ISSN:1306-3111
e-Journal of New World Sciences Academy
2011, Volume: 6, Number: 3, Article Number: 3A0036

PHYSICAL SCIENCES
Received: March 2011
Accepted: July 2011
Series : 3A
ISSN : 1308-7304

Pelin Kasap
Barış Ergül
Arzu Altın Yavuz
Ankara University
pkasap@science.ankara.edu.tr
Ankara-Turkey

EVALUATION OF TURKEY'S HYDROELECTRIC POTENTIAL BY USING STATISTICAL TECHNIQUES

ABSTRACT

In today's world, known energy resources are rapidly consumed; therefore, seeking alternative energy resources is of crucial importance. Recurrent, renewable and raw material-free energy resources are now the focus of investment in many countries. Hydroelectric energy is one of the renewable energy resources in which many countries are investing, especially the European Union member states. Despite having a high potential for hydroelectricity, Turkey still makes use of only part of this potential. To use and scientifically evaluate this potential, this study statistically examined the effect of variables such as installed capacity, operation time and the amount of incoming water on hydroelectric energy generation in Turkey. Least Squares and Ridge Regression techniques were used to determine the relation between the variables.
Keywords: Ridge Regression, Least Squares, Multicollinearity, Multiple Linear Regression, Hydroelectric Energy

TÜRKİYE'NİN HİDROELEKTRİK POTANSİYELİNİN İSTATİSTİKSEL TEKNİKLER KULLANILARAK DEĞERLENDİRİLMESİ

ÖZET

Bilinen enerji kaynaklarının hızla tükendiği günümüzde, alternatif enerji kaynakları arayışı giderek hayati bir önem taşımaya başlamıştır. Kendini sınırsız tekrarlayan, yenilenebilir ve hammadde bağımlısı olmayan enerji kaynakları birçok ülkenin enerji yatırım odağını oluşturmaktadır. Yenilenebilir enerji kaynaklarından olan hidroelektrik enerji başta Avrupa Birliği üye ülkeleri olmak üzere birçok ülkenin başlıca yatırım yaptığı bir enerji kaynağıdır. Hidroelektrik enerji potansiyeli bakımından çok zengin olan ülkemiz, hala bu potansiyelin çok az bir kısmını değerlendirebilmektedir. Bu potansiyelin kullanılabilmesi ve bilimsel olarak değerlendirilebilmesi amacıyla, bu çalışmada kurulu güç, çalışma süresi ve gelen su miktarı değişkenlerinin Türkiye'nin hidroelektrik enerji üretimi üzerindeki etkisi istatistiksel olarak incelenmiştir. Değişkenler arasındaki ilişkileri belirlemek amacıyla En Küçük Kareler ve Ridge Regresyon tekniklerinden yararlanılmıştır.
Anahtar Kelimeler: Ridge Regresyon, En Küçük Kareler, Çoklu Bağlantı, Çoklu Lineer Regresyon, Hidroelektrik Enerji

1. INTRODUCTION (GİRİŞ)

Energy is one of the most important factors necessary for the development of a country. Technological developments and population growth lead to increased energy consumption [10]. Therefore, investment in the energy generation sector has intensified in order to meet growing energy needs. However, a large proportion of greenhouse gas emissions (80-90%), which causes global climate change, are the result of energy generation activities [5]. Scientific data show that it is necessary to protect the natural balance in all phases of energy generation [10]. In this case, renewable energy resources have gained importance as a means to protect natural balance and to counteract climate change. Turkey may be regarded as poor in terms of primary energy resources such as oil, natural gas and coal, but has a rich potential for solar, wind and water resources [2].

Compared to wind and solar energy, hydroelectric generation is the most economical and technically most appropriate renewable energy resource in Turkey. Moreover, hydroelectric energy is an important alternative energy resource that does not produce heated water, air pollution emissions, ash and radioactive waste as thermal and nuclear plants do. According to the provisions of The Framework Convention on Climate Change, hydroelectric plants are the most environmentally friendly energy generation plants regarding the importance of energy in the national economy and environmental protection [5]. In addition, it is possible to use water for other purposes after generating energy. For this reason, the importance of water and energy obtained from water is increasing daily [10].

While determining the hydroelectric potential of a country, gross theoretical, economically viable and technically viable hydroelectric potentials should be determined. The *gross theoretical hydroelectric potential* of a country is the hydroelectric potential calculated under the assumption that all natural flows from the borders to the seas will be used with 100% efficiency. However, even the latest technologies can not make maximum use of this potential. Therefore, the maximum potential that can be utilized with existing technologies is referred to as the *technically viable hydroelectric potential*. Nevertheless, not every technically viable utility is economically viable. Thus, the portion of the technically viable potential that can be realized under existing and expected local economic conditions is referred to as *the economically viable hydroelectric potential*. Turkey's theoretical hydroelectric potential is 1% of the global total and 16% of that of Europe [6].

The gross theoretical viable hydroelectric potential in Turkey is 433 billion kWh, the technically viable potential is 216 billion kWh and the present economically viable potential is 150 billion kWh [15].

Although Turkey has a high hydroelectric potential, only 35% of the existing potential is used. This level of utilization is insufficient; therefore, maximizing the use of this potential will play an important role in meeting Turkey's energy deficit. The tax deductions and subsidy policies available for green energy within the European Union will facilitate efforts to increase the economically viable potential of hydroelectricity [1].

This article statistically examined a range of variables with an effect on energy generation by addressing total annual hydroelectric energy generation in Turkey between 1970 and 2007.

2. RESEARCH SIGNIFICANCE (ÇALIŞMANIN ÖNEMİ)

This study attempted to determine the relationship between the variables such as annual total hydroelectric energy generation in

Turkey (GWh), installed capacity (MW), average operation time (h) and the amount of incoming water (million m^3) by means of a multiple linear regression model. Data for the period 1970-2007 was obtained from the General Directorate of Electric Power Resources Survey and Development Administration (EIE) and the Turkish Electricity Transmission Corporation (TEIAŞ).

3. METHOD (METOT)

Multiple linear regression analysis attempts to explain the relations between the independent variables (installed capacity, average operation time and the amount of incoming water) and a dependent variable (annual total hydroelectric energy generation) with a straight-line fit to the data. When there are q independent variables $X_1, X_2, \ldots, X_q$; the linear multiple regression equation is in the general form

$$Y = X\theta + \varepsilon . \tag{3.1}$$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1q} \\ 1 & x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nq} \end{bmatrix} \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_q \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Y is the dependent variable; $X_1, X_2, \ldots, X_q$ are the independent variables (explanatory variables); θ_i $(0 \le i \le q)$ are the regression coefficients; ε_i are random errors.

Least Squares (LS) estimators of θ_i regression coefficients were obtained from the following equation:

$$\hat{\theta} = \left(X'X \right)^{-1} X'Y . \tag{3.2}$$

In order to apply the LS technique, certain assumptions are required to be met, as follows:

- Error terms have mean zero.
- Error terms are uncorrelated, and have common variance σ^2 (homogeneous variance).
- There is no correlation between the errors and dependent variable.
- The error terms and independent variables are uncorrelated.
- There is no significant relation between independent variables [7].

In cases where one of the above-mentioned assumptions was not met, the LS estimators would be biased and inconsistent and no more effective. These estimators were considered as Best Linear Unbiased Estimators (BLUE) as long as they met the first and second assumptions, above.

Another common problem encountered in multiple regression analysis is that of multicollinearity, in which the independent variables were correlated. Multicollinearity causes problems such as increased variances and covariances of the regression coefficients and infinite standard errors of these coefficients. Moreover, the R^2 value is high and some of the independent variables are not significant according to the t-test results. If the independent variables are correlated, then becomes necessary to eliminate the multicollinearity or to reduce its effect. To eliminate such a problem, many researchers tried to achieve a solution by removing one or more of the multicollinear variables from the model. However, such a solution may produce inaccurate results, as the variable selection methods are also affected by the multicollinearity [12].

If there are multicollinearity between the independent variables, it will not be appropriate to solve the problem by means of the LS technique. The following methods can be employed to determine whether or not variables show multicollinearity [7, 11 and 12]:

- A simple correlation matrix is examined. Superficially, if the simple correlation coefficient between two independent variables is highly significant (r > 0.75), then this may cause a multicollinearity problem.
- If one or more of the eigenvalues of the $X'X$ matrix is zero or close to zero, then this may indicate multicollinearity.
- Variance Inflation Factor (VIF) can be examined.
- Condition Number (CN) or Condition Index (CI) values can be calculated. These values are calculated by using the eigenvalues of the $X'X$ matrix as follows:

$$CN = [Maximum\ Eigenvalue / Minimum\ Eigenvalue]$$

$$CI = [Maximum\ Eigenvalue / Minimum\ Eigenvalue]^{1/2}$$

A CI value between 10 and 30 indicates moderate multicollinearity and a value above 30 indicates strong multicollinearity; A CN value between 100 and 1000 indicates moderate multicollinearity and a value above 1000 indicates strong multicollinearity.

- *t*-test results for independent variables and R^2 value can be used.

Each of the approaches mentioned above to determine the multicollinearity problem are found to have some disadvantages. Moreover, there is no suggestion about which approach should be employed in which situation [7].

When a multicollinearity problem is determined, one of the ways to solve this problem is the Ridge Regression (RR) technique.

Estimations using the LS technique will be biased and efficient if the assumptions of this technique are met. However, when the existence of multicollinearity problem is determined, the variance and covariance values of the regression coefficients increase. This causes the standard errors of the regression coefficients to be high and also the t-tests of the regression coefficients of these variables give statistically insignificant results. In case of multicollinearity, important changes occur in the partial regression coefficients when the data for any of the independent variables or observations are added to or removed from the model. Furthermore, in case of multicollinearity, the signs of the partial regression coefficients may be different than expected. In brief, standardized regression coefficients calculated using multicollinear data lose their stability and determination. The RR technique help to reduce the variance by adding a small bias constant to the estimations. Generally, except for adding a small bias constant (k^*) to the diagonal elements of a correlation matrix, the operation of the RR and LS techniques are the same [4, 8, 9 and 14].

In the RR technique, the biased standardized regression coefficients are calculated as follows:

$$\hat{\theta}_R = \left(X'X + k^* I\right)^{-1} X'Y$$
$$\hat{\theta}_R = \left(X'X + k^* I\right)^{-1} X'X\hat{\theta} \qquad (3.3)$$
$$\hat{\theta}_R = Z\hat{\theta}$$

Here, $\left(X'X + k^* I\right)^{-1} \cdot X'X = Z$ and k^* is a positive value less than 1 [8].

For a specified k^* value, the mean square errors (MSE) of the biased estimations are less than the MSE values of the estimators obtained by using LS technique. However, the optimum k^* constant depends on the real regression coefficients. There is no approach, to date, that guarantees the optimum solution. Here, k^* is between 0 and 1. As the k^* value is close to 1, the bias of the estimations increases while the variances decreases.

To investigate the optimum k^* constant, it is possible to use Ridge Trace graphs, calculated between standardized regression coefficients and k^* [9]. In these graphs, the regression coefficients are shown as a function of k^*. The optimum value of k^* is chosen from the area where the regression coefficients are stable. In general, standardized regression coefficients change abnormally with small values of k^* and then become stable. In this area where the regression coefficients are stable, the smallest possible k^* value is chosen as the optimum k^* value. Other criteria used in the selection of optimum k^* value is the k^* constant approaches, minimizing the sum of square errors of the coefficients; and providing the minimum VIF's (VIF values closing to 1 for the independent variables) [13].

4. FINDINGS (BULGULAR)

This study examined the relationship between the variables such as annual total hydroelectric energy generated in Turkey (GWh), installed capacity (MW), average operation time (h) and the amount of incoming water (million m^3) by means of a multiple linear regression model. The variables used were:

Y : Annual generation (YU),

X_1 : Average operation time (CS),

X_2 : Amount of incoming water (GS) and

X_3 : Installed capacity (KG).

To provide the constant variance and remove the seasonality effect, logarithmic values of the data were taken and the analysis used logarithmic values.

Multicollinearity between the variables was shown through various techniques. First, the correlation matrix of the independent variables was obtained as follows:

Table 1. Correlations
(Tablo 1. Korelasyonlar)

Correlations	logcs	loggs	logkg
Logcs	1,00	0,14	-0,30
Loggs	0,14	1,00	0,79
Logkg	-0,30	0,79	1,00

When the simple correlation matrix was examined, the simple correlation coefficient between two independent variables (loggs and logkg) was found to be highly significant ($r > 0.75$), thus this could cause a multicollinearity problem.

Then the eigenvalues and CI values were found.

Table 2. Eigenvalue and condition ındex
(Tablo 2. Özdeğer ve koşul indeksi)

Dimension	Eigenvalue	CI
1	3,989000	1,000
2	0,008903	21,169
3	0,001490	51,750
4	0,000190	144,834

As shown in Table 2, one or more of the eigenvalues of the $X'X$ matrix were zero or close to zero, so the multicollinearity problem might occur. In addition, the CI value was calculated as 144.834. As this value was higher than 30, it indicated multicollinearity.
The table for the VIF values was found as follows.

Table 3. VIF values
(Tablo 3. VIF değerleri)

Parameter	Estimate	VIF
constant	-2,99402	
logcs	0,984912	1,925
loggs	0,005615	4,720
logkg	0,996507	5,082

It was observed from this table that the VIF values were high.
The results of regression analysis and variance analysis were as follows:

Table 4. Model results for multiple regression
(Tablo 4. Çoklu regresyon modeli için sonuçlar)

Parameter	Estimate	Standard Error	T Statistic	P-Value
constant	-2,94402	0,0362196	-81,2825	0,0000
logcs	0,984912	0,0106692	92,314	0,0000
loggs	0,00561517	0,0109621	0,512233	0,6118
logkg	0,996507	0,00433445	229,904	0,0000

Table 5. Analysis of variance
(Tablo 5. Varyans analizi)

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value	R-Sq.
Model	5,33575	3	1,77858	82048,61	0,0000	0,9998
Residual	0,000737024	34	0,0000216772			
Total (Corr.)	5,33649	37				

The R^2 value was 0.9998 and the result of the t-test for the independent variables showed that some of the coefficients were not significant.

According to these calculations, it was clear that there was multicollinearity between the independent variables. Therefore, if the LS technique had been used, it would have been possible for parameter estimations to give inaccurate results and an unreliable model. Thus, the use of the RR technique could remove the effect of multicollinearity, thus providing more accurate estimators and a more reliable model.

When the RR technique was applied for this dataset by means of Statistica (Release 5.0) and Statgraphics (Version 15.2.11) statistical programs, the standardized regression coefficients and the VIF values were found as follows:

Table 6. Standardized regression coefficients for k^*
(Tablo 6. k^* için standartlaştırılmış regresyon katsayıları)

k^*	β_1	β_2	β_3	R-Sq.(%)
0,00	0,258	0,002	1,040	99,986
0,01	0,233	0,048	0,991	98,890
0,02	0,213	0,085	0,947	97,899
0,03	0,195	0,116	0,909	96,987
0,04	0,180	0,140	0,876	96,140
0,05	0,167	0,164	0,847	95,345
0,06	0,156	0,183	0,822	94,592
0,07	0,146	0,199	0,799	93,877
0,08	0,137	0,213	0,779	93,193
0,085	0,133	0,220	0,769	92,861
0,086	0,132	0,221	0,767	92,795
0,087	0,132	0,222	0,765	92,729
0,088	**0,131**	**0,224**	**0,764**	**92,664**
0,09	0,129	0,226	0,760	92,535
0,1	0,122	0,237	0,743	91,900

Table 7. VIF values for k^*
(Tablo 7. k^* için VIF değerleri)

k^*	VIF_1	VIF_2	VIF_3
0,00	1,925	4,720	5,082
0,01	1,712	3,942	4,231
0,02	1,549	3,352	3,586
0,03	1,419	2,895	3,087
0,04	1,314	2,533	2,691
0,05	1,227	2,242	2,372
0,06	1,154	2,003	2,113
0,07	1,092	1,806	1,898
0,08	1,039	1,639	1,717
0,085	1,015	1,567	1,638
0,086	1,010	1,553	1,623
0,087	1,005	1,539	1,608
0,088	**1,000**	**1,525**	**1,593**
0,09	0,992	1,499	1,564
0,1	0,951	1,379	1,434

The parameter estimations obtained for different k^* values and the R^2 values are shown in Table 6. As seen from Table 6, when k^* is 0.088, β_1, β_2 and β_3 values were 0.131, 0.224 and 0.764, respectively, and the corresponding R^2 value was found 92.664. As shown in Tables 6 and 7, 0.088 value –at which the standardized regression coefficients and the VIF values became stable– was determined as k^*.

The Ridge Trace and VIF graphs based on the standardized regression coefficients and VIF values are shown below:

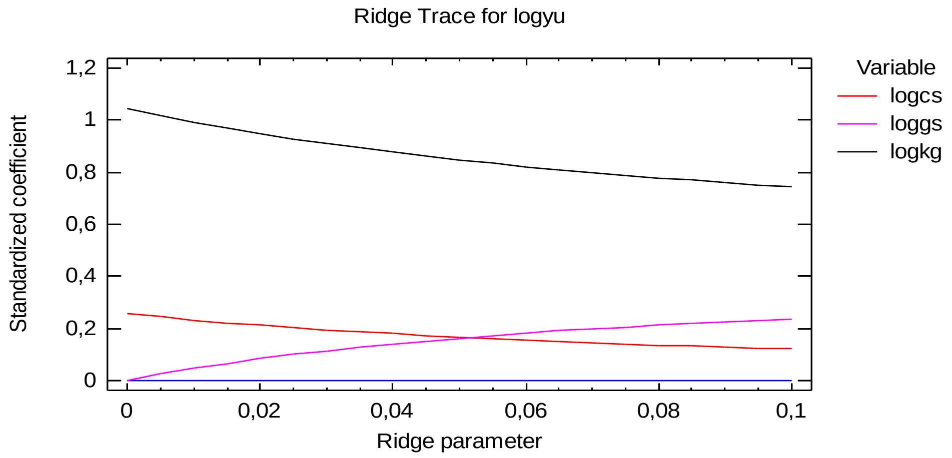

Figure 1. Ridge trace graph
(Şekil 1. Ridge trace grafiği)

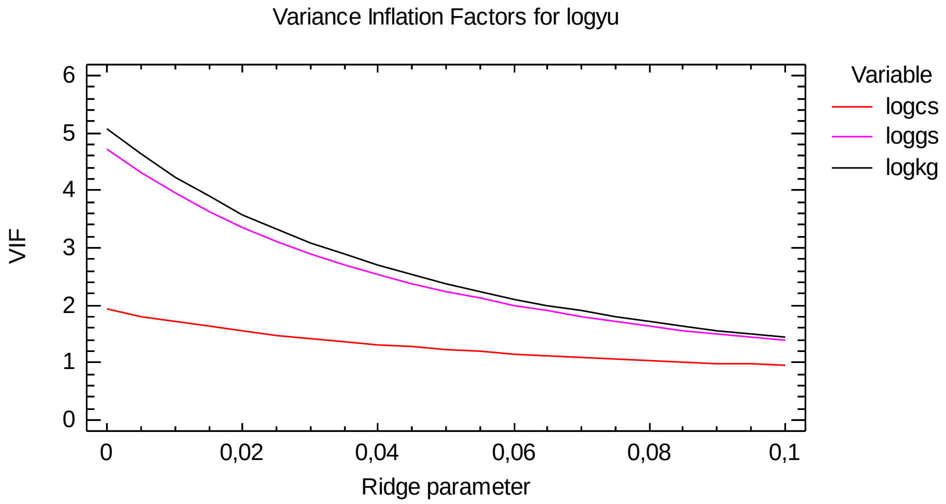

Figure 2. VIF graph
(Şekil 2. VIF grafiği)

According to these, the RR analysis results for a value of $k^* = 0.088$ were as follows:

Table 8. Model results when $k^* = 0,088$
(Tablo 8. $k^* = 0,088$ olduğunda model için sonuçlar)

Parameter	Estimate	VIF	R-Sq.
constant	-1,30542		0,9266
logcs	0,49883	1,00095	
loggs	0,5596	1,52543	
logkg	0,728504	1,59337	

When the RR technique was applied to the data, the following model equation estimating the total annual hydroelectric generation was obtained:

$$\log yu^{*} = -1.305 + 0.499*\log cs + 0.560*\log gs + 0.729*\log kg \ . \quad (4.1)$$

According to the equation, the installed capacity can be said to have the strongest impact on annual hydroelectric energy generation.

5. RESULT AND SUGGESTIONS (SONUÇ VE ÖNERİLER)

Hydroelectric energy resources are entirely domestic, not dependent on foreign countries and generated at plants established on the rivers, i.e. the natural resources of a country. As long as Turkey maximizes the use of hydroelectric energy capacity, dependence on energy sources obtained from foreign countries may be reduced. One of the most important advantages of hydroelectric plants is their energy storage potential. In Turkey, the energy storage capacity of the existing hydroelectric plants is half of the total annual generation capacity [2]. This is an important advantage for Turkey, which mostly depends on foreign energy resources. Also, both the construction and operation of hydroelectric plants have positive impacts on the regional economy. Generally, the areas in which hydroelectric plants are established are economically undeveloped rural regions. Utilization of hydroelectric energy potential of the region by external actors will increase the welfare of the local community. In addition, dam construction will create new employment opportunities.

In Turkey, hydroelectricity represented 38.5% of renewable energy resources for electricity generation in 1997; In 2010, this proportion had declined to 24.6%.

Moreover, the economic lifespan of hydroelectric plants are longer than other types of plants (75 years). After the first period, these plants can continue to generate electricity for second, third and fourth 75-year periods with just minor investments. Hydroelectric plants have the lowest operating cost and have no fuel cost [3].

This study attempted to determine the relationship between variables such as annual total hydroelectric energy generation in Turkey (GWh), installed capacity (MW), average operation time (h) and the amount of incoming water (million m^3) by means of a multiple linear regression model. However, during the examination of independent variables, it was seen that there was a problem of multicollinearity. Therefore, this problem was examined using the RR technique to develop a model of the relevant variables.

Finally, the amount of total hydroelectric energy produced in a year (y) is explained by the independent variables called as the operation time (x_1), the amount of incoming water (x_2) and the installed capacity (x_3). Annual total hydroelectric energy produced in Turkey can be estimated as 448.2 billion kWh when we take the independent variables as x_1=6000 hours, x_2=200 million m^3 and x_3=15000 MW. As it can be seen from the equation (4.1), annual total hydroelectric energy produced in Turkey increases as the amount of installed capacity, the amount of incoming water and the operation time increase.

REFERENCES (KAYNAKLAR)

1. Alemdaroğlu, N., (2007). Enerji Sektörünün Geleceği Alternatif Enerji Kaynakları ve Türkiye’nin Önündeki Fırsatlar, İTO Yayınları, Yayın No: 29, İstanbul.
2. Bakır, N.N., (2001a). Türkiye’nin Hidroelektrik Potansiyelinin Yeniden Değerlendirilmesi.
3. Bakır, N.N., (2001b). Hidroelektrik Perspektifinden Türkiye ve AB Enerji Politikalarına Bakış, www.ere.com.tr

4. Darlington, R.B., (1978). Reduced Variance Regression, Psychological Bulletin, 85, pp. 1283- 1255.
5. Eroğlu, V., (2010). İklim Değişikliği ile Mücadelede Yenilenebilir Enerji Olarak Hidroelektrik Enerji Üretiminin Artırılması ve Diğer DSİ Faaliyetleri, www.tasam.org/images/tasam/Veysel_Eroglu.pdf *Erişim Tarihi: Aralık, 2010.*
6. Gençoğlu, M.T. ve Cebeci, M., (2001). Büyük Hidroelektrik Santrallar İle Küçük Hidroelektrik Santralların Karşılaştırılması, YEKSEM '2001 Yenilenebilir Enerji Kaynakları Sempozyumu, 265-271, İzmir.
7. Gujarati, D.N., (1995). Basic Econometrics. 3rd ed. New York: McGraw-Hill.
8. Hoerl, A.E. and Kennard, R.W., (1970a). Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics, 12, pp. 55-67.
9. Hoerl, A.E. and Kennard, R.W., (1970b). Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics, 12, pp. 69-82.
10. Küçük, İ., (1996). Türkiye'de Hidroelektrik Potansiyeli Üzerine Bir Değerlendirme, TMMOB 1. Enerjı Sempozyumu-12-14 Kasım 1996 Ankara.
11. Montgomery, D.C. and Peck, E.A., (1982). Introduction to Linear Regression Analysis, John Wiley and Sons, New York.
12. Myers, R.H., (1990). Classical and Modern Regression with Applications, Massachusetts: PWS-Kent Publishing Company, Boston.
13. Orhunbilge, N., (2000). Uygulamalı Regresyon ve Korelasyon Analizi, Avcıol-Basım Yayın, İstanbul.
14. Price, B., (1979). Ridge Regression: Application to Nonexperimental Data, Psychological Bulletin, 84, pp. 759-766.
15. Türkiye Enerji Raporu, Dünya Enerji Konseyi Türk Milli Komitesi, (2007-2008) Erişim Tarihi: Aralık, 2010. http://www.dektmk.org.tr/upresimler/2008_enerji_raporu.pdf