The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2022

Volume 21, Pages 11-17

IConTES 2022: International Conference on Technology, Engineering and Science

Evaluation of Electricity Consumption in Turkey via Cubic Regression Analysis and One Way Anova Test

Ezgi GULER
Bilecik Seyh Edebali University

Suheyla YEREL-KANDEMIR
Bilecik Seyh Edebali University

Abstract: In the globalizing world, the importance of energy is increasing day by day. Countries can make investment plans within the analysis of production and consumption data for electrical energy. In this study, sectoral electricity consumption data (State Agency, Industrial Enterprise, Business, Residence, Agricultural Irrigation, Street Lighting) in Turkey were analyzed with statistical techniques. Using the sectoral data between 2000 and 2020, the consumption data for the next 5 years was estimated via Cubic Regression Analysis. The success of the Cubic Regression models was evaluated with the success criteria of MAPE (Mean Absolute Percent Error), RMSE (Square Root Mean Errors), and R^2 (Coefficient of Determination). In addition, the One-Way ANOVA technique was used in the study to make sector comparisons for electricity consumption data. It was investigated whether there was a significant difference between the averages of the sector groups, the suitability of variance was examined and the Tamhane T2 test, one of the Pairwise Comparison Tests (Post Hoc), was used. As a result, in this study, electricity consumption data was analyzed with different statistical techniques, and information was produced for both current and future periods.

Keywords: Cubic regression analysis, Electricity consumption, Forecasting

Introduction

Environmental degradation, global warming, and the need for energy are among the biggest challenges for societies in today's world. Given the expansion of the Internet of Things, some estimates indicate that the energy supply in 2040 will exceed the expected energy demand for 2040 (Duranton et al., 2022; Gellert et al., 2022). In this context, energy modeling has become a subject of interest for researchers interested in energy production and consumption. Modeling, analyzing, and estimating electricity consumption, which is one of the important energy resources, is also of critical importance for national and international maintenance and planning activities (Grigoryan, 2021). Electricity consumption forecasting refers to the forecast of future electricity sales by collecting and analyzing historical consumption data (Duan et al., 2019). There are many studies in the literature that include electricity consumption estimation and statistical analysis. Some of these studies are summarized:
Saranj and Zolfaghari proposed a hybrid approach to estimating electricity consumption based on wavelet transfer. Forecasting models are based on autoregressive integrated moving average with adaptive WT (AWT)-long short-term memory (LSTM) and explanatory variable (ARIMAX)-generalized autoregressive conditional variable variance (GARCH) type models (Saranj and Zolfaghari, 2022). Tyagi and Singh compared prediction models such as support vector regression (SVR), gradient boost decision tree (GBDT), artificial neural network (ANN), random forest (RF), and extreme gradient boost (XGBoost) for electricity consumption estimation. They used the extreme gradient boosting method to predict future building electricity consumption (Tyagi and Singh, 2022). Soyler and Izgi measured the annual electrical energy use of 23 public hospitals with over 100 beds in Istanbul, and after determining the monthly peak loads, they created two new forecasting models using

regression techniques for maximum demand forecasting. A positive result was obtained from the linear regression technique, which is one of the basic regression techniques, and a new design factor was determined in the light of the determined data, and the maximum power requirement of the hospital was estimated (Soyler and Izgi, 2022). Nsangou et al. evaluated the performance of electricity consumption factors using quantitative regression, decision tree, and neural network models for prediction. They used a data sample from a household electricity consumption survey in Cameroon to apply these three models through a comparative analysis (Nsangou, et al.,2022).

In this study, electricity consumption in different sectors/areas in Turkey is analyzed statistically. Consumption data for the years 2000-2020 are obtained from TURKSTAT (URL-1). 5-year estimation data are obtained by Cubic Regression Analysis. One Way ANOVA analysis is conducted to determine the consumption relations between sector groups. The difference between groups was tested with the Tamhane T2 test.

Method

In this study, the electricity consumption in Turkey data for the next 5 years was estimated via Cubic Regression Analysis. The success of the Cubic Regression models was evaluated with the success criteria of MAPE, RMSE, and R^2. In addition, the One-Way ANOVA technique was used in the study to make sector comparisons for electricity consumption data. It was investigated whether there was a significant difference between the averages of the sector groups, the suitability of variance was examined and the Tamhane T2 test, one of the Pairwise Comparison Tests (Post Hoc), was used.

Cubic Regression Analysis

Polynomial Regression Analysis is essentially a special case of multivariate linear regression (Meyers et al., 2016). Only one independent variable (x) is considered in Polynomial Regression Analysis. The variable x is “period (day, month, year, etc.)” for time series. In this study, the variable x is taken as "year". For the polynomial regression model, when the degree of the independent variable x is “3”, the related model is called the “Cubic Regression Model”. The cubic regression model is given in Eq. (1) (Guler and Kandemir, 2022).

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \mathcal{E} \quad (1)$$

where y_i is the dependent variable, β_0 is the regression constant, β_1, β_2, β_3 are the regression coefficients, x_i is the independent variable, $\mathcal{E}$ is the error term.

Model Evaluation Criteria

MAPE (Mean Absolute Percent Error)

MAPE is an error metric that represents the percent error between actual and estimated values (Ozkaya et al., 2022). The MAPE value is high when the estimated values do not reflect or overlap the actual values. When the MAPE value is below 10%, the accuracy of the prediction model is stated as very high (Gonultas et al., 2020). The MAPE formula giving the absolute error percentage is given in Eq. (2).

$$MAPE = \frac{100}{N} \sum_{j=1}^{N} \left| \frac{y_j - \hat{Y}_j}{y_j} \right| \quad (2)$$

Where, y_j and $\hat{Y}_j$ are actual and predictive values, N is sample size or the number of values.

RMSE (Square Root Mean Errors)

The formula of RMSE in comparison to regression equations is given in Eq. (3) (Tatliyer, 2020).

$$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (y_j - \hat{Y}_j)^2} \quad (3)$$

Where, y_j and $\hat{Y}_j$ are actual and predictive values, N is sample size or the number of values.

R^2 (Coefficient of Determination)

The R^2 calculated in the regression analysis measures how well the relationship between the predictive values and actual values fits a linear curve. A high R^2 value is a desirable feature for the regression model (Yerel and Ersen, 2013). The formula for R^2 is given in Eq. (4) (Piepho, 2019).

$$R^2 = 1 - \frac{\sum_{j=1}^{N}(y_j - \hat{Y}_j)^2}{\sum_{j=1}^{N}(y_j - \bar{Y}_j)^2} \quad (4)$$

Where, y_j and $\hat{Y}_j$ are actual and predictive values, $\bar{Y}_j$ is the mean value of the dataset, N is the sample size or the number of values.

One-Way ANOVA

If there are more than two independent groups and the data in these groups are quantitative data, it is decided by ANOVA of variance whether there is a statistically significant difference between the group means (URL-2). H_0 and H_1 hypotheses are established in the One-Way ANOVA test. If the H_0 hypothesis is rejected, there can be a statistically significant difference between the means of the groups. However, it cannot be determined between which averages there is a difference. For this, it is necessary to look at multiple comparison tests. If the variances are homogeneous, tests such as Tukey are used. If there is no homogeneity of variance, the Tamhane T2 test can be used.

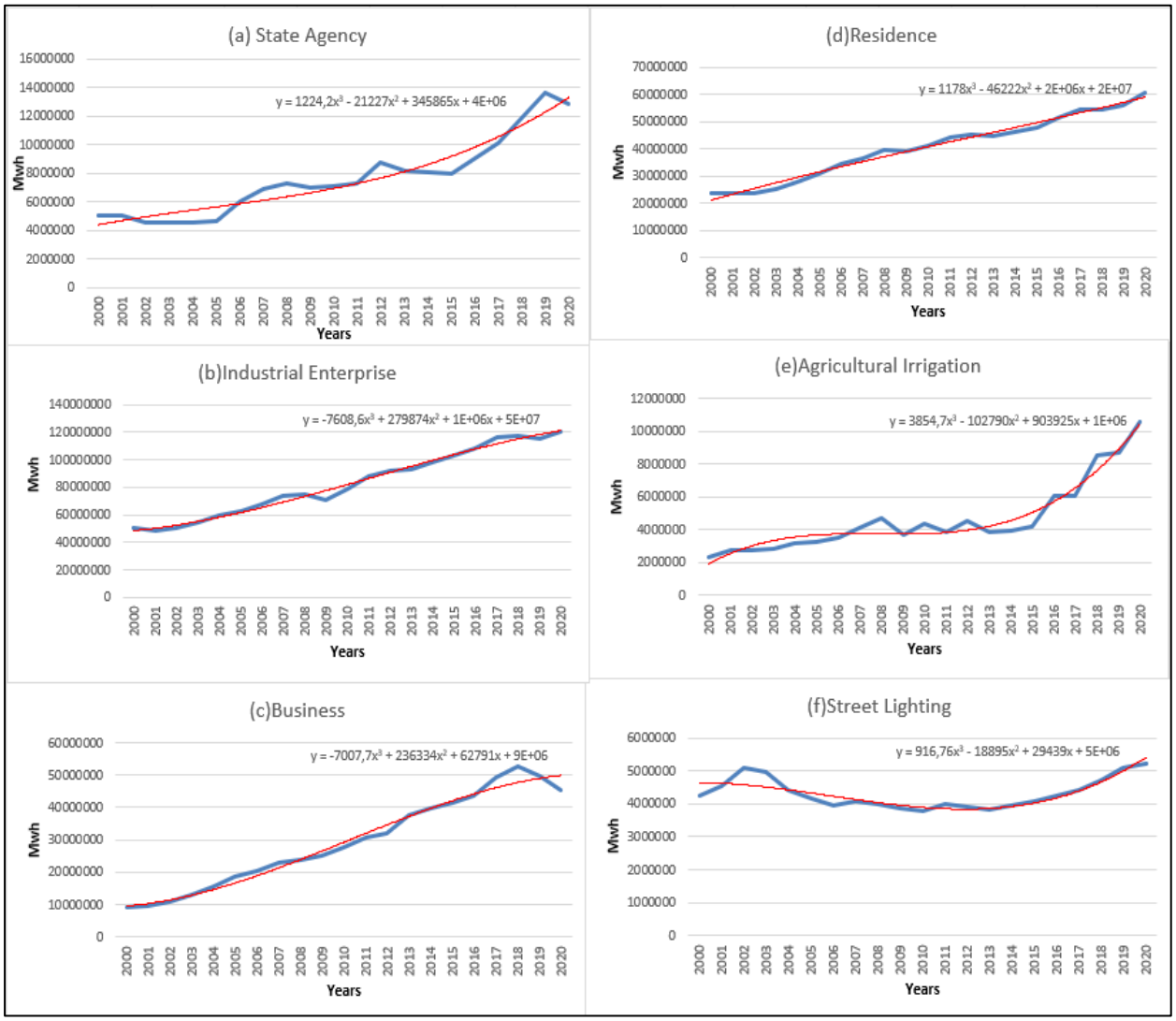

Figure 1. Forecasting model equations and graphs for sectors

Results and Discussion

In this study, Turkey's electricity consumption data between the years 2000-2020 has been analyzed. A cubic regression model was established for each sector group. Forecasting model equations and graphs obtained via cubic regression analysis are given in Figure 1. The values of the model evaluation criteria calculated as a result of the analysis of the cubic regression models are given in Table 1.

Table 1. Cubic regression models' evaluation criteria for sector groups

		Sector Groups					
		State Agency	Industrial Enterprise	Business	Residence	Agricultural Irrigation	Street Lighting
Model Evaluation Criteria	MAPE	0.09	0.04	0.05	0.08	0.10	0.09
	RMSE	722740.85	3804307.07	1993084.46	3896972.73	522048.55	420659.21
	R^2	0.93	0.99	0.98	0.98	0.94	0.80

According to Table 1, the cubic regression model accuracies for all groups are "quite good". With a MAPE value of 0.10, the regression model created for the Agricultural Irrigation group can be evaluated between "very good" and "good". The other calculated values also showed that the model accuracies can be predicted strongly. Estimated electricity consumption values for the next 5 years obtained via cubic regression models are given in Table 2.

Table 2. Estimated electricity consumption values for sector groups

		Sector Groups					
		State Agency	Industrial Enterprise	Business	Residence	Agricultural Irrigation	Street Lighting
Years	2021	14370443.60	126442643.20	50149068.40	54171896.00	12180835.60	6264138.48
	2022	15620653.40	128479509.80	50202193.10	55881288.00	14314499.90	6835860.92
	2023	16997348.80	130026137.60	49760923.20	57660800.00	16774532.80	7496306.24
	2024	18507875.00	131036875.00	48783212.50	59517500.00	19584062.50	8250975.00
	2025	20159577.20	131466070.40	47227014.80	61458456.00	22766217.20	9105367.76

In this study, One-Way ANOVA test was used as a different statistical analysis. It has been investigated whether there is a significant difference between the averages of electricity consumption data in terms of sector groups. In the analysis, firstly, homogeneity of variance was investigated. Table 3 contains the variance homogeneity table.

Table 3. Test of homogeneity of variances

Levene Statistic	df1	df2	Sig.
40.71	6	140	.00

Since the significance level (Sig.) is less than 0.05, the H_0 hypothesis, which accepts the variance of the groups as equal, is rejected. That is, group variances are not homogeneous. One-Way ANOVA test results are given in Table 4.

Table 4. Results of One-Way ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	107235640463489696.00	6	17872606743914950.00	131.98	.000
Within Groups	18958480213475988.00	140	135417715810542.83		
Total	126194120676965696.00	146			

Since the significance level (Sig.) is less than 0.05, the H_0 hypothesis, which accepts the mean of the groups as equal, is rejected. That is, group variances are not homogeneous. That is, there is a statistically significant difference between the average of the sector groups.

In this study, the Tamhane T2 test was used to analyze the difference between the averages of which sector groups. (yazim formati farkli) Results of Tamhane T2 test are given Table 5.

Table 5. Results of Tamhane T2 test

(I) Sector Group	(J) Sector Group	Mean Difference (I-J)	Std. Error	Sig.
State Agency	Industrial Enterprise	-75462742.81*	5356512.88	.00
	Business	-21800143.57*	3155278.62	.00
	Residence	-32885684.05*	2605760.31	.00
	Agricultural Irrigation	2996483.48*	757171.95	.01
	Street Lighting	3345061.57*	597360.44	.00
Industrial Enterprise	State Agency	75462742.81*	5356512.88	.00
	Business	53662599.24*	6160755.09	.00
	Residence	42577058.76*	5898229.13	.00
	Agricultural Irrigation	78459226.29*	5345318.46	.00
	Street Lighting	78807804.38*	5325031.42	.00
Business	State Agency	21800143.57*	3155278.62	.00
	Industrial Enterprise	-53662599.24*	6160755.09	.00
	Residence	-11085540.48	4006577.07	.17
	Agricultural Irrigation	24796627.05*	3136236.98	.00
	Street Lighting	25145205.14*	3101533.90	.00
Residence	State Agency	32885684.05*	2605760.31	.00
	Industrial Enterprise	-42577058.76*	5898229.13	.00
	Business	1108554.48	4006577.07	.17
	Agricultural Irrigation	35882167.52*	2582670.32	.00
	Street Lighting	36230745.62*	2540416.53	.00
Agricultural Irrigation	State Agency	-2996483.48*	757171.95	.01
	Industrial Enterprise	-78459226.29*	5345318.46	.00
	Business	-24796627.05*	3136236.98	.00
	Residence	-35882167.52*	2582670.32	.00
	Street Lighting	348578.09	486866.23	1.00
Street Lighting	State Agency	-3345061.57*	597360.44	.00
	Industrial Enterprise	-78807804.38*	5325031.42	.00
	Business	-25145205.14*	3101533.90	.00
	Residence	-36230745.62*	2540416.53	.00
	Agricultural Irrigation	-348578.09	486866.23	1.00

*The mean difference is significant at the 0.05 level.

According to the results in Table 5, there is no significant difference between the average of Business and Residence electricity consumption. Since the sig value is greater than 0.05, there is no significant difference between the average of electricity consumption from Agricultural Irrigation and Street Lighting. The difference between the means of the other groups is statistically significant.

Conclusion

Analyzing electricity production and consumption data is the most important factor facilitating planning in this area. In this study, electricity consumption data in Turkey were analyzed statistically. Data between 2000 and 2020 were obtained and cubic regression analysis was applied for the electricity consumption values of 6 different sector groups. The success of the prediction models was analyzed and verified by calculating MAPE, RMSE and R^2 values. Consumption data for the next 5 years are estimated with the created forecast models. Different from the studies in the literature, additional analysis was made in the study and the electricity consumption averages between the groups were examined. The variance and differences of the groups were examined with the One-Way ANOVA test. Differentiating groups were evaluated using the Tamhane T2 test. It is important for researchers and decision makers that the study can create statistically different outcomes.

Recommendations

For future studies, the scope of the study can be expanded by including different dependent and independent variables in the analysis. Comparisons can be made using different regression techniques. Apart from regression analysis, different estimation techniques can be used.

Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to authors.

Acknowledgements or Notes

* This article was presented as an oral presentation at the International Conference on Technology, Engineering and Science (www.icontes.net) held in Antalya/Turkey on November 16-19, 2022.

References

Duan, J., Tian, X., Ma, W., Qiu, X., Wang, P., & An, L. (2019). Electricity consumption forecasting using support vector regression with the mixture maximum correntropy criterion. *Entropy, 21(7),* 707.

Duranton, M., Bosschere, K.D., Gamrat, C., Maebe, J., Munk, H., Zendra, O. (2022, Jul 30). *The HIPEAC Vision 2017*, Retrieved from https://www.hipeac.net/vision/2017/

Gellert, A., Fiore, U., Florea, A., Chis, R., & Palmieri, F. (2022). Forecasting electricity consumption and production in smart homes through statistical methods. *Sustainable Cities and Society, 76*, 103426.

Gonultas, H., Kizilaslan, H., & N. Kizilaslan. (2020). Projections of effects of global warming on rainfall regime in some provinces; Ankara, Rize, Aydin and Hakkâri provinces example. *Turkish Journal of Agriculture-Food Science and Technology*, *8*(10), 2156-2163.

Grigoryan, H. (2021, November). Electricity consumption prediction using energy data, Socio-economic and weather indicators. A case study of Spain. In *2021 9th International Conference on Control, Mechatronics and Automation (ICCMA)* (pp. 158-164). IEEE.

Guler, E., & Kandemir, S. Y. (2022). Lineer ve kubik regresyon analizleri kullanilarak OECD ülkelerinin CO_2 emisyonlarinin tahminlemesi. *European Journal of Science and Technology, (34),* 175-180.

Meyers, L.S., Gamst, G., & Guarino, A.J. (2016). *Applied multivariate research: Design and interpretation*, 1st ed., New York, USA: SAGE Publications.

Nsangou, J. C., Kenfack, J., Nzotcha, U., Ekam, P. S. N., Voufo, J., & Tamo, T. T. (2022). Explaining household electricity consumption using quantile regression, decision tree and artificial neural network. *Energy, 250*, 123856.

Ozkaya, U., & Ozturk S. (2022). Gaussian regression models for day-level forecasting of COVID-19 in European countries. Understanding COVID-19: The role of computational ıntelligence. *Springer, Cham*, 339-356.

Piepho, H. P. (2019). A coefficient of determination (R2) for generalized linear mixed models. *Biometrical Journal*, *61*(4), 860-872.

Saranj, A., & Zolfaghari, M. (2022). The electricity consumption forecast: Adopting a hybrid approach by deep learning and ARIMAX-GARCH models. *Energy Reports*, *8*, 7657-7679.

Soyler, I., & Izgi, E. (2022). Electricity demand forecasting of hospital buildings in Istanbul. *Sustainability, 14(13),* 8187.

Tatliyer, A. (2020). Turkiye'deki tiftik uretimi degisiminin regresyon analizi ile incelenmesi. *Turkiye Tarimsal Arastirmalar Dergisi, 7(3),* 321-326.

TURKSTAT, (2022). *Energy Report*. https://www.tuik.gov.tr/

Tyagi, S., & Singh P. (2022). Short term and long-term building electricity consumption prediction using extreme gradient boosting. *Recent Advances in Computer Science and Communications, 15,* 1082-1095.

URL-1: TURKSTAT (2021) Electricity comsumption data.

URL-2: https://avys.omu.edu.tr/storage/app/public/burcinseyda.corba/122288/5.HAFTA.pdf

Yerel S. & Ersen T. (2013). Prediction of the calorific value of coal deposit using linear regression analysis. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 35(10),* 976-980.

Author Information

Ezgi Guler
Bilecik Seyh Edebali University
Bilecik, Turkey
Contact e-mail: *ezgi.guler@bilecik.edu.tr*

Suheyla Yerel Kandemir
Bilecik Seyh Edebali University
Bilecik, Turkey

To cite this article:

Guler, E. & Yerel Kandemir, S. (2022). Evaluation of electricity consumption in Turkey via cubic regression analysis and One-Way ANOVA test. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 21*, 11-17.