*, 2016, 14, 92-102. doi: 10.1515/chem-2016-0009
- [9] Yergin D., The quest: energy, security, and the remaking of the modern world, The Penguin Press, New York, 2011.
- [10] Awerbuch S., Portfolio-based electricity generation planning: policy implications for renewables and energy security, *Mitigation and Adaptation Strategies for Global Change*, 2006, 11(3), 693-710. doi: 10.1007/s11027-006-4754-4
- [11] Sovacool B.K., Brown M.A., Competing dimensions of energy security: an international perspective, *Annual Review of Environment and Resources*, 2010, 35, 77-108. doi: 10.1146/annurev-environ-042509-143035

- [12] Cherp A., Jewell J., The three perspectives on energy security: intellectual history, disciplinary roots and the potential for integration, *Current Opinion in Environmental Sustainability*, 2011, 3(4), 202-212. doi: 10.1016/j.cosust.2011.07.001
- [13] Belkin P., The European Union's energy security challenges, *Connections*, 2008, 7(1), 76-102. <http://www.jstor.org/stable/26323321>
- [14] Ang B.W., Choong W.L., Ng T.S., Energy security: definitions, dimensions and indexes, *Renewable and Sustainable Energy Reviews*, 2015, 42, 1077-1093. doi: 10.1016/j.rser.2014.10.064
- [15] Iyit N., Yonar H., Genc A., Generalized linear models for European countries energy data, *Acta Physica Polonica A*, 2016, 130(1), 397-400. doi: 10.12693/APhysPolA.130.397
- [16] Sajdak M., Piotrowski O., C&RT model application in classification of biomass for energy production and environmental protection, *Open Chemistry*, 2013, 11(2), 259-270. doi: 10.2478/s11532-012-0147-6
- [17] Er Z., A study of evaluation of solar energy simulation and modeling systems, *Acta Physica Polonica A*, 2016, 130(1), 72-77. doi: 10.12693/APhysPolA.130.72
- [18] Iliashenko R., Zozulia O., Doroshenko A., High stokes shift long-wavelength energy gap regulated fluorescence in the series of nitro/dimethylamino-substituted ortho-analogs of POPOP, *Open Chemistry*, 2011, 9(6), 962-971. doi: 10.2478/s11532-011-0092-9
- [19] Dokuzoglu D., Purutcuoglu V., Comprehensive analyses of gaussian graphical model under different biological networks, *Acta Physica Polonica A*, 2017, 132(3), 1106-1111. doi: 10.12693/APhysPolA.132.1106
- [20] Loomis D.G., Hayden J., Noll S., Payne J.E., Economic impact of wind energy development in Illinois, *Journal of Business Valuation and Economic Loss Analysis*, 2016, 11(1), 3-23. doi: 10.1515/jbvela-2015-0008
- [21] McCullagh P., Nelder J.A., *Generalized linear models*, 2nd ed., Chapman and Hall, London, UK, 1989.
- [22] Agresti A., *Foundations of linear and generalized linear models*, John Wiley&Sons, Inc., Hoboken, New Jersey, 2015.
- [23] Heck R.H., Thomas S.L., Tabata L.N., *Multilevel modeling of categorical outcomes using IBM SPSS*, Quantitative Methodology Series, Taylor&Francis Group, New York, 2012.
- [24] De Jong P., Heller G.Z., *Generalized linear models for insurance data*, Cambridge University Press, Cambridge, 2008.
- [25] Hilbe J.M., Robinson A.P., *Methods of statistical model estimation*, CRC Press, Boca Raton, FL, 2013.
- [26] Ozaltin O., Iyit N., Modelling the US diabetes mortality rates via generalized linear model with the Tweedie distribution, *International Journal of Science and Research (IJSR)*, 2018, 7(2), 1326-1334. <https://www.ijser.net/archive/v7i2/ART2018368.pdf>
- [27] Pearson K., LIII. On lines and planes of closest fit to systems of points in space, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1901, 2(11), 559-572. doi: 10.1080/14786440109462720
- [28] Hoerl A.E., Kennard R.W., Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, 1970, 12(1), 55-67. doi: 10.1080/00401706.1970.10488634
- [29] Tibshirani R., Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B (Methodological)*, 1996, 58(1), 267-288. <http://www.jstor.org/stable/2346178>