Application of a hybrid model based on multiple linear regression -principle component analysis (MLR-PCA) for electricity export forecasting

Maryam Doroodi[1], Mohammad Reza Amin-Naseri[2], Bakhtiar Ostadi[3,*]

[1] Ph.D. Student, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran. Iran, M.doroodi@modares.ac.ir

[2] Professor, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran. Deceased July 2021

[3] Associate Professor, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran, P.O. Box: 14115-111, bostadi@modares.ac.ir, +98 912 579 5141, +98 21 82884385

Abstract:

International electricity trade as a strategic commodity plays a prominent role in the foreign trade market of countries. Electricity export forecasting leads to better production planning, supply security, blackouts reduction, and obligations fulfillment. This paper aimed to provide a model for electricity export forecasting. In this regard, electricity consumption in different consumer sectors, gas consumption, population, GDP, and electricity prices have been entered into the multiple regression model as predictor variables. Although $R^2 = .976$, $F=66.110$, and $SIG<.05$ indicate the model appropriateness, the high correlation between the predictor variables created collinearity. In other words, Tolerance, VIF (variance inflation factor), Eigenvalue and the Condition Index are less than .2, more than 10, close to zero, and more than 15 respectively. To solve this problem, two hybrid methods of Multiple Regression-First Difference Function and Multiple Regression-PCA have been used. In the first hybrid method ($R^2 = .553$) the Tolerance and VIF index still show the presence of collinearity. In the second hybrid method ($R^2 = .936$, $F=169.9$, $SIG<.05$) due to all the mentioned indicators, the collinearity has been completely resolved. So, the MLR-PCA method is the most appropriate model for electricity export forecasting. The data collected from Iran have been used to illustrate the model.

Keywords: Electricity export, Forecasting, Regression, Principle Component Analysis, MLR-PCA

1. Introduction

Due to differences in geographical and spatial conditions of different countries as well as climate differences in different seasons of the year, there are usually daily differences in peak times of electricity consumption between neighboring countries. In addition, the cost of electricity production is not the same in each country at different times of the year. Therefore, instead of investing heavily in power plants to meet peak demand, countries can use the capabilities and capacities of neighboring countries, especially those with low electricity production costs, to meet their peak electricity demand. Therefore, by planning for proper trade between the countries of the region, the time and cost of blackouts and investment in electricity generation and meeting demand can be reduced [1].

Iran has the first rank of gas reserves, the fourth rank of oil reserves, and also has various sources of renewable energy. Utilization of these resources in its development seems necessary. Energy exports are an integral part of national policy to boost economic income, which also provides opportunities for progress in other fields. It has long played a significant role in international energy markets, and with its significant natural reserves, this is expected to continue in the future [2]. While the energy sector in Iran is sanctioned and trade between the country and the West is very limited, the electricity sector, which is not easily sanctioned, may become an alternative energy source and a source of regional influence for Iran [3]. Exports of electricity generated from primary and secondary energy sources have increased in recent years in Iran and will attract more attention in the future [2]. But it should be borne in mind that the development of the electricity trade sector is impossible without identifying the factors and variables that affect it. It can be said that determining the factors affecting the electricity trade and determining the extent of their impact is the first step in the development of the electricity trade. On the one hand, electricity exports have an economic justification and can be a factor of growth and development, and on the other hand, it affects fossil fuel sources, the environment, how to respond to domestic electricity demand, etc. Therefore, forecasting the amount of electricity exported can lead to more accurate planning in this area. Export forecasting is important in production and supply management and helps plan to achieve the strategic goals of Iran's electricity sector, including becoming a regional electricity hub, increasing supply security, and achieving sustainable development in the electricity sector.

Power consumption modeling is a topic of interest for many researchers. The related literature can be divided into two parts: Univariate modeling and Multivariate modeling [4],[5]. Univariate modeling is based on time-series data such as Grey models(GM) [6,7] and AutoRegressive Integrated Moving Average (ARIMA) [8]. Multivariate modeling focuses on the effect of different variables on power consumption. For example, artificial neural network (ANN) models [9,10], FUZZY [11], and Support vector regression (SVR) model [12,13]. A benefit of Univariate models is that there is no need to determine other influential variables [4]. Instead, the ability to consider two or more time series at the same time, which is needed in some real cases, is the advantage of multivariate models [14]. So, in such a situation, Multivariate models should be used.

Among the statistical approaches, regression techniques have attracted more attention because they are relatively simple to implement and require less computational power than other statistical methods such as genetic algorithms, neural networks, and support vector machines. In addition, they can predict satisfactorily [15]. Table 1 summarizes the studied articles on multivariate methods focusing on regression models that are considered in this article. Relatively simple performance, acceptable accuracy of prediction, and the need for less computing power than some other statistical methods are some advantages of regression techniques [15].

Mohamed and Bodger, 2005, used multiple regression to examine annual electricity consumption in New Zealand. For this purpose, the impact of economic and demographic factors on electricity consumption was considered. The authors included the variables in the model from 1965 to 1999 as predictor variables .The results showed that all variables significantly affect power consumption[16]. Bianco et al., 2009, used different linear regression models to predict electricity consumption in Italy. Their main purpose was to study the impact of GDP and electricity prices on household and non-household electricity consumption. The results showed that in Italian electricity consumption forecasting models; the price of electricity should not be included in the model as an explanatory variable. The results of this study were consistent with the existing official predictions[17]. In 2009, Choi et al. Forecast long-term electricity demand in Ontario, Canada. For this purpose, three autoregressive models, simple linear regression and multiple linear regression were compared. In this research, multiple regression has been introduced as the most accurate model [18]. In 2013, a study was conducted to predict the power consumption of Italy with the help of single and multiple regression models. In this study, GDP and population were entered into the model as independent variables and the results showed the high impact of these variables on electricity consumption. Household and non-household electricity consumption were modeled in this research with a reliability coefficient of .99 and .961 Respectively [19]. Custer et al. Studied the models used to predict electric load and compared them with several criteria. According to the results of this study, regression and multiple regression models are widely used in the field of electricity forecasting and are very efficient in long-term forecasting [20]. Dudic et al, 2020, used multiple regression to investigate the reason why Germany's electricity consumption did not grow in the last decade [21]. Another study in 2020, predicted the electricity consumption in India with regression. Results demonstrated that sector-wise Net Domestic Savings was the most effective variable on future electricity demand in the domestic sector [22]. A study was conducted in 2021 to predict electricity consumption in various consumer sectors in India. In this research, electricity consumption in different consumer sectors including industry, agriculture, trade, home, and rail lines was formulated using four regression models. Finally, the best model was determined according to R^2 and MAPE indices [23].

In multivariate models, sometimes there is a correlation between independent variables, to solve this problem methods like PCA are used. An advantage of this method is to reduce a large number of inter-correlated variables to uncorrelated components [24].

Given the extensive studies that have been predicted in the field of electricity; So far, one study has not specifically focused on forecasting electricity exports. However, few studies have been conducted to forecast the export of some energy carriers [25–29]. Different variables affect the export of electricity and it is important to study how these variables affect and how they relate to the amount of export, so the use of regression models to conduct such a study seems appropriate. In this research, the multiple regression model, regression-principal factor analysis, and regression-first difference function are compared to forecast the amount of electricity export.

The structure of this article is as follows: In the next Section, the research methodology is examined. Section 3 provides a numerical example and analysis of the results, and finally, the fourth section is devoted to the conclusions and summaries obtained.

2. Research Methodology

In this paper, to provide a model for forecasting Iran's electricity exports, the linear multiple regression method has been used. In this model, the amount of electricity exports is considered as the response variable. Since the consumption of any goods and services within a country will affect the amount of export and import of those goods or services; Electricity consumption in different consumer sectors, including household-public-commercial, industry, transportation, and agriculture, are considered predictor variables. On the other hand, electricity production in Iran is significantly dependent on natural gas, so the amount of natural gas consumption has entered the model as another independent variable. In addition, the population of a country as well as GDP will directly affect the consumption, export, and import of goods and services in each country, so these two variables in the forthcoming article are considered explanatory or independent variables. Because the price of goods and services is thought to be involved in consumption and, consequently, in exports, this variable has also entered the model. One of the problems that may arise in the use of regression is the collinearity due to the correlation between the independent variables. The collinearity makes the regression model unreliable [30] and affects the parameters estimated by the least-squares method. There are several ways to detect collinearity. The high correlation in the correlation matrix, high VIF (variance inflation factors) index, and condition number may indicate correlated variables. To solve the collinearity, several methods have been introduced in the literature, including the selection of larger sample sizes, deletion or conversion of variables, as well as PCA and ridge regression methods [31]. Two of these methods that are considered in this article are the first difference function and the principal factor analysis method. In the following, the methods used in this article are briefly mentioned.

2.1 Iran's electricity exports and imports situation

Figure 1 shows a graph of electricity exports to different parts of the world in 2017[32].

Iran exports electricity to Nakhchivan, Armenia, Azerbaijan, Turkmenistan, Armenia, Afghanistan, Pakistan, and Iraq, and imports electricity from Nakhchivan, Armenia, Azerbaijan, Turkmenistan, and Turkey. Figure 2 shows the trend of electricity exports and imports in Iran [32].

At present, Iran has electrical connections to countries with which it shares a land border. The regional electricity cooperation plan can also pave the way for the transfer of Iranian electricity to countries that are not in Iran's neighborhood. This is a precondition for connecting Iran's electricity to the European electricity grid. One of the priorities of the Ministry of Energy is to maintain and improve the country's position in playing the role of an energy bridge and becoming an electricity hub in the region [33]. Achieving these goals is not possible without careful planning to identify the factors affecting exports. Regression-based methods can be used to identify the variables affecting the export environment and to estimate the relationships between the variables and how they affect the amount of export. Multiple regression is used in this direction because several variables affect electricity exports.

2.2 Regression model

Regression is a simple statistical model that allows observing the relationship between variables. In the regression model, the response or dependent variable can be defined with the help of variables called independent, predictor, or explanatory variables. Multiple regression is still widely useful and used for long-term predictions [20].

2.3 The first difference regression model

In this method, Δx is used instead of any x.

$$\Delta x = x_t - x_{t-1} \quad (1)$$

Then a regression model with modified data is formed.

2.4 Principle component analysis (PCA)

One of the proposed methods to eliminate collinearity in the regression model, which occurs as a result of the correlation between independent variables, is principal factor analysis. In this method, the initial independent variables are converted into new and independent components. New components are linear combinations of primary variables. In this method, because all variables are used in the formation of components; the information of the primary variables is provided with the help of components with the least losses. In other words, the information is not lost[34]. PCA is fully described in [34]. Before implementation, it must be ensured that the available data is suitable for this method. Is the number of data (sample size and relationship between variables) appropriate for component analysis? For this purpose, the KMO (Kaiser-Meyer-Olkin) index and Bartlett test are used. KMO index is in the range of zero to one, and the larger its value, the more appropriate the data for PCA, and otherwise (usually less than 6.) PCA is not suitable for the data under study. If the significance level of the Bartlett test is less than *5%*, PCA is appropriate.

2.4.1 Combined multiple regression and principal component analysis (MLR-PCA)

By determining the main and appropriate components in the PCA method, these components are considered as independent variables of multiple regression and the most appropriate regression can be generalized to the data and the MLR-PCA hybrid model is formed. In this step, to prevent correlation between independent variables, we must first calculate the mean difference of each data (Centered) for the independent variables (factors) and then form the regression model. Figure 3 shows the methodology of this research.

3. The Numerical example and analysis of results

In this section, a suitable model is presented to predict the amount of Iran's electricity exports, using the mentioned methodology.

3.1 Research data

In this study, data from Iran's energy balance sheets, published by the Ministry of Energy, have been used. Table 2 contains the variables under consideration. Data from 1993 to 2017 have been used in this study. In this research, IBM SPSS statistics version 22 and MINITAB 16 software were used.

One of the regression assumptions is the normality of the dependent variable distribution; the normality of the Y variable distribution was investigated using Kolmogorov-Smirnov and Shapirovik tests. According to Table 3, the *SIG* index in both tests is less than 5% and this indicates that this variable does not follow the normal distribution. Therefore, to solve this problem, all calculations of this research have been done with the natural logarithm of this variable.

$$Y_1 = LN(Y) \qquad (2)$$

3.2 Multiple regression implementation

In this section, a multiple regression model has been developed to predict the amount of electricity export using independent variables. According to Table 4, only the variable X5, which represents GDP, has a *p-value* less than *.05* and should be included in the model. In other words, according to the regression output for the available data, the amount of electricity exports in Iran is only related to the GDP variable and has shown a positive relationship with this variable. But can these results be trusted? This article is discussed in more detail below. In the following, this issue is addressed by further interpreting the outputs of the presented regression model.

According to Table 5, the coefficient of determination (*R2*) and also the adjusted coefficient of determination (*adj* R^2) show suitable conditions.

Table 6 shows model errors and residuals. In addition, due to the high F-statistic and the zero *P-value* represented by *sig*, the existence of a regression relationship is confirmed and it can be said that at least one independent variable has a non-zero coefficient.

Figure 4 shows the diagrams for the residues. As shown in the figure, in the normal probability plot, the residues are around and close to the *45*-degree line and no major deviations from the straight line are seen and it can be said that they have a normal distribution [31]. The versus fits diagram, residual against the predicted values, does not have a specific pattern and there is no indication that the model is abnormal, so the assumption of linearity and the assumption of constant variance of errors is confirmed[31].

But the main problem of this model can be seen in the two columns at the end of Table 4. The lower the tolerance index, the less information about the variables and the difficulty in the regression model. The variance inflation factor (VIF) is also the inverse of tolerance. The higher the variance, the greater the variance of the regression coefficient, making the model unsuitable for forecasting. In general, tolerances less than *.1* and VIF greater than 10 indicate the probability of collinearity [35]. In the present model, very low tolerance and very high variance inflation confirm the existence of collinearity as well as the inadequacy of the model. In addition, Table 7 shows the Eigenvalue and the Condition Index. When the eigenvalues are near zero, the correlation is high also small changes in the data values lead to large changes in the regression coefficients. The condition Index with a value greater than *15* also indicates the probability of collinearity between the independent variables, and a value greater than 30 indicates crucial trouble in regression [36]. According to this table, the eigenvalues and the status index of this model do not show suitable conditions and there is a correlation between the independent variables.

In this study, to solve the collinearity, the first difference regression and the regression-principal component analysis method have been used and compared.

3.3 First difference regression implementation

After creating the first difference regression, according to Table 8, only the variable $\Delta x1$, i.e. the first difference of the variable of electricity consumption in the Commercial- public- household sector, can enter the model. In the following, other results related to this model are interpreted. As shown in Table 9, the adjusted R^2 is *.357* and is not a desirable value.

According to Table 10, which is related to the analysis of variance of the proposed model, since *sig* is less than *5%*, it can be claimed that there is a regression relationship and at least one independent variable has a non-zero coefficient.

Unlike the multiple regression model (initial model) which had a very low tolerance index and a very high VIF, according to Table 8, in this model, more favorable conditions are observed than in the previous model. However, the tolerance index is still less than *.1* and the VIF is more than 10, and it can be concluded that there is a possibility of collinearity in the model. According to the Eigenvalue and the Condition Index in Table 11, the current model is more acceptable than the previous model. There is also no condition index with a value greater than *15* that indicates the probability of collinearity between the independent variables. According to these two indicators, it seems that the method of the first difference regression has solved the existing collinearity. However, the coefficient of determination of the model (R^2) is not favorable.

3.4 Implementation of the combined multiple linear regression-principal component analysis hybrid method (MLR-PCA)

To form an MLR-PCA hybrid model, the suitability of the data for PCA implementation must first be examined. According to Table 12, the KMO index is .818 and the significance level of the Bartlett test is less than *5%*, so the research data are suitable for PCA.

In this step, the Eigenvalue vectors must be selected. According to Table 13, the first component explains 93.234% of the total variance in the data, in addition, according to Figure 5, which shows the changes in eigenvalues against the number of components, for components larger than 1, a line with a slope close to zero is created. Therefore, at this stage, only the first component is selected. According to Table 14, the first factor is as follows.

$$PC = (.990 * X_1) + (.992 * X_2) + (.883 * X_3) + (.988 * X_4) + (.991 * X_5) + (.955 * X_6) + (.955 * X_7) \quad (3)$$

At this point, the PC enters the regression model as an independent variable and the appropriate regression model is the Grade 2 polynomial model. In polynomial regression, to prevent correlation between independent variables, the mean difference of each data for the independent variables is calculated (Centered) and then entered into the model as an independent variable.

$$PC_1 = PC - Mean(PC) \quad (4)$$

According to Table 15, the calculated regression model is as equation 4.

$$Y_1 = -1.468 + (1.990 * (10 \wedge -7)) * PC_1 + (-7.376 * (10 \wedge -15)) * (PC_1) \wedge 2 \quad (5)$$

Other outputs will be examined below. According to Table 16, the proposed model has an adjusted coefficient of determination of .936 which is a desirable amount.

According to the results of the analysis of variance in Table 17, the regression relationship is significant and *p_value* or *SIG* is less than *5%*. F statistics show that there is at least one independent variable with a non-zero coefficient in the model.

According to Figure 6, the residuals of the MLR-PCA model have a normal distribution and the assumption of linearity and constant variance of errors is also confirmed.

According to Table 15, in the combined MLR-PCA model, the tolerance indices and VIF show good condition and there is no tolerance index less than *.1* and VIF greater than 10. Therefore, it can be concluded that the collinearity in the model has been eliminated by combining the regression method with principal component analysis. In addition, according to Table 18, the Eigenvalue and the Condition Index confirm the absence of collinearity in the hybrid model.

4. Conclusions and future studies

In this research, a model for electricity export forecasting is presented. For this purpose, the multiple regression method and two other hybrid methods including the multiple regression-first difference function and the multiple regression-PCA has been investigated.

To compare the models, indicators such as coefficient of determination, adjusted coefficient of determination, the sum of squares of residual error, statistical distribution (normality) of residuals, and the presence or absence of collinearity have been used. The collinearity has been checked according to indicators such as Tolerance, VIF, Eigenvalue, and the Condition Index. According to Table 19, the initial multiple regression model has the highest coefficient of determination and the lowest error rate. However, the existence of collinearity in this model, which is due to the correlation between independent variables, cannot be ignored and leads to an inappropriate model. In this model, Tolerance is less than *.2*, VIF is more than 10, Eigenvalue is close to zero, and the condition index is more than *15* .All these values show the presence of collinearity. To eliminate the existing collinearity, the first difference regression and regression-principal component analysis method have been implemented. the first difference regression model has succeeded to some extent in eliminating the collinearity. The Tolerance and VIF still prove the presence of collinearity in the model. In addition due to the not very suitable coefficient of determination ($R^2=.553$), This model is not introduced as a desirable model. The combined MLR-PCA method, unlike the two previous cases, has completely resolved the collinearity and has favorable conditions in terms of error rate and coefficient of determination. Considering that the coefficient of determination of the first model is higher than the hybrid MLR-PCA model and has shown less error ($SSE=1.340$) but according to the collinearity results the hybrid MLR-PCA model is preferable. So the hybrid MLR-PCA model is introduced as the most appropriate model in this research. Contrary to the study of WU et al., Who concluded that the use of a combination of PCA and multiple regression is not very useful and efficient [37]. The combined model presented in the present study can be introduced as the best model of this research to predict the amount of electricity export according to the desired independent variables. The results of this study are consistent with the studies of some other researchers [38,39].

Electricity export forecasting can be effective for better decision-making and planning in line with electricity production and distribution. Forecasting will lead to the fulfillment of domestic and foreign commitments and will bring increased satisfaction. Managers and decision-makers of the electricity industry, for better management, decision-making, and planning, can increase their insight into the future trend of exports by using forecasting models. They can also predict how this trend will change due to the change of the effective variables and be able to control the future situation.

Considering the claim that the ridge regression method is more common than PCA in collinearity elimination[30] a comparison of these two methods is suggested in future studies.

Nomenclature

Phrase	Abbreviation
Multiple linear regression	MLR
Principle component analysis	PCA
Grey models	GM
AutoRegressive Integrated Moving Average	ARIMA
Support vector regression	AVR
Suppotr vector machine	AVM
variance inflation factors	VIF
Kaiser-Meyer-Olkin	KMO

References

1. Shakouri, H. G., Eghlimi, M., and Manzoor, D., “Economically optimized electricity trade modeling : Iran – Turkey case”, *Energy Policy*, **37**(2), pp. 472–483 (2009).
2. Alipour, M., Hafezi, R., Ervural, B et al., “Long-term policy evaluation: Application of a new robust decision framework for Iran’s energy exports security”, *Energy*, **157**(July 2015), pp. 914–931 (2018).
3. Mirsaeedi-glossner, S., “Electricity Exports as a Loophole to Sanctions”, *Sci. Dipl.*, **2**(3), pp. 2–11 (2013).
4. Akdi, Y., Gölveren, E., and Okkaoğlu, Y., “Daily electrical energy consumption: Periodicity, harmonic regression method and forecasting”, *Energy*, **191**(jan 15), pp. 116509–116524 (2020).
5. Doroodi, M. and Mokhtar, A., “Comparison of time series approaches for prediction of energy consumption focusing on greenhouse gases emission in Iran”, *Energy Sect. Manag.*, (mar 6) (2019).
6. Ding, S., Hipel, K. W., and Dang, Y. guo, “Forecasting China’s electricity consumption using a new grey prediction model”, *Energy*, **149**(app 15), pp. 314–328 (2018).
7. Wang, Z. X., Li, Q., and Pei, L. L., “A seasonal GM(1,1) model for forecasting the electricity consumption of the primary economic sectors”, *Energy*, **154**(jul 1), pp. 522–534 (2018).
8. Cabral, J. de A., Legey, L. F. L., and Freitas Cabral, M. V. de, “Electricity consumption forecasting in Brazil: A spatial econometrics approach”, *Energy*, **126**(May 1), pp. 124–131 (2017).

9. Rahman, A., Srikumar, V., and Smith, A. D., "Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks", *Appl. Energy*, **212**(FEB 15), pp. 372–385 (2018).
10. Kim, M. K., Kim, Y. S., and Srebric, J., "Predictions of electricity consumption in a campus building using occupant rates and weather elements with sensitivity analysis: Artificial neural network vs. linear regression", *Sustain. Cities Soc.*, **62**(NOV 1), p. 102385 (2020).
11. Kucukali, S. and Baris, K., "Turkey's short-term gross annual electricity demand forecast by fuzzy logic approach", *Energy Policy*, **38**(5), pp. 2438–2445 (2010).
12. Shen, M., Lu, Y., Wei, K. H et al., "Prediction of household electricity consumption and effectiveness of concerted intervention strategies based on occupant behaviour and personality traits", *Renew. Sustain. Energy Rev.*, **127**(JUL 1), p. 109839 (2020).
13. Kavaklioglu, K., "Modeling and prediction of Turkey's electricity consumption using Support Vector Regression", *Appl. Energy*, **88**(1), pp. 368–375 (2011).
14. Succetti, F., Rosato, A., Araneo, R et al., "Deep Neural Networks for Multivariate Prediction of Photovoltaic Power Time Series FEDERICO", *IEEE Access*, **8**(NOV 20), pp. 211490–211505 (2020).
15. Fumo, N. and Rafe Biswas, M. A., "Regression analysis for prediction of residential energy consumption", *Renew. Sustain. Energy Rev.*, **47**(JUL 1), pp. 332–343 (2015).
16. Mohamed, Z. and Bodger, P., "Forecasting electricity consumption in New Zealand using economic and demographic variables", *Energy*, **30**(10), pp. 1833–1843 (2005).
17. Bianco, V., Manca, O., and Nardini, S., "Electricity consumption forecasting in Italy using linear regression models", *Energy*, **34**(9), pp. 1413–1421 (2009).
18. Chui, F., Elkamel, A., Surit, R et al., "Long-term electricity demand forecasting for power system planning using economic, demographic and climatic variables", *Eur. J. Ind. Eng.*, **3**(3), pp. 277–304 (2009).
19. Bianco, V., Manca, O., and Nardini, S., "Linear regression models to forecast electricity consumption in Italy", *Energy Sources, Part B Econ. Plan. Policy*, **8**(1), pp. 86–93 (2013).
20. Kuster, C., Rezgui, Y., and Mourshed, M., "Electrical load forecasting models: A critical systematic review", *Sustain. Cities Soc.*, **35**(NOV 1), pp. 257–270 (2017).
21. Dudic, B., Smolen, J., Kovac, P et al., "Electricity Usage Efficiency and Electricity Demand Modeling in the Case of Germany and the UK", *Appl. Sci.*, **10**(7), p. 2291 (2020).
22. Wani, T. A. and Shiraz, M., "Electricity Demand forecasting using regression techniques", *Adv. Energy Built Environ.*, **36**(JUNE 30), pp. 111–121 (2020).
23. Rekhade, R. and Sakhare, D. K., "Forecasting sector-wise electricity consumption for India using various regression models", *Curr. Sci.*, **121**(3), pp. 365–371 (2021).
24. Zhang, J., Yang, X. Y., Shen, F., Li et al., "Principal component analysis of electricity consumption factors in China", *Energy Procedia*, pp. 1913–1918 (2012).
25. Atsalakis, G., Frantzis, D., and Zopounidis, C., "Energy's exports forecasting by a neuro-fuzzy controller", *Energy Syst.*, **6**(2), pp. 249–267 (2015).
26. Windarto, A. P., Dewi, L. S., and Hartama, D., "Implementation of Artificial Intelligence in Predicting the Value of Indonesian Oil and Gas Exports With BP Algorithm", *Ind. J. Recent trends Eng. Res.*, pp. 1–12 (2017).
27. Febriadi, B., Zamzami, Z., Yunefri, Y et al., "Bipolar function in backpropagation algorithm in predicting Indonesia's coal exports by major destination countries", *IOP Conf. Ser. Mater. Sci. Eng.*, p. 012087 (2018).
28. Wanto, A., Herawan Hayadi, B., Subekti et al., "Forecasting the Export and Import Volume of Crude Oil, Oil Products and Gas Using ANN", *J. Phys. Conf. Ser.*, **1255**(1), p. 012016 (2019).
29. Febriyati, N. A. and Gs, A. D., "Analysis of Backpropagation Algorithm Using the Traingda Function for Export

Prediction in East Java", *internatioal J. Inf. Syst. Technol.*, **4**(36), pp. 550–558 (2020).
30. Morgül Tumbaz, M. N. and İpek, M., "Energy Demand Forecasting: Avoiding Multi-collinearity", *Arab. J. Sci. Eng.*, **46**(2), pp. 1663–1675 (2021).
31. C.Montgomery, D., A.Peck, E., and Vining, G. G., *Introduction Linear Regression Edition Fifth Edition*, 4th Edn., Wiley series in probability and statistics (2012).
32. ministry of energy, I., *Balance Sheet* (2018).
33. Ministry of energy, I., *Energy Balance Sheet* (2017).
34. Brooks, C., *Introductory Econometrics for Finance* (2008).
35. Landau, S. and S.Everitt, B., *A Handbook of Statistical Analysis Using SPSS* (2004).
36. Kim Jong Hae, "Multicollinearity and misleading statistical results", *Korean J. Anesthesiol.*, **72**(6), pp. 558–569 (2019).
37. Wu, J. X., Reddy, T. A. A., and Claridge, D. E. E., "Statistical Modeling of Daily Energy Consumption in Commercial Buildings Using Multiple Regression and Principal Component Analysis", *8th Symp. Improv. Build. Syst. Hot Humid Clim.*, pp. 155–164 (1992).
38. Reddy, T. A. and Claridge, D. E., "Using synthetic data to evaluate multiple regression and principal component analyses for statistical modeling of daily building energy consumption", *Energy Build.*, **21**(1), pp. 35–44 (1994).
39. Ruch, D., Chen, L., Haberl, J. S et al., "A change-point principal component analysis (CP/PCA) method for predicting energy usage in commercial buildings: The PCA model", *J. Sol. Energy Eng. Trans. ASME*, **115**(2), pp. 77–84 (1993).
40. Egelioglu, F., Mohamad, A. A., and Guven, H., "Economic variables and electricity consumption in Northern Cyprus", *Energy*, **26**(4), pp. 355–362 (2001).
41. Tsekouras, G. J., Dialynas, E. N., Hatziargyriou, N. D et al., "A non-linear multivariable regression model for midterm energy forecasting of power systems", *Electr. Power Syst. Res.*, **77**(12), pp. 1560–1568 (2007).
42. Samuel, I. a, F, F. C., A, A. A., and Awelewa, A. a, "Medium-Term Load Forecasting Of Covenant University Using The Regression Analysis Methods", *J. Energy Technol. Policy*, **4**(4), pp. 10–17 (2014).
43. Koen, R. and Holloway, J., "Application of multiple regression analysis to forecasting South Africa's electricity demand", *J. Energy South. Africa*, **25**(4), pp. 48–58 (2014).
44. Liang, Y., Niu, D., Cao, Y et al., "Analysis and modeling for China's electricity demand forecasting using a hybrid method based on multiple regression and extreme learning machine: A view from carbon emission", *Energies*, **9**(11), p. 941 (2016).
45. Hameed, L., Ali Khan, A., and Author, C., "POPULATION GROWTH AND INCREASE IN DOMESTIC ELECTRICITY CONSUMPTION IN PAKISTAN: A CASE STUDY OF BAHAWALPUR CITY", *Explor. Islam. J. Soc. Sci.*, **2**(1), pp. 2411–132 (2016).
46. Usman, O. Y., Abdullah, M. K., and Mohamme, A. N., "Estimating Electricity Consumption in the Commercial Sector of Nigeria's Economy d", *Int. J. Recent Technol. Eng.*, **7**(6s5), pp. 1594–1600 (2019).
47. González-Briones, A., Hernandez, G., Corchado, J. M et al., "Machine Learning Models for Electricity Consumption Forecasting: A Review", *2nd Int. Conf. Comput. Appl. Inf. Secur. ICCAIS 2019*, pp. 1–6 (2019).
48. Ojonimi, Y. U., Kamil, A., and Akmal, N. M., "Modeling and Forecast of Residential Electricity Consumption in Nigeria", *Adv. Res. Energy Eng.*, **1**(2), pp. 1–7 (2019).
49. Ertuğrul, Ö. F., Tekin, H., and Tekin, R., "A novel regression method in forecasting short-term grid electricity load in buildings that were connected to the smart grid", *Electr. Eng.*, **103**(1), pp. 717–728 (2021).

Figure captions:

Figure 1. The amount of electricity exported in different parts of the world.

Figure 2. Iran’s import-export trend.

Figure 3. Research methodology.

Figure 4. Residual plots of MLR model.

Figure 1.Eigenvalue versus the component number.

Figure 6. Residual plots of PCA-MLR model.

Table captions:

Table 1. Summary of the reviewed papers

Table 2. Research variables.

Table 3. Response variable normality investigation.

Table 4. Regression coefficients.

Table 2. Determination coefficients.

Table 3. ANOVA table of MLR model.

Table 4. Collinearity diagnostics table of MLR model.

Table 5. Regression coefficients.

Table 6. Determination coefficient of the first differences regression model.

Table 7. ANOVA table of first differences regression model

Table 8. Collinearity diagnostics table of first differences regression model

Table 9. Bartlett's test and KMO measure.

Table 10. Total variance explained.

Table 11. Component matrix.

Table 12. The regression coefficient of the MLR-PCA model.

Table 13. Determination coefficient of PCA-MLR model.

Table 14. ANOVA table of PCA-MLR model.

Table 15. Collinearity diagnostics of PCA-MLR model.

Table 19. Comparison of MLR, first differences model, and PCA-MLR model

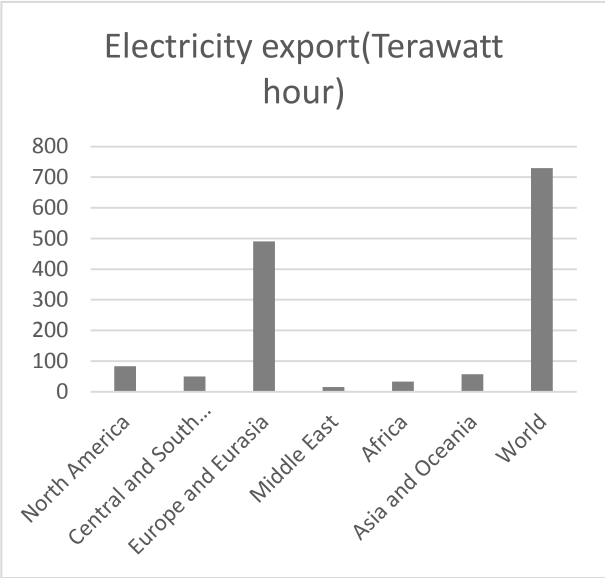

Figure2 .The amount of electricity exported in different parts of the world.

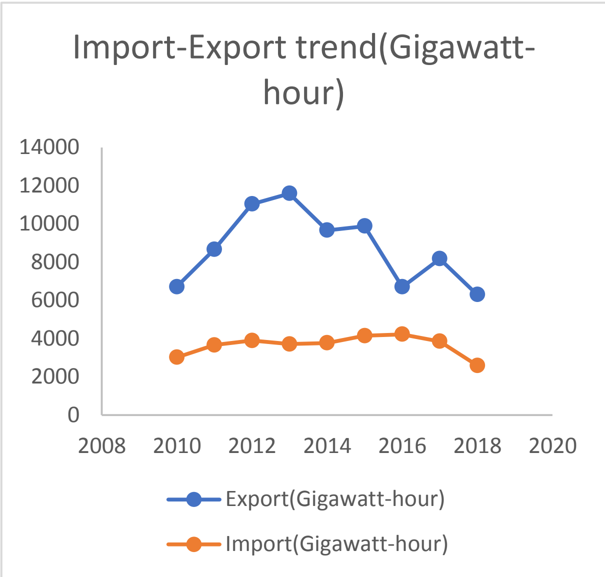

Figure3 .Iran's import-export trend.

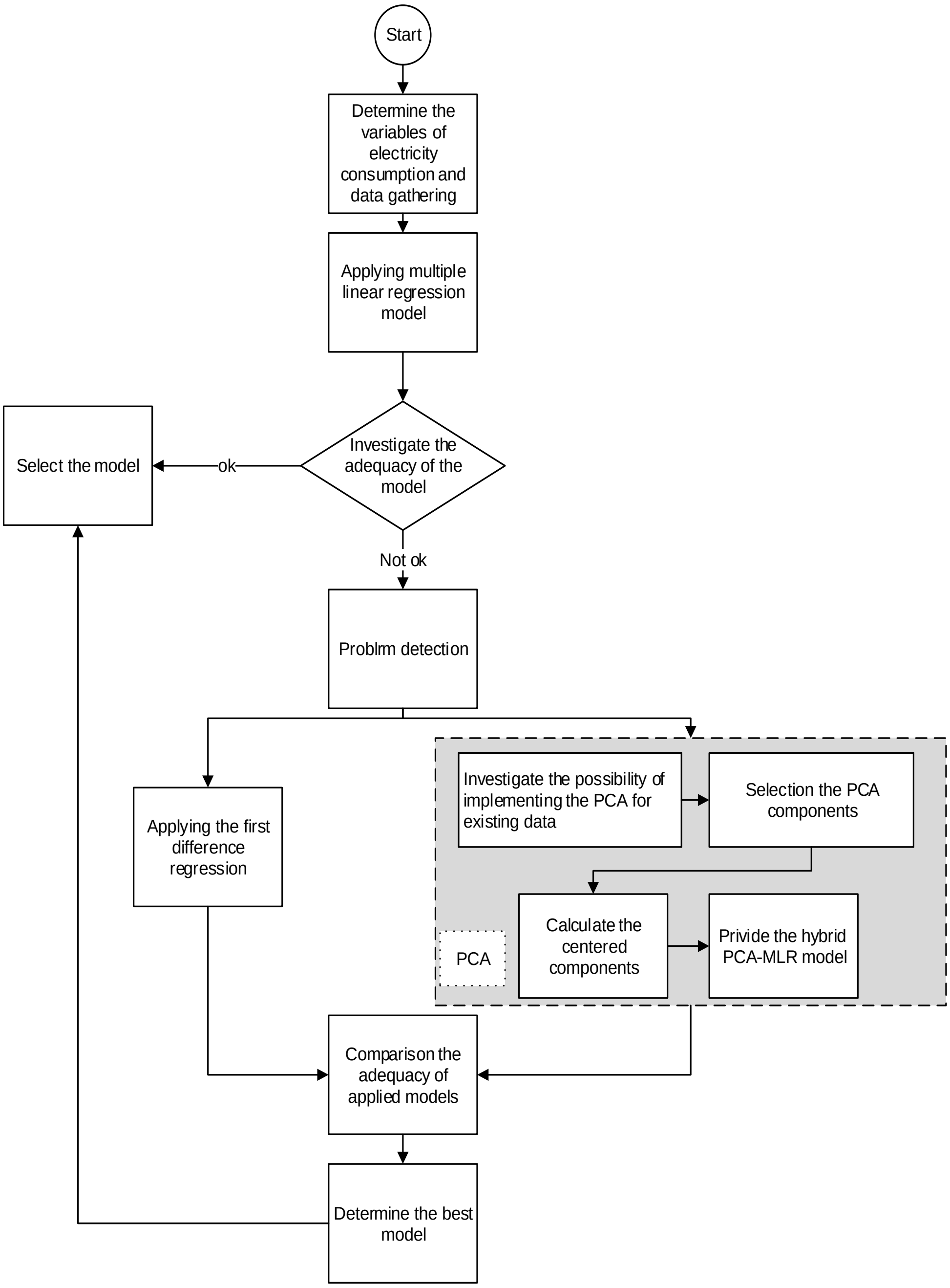

Figure 4.Research methodology.

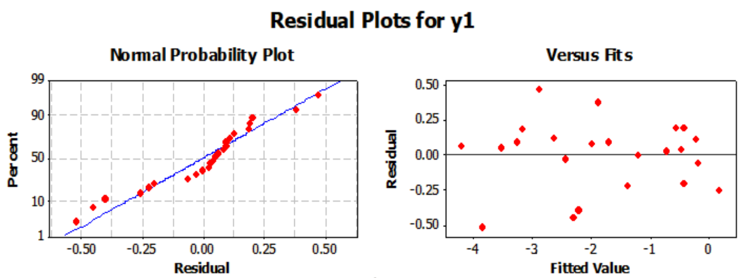

Figure5.Residual plots of MLR model.

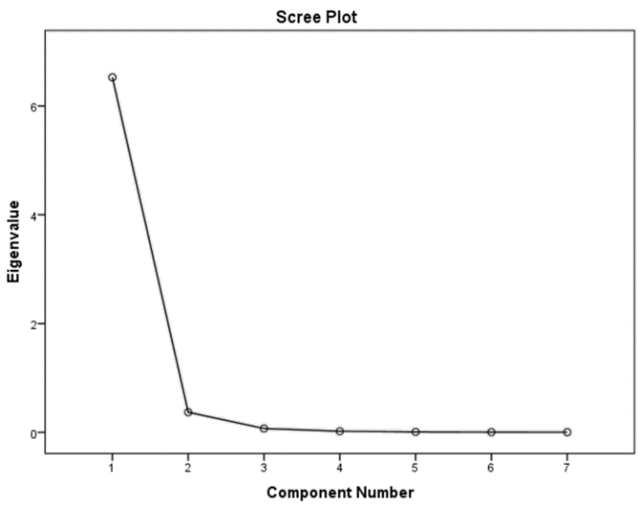

Figure 6.Eigenvalue versus the component number.

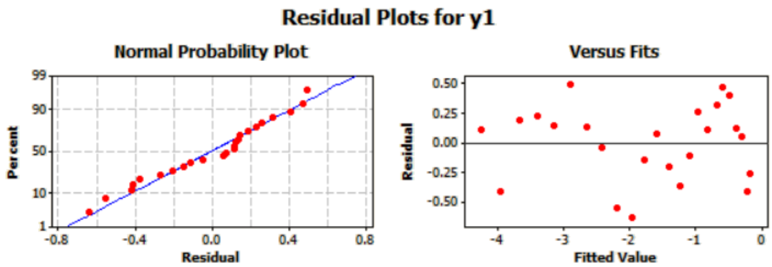

Figure 6. Residual plots of PCA-MLR model.

Table 16. Summary of the reviewed papers

reference	purpose	Method	Predictor variables	Main results
[40]	Examine The effect of economic factors on electricity consumption in Cyprus	multiple regression analyses	electricity price number of tourists number of customers	The mentioned variables as independent variables have a high ability to predict the amount of annual electricity consumption.
[16]	Forecasting electricity consumption in New Zealand	multiple linear regression	GDP Electricity average price Population	electricity consumption correlated effectively with all variables
[41]	midterm energy forecasting of the Greek power system	non-linear multivariable regression method	Annual energy of previous year Number of hot also cold days Gross national product Indices of oil and coal yields, Essential metals production,	The presented method provides satisfactorily results and can be used for mid-term load also energy predicting

reference	purpose	Method	Predictor variables	Main results
			Ultimate metal yield production, Paper- paper yields, Chemical yields, food and beverages, Durable consumption products, Non-durable consumption products Customers number	
[18]	Long-term Ontario's electricity demand forecasting	Autoregressive simple and multiple linear regression	economic variables demographic variables climatic variables	multiple linear regression was the most accurate model for depicting Ontario's electricity demands
[17]	Investigating the effect of economic and demographic variables on the annual electricity consumption in Italy	Different regression models	GDP GDP per capita population	Proposed regression models are compliant with the official projections, with 1% -11% deviations for the best and worst-case respectively.
[24]	Investigate the effect of 10 economic factors on electricity consumption	PCA- regression	gross domestic product, Earnings, industrial output value, product exports, added services industry value, household numbers, village population, price index product imports and efficiency of electricity	Unlike other variables, enhancement in product imports also the efficiency of electricity decreases electricity consumption.
[19]	Forecast Electricity Consumption in Italy	Multiple and single regression models	GDP GDP per capita population	with ±5% error regression models are agreeing with the official data
[42]	Medium-Term Load Forecasting Of a Nigerian educational institution	Three models based on the linear, compound-growth, and cubic methods of regression analysis	the number of months	Du to MAPE and RMSE, the linear regression model was the most appropriate.
[43]	Predicting South Africa's electricity demand	multiple regression models	GDP Population Final Consumption Expenditure by Households Rail freight Platinum and Coal production volume index	the presented methodology can be used to planning into a corresponding electricity demand
[44]	Forecasting the electricity demand in China	Combined model of multiple regression and Extreme Learning Machine	GDP Population Co_2 emission per GDP Energy Consumption Per GDP	The introduced method outperformed some monomial predicting methods, especially in improving instability.
[45]	Investigate the effect of population growth on electricity consumption in Bahawalpur City of the Punjab province of Pakistan	Regression model	Population Monthly income	Population growth is leading to a significant increase in the use of electrical appliances and causes a lot of changes in household electricity consumption
[46]	Identifying variables influencing the amount	Stepwise multiple linear regression	Temperature Rainfall,	rainfall, total electricity generated, total primary energy and

reference	purpose	Method	Predictor variables	Main results
	of electricity consumption in the commercial sector of Nigeria		Total electricity delivered Total primary energy Relative humidity	population indicate a huge influence on commercial electricity consumption
[47]	Study machine learning to electricity use prediction in a shoe market.	Linear Regression; Support Vector Regression; K-nearest Neighbours; Random Forest; Decision Tree	date, day of the week, day of the year, week, weekend, Previous day electricity consumption, electricity consumption	Linear Regression also Supports Vector Regression was introduced as the best model.
[48]	Residential Electricity Consumption forecast in Nigeria	Multiple linear regression analysis	Gross domestic product per capita Electricity consumption per capita Electricity access christened Connectivity rate Percentage of urban access to electricity Household size Temperature Electricity price	Gross domestic product per capita, Electricity consumption per capita, Electricity access christened Connectivity rate and Household size was determined as the most significant variable. The annual growth rate of electricity was forecasted.
[49]	buildings short-term electricity load predicting.	A novel regression method, the recurrent linear regression method (R-LR)	Month day week Month Day Weekdays/weekend Hour the energy of Solar panel Energy production and consumption	Demonstrates good results in system dynamic modeling and desirable results in predicting with time-series data.
[23]	Predicting electricity consumption in different customer sectors in India	Linear regression, logarithmic, power and exponential regression models	Time	The exponential regression model has the highest R^2 and lowest MAPE with the best accuracy.

Table 17.Research variables.

Variable		Mean	Standard deviation
electricity exports (Million tons oil equivalent)	Y	0.328950765	0.308290057
Commercial- public- household electricity (Million tons oil equivalent)	X1	6.178274409	2.235184077
Industry electricity consumption (Million tons oil equivalent)	X2	4.034256758	1.602126691
Agriculture electricity consumption (Million tons oil equivalent)	X3	2.097351123	2.167235775
Gas consumption (Million tons oil equivalent)	X4	56.54774693	29.82561859
Population	X5	71133027.13	5958457.528
Gross domestic product(GDP) (Billion Rials based on base year prices in 2004)	X6	1657403.226	412619.9802
Electricity price(Rials)	X7	234.28125	205.3743444

Table 18.Response variable normality investigation.

	Kolmogorov-Smirnov[a]			**Shapiro-Wilk**		
	Statistic	*df*	*Sig.*	*Statistic*	*df*	*Sig.*
Y	*.216*	*24*	*.005*	.846	*24*	*.002*

Table 19.Regression coefficients.

Model		**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**	**Collinearity Statistics**	
		B	***Std. Error***	***Beta***			***Tolerance***	***VIF***
1	(Constant)	*-27.567*	*7.037*		*-3.918*	*.001*		
	X1	*-.493*	*.295*	*-.852*	*-1.670*	*.114*	*.008*	*124.701*
	X2	*.206*	*.674*	*.256*	*.306*	*.763*	*.003*	*333.799*
	X3	*-.017*	*.133*	*-.028*	*-.125*	*.902*	*.042*	*23.775*
	X4	*.035*	*.024*	*.806*	*1.455*	*.165*	*.007*	*146.916*
	X5	*4.175E-7*	*.000*	*1.925*	*3.225*	*.005*	*.006*	*170.486*
	X6	*-1.514E-6*	*.000*	*-.484*	*-1.504*	*.152*	*.020*	*49.505*
	X7	*-.005*	*.002*	*-.750*	*-1.985*	*.065*	*.015*	*68.418*

Dependent Variable: y1

Table 20. Determination coefficients.

Model	***R***	***R Square***	***Adjusted R Square***
1	*.983*[a]	*.967*	*.952*

a. Predictors: (Constant), X7, X6, X3, X1, X4, X5, X2
b. Dependent Variable: y1

Table 21. ANOVA table of MLR model.

Model		***Sum of Squares***	***df***	***Mean Square***	***F***	***Sig.***
1	Regression	*38.748*	*7*	*5.535*	*66.110*	*.000*[b]
	Residual	*1.340*	*16*	*.084*		
	Total	*40.087*	*23*			

a. Dependent Variable: y1
b. Predictors: (Constant), X7, X6, X3, X1, X4, X5, X2

Table 22. Collinearity diagnostics table of MLR model.

Model	**Dimension**	**Eigenvalue**	**Condition Index**
1	*1*	*7.385*	*1.000*
	2	*.528*	*3.740*
	3	*.067*	*10.537*
	4	*.018*	*20.504*
	5	*.001*	*70.381*

	6	.001	88.558
	7	.000	126.408
	8	2.503E-5	543.133

Table 23. Regression coefficients.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-3.180	.557		-5.706	.000		
	ΔX1	2.174	.970	1.032	2.241	.040	.132	7.583
	ΔX2	2.591	2.264	.674	1.144	.269	.081	12.385
	ΔX3	.031	.348	.019	.088	.931	.579	1.726
	ΔX4	.103	.098	.330	1.049	.310	.282	3.542
	ΔX5	-1.219E-7	.000	-1.142	-1.700	.108	.062	16.135
	ΔX6	-6.390E-6	.000	-1.042	-1.602	.129	.066	15.138
	ΔX7	.010	.008	.353	1.343	.198	.406	2.464

a. Dependent Variable: y1

Table 24. Determination coefficient of the first differences regression model.

Model	R	R Square	Adjusted R Square
1	.743[a]	.553	.357

a. Predictors: (Constant), ΔX7, ΔX5, ΔX3, ΔX4, ΔX1, ΔX2, ΔX6

b. Dependent Variable: y1

Table 25. ANOVA table of first differences regression model

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22.153	7	3.165	2.823	.041[b]
	Residual	17.935	16	1.121		
	Total	40.087	23			

a. Dependent Variable: y1

b. Predictors: (Constant), ΔX7, ΔX5, DX3, ΔX4, ΔX1, ΔX2, ΔX6

Table 26. Collinearity diagnostics table of first differences regression model

Model	Dimension	Eigenvalue	Condition Index
1	1	4.967	1.000
	2	1.534	1.799
	3	.724	2.620
	4	.503	3.142
	5	.152	5.721

	6	.052	9.797
	7	.037	11.567
	8	.032	12.482

Table 27. Bartlett's test and KMO measure.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.818
Bartlett's Test of Sphericity	Approx. Chi-Square	439.637
	df	21
	Sig.	.000

Table 28. Total variance explained.

Component	Initial Eigenvalues	
	Total	% of Variance
1	6.526	93.234
2	.369	5.275
3	.070	.997
4	.020	.288
5	.008	.118
6	.004	.059
7	.002	.029

Table 29. Component matrix.

	Component
	1
X1	.990
X2	.992
X3	.883
X4	.988
X5	.991
X6	.955
X7	.955

Table 30. The regression coefficient of the MLR-PCA model.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-1.468	.100		-14.651	.000		
	PC1	1.990E-7	.000	.968	18.302	.000	.991	1.009
	PC1^2	-7.376E-15	.000	-.210	-3.965	.001	.991	1.009

a. Dependent Variable: y1

Table 31. Determination coefficient of PCA-MLR model.

Model	R	R Square	Adjusted R Square
1	.970[a]	.942	.936

a. Predictors: (Constant), PC1, PC1^2
b. Dependent Variable: y1

Table 32. ANOVA table of PCA-MLR model.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	*37.755*	*2*	*18.877*	*169.951*	*.000[b]*
	Residual	*2.333*	*21*	*.111*		
	Total	*40.087*	*23*			

a. Dependent Variable: y1
b. Predictors: (Constant), PC1, PC1^2

Table 33. Collinearity diagnostics of PCA-MLR model.

Model	Dimension	Eigenvalue	Condition Index
1	*1*	*1.735*	*1.000*
	2	*1.000*	*1.317*
	3	*.265*	*2.561*

Table 19. Comparison of MLR, first differences model, and PCA-MLR model

Model	R^2	adjusted R^2	SSE	Normality of residues	Collinearity	
					Tolerance & VIF	Eigenvalue & the Condition Index
Multiple regression	*.967*	*.952*	*1.340*	yes	Collinearity existence	Collinearity existence
First difference regression	*.553*	*.357*	*17.935*	yes	Collinearity existence	Collinearity absence
MLR-PCA	*.936*	*.942*	*2.333*	yes	Collinearity absence	Collinearity absence

Biographies

Maryam Doroodi is a PhD Candidate at Tarbiat Modares University. She received MScdegrees in Industrial Engineering from K.N. Toosi Universityof Technology. Her research interest includes prediction modeling and system dynamics in the field of Energy and GHG emission.

Mohammad Reza Amin-Naseriis was a FullProfessor of the Industrial Engineering Department at Tarbiat Modares University, Tehran, Iran. He received his PhD degree in Industrial and Systems Engineering from West Virginia University, WV, U.S.A., and MSdegree in Operations Research from Western MichiganUniversity, and a BS degree in Chemical and Petro-chemical Engineering from Amir-Kabir University of Technology (Tehran Polytechnic). Dr. Amin-Naseri's research and teaching interests include computational intelligence, artificial neural networks, and their applications to forecasting, meta-heuristics and their applications to various optimization problems.

Bakhtiar Ostadi is an Associate Professor of Industrial Engineering at Tarbiat Modares University. He received his PhD in Industrial Engineering from the School of Engineering at Tarbiat Modares University. In addition, He received his MSc from the University of Tehran in the Industrial Engineering and his BSc in Applied Mathematics from K.N. Toosi University of Technology. His main areas of research interests include “Data-Driven Risk Assessment”, “Pricing Strategies for Bidding”, “Resilience engineering”, “Quality Engineering and Management”, “Process

Management, Modeling and Simulation”, “Reliability Engineering”, and “Customer Value, Customer Experience, and Marketing”. He has published several articles in international academic journals and conferences including Int. J. Production Research, Computers & Industrial Engineering, Energy Policy, Energy, Journal of biomedical informatics, Applied Mathematical Modelling, Journal of Statistical Computation and Simulation, Journal of Risk and Reliability, Applied Mathematics and Computation, Scientia Iranica, Iranian Journal of Management Studies, Int. J. Productivity and Quality Management, International Journal of Business Innovation and Research, etc.