Gazi University Journal of Science
GU J Sci
29(2):237-244 (2016)

Comparing Energy Demand Estimation Using Various Statistical Methods: The Case of Turkey

Y. Murat BULUT[1,♠], Zeki YILDIZ[2]

[1]*Eskişehir Osmangazi University, Faculty of Science and Letters, Department of Statistics, Eskişehir, Turkey, e-mail: ymbulut@ogu.edu.tr*

[2]*Eskişehir Osmangazi University, Faculty of Science and Letters, Department of Statistics, Eskişehir, Turkey, e-mail: zyildiz@ogu.edu.tr*

Received: 17/04/2014 *Revised: 30/09/2015* *Accepted: 11/01/2016*

ABSTRACT

Many engineers and scientists concern with future energy demand. They use many different statistical methods to estimate future energy demand such as multiple linear regression, neural networks, genetic algorithms and so on. In this paper, we propose ridge regression (RR) and partial least squares regression (PLSR) methods to estimate future energy demand. Because of the fact that variables, which are used in energy demand, are very collinear, ridge regression and partial least squares regression methods give more realistic results than least squares regression method. So, energy demand equations are developed based on RR and PLSR methods. Since, RR give better estimation, we estimate Turkey's future energy demand based on RR method.

Keywords: *Energy Demand, Energy Modeling, Biased Regression*

1. INTRODUCTION

Energy is of great importance for countries' social, economic and technological progress. Thus, energy sources are very important for all the countries of the world. Parallel to the rapid development in industry, agriculture, transport and other sectors, global energy demand has continuously increased. Indeed, together with the rapid development in these sectors, the increase in countries' populations, gross national products and their import and export figures has led to an increase in the demand for energy (Ceylan and Öztürk (2004)). For this reason, estimations are made regarding the amounts of energy that each country will need in the future and plans are formulated based on these estimations.

Therefore, accurate estimation is very important. Hence, widespread energy estimation modeling is a subject of current interest among practitioners and academicians concerned with problems of energy production and consumption (Sözen et al. (2005), Kıran et al. (2012)).

Turkey, with its rapidly growing young population, strong economic growth and urbanization, is one of the largest markets for energy in the world. Although Turkey's economy has a dynamic structure induced by these factors (Kıran et al. (2012)), Turkey is highly dependent on imports to satisfy its energy needs. Due to a lack of fossil fuel resources, Turkey's dependency level is approximately 70%, and it may rise to over 80% by 2030. In terms of primary energy consumption, Turkey is 21st; in terms of natural gas consumption, it is

♠Corresponding author, e-mail: *ymbulut@ogu.edu.tr*

24th; furthermore, it is 12th in its hydroelectric consumption and 15th in its coal consumption (the annual report of the Ministry website). Additionally, Turkey imports the most energy from Russia. Currently, Turkey meets 40% of its energy needs by means of oil, and 90% of its oil supplies are imported from the Middle East (Saudi Arabia, Iran, Iraq and Syria) and the Russian Federation (Ünler (2008)).

Because the demand for energy is rapidly increasing, the estimations of future energy demand are gaining importance. The first studies about estimating the energy demand in Turkey used the simple regression method of the State Planning Organization. In 1984, econometric models to estimate future energy demand began to be used in our country. One of these econometric models is the MAED model (the Model for the Analysis of the Energy Demand), which was developed by the International Atomic Energy Agency and is used by the Ministry of Energy and Natural Resources (Toksarı (2007)). Specifically, MAED was used to estimate medium-and long-term energy demands. This model provides estimates of energy demand by using changes in the GNP (Gross National Product) and population. As an alternative to MAED, OREM (the Optimal Renewable Energy Mathematical Model) and ESM (the Energy Simulation Model), which are based on changes in the population and GNP, are also used (Ceylan and Öztürk (2004)). Based on the literature, there is a relationship between energy consumption and GNP (Öztürk and Ceylan (2005)). Ebohan (1996) has revealed that there is a relationship between energy consumption and the economic growth of two countries. Dinçer and Dost (1997) have also demonstrated that there is a relationship between energy consumption and GDP (Gross Domestic Product). In recent years, many methods have been used to forecast energy demand, such as the cyclic patterns method (Ediger and Tatlıdil (2002)), time series analysis (Aras and Aras (2004), Ediger et al. (2006)), neural networks (Sözen et al. (2005), Murat and Ceylan (2006)), genetic algorithms (Ceylan and Öztürk (2004), Canyurt et al. (2004), Haldenbilen and Ceylan (2005), Ceylan et al., (2005), Öztürk et al. (2004)), ant colony optimization (Toksarı (2007)) and particle swarm optimization (Ünler (2008)).

Multiple linear regression analysis is a method that is commonly used to estimate energy demand (Yumurtacı and Asmaz (2004), Abdel-Aal et al. (1997), Al-Garni et al. (1994), Jyoti et al. (2007)). In this method, when the explanatory variables are related to one another, the estimates made will be biased and inconsistent. This relationship is called multicollinearity. Indeed, the variables in energy demand studies are often related variables, and various methods have been developed to overcome the problem of multicollinearity. In this study, because of the problem with multicollinearity, Partial Least Squares (PLSR) and Ridge Regression (RR) methods were used to establish a model. PLSR and RR, which are compositions of regression analysis, are statistical tools that have been specifically designed to address multicollinearity problems where the number of observations is limited and the correlations between the predictor variables are high. These characteristics of PLSR and RR have been demonstrated in many studies (Garthwaite (1994), Bulut (2011)). The PLSR and RR approaches are new in the field and have not previously been used to model the energy demand. Many factors influence the energy demand, such as the GNP, population, imports and exports of a country.

In the following section, the concepts of PLSR and RR are explained. In Section 3, the energy demand forecasting model is developed with PLSR and RR for Turkey. The results of energy demand forecasting and future projections are presented in Section 4. Then in Section 5, we conclude our paper.

2. METHOD: RIDGE REGRESSION, PARTIAL LEAST SQUARES REGRESSION AND MODEL DEVELOPMENT

In linear regression analysis, the case where one dependent variable has more than one explanatory variable is called multiple linear regression analysis. In the multiple linear regression method, the parameters are traditionally calculated by the ordinary least squares (OLS) method. To use an OLS estimator, it is assumed that there is not a linear relationship among the explanatory variables. However, this assumption is not always valid for the data sets that are actually used. That is, there may be relationship between the explanatory variables. This condition is known in the literature as a multicollinearity problem.

The problem of multicollinearity causes variance values of the least squares estimator obtained from the analysis to be higher than they should be, and furthermore, the regression coefficients diverge from their real values. When faced with this problem, the least squares estimator gives unreliable results for the regression coefficients (Gunst and Mason (1980)). Biased regression methods have been developed for multicollinearity. In this study, RR and the PLSR methods, which are biased methods, were used.

2.1. Ridge Regression

When the ridge regression method is employed due to multicollinearity, this process involves the most common applications of biased estimators. In 1970, this estimator was proposed by Hoerl and Kennard for the first time. Hoerl and Kennard (1970) obtained Ridge estimators by adding a small constant to the diagonal elements of the matrix $X'X$ as shown in equation (1).

$$\hat{\beta}_R = (X'X + kI)^{-1}X'y \tag{1}$$

where, k $(0 < k < 1)$ is called the bias constant or the shrinkage parameter, selecting the constant k affects the performance of the Ridge estimator. If $k = 0$, the Ridge estimator becomes the OLS estimator (Hoerl and Kennard (1970)). The expected value of the Ridge estimator given in equation (1) and its bias are calculated as in equations (2) and (3).

$$E(\hat{\beta}_R) = (X'X + kI)^{-1}X'X\beta \tag{2}$$

$$Bias(\hat{\beta}_R) = -k(X'X + kI)^{-1}\beta \tag{3}$$

where, β represents the regression coefficients calculated with the OLS method.

The selection of k (which is a bias parameter) determines to what extent it will address the problem of the multicollinearity of the Ridge estimator, and hence, it is an important subject. Moreover, because the parameter k determines how the estimator can be biased, this parameter is often chosen to be close to zero. Many methods have been offered for the selection of k and these methods can be classified into two groups: graphical methods and analytic methods. In this study, the Ridge trace method, which is one of the graphical methods, will be used.

2.2. Partial Least Squares Regression

Partial Least Squares Regression (PLSR) is a regression method that is a combination of principle component analysis and multiple linear regression (Abdi (2010)). This method, which was developed by Herman World in the 1960s, is a statistical method that is typically used when the problem of multicollinearity is encountered or when the number of the variables is more than the observation can account for.

In PLSR, new explanatory variables are defined with the help of explanatory variables that have multicollinearity. When these variables are defined, the change in the dependent variables is taken into account (Garthwaite (1994)). In the literature, there are many algorithms defined for this method. Some of them are as follows: NIPALS as suggested by Wold (1975), UNIPALS as suggested by Glen (1989), the Kernel algorithm suggested by Lindgren and et al. (1993) and the SIMPLS algorithms suggested by De Jong (1993). In this study, the NIPALS algorithm, which is known as the classical algorithm, will be used.

In PLSR, before the calculation of the β_{PLS} coefficients, the X data matrix is separated as follows

$$X = t_1 p_1' + \cdots + t_p p_p' = \sum_{i=1}^{p} t_i p_i' = TP' \quad (4)$$

where, $t_i{}'s$ represents linear combinations of X, the $p_i{}'s$ represents loads. To provide orthogonality, the $t_i{}'s$ surplus matrix are calculated as combinations of $E_i{}'s$ as in equation (5).

$$t_i = E_{i-1} w_i, E_i = X - \sum_{j=1}^{i} t_j p_j', E_0 = X \quad (5)$$

where, the $w_i{}'s$ are orthonormal vectors. After the $t_i{}'s$ are calculated, the $p_i{}'s$ are calculated by performing a regression of X on $t_i{}'s$. Then, the factor m, which has the maximum information for the estimation model that is catching the majority of the changes in the matrix X, is calculated. Hence, the relationships given in equations (6)-(8) are obtained.

$$T_m = XR_m \quad (6)$$

$$P_m = X'T_m(T_m'T_m)^{-1} \quad (7)$$

$$R_m = W_m(P_m'W_m)^{-1} \quad (8)$$

where the matrix T_m has the form of $[t_1, t_2, \dots, t_m]$. The same situation is also valid for other matrices.

By means of these equations, $\hat{\beta}_{PLS}^{m}$ is calculated as in equation (9) (Phatak and De Jong (1997)).

$$\hat{\beta}_{PLS}^{m} = R_m(R_m'X'XR_m)^{-1}R_m'X'X\hat{\beta}_{OLS} \quad (9)$$

3. The Application of the RRED and PLSRED Models

The data used in this application are derived from various resources. The GNP is collected from the Central Bank of Turkey (CBT). The observations are collected from the World Energy Council Turkish National Committee (WECTNC). General import and export figures of Turkey are obtained from National Statistics (NS). Energy consumption, GNP, population, and import and export data for Turkey between 1979 and 2011 are shown in Table 1.

Table 1. Energy demand, GNP, population, and import and export data between 1979 and 2011

Year	Energy Demand	GNP	Population	Imports	Exports

1979	30.71	82	45.53	5.07	2.26
1980	31.97	68	44.44	7.91	2.91
1981	32.05	72	45.54	8.93	4.7
1982	34.39	64	46.69	8.84	5.75
1983	35.7	60	47.86	9.24	5.73
1984	37.43	59	49.07	10.76	7.13
1985	39.4	67	50.31	11.34	7.95
1986	42.47	75	51.43	11.1	7.46
1987	46.88	86	52.56	14.16	10.19
1988	47.91	90	53.72	14.34	11.66
1989	50.71	108	54.89	15.79	11.62
1990	52.98	151	56.1	22.3	12.96
1991	54.27	150	57.19	21.05	13.59
1992	56.68	158	58.25	22.87	14.72
1993	60.26	179	59.32	29.43	15.35
1994	59.12	132	60.42	23.27	18.11
1995	63.68	170	61.53	35.71	21.64
1996	69.86	184	62.67	43.63	23.22
1997	73.78	192	63.82	48.56	26.26
1998	74.71	207	65	45.92	26.97
1999	76.77	187	66.43	40.67	26.59
2000	80.5	200	67.42	54.5	27.78
2001	75.4	146	68.37	41.4	31.33
2002	78.33	181	69.3	51.55	36.06
2003	83.84	239	70.23	69.34	47.25
2004	87.82	299	71.15	97.54	63.17
2005	91.58	361	72.97	116.77	73.48
2006	99.59	483	72.97	139.58	85.54
2007	107.625	531	70.59	170.06	107.27
2008	106.273	648	71.13	201.96	132.03
2009	106.138	730	73.23	140.93	102.14
2010	109.266	615	74.47	185.54	113.88
2011	114.48	731	74.72	240.84	134.91

Using the data between 1979 and 2005, the RR and PLSR Models were obtained in equations (10) and (11), respectively. The data between 2006 and 2011 were used for testing the validity of the model.

$$\hat{y}_{RRED} = -47.2196 + 0.034889X_1 + 1.69228X_2 + 0.0963728X_3 - 0.0717829X_4 \tag{10}$$

$R^2 = 0.987$

When the Ridge regression coefficients were calculated, parameter k was chosen with the help of the Ridge trace graph given in Figure 1.

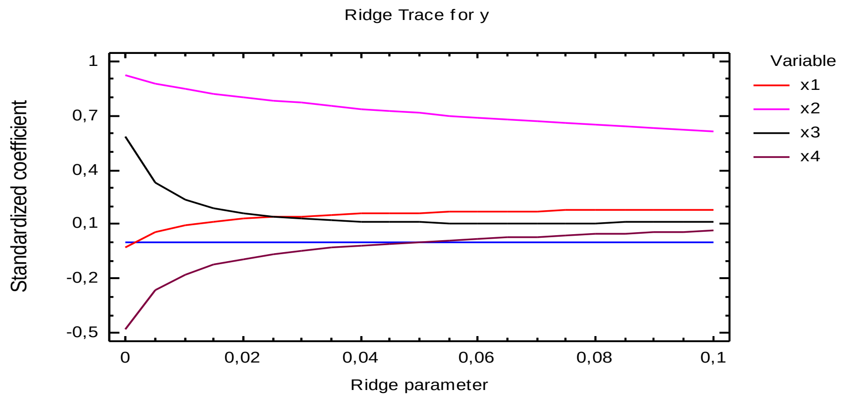

Figure 1. A ridge trace graph

From figure 1, the smallest value with which the coefficients come close to stopping was chosen, and this value is 0.025. Using this value, the regression equation was established.

In the PLSR regression before establishing a model, the number of variables needed for the model is decided. For this purpose, the graphs for the variance's explanatory ratios against the number of components needed in the model were given in Figure 2.

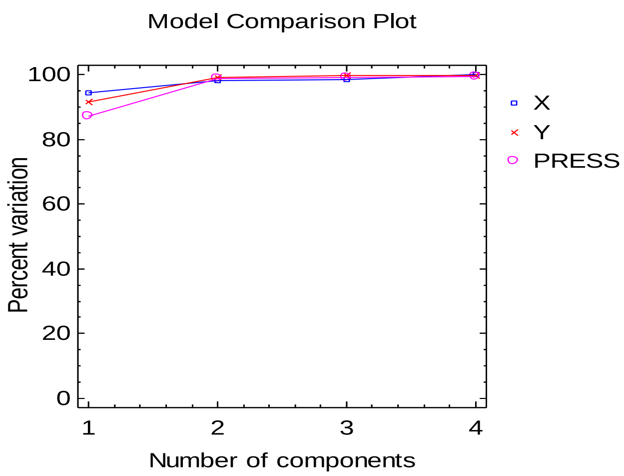

Figure 2. A graph of the variance's explanatory ratios against the number of components.

When Figure 2 examined, it seems that a two-component model is convenient. In this case, the number of components necessary in the model was taken as two. The two-component PLSRED model is as stated in equation (11).

$$\hat{y}_{PLSRED} = -55.8452 + 0.0412233X_1 + 1.86044X_2 - 0.000655213X_3 - 0.0196868X_4 \quad (11)$$

$R^2 = 0.989$

Using the Ridge regression energy demand (RRED) and Partial Least Squares regression energy demand (PLSRED) models, the energy demand between 2006 and 2011 was calculated, and these values were compared to the expected values. The values calculated by the RRED and PLSRED models, the observed values and the relative error (RE) and mean squared error (MSE) value of the model are given in Table 2.

Table 2. Testing of the RRED and PLSRED Models.

Year	ED(MTOE)	RRED(MTOE)	R.E(%)	PLSRED(MTOE)	R.E(%)
2006	99.59	100.44	-0.84	98.05	1.53
2007	107.63	99.44	8.81	95.14	13.44
2008	106.28	105.71	0.60	100.43	6.21
2009	106.14	108.43	-2.44	108.40	-2.40
2010	109.27	109.95	-0.75	105.67	3.93
2011	114.48	118.27	-4.34	110.50	4.56
		Mean	0.17		4.55
		MSE	14.68		37.72

4. ESTIMATION OF FUTURE ENERGY DEMAND

To estimate future energy demand, the RRED model, which gives better estimations for the energy demand, is used. Parallel to the study by Toksarı (2007), the following three scenarios were established based on the data from past years.

Scenario 1: It is assumed that between 2012 and 2025, the GDP will increase by 6%, the population will increase by 0.17%, imports will increase by 4.5% and exports will increase by 2%.

Scenario 2: It is assumed that between 2012 and 2025, the GDP will increase by 5%, the population will increase by 0.15%, imports will increase by 5%and exports will increase by 4.5%

Scenario 3: It is assumed that between 2012 and 2025, the GDP will increase by 4%, the population will increase by 0.18%, imports will increase by 4.5% and exports will increase by 3.5%.

In Table 3, according to each scenario with the RRED model, the energy demand estimation is given for the years from 2012 to 2025.

Table 3. Future estimations of the total energy demand according to the above scenarios

		Scenario 1	Scenario 2	Scenario 3
Year	ED(MTOE)	RRED(MTOE)	RRED(MTOE)	RRED(MTOE)
2012	N/A	112.48	112.09	111.85
2013	N/A	114.58	113.74	113.26
2014	N/A	116.78	115.48	114.70
2015	N/A	119.11	117.28	116.20
2016	N/A	121.56	119.17	117.75
2017	N/A	124.15	121.15	119.35
2018	N/A	126.89	123.21	121.00
2019	N/A	129.78	125.37	122.71
2020	N/A	132.83	127.63	124.48
2021	N/A	136.05	129.98	126.32
2022	N/A	139.45	132.45	128.21
2023	N/A	143.05	135.03	130.18
2024	N/A	146.85	137.74	132.21
2025	N/A	150.86	140.56	134.32

5. CONCLUSION AND DISCUSSION

Although regression methods are widely used in energy demand estimation, the results are far from the actual values when there is multicollinearity among the data. In this case, the PLSR and RR methods can be used as an alternative. This study shows that the PLSR and RR methods can be applied in studies of energy demand estimation. When these methods are used, more consistent estimates can be obtained than are obtained from classical regression.

Because there was believed to be a relationship between economic development and energy demand, the GNP, population, import and export figures were used to estimate the energy demand in Turkey. Furthermore, models to explain the demand for energy have been established with the help of these variables. Additionally, the validity of the models was tested using the data between 2006 and 2011. When the validity of the model was examined as in Table 2, it was observed that the mean relative error (0.17) was smaller than the error (4.55) of the model established with the help of PLSR. Except for the relative error values, the calculated values of the MSE were acceptable, and similar results to the relative errors were obtained. The MSE values are smaller, when the model developed by RR was used for estimating the future energy demand. In addition, the three different scenarios reflecting the GNP, the population, the import and export figures are considered for future energy demand estimation and the energy demand has been estimated according to these scenarios in the coming years.

As a result, in this study we observed multicollinearity among the variables used in the estimation of the energy demand, and the inconvenience of the widely used classical regression method was stated. It is suggested that researchers in this field use RR or PLSR methods as alternatives to classical methods if there is multicollinearity among the explanatory variables that are related to the energy demand.

CONFLICT OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Ceylan, H and Öztürk, H. K. 2004. Estimating energy demand of Turkey based on economic indicators using genetic algorithm approach. Energy Conversion and Management 45: 2525-2537.

[2] Sözen, A., Arcaklıoğlu, E., Özalp, M. and Çağlar, N. 2005. Forecasting based on neural network approach of solar potential in Turkey. Renewable Energy 30: 1075-1090.

[3] Kıran, M. S., Özceylan, E., Gündüz M. and Paksoy, T. 2012. A novel hybrid approach based on Particle Swarm Optimization and Ant Colony Algorithm to forecast energy demand of Turkey. Energy Conversion and Management 53: 75-83.

[4] Ünler, A. 2008. Improvement of energy demand forecast using swarm intelligence: The case of Turkey with projections to 2025. Energy Policy 36: 1937-1944.

[5] Toksarı, M. D. 2007. Ant colony optimization approach to estimate energy demand in Turkey. Energy Policy 35: 3984-3990.

[6] Öztürk, H. K. and Ceylan H. 2005. Forecasting total and industrial sector electricity demand based on genetic algorithm approach: Turkey case study. International Journal of Energy Research 29: 829-840.

[7] Ebohan, O. J. 1996. Energy, economic growth and casuality in developing countries: a case study of Tanzania and Nigeria. Energy Policy 24: 447-453.

[8] Dinçer, I. and Dost, S. 1997. Energy and GDP. International Journal of Energy Research 21(2): 153-167.

[9] Ediger, V. Ş. and Tatlıdil, H. 2002. Forecasting the primary energy demand in Turkey and analysis of cyclic patterns. Energy Conversion and Management 43: 473-487.

[10] Aras, H. and Aras, N. 2004. Forecasting residential natural gas demand. Energy Sources 26: 463-472.

[11] Ediger, V. Ş., Akar, S. and Uğurlu, B. 2006. Forecasting production of fossil fuel sources in Turkey using a comparative regression and ARIMA model. Energy Policy 34: 3836-3846.

[12] Sözen, A., Arcaklıoğlu, E. and Özkaymak, M. 2005. Modelling of the Turkey's net energy consumption using artificial neural network. International Journal of Computer Applications in Technology 22(2/3): 130-136.

[13] Murat, Y. S. and Ceylan, H. 2006. Use of artificial neural Networks for transport energy demand modelling. Energy Policy 34: 3165-3172.

[14] Canyurt, O. E., Ceylan, H., Öztürk, H. K. and Hepbaşlı, A. 2004. Energy demand estimation based on two-different genetic algorithm approaches. Energy Sources 26: 1313-1320.

[15] Haldenbilen, S. and Ceylan, H. 2005. Genetic algorithm approach to estimate transport energy demand in Turkey. Energy Policy 33: 89-98.

[16] Ceylan, H., Öztürk, H. K., Hepbaşlı, A. and Utlu, Z. 2005. Estimating energy and exergy production and consumption values using three different genetic algorithm approaches. Part 2: Application and scenarios. Energy Sources 27: 629-639.

[17] Öztürk, H. K., Canyurt, O. E., Hepbaşlı, A. and Utlu, Z. 2004. Residential-commercial energy input estimation based on genetic algorithm

approaches: an application of Turkey. Energy and Building 36(2): 175-183.

[18] Yumurtacı, Z. and Asmaz, E. 2004. Electric energy demand of Turkey for the year 2050. Energy Sources 26: 1157-1164.

[19] Abdel-Aal, R. E., Al-Garni, A. Z. and Al-Nassar, Y. N. 1997. Modeling and forecasting monthly electric energy consumption in Eastern Saudi Arabia using abductive Networks. Energy 22: 911-921.

[20] Al-Garni, A. Z., Zubair, S. M. and Nizami, J. S. 1994. A regression model for electric-energy consumption forecasting in Eastern Saudi Arabia. Energy 19: 1043-1049.

[21] Jyoti, P., Pallav, P. and Pallavi, M. 2007. Demand projections of petroleum products and natural gas in India. Energy 32: 1825-1837.

[22] Garthwaite, P. H. 1994. An interpretation of Partial Least Squares. Journal of the American Statistical Association 89: 122-127.

[23] Bulut, Y. M. 2011. Comparison of Partial Least Squares Regression and its Alternative Methods When Multicollinearity Exists. M. Sc. Thesis, Eskisehir Osmangazi University, Institute of Natural Science, 79 p. (Unpublished).

[24] Gunst, R. F. and Mason, R. L. 1980. Regression Analysis and its Application: A Data-oriented Approach. New York: Marcel Dekker.

[25] Hoerl, A. E. and Kennard, R. W. 1970. Ridge regression: Biased Estimation for Non-orthogonal Problems. Technometrics 55-67.

[26] Abdi, H. 2010. Partial least square regression, projection on latent structure regression, PLS-Regression. Wiley Interdisciplinary Reviews: Computational Statistics 2: 97-106.

[27] Wold, H. 1975. Soft Modelling by Latent variables: the nonlinear iterative partial least squares approach. In Gani, J., editor, Perspectives in Probability and Statistics, Papers in Honour of M.S. Bartlett, London, Academic Press.

[28] Glen, W. G., Sarker, M., Dunn III, W. J. and Scott, D, R. 1989. UN1PALS: Software for principal components analysis and partial least squares regression. Tetrahedron Comput. Methodol., 2(6): 377–396.

[29] Lindgren, F., Geladi, P. and Wold, S. 1993. The kernel algorithm for PLS., J. Chemometrics, 7: 45–59.

[30] De Jong, S. 1993. SIMPLS: An Alternative Approach to Partial Least Squares Regression. Chemometrics and Intelligent Laboratory Systems 18: 251-263.

[31] Phatak, A. and De Jong, S. 1997. The Geometry of Partial Least Squares. Journal of Chemometrics 11: 311-338.

[32] CBT, Central Bank of Turkey, http//.www.tcmb.gov.tr (in Turkish); 2013.

[33] WECTNC, World Energy Council Turkish National Committee, Energy Report, Ankara (in Turkish) 2011.

[34] WECTNC, World Energy Council Turkish National Committee, Energy Report, Ankara (in Turkish) 2012.

[35] NS, National Statistics http//.www.tuik.gov.tr (in Turkish); 2013.