DOI: https://doi.org/10.53486/cike2024.16
UDC: 620.92:[504.05:551.588.74](560)

FORECASTING AND DISTRIBUTION OF THE GREENHOUSE GAS EMISSIONS OF TÜRKİYE

MEHMET NIYAZI ÇANKAYA
PhD in Statistics,
Uşak University, Faculty of Applied Science,
E-mail: mehmet.cankaya@usak.edu.tr
ORCID ID:0000-0002-2933-857X

ERCAN ÖZEN
PhD in Finance and Banking,
Uşak University, Faculty of Applied Science,
E-mail: ercan.ozen@usak.edu.tr
ORCID ID: 0000-0002-7774-5153

Abstract: A simple linear regression model is used to predict and forecast total greenhouse gas emissions. The parameters of the regression model are determined using maximum likelihood estimation method, assuming that the error terms follow an exponential power distribution. The grid search algorithm efficiently determines the shape parameter of this distribution, with a calculated value of 111.92, suggesting a tendency towards a uniform distribution of the observed values. The convergence of the exponential distribution of carbon emission data to a uniform distribution indicates collective behavior in the consumption of natural resources. Forecasts for the years 2023 and 2035 are provided for the total greenhouse gas emissions in Türkiye.
Keywords: economy, greenhouse gas emissions, energy, statistics, inference.
JEL Classification**:** Q50, Q54, H54

Introduction

Greenhouse gas (GHG) emissions have become a focal point in the global effort to combat climate change, as these emissions significantly contribute to the warming of our planet. Accurate forecasting of GHG emissions is essential for policymakers, researchers, and industries aiming to mitigate their environmental impact and comply with international climate agreements.

In recent years, various statistical and machine learning methods have been developed and refined to improve the accuracy of GHG emissions forecasts. These methods include traditional time series models like ARIMA (Auto Regressive Integrated Moving Average) and ETS (Exponential Smoothing State Space Model), which are commonly used for their robustness in handling seasonal variations and trends in data. More advanced techniques, such as the Prophet model and TBATS (Trigonometric, Box-Cox transform, ARMA errors, Trend, and Seasonal components), offer enhanced capabilities in capturing complex seasonal patterns and non-linearities, making them particularly useful in GHG emissions forecasting (Hyndman et al., 2008; Hyndman and Athanasopoulos, 2021; Brownlee, 2021; Taylor and Letham, 2017; Xu, etal., 2019). Additionally, the exponential power distribution can be applied to model the error terms in these forecasts, providing a more flexible approach to handling the distribution of residuals, especially in the presence of outliers or heavy-tailed distributions (Mineo, and Ruggieri, 2005).

In this study, we estimate the regression parameters by first identifying the distribution of the observed data. GHG emissions, in particular, play a crucial role in maintaining, fixing, etc. that can inform policy decisions and contribute to the global efforts in reducing GHG emissions. The primary findings and forecasts are presented in Sections 1 and 2, respectively, while the final section offers discussions and conclusions based on the analysis in the paper. The results of this study will not only demonstrate the effectiveness of the method suggested but also offer insights into the future trajectory of GHG emissions in the region.

Section 1: Research methodology and computational process

The regression model can be used to forecast the energy consumption in the next years. The regression model is given by the following form:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1,2,\dots,n,$$

y and x are dependent and independent variables, respectively. ε_i is error term of the regression model and n is the number of sample size of data set.

If the observations from $y_1, y_2, \dots, y_n$ are assumed to be a member of a parametric model such as normal, Student t, etc. distributions, then the statistics which are functions of random variables are obtained according to these chosen parametric models. Likelihood function for a parametric distribution, i.e. f, the maximum likelihood estimation method is given by the following form for the regression case:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f(y_i - x_i\boldsymbol{\beta}),$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)$ is a vector of parameters β_0 and β_1 in the regression model. The maximization of $L(\boldsymbol{\beta})$ the according to the parameters β_0 and β_1 will produce the maximum likelihood estimators $\widehat{\beta_0}$ and $\widehat{\beta_1}$ (Casella, G., and Berger, 2021; Mineo, and Ruggieri, 2005).

Since we are applying to the real data set, it is impossible to know what the true values of the parameters are. In order to solve the problem of the determination of the shape parameter of the exponential power distribution, the distance between the values of the observed variables and the predicted variables must be reduced as much as we can do. In order to obtain the reduced value for this distance, we can adjust the value of the tuning parameter and thus obtain the predicted values for the dependent variable of the regression model. We can use the method of grid search, which will be introduced in the research methodology in an algorithmic way (Woodward et. al., 2017; Strickland, 2015; Ghatak, 2019):

1. Set values for the constant
2. Set a sequence of values for the constant defined previously
3. Try the values provided by the sequence in the Step 2
4. Get values of the errors defined as difference between predicted and observed values
5. Find the minimum values among the error values obtained by the tried values of the shape parameter p.

6. Terminate the procedure and provide the value of p determined according to the minimum value of errors

It should be noted that focusing only on the regression model is not a sufficient approach for forward-looking forecasting. At the same time, focusing on the distribution of observations, i.e. the distribution of error terms representing the dependent variable y, also improves the success of forward-looking forecasts. In this study, we have chosen to focus on the distribution of observations in favor of the distribution of errors.

Section 2: Main results

This section presents the scatterplot, regression model parameter estimates and statistical significance of the full regression model. A scatter plot comparing the independent variable "years" with the dependent variable "GHG emissions" is a valuable tool for identifying suitable regression model to establish the relationship between these variables (refer to Figure 1, Internet access, 08/29/2024).

Once the appropriate regression model is selected, its parameters should be estimated using the maximum likelihood estimation method, particularly when the error terms follow an exponential power distribution. The overall statistical validity of the regression equation is then assessed through ANOVA (refer to Table 2).

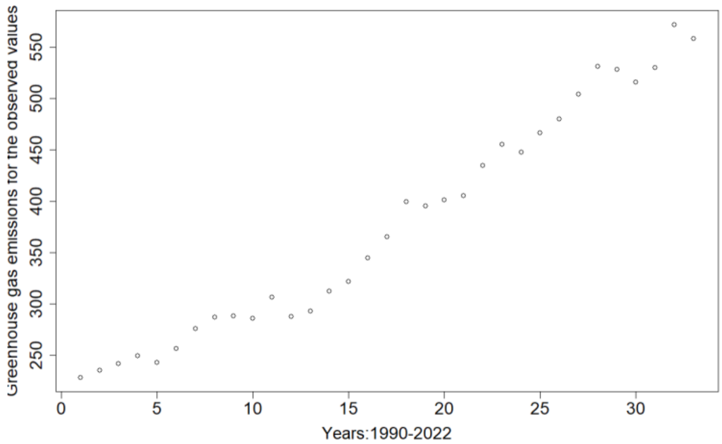

Figure 1. Years and GHG emissions for the observed values from official website of Turkish statistical institute (Internet access, 08/29/2024)

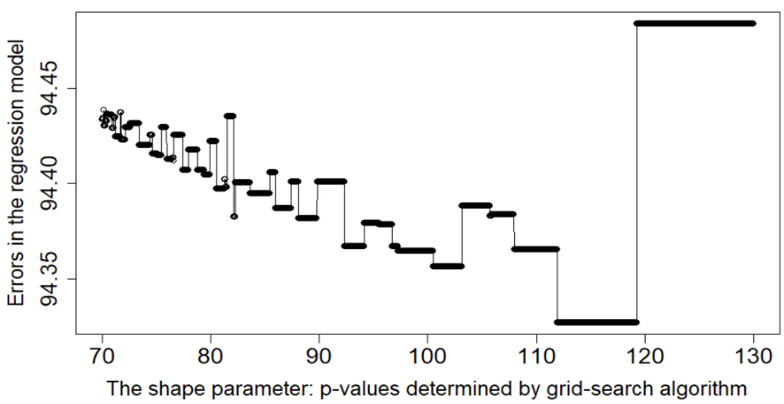

Figure 2. The absolute value of the errors defined as a difference between the predicted values of y and the observed values of y

Figure 2 illustrates the outcomes of the grid-search algorithm applied to identify the optimal value for the shape parameter of the exponential power distribution. In other words, when Figure 2 is analysed, it is seen that the value of the shape parameter p takes the smallest error value around 111. Note that other p-values were also tried, however, the smallest error value in the simulation run was found at the respective p-values at the interval [70,130] at Figure 2, according to the updated p-value.

Table 1. The estimates and ANOVA for full model

	The of estimates of parameters		Sum of Squares	Mean of Squares	F-test and its p-value
$\widehat{\beta_0}$	372.6	Regression	SSR: 373053	MSR: 373053	MSR/MSE: 983.49 and 2.2 10^{-16}
$\widehat{\beta_1}$	614.1	Error	SSE: 11759	MSE: 379	

Table 1 shows that the full model is statistically significant. It is reasonable to observe the simple linear regression which can be illustrated by Figure 1 with error terms, because when Figure 1 showing the scatter plot of years GHG emissions is examined, it is observed that there can exist a general linear trend.

Section 2.1: Forecasting of the Greenhouse gas emissions

Since the simple linear regression model has been used to fit the observed data, the predicted values show that there is a linear trend in the future prediction. Note that since the main motivation for the prediction should also be the distribution of the error terms in the regression model, different regression models cannot be preferred for forecasting. On the other hand, it is important to note that the assumed regression model for the fitted dataset cannot play a more important role in some sense; in our research methodology, we have focused on the distribution of error terms; thus, the sum of errors of the predictions has been tried to be decreased until the optimal value of shape parameter p is determined. Furthermore, if the distribution of errors can be identified, then the general tendency or distribution of the observations can be clarified as well.

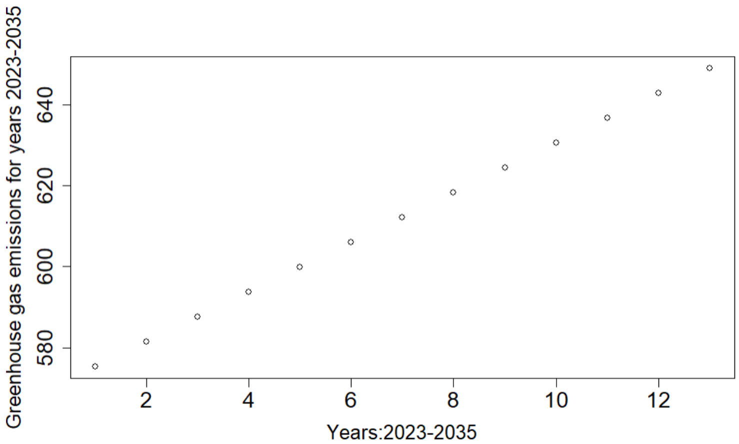

Figure 3. Forecasted values for the years 2023-2025

According to Figure 3, it is observed that there is a linear increment and the value at year 2025 will be 648.9709.

Discussion and conclusions

The simple linear regression model has been used to predict and forecast total GHG emissions. The parameters of the regression model could be determined by using the maximum likelihood estimation method when the distribution of the error terms is assumed to be exponential power distribution. The proposed method, which uses the grid search algorithm, provides a method for determining the value of the shape parameter of the exponential power distribution. The determined value of the shape parameter is 111.92, which shows that the distribution of observed values generally tends to uniform distribution. In the social part of science, we can see from the convergence of the exponential distribution of carbon emission data to a uniform distribution that people act in concert in consuming natural resources at will. The estimates of regression model were used to forecast the years 2023 and 2035 for total GHG emissions in Türkiye.

References

1. Hyndman, R. J., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2008). Forecasting with Exponential Smoothing: The State Space Approach. Springer. https://doi.org/10.1007/978-3-540-71918-2
2. Hyndman, R. J., and Athanasopoulos, G. (2021). Forecasting: principles and practice (3rd ed.). OTexts. https://otexts.com/fpp3/
3. Brownlee, J. (2021). How to Create an ARIMA Model for Time Series Forecasting in Python. MachineLearningMastery.com. Retrieved from https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python
4. Taylor, S. J., and Letham, B. (2018). Forecasting at scale. The American Statistician, 72(1), 37-45.
5. Casella, G., and Berger, R. L. (2021). Statistical inference. Cengage Learning.
6. Mineo, A., and Ruggieri, M. (2005). A software tool for the exponential power distribution: The normalp package. *Journal of Statistical Software*, *12*, 1-24.
7. Du, Y.B., Yi, Q., Li, C.B., Liao, L., (2015). Life cycle oriented low-carbon operation models of machinery manufacturing industry. *J. Clean. Prod.*, 91, 145-157.
8. Ghatak, A. (2019). Deep learning with R (Vol. 245). Singapore: Springer.
9. Xu, N., Ding, S., Gong, Y., & Bai, J. (2019). Forecasting Chinese greenhouse gas emissions from energy consumption using a novel grey rolling model. *Energy*, *175*, 218-227.
10. https://data.tuik.gov.tr/Kategori/GetKategori?p=Cevre-ve-Enerji-103, Web, 08/29/2024.
11. Strickland, J. (2015). Predictive analytics using R. Lulu. com.
12. Woodward, W. A., Gray, H. L., & Elliott, A. C. (2017). Applied time series analysis with R. CRC press.