Turkish Journal of Engineering

https://dergipark.org.tr/en/pub/tuje

e-ISSN 2587-1366

A lasso regression-based forecasting model for daily gasoline consumption: Türkiye Case

Ertugrul Ayyıldız *1, **Mirac Murat** 1

1 *Karadeniz Technical University, Department of Industrial Engineering, Türkiye, ertugrulayyildiz@ktu.edu.tr, miracmurat@ktu.edu.tr*

Cite this study: Ayyıldız, E., & Murat, M. (2024). A lasso regression-based forecasting model for daily gasoline consumption: Türkiye Case. Turkish Journal of Engineering, 8 (1), 162-174

Keywords
Gasoline consumption
Forecasting
Lasso regression
Ridge regression

Research Article
DOI: 10.31127/tuje.1354501

Received:03.09.2023
Revised: 03.01.2024
Accepted:04.01.2024
Published:16.01.2024

check for updates

Abstract

Gasoline is one of the most sought-after resources in the world, where the need for energy is indispensable and continuously increasing for human life today. A shortage of gasoline may negatively affect the economies of countries. Therefore, analysis and estimates about gasoline consumption are critical. Better forecast performance on gasoline consumption can serve the policymakers, managers, researchers, and other gasoline sector stakeholders. This study focuses on forecasting daily gasoline consumption in Türkiye using a lasso regression-based methodology. The methodology involves three main stages: cleaning data, extracting/selecting features, and forecasting future consumption. Additionally, Ridge Regression is employed for performance comparison. Results from the proposed methodology inform strategies for gasoline consumption, enabling more accurate planning and trade activities. The study emphasizes the importance of daily forecasts in deciding import quantities, facilitating timely planning, and establishing a well-organized gasoline supply chain system. Application of this methodology in Türkiye can pave the way for globally coordinated steps in gasoline consumption, establishing efficient gasoline supply chain systems. The findings provide insights for establishing a smooth and secure gasoline collection/distribution infrastructure, offering effective solutions to both public and private sectors. The proposed forecasting methodology serves as a reference for ensuring uninterrupted gasoline supply and maximizing engagement between customers and suppliers. Applied and validated for Türkiye, this methodology can guide global efforts, fostering planned approaches to gasoline consumption and enhancing supply chain systems.

1. Introduction

Türkiye's economy is an emerging market economy, as defined by the International Monetary Fund (IMF) [1]. Türkiye has a substantially improved economy. Over the past 20 years, Türkiye globalizes its economy with emerging technologies. The gross domestic product (GDP), which was 272 billion dollars in 2000, tripled in 2022 to 906 billion dollars. Simultaneously, the GDP per capita increased 2.9 times in 2022 compared to 2002, from $ 3,581 to $ 10,661 [2]. The change of the GDP per capita and purchasing power parity by years are given in Figure 1 for Türkiye [3]. Türkiye's economy is the world's 19th, and Europe's 7th largest economy in 2019 [2].

While various energy sources are vital for economic prosperity and technological competitiveness [4], oil and natural gas stand out as the most sought-after resources globally [5], particularly as the world's need for energy is continuously increasing. It is clear that if there is a shortage in the provision of these two resources, the economies of the country, and therefore the world economy, will be affected. Since the late 19th century, oil, as a primary energy source, has been one of the main factors consumed in parallel with the development of industry and has accelerated its growth, alongside electricity consumption, which has also played an influential role in the growth of economies [6] and has been recognized as another key energy type for end-users. With its much wider usage in the 20th century, oil is indispensable for human life today. Research shows that oil will remain the essential energy source in the medium term.

Three main views are accepted in measuring the linear relationship between oil consumption and economic growth. According to the first view, oil is one of the most important production factors, and oil demand is the primary economic development dynamic. The second view is that both oil consumption and economic growth affect each other, and there is bi-directional variability between the existing variables. According to the third view, it is argued that there is no linear relationship between oil consumption and economic growth [7].

The growing economy and developing technology have increased the need for oil in Türkiye. The demand for oil is increasing daily because of the rapid growth of the economy, particularly within the context of the oil market. Oil is used for transportation, energy, etc. Indeed, in 2019, the amount of oil consumed in Türkiye is over thirty million liters [8]. More than 90 percent of Türkiye's oil is imported from different countries as Iraq, Russia, Iran, etc. [9]. TÜPRAŞ's (the biggest petroleum refinery holding of Türkiye) yearly amount of exported crude oil and domestic production is given in Figure 2 [10].

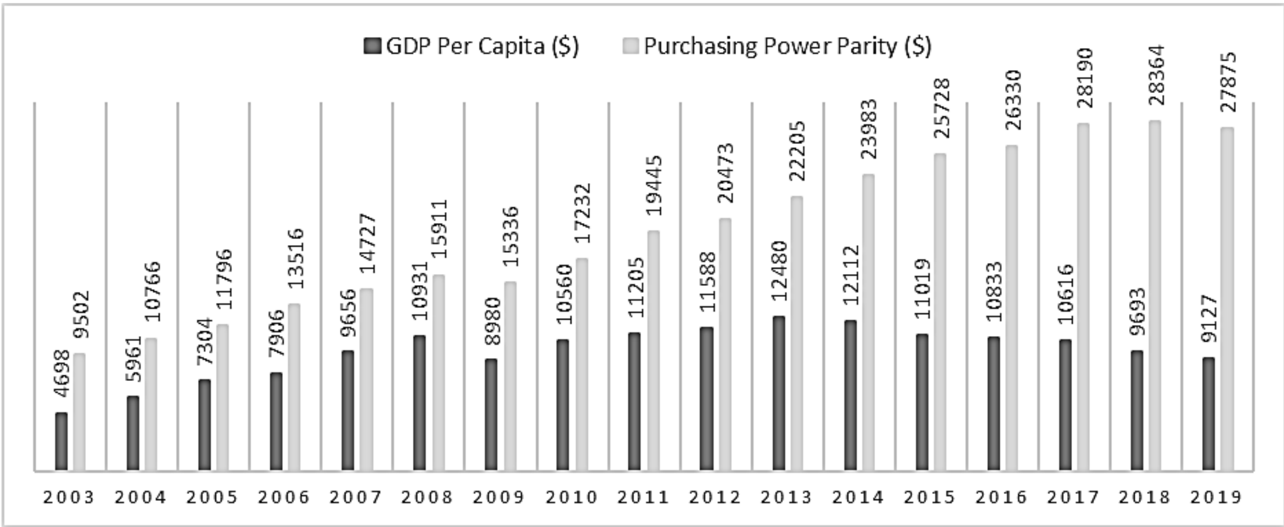


Figure 1. Türkiye's GDP and purchasing parity by years.

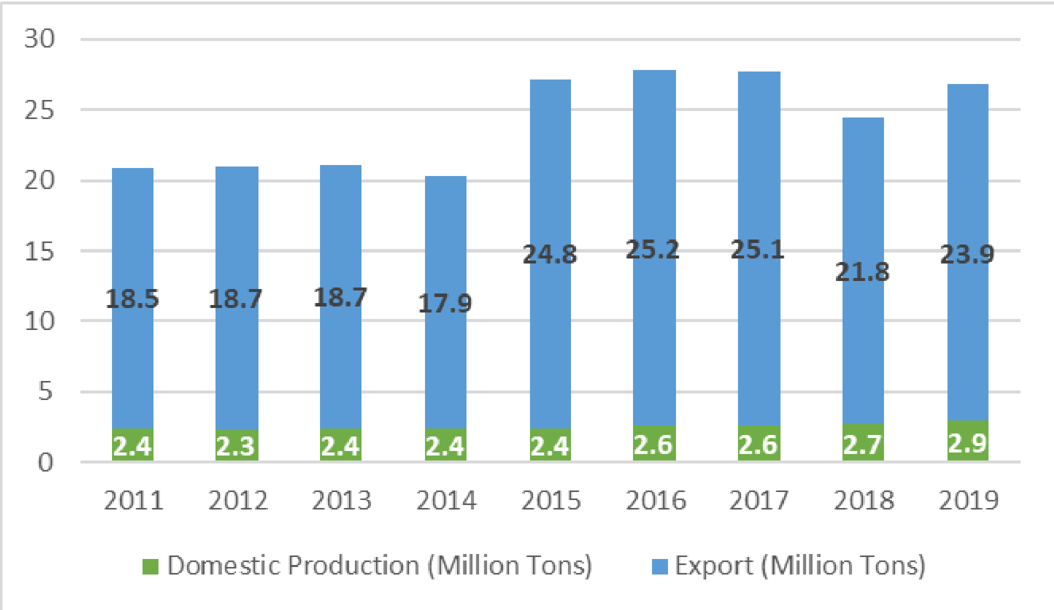


Figure 2. TÜPRAŞ's yearly amount of crude oil supplies by type.

Accurate forecasts help maintain a balance between gasoline supply and demand, allowing for better economic planning. This is important to avoid shortages or surpluses that could disrupt economic activity. Forecasts influence import strategies and trade planning by helping to determine the amount of gasoline to be imported. This is important for Türkiye's economic relations with other countries. Price stability is achieved by understanding and forecasting gasoline consumption trends. This contributes to the Turkish economy, where most of the gasoline consumed is imported, and is particularly important for consumers, businesses, and policymakers in managing inflation and economic fluctuations. Daily forecasts of gasoline consumption provide a basis for the formulation of energy policies, helping governments set targets and develop strategies to balance energy needs with environmental sustainability. Better forecast performance on oil consumption can serve the policymakers, managers, researchers, and other stakeholders of the oil sector in both Türkiye and the world [11–13].

Researchers and theoreticians have focused on the oil or gasoline consumption forecasting problem and its extensions over the years. Nel and Cooper analyze the relationship between oil consumption and economic growth for countries. They use a logistic curve to predict oil demand for China. They construct projections until 2030 [14]. Azadeh et al. concentrate on forecasting oil prices in various settings, including noisy, uncertain, and complex environments. Artificial Neural Network (ANN) is integrated with fuzzy regression to determine the best oil price forecast. Different statistical tests are performed to show the validity of the proposed methodology [15]. Narayan and Wong examine the components that affect oil consumption for the territory, located in Australia for twenty years. They determine oil price and income as the most active factors. Income is then estimated to be significantly important for oil consumption, while oil prices are not [16]. Yang et al. propose a forecasting model that includes both a genetic algorithm and the backpropagation neural network to estimate China's oil consumption. They also use a gray model to take advantage of both nonlinear and linear models [17]. Li et al. develop 26 different combination models using a traditional combination method to handle overfitting and improving prediction accuracy. They test regression models, gray theory, and a hybrid model to predict China's oil consumption [18]. Assareh et al. use the genetic algorithm and particle swarm optimization techniques to predict Iran's oil demand. They predict yearly demand using data of 35 years [19]. Lin and Xie focus on oil demand forecasting for the transport sector in China. Multiple factors such as gross domestic product, road condition, oil price, and labor performance are determined, and the forecasting model is proposed. Monte Carlo simulation is also used for risk analysis of the proposed model [20]. Rao and Parikh use the forecasting model based on a time series to forecast the oil demand for India. They perform forecasting for different oil products, such as gasoline, diesel oil, and liquid petroleum. Six different variables are used in the model [21]. Wang and Song examine the different versions of the gray model and propose a novel gray model based on time series. Yearly demand between 1990 and 2017 is used to forecast the next ten years for oil consumption in China [22]. Duan et al. extend the SIGM model with gray theory to forecast China's crude oil consumption. To handle conforming class ratio check, least squares estimation is performed. Particle swarm optimization techniques are also used to achieve better forecasting performance. They use yearly data from 2002 to 2014 to forecast between 2015 and 2020 [23]. Behrang et al. present a novel approach to forecast oil consumption. Gross domestic product, import, export, and population are determined as variables in the forecasting model. Gravitational Search Algorithm is used to forecast oil consumption in Iran, then the Genetic Algorithm and Particle Swarm Optimization are used to compare the proposed model [24]. Minniear evaluates global petroleum product production. Saudi Arabia, Kuwait, Iran, Venezuela, United Arab Emirates, and Iraq are analyzed, and future prediction is made [25]. Al-Qaness et al. improve the adaptive neuro-fuzzy inference system (ANFIS) with a sine-cosine algorithm for forecasting oil consumption. Monthly oil consumption data from 2007 to 2017 from three different countries, as Canada, Germany, and Japan, is used to assess the performance of the proposed methodology. Extensions of ANFIS are used for comparison [26]. Fatima et al. predict oil demand in China using oil prices, oil reserves, oil consumption, and economic growth. Time-series techniques are used to investigate the linkages between parameters and oil demand in this paper [27]. Yu et al. investigate the relationship between parameters of oil consumption. Popular forecasting methods are applied to forecast oil consumption using Google trends, including big data [28]. Keshavarzian et al. make projections of oil consumption for the road transportation sector [29]. First, vehicle ownership is forecasted using a time series via a nonlinear Gompertz model. Then, specific geographic and demographic variables are implemented into the model for different countries. Yearly data from 1972 to 2008 in 154 countries are used in the study.

Sadri et al. aim to determine gasoline consumption for a period of ten-year using vehicle population, traffic volume, average private vehicles kilometers traveled/capita as input variables with different artificial neural network models [30]. Melikoglu employs semi-empirical models to predict Türkiye's yearly demand for gasoline, diesel, liquefied petroleum gas bioethanol, and biodiesel [31]. Azadeh et al. present a fuzzy mathematical programming-analysis of variance approach to forecasting yearly gasoline consumption of Iran, Kuwait, USA, Canada, and Japan [32]. Sapnken et al. focus on the gasoline consumption of Cameroon. They use linear regression-based forecast methodologies to determine yearly gasoline consumption for two years [33]. Anggarani and Watada aim to develop an accurate harmony search-based forecast model for gasoline consumption forecasting [34]. Chen et al. present a gray fractional FGM (1, 1) model to predict six years' energy consumption for natural gas, crude oil, gasoline, and diesel [35]. Güngör et al. investigate Türkiye's gasoline consumption after pandemic with Autoregressive Integrated Moving Average (ARIMA) models [36]. Wang et al. use single-linear, hybrid-linear, and non-linear time series models to forecast the yearly energy demand of China and India [37]. Wang et al. employ hybrid ARIMA and metabolic nonlinear gray model for forecasting the monthly gas production in the United States [38]. Wang et al. analyze COVID-19 impact on the United States' oil consumption via developed novel ARIMA based forecasting model [39]. Wang et al. investigate the pandemic impact on energy consumption in China via an ARIMA model and show the reduction in energy consumption [40]. Wang et al. employ ARIMA based methodologies to analyze the pandemic impact on carbon emissions [41].

As seen in the studies above, the problem of accurate forecast on oil and its derivatives consumption is one of the - studied topics in the forecasting literature. But there are a limited number of studies on daily gasoline consumption forecasting. Therefore, apart from other studies in the literature, the problem of the forecasting of daily gasoline consumption is discussed in this study. Accurate forecasting performance plays a key role for decision-makers.

This study aims to develop a forecasting strategy for daily gasoline consumption, which is scarce in the literature. For this purpose, a lasso regression-based forecasting methodology is proposed. To the best of our knowledge, lasso regression is used to forecast gasoline consumption for the first time with this study.

This study stands out from the current academic perspective by addressing a specific gap in daily gasoline consumption forecasting research. The primary contribution lies in introducing a forecasting technique based on lasso regression, specifically designed for daily gasoline consumption. This represents a significant innovation in the region, filling existing gaps and providing decision makers with a new and practical tool for effective planning and strategy development in the petroleum sector. Another distinguishing point is that this distribution is the first to survive lasso regression in predicting gasoline consumption. The paper also expands the repertoire of forecasting techniques used in the existing literature to predict gasoline consumption. Beyond the methodological advances, the article emphasizes that the proposed forecasting strategies are presented in a practical manner for decision makers. Accurate daily gasoline consumption forecasts are positioned as indispensable for effective decision making, drawn to the extent of increasing the distribution of oil imports and providing guidance on production.

2. Method

Lasso regression is selected as forecasting methodology because it fits the nature of the daily gasoline consumption forecasting problem. Lasso regression is particularly suited to situations with a large number of potential predictor variables because it introduces a penalty term that forces some regression coefficients to be exactly zero to encourage sparse solutions [42]. In the realm of forecasting daily gasoline consumption, numerous factors intrinsic to the components of the time series have influence over the results. It becomes imperative to discern the most pertinent features in this context. Lasso regression is compatible with the goal of isolating the most influential variables in daily gasoline consumption time series data by automatically performing variable selection and feature extraction [43].

Variables for energy consumption prediction often exhibit multicollinearity, i.e. they are highly correlated. Lasso regression handles multicollinearity by selecting one variable from a set of highly correlated variables, making the model more robust [44]. Given the complex dependencies inherent in energy consumption data, this method is well suited to ensure the reliability of the predictive model. The penalty term of lasso regression not only assists in variable selection but also prevents overfitting [45]. Reducing the risk of overfitting provides a more accurate and feasible daily gasoline consumption prediction model [46]. Moreover, Lasso regression but also provides interpretability in terms of variable importance [47]. This is critical in the context of daily gasoline consumption, as policymakers, managers, and researchers need insight into which factors contribute significantly to observed changes. The transparent nature of lasso regression meets the need for an understandable and actionable forecasting model.

The time series forecast in this study is based on extracting prospective features from data. Features are generated by mining both the observations and the time points. One of these features is "lag" that describes earlier observations in the series [48]. The lag is based on the concept of autocorrelation, which is the similarity between the series and observations as a function of the time lag. Another feature is the tuple of date and/or time components. With these types of features, the unit time in the series can be transferred into ones of a minute of an hour, an hour of a day, a day of a week, a month of a year, etc. The transformation can also be made by considering if a day is a holiday or if there is a special event. Hence, each matching unit time is coded in the same way to capture possible patterns of the series. The last feature used for forecasting is the target encoding that generally calculates averages of identical date features [49]. This distinguishes a forecasting model from any other time series that involves the same time interval and creates an exclusive time series forecasting.

In this study, the forecasting approach used for daily gasoline consumption consists of three main stages: i) preprocessing the data ii) extracting and selecting features iii) forecasting the future of daily gasoline consumption time series via the proposed models, and six basic steps described in following. A detailed description of each step is provided in the following three subsections.

Step 1. Gathering the data for gasoline consumption from the related sources and reformatting the data points to enable forecasting.

Step 2. Deleting and then replacing excessively oscillating and missing values with interpolated values.

Step 3. Evaluating if the time series is stationary or not to form an idea related to possible forecasting performance.

Step 4. Extracting features that depend on the series itself. Creating various historical time series, which are totally-dependent features; and date components, which are semi-dependent, from the original data.

Step 5. Developing an effectively forecasting gasoline consumption model by regularizing it and selecting significant features using Lasso Regression.

Step 6. Forecasting future gasoline consumption through the selected features and developed model.

2.1. Extracting features from data

The data used in this study was gathered from the weekly official reports of the Turkish Ministry of Energy and Natural Resources [50]. It has daily observations of the gasoline consumption in Türkiye. Since the data exhibits inconsistencies, such as variations in format, unit discrepancies, and gaps in information sharing because of weekly releases, the data had to be organized. Missing and abnormal observations in the data were rectified by updating them with new values calculated through an interpolation method, considering assumptions about the underlying continuous function.

The forecasting process for gasoline consumption relies on features extracted from the series itself, requiring diverse features that are not necessarily independent of the series. These features can be divided into two groups: totally-dependent features and semi-dependent features. The values of features included in the group of totally-dependent are calculated from the values of the series itself considering the relevant dates. On the other hand, the values of the features in the semi-dependent group are contingent upon the time points within the time span of the series. These values would be identical if the time spans of two dissimilar time series are the same. Week of year for the fifth time points, for instance, of two time series gasoline consumption vs coal consumption in the same year, would have the same value.

The features extracted as date components from the series fall under the category of semi-dependent. Decomposing the date into independent parts, such as year, month, week, day of the week, and week of the year, has the potential to generate robust predictors. However, these features require systematic transformation owing to their inherent ordered and cyclic nature, since the initial and final data points, akin to the intermediate ones, are interconnected in a sequence, demonstrating the continuous and cyclical nature of time in data analysis. These types of data generally are transformed into two variables that swing back and forth [51].

Calculating average values of the same time units and assigning them for corresponding units creates eventual distinguishing features that fall under totally-dependent. These features that represent data regarding the observations itself, along with their temporal component, are exclusive to the series. Hence, each observation in the series is enriched by average values associated with the time units, such as weekday, week, or month to which it belongs. For instance, if a value named *monthaverage* is to be determined for an observation, this value is calculated by taking the average of the data points only from the months in which that observation is present.

An advance transformation for the date feature is to determine if any observation of the series belongs to an exclusive time of the year, such as a weekend or holiday. This helps to understand the cause for the unusual deviations (within allowed limits) from prior and posterior observations. Such features are also included in the semi-dependent group.

Another group of features is *lags* that represent the relationship between the current and historical observations of the series. Lags, a delayed rendition of the series, are derived exclusively through time shifting, excluding the direct utilization of time values or observation averages from the series. To interpret the structure of a time series, autocorrelation with various lags serves as a valuable criterion. A plot generated through autocorrelations can aid in acquiring a better understanding of the structure of the series, such as the magnitude of seasonality. The Autocorrelation Function (ACF) plot encompasses correlations with all feasible lags, contributing to a comprehensive analysis.

2.2. Model development and feature selection

The basic regularization techniques used to forecast gasoline consumption alongside linear regression are ridge regression (RR) and lasso regression (LR).

The Equation 1 for the linear regression model that has an error term (e) with an expected value of 0 ($E[e] = 0$) and constant variance ($Var(e_i) = \sigma^2$) is:

$$y = \beta_0 I + X\beta + e \qquad (1)$$

where $y = (y_1, y_2, \dots, y_N)^T$ is the vector of N observations, I is the unit matrix, and β_0 is the intercept. X is a $N \times p$ matrix of features; β is the vector of the regression coefficients of p features.

Linear regression may cause a model that has poorly determined coefficients with high variance due to complex relationships caused by a large number of features [52]. RR is ideal to solve the issue of determining adequate coefficients and to develop high-quality models with numerous predictors, each having a small effect. RR overcomes this issue by shrinking the coefficients of correlated predictors towards zero and improves the forecasting performance [42].

RR minimizes the loss function given in Equation 2.

$$\underset{\beta}{\text{minimize}}\left\{\frac{1}{2N}\|\mathrm{y} - \mathrm{X}\beta\|_2^2 + \lambda\|\beta\|_2^2\right\}, \qquad \lambda \geq \qquad (2)$$

where $y = (y_1, y_2, \dots, y_N)^T$ is a $Nx1$ dimensional vector that stands for the dependent variable, X is a Nxp dimensional matrix whose elements stand for p independent variables and β is the $px1$ dimensional vector of coefficients of independent variables. $\|.\|_2$ indicates the Euclidean distance. $\|\beta\|_2^2$ is the l_2 norm penalty on β with the tuning parameter of λ that differs LR from ordinary regression analysis [53]. The optimal values for parameters β and λ are determined according to the performance measurement of mean squared error (MSE) acquired cross validation procedure.

The process of feature extraction can produce a large number of so-called features. Some of these features might have a low or no effect on the forecasting performance or might cause overfitting. Therefore, there is a need for a feature selection/reduction to apply effective forecasting [54]. Here, another regularization technique that provides an opportunity for feature selection/reduction, as well as model simplicity and prevention of overfitting, is Least Absolute Shrinkage and Selection (Lasso), i.e. LR [55].

LR was essentially developed to eliminate the disabilities of the least squares method by Tibshirani [56]. However, it recently stands out as one of the methods that improve the forecasting performance in analyzing high-dimensional data [53]. LR is an inbuilt variable selection method, it has the mastery to control the parameters leading to revealing ineffective variables. It shrinks some regression coefficients to make them zero according to Equation 3 given in the Lagrange form [57]. Therefore, the accuracy of predictions and ease of interpretation are enhanced.

$$\underset{\beta}{\text{minimize}}\left\{\frac{1}{2N}\|\mathrm{y} - \mathrm{X}\beta\|_2^2 + \lambda\|\beta\|_1\right\}, \qquad \lambda \geq \qquad (3)$$

where $\|.\|_1$ is Manhattan distance and $\|\beta\|_1$ indicates l_1 norm penalty on β.

Lasso regularization forces some coefficients to be zero to vanishing, thus some predictors may completely leave out from the forecasting of the series. The non-zero coefficients sign the effective variables to forecast the time series. The remaining variables of interest are more powerful to represent the relationship between the parts of the series and itself. Hence, LR produces a more interpretable regression model with only the most relevant and predictive subset of features while the forecasting performance for the time series increases.

2.3. Forecasting daily gasoline consumption using the proposed models

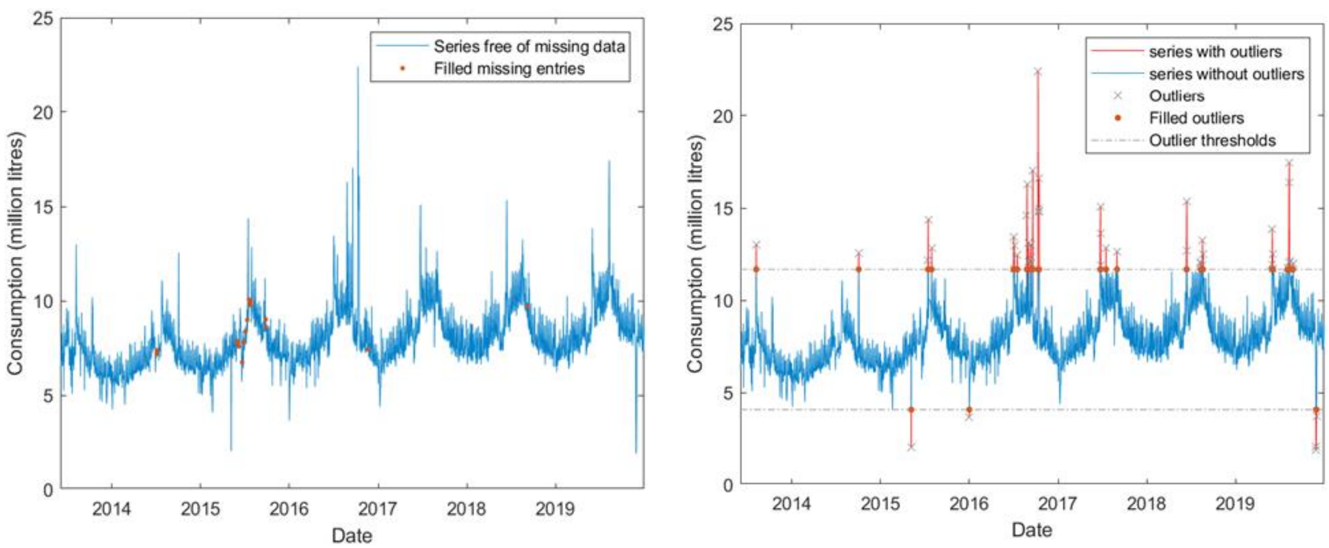
The top-performing models from each regression technique are employed to predict the daily gasoline consumption for the test data period, corresponding to the final year of the series. The performance of each model is compared according to the forecasting error. Mean absolute percentage error (MAPE) is a simple comparative measurement that simplifies the intuitive interpretation of the error. Equation 4 gives MAPE between the actual values (A), comprising n observations, and their corresponding forecasted values (F).

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{(|A_t - F_t|)^2}{A_t} \qquad (4)$$

3. A real case application for Türkiye

The raw time series used in this study is a six-and-a-half-year time series that has daily observations recorded in the interval between 2013-06-03 and 2019-12-31 as shown in Figure 3. The series contained missing values, and the missing values were filled using the Piecewise Cubic Hermite Interpolating Polynomial (Pchip) interpolation technique. Pchip proved to be the optimal interpolation method after extensive testing against cubic spline and Akima piecewise cubic Hermite interpolation across 1000 time series, each artificially designed to include missing values. In every iteration, the three techniques were simultaneously assessed for their ability to interpolate the so-called missing values, which were randomly assigned within the longest section of the time series (from 2016-11-21 to 2018-09-07) devoid of missing data. After 1000 trials, Pchip stood out for achieving the majority of best fit performances, especially excelling with non-smooth and widely spaced data [58].

In the series, instances of excessive consumption data were observed during periods of peak consumption, particularly around summer. While it is not believed that these data were inaccurately recorded, a further evaluation was conducted using the quartiles method, which specifically, data points beyond 1.5 interquartile ranges are identified outliers. Figure 3b illustrates the outlier data with the red line. These data points have been replaced with values at the specified thresholds. Consequently, the blue line in Figure 3b is the adjusted time series used to develop forecasting models.



(a) (b)

Figure 3. a) Gasoline consumption time series with missing data filled by Pchip technique, b) time series without outliers

The forecasting performance of a time series significantly depends on the presence of stationarity, denoting a consistent mean and variance over time that enhances predictability. In this study, the adjusted time series exhibits non-stationarity, as indicated by the Dickey-Fuller test [59] result with a p-value of 0.3990 at a significance level of 0.05. This non-stationary characteristic, as illustrated in Figure 4, while posing a challenge to conventional methodologies of forecasting, the approach in this study is designed to operate effectively without the assumption of existing stationarity.

Figure 5 illustrates the correlation values between the series and all possible lags. The plot has sine-like waves, indicating cycles of strong negative and positive correlation, which is a strong sign of seasonality in the time series. Approximately 358 lags after the initial lag, a relatively high correlation of 0.634 occurs again. Thus, it can be simply captured that there is a relationship of a point of time series with past points in the same season that is around a year.

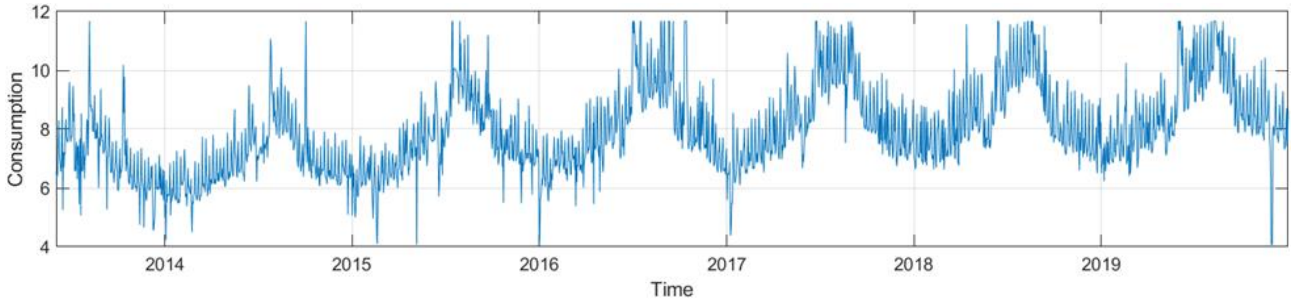


Figure 4. Non-stationary time series of consumption.

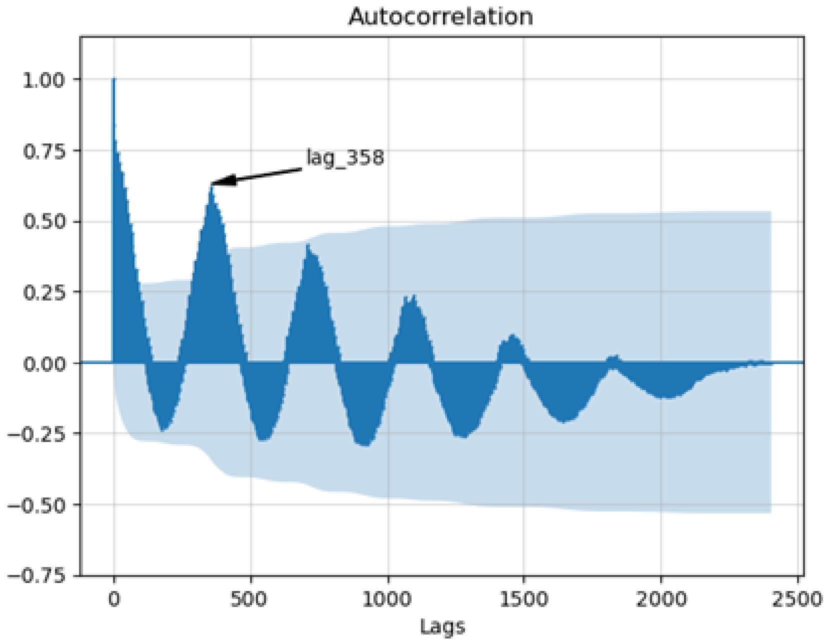


Figure 5. The Autocorrelation plot of time series

Once organizing the time series and preparing it for forecasting, a non-random partitioning is applied. Time points corresponding to almost the first year (2013-06-03 and 2014-05-26) are excluded from data due to 358 lags. Subsequently, the preprocessed time series data, now complete and free of missing values, has been divided into three sets: training, validation, and test data. Approximately 20% of the series, representing the last year (data points between 2019-01-01 and 2019-12-31), has been reserved as test data. The training data comprises the initial 75% of the remaining series (data points between 2014-05-27 and 2018-08-04), while the validation data consists of the consecutive 25% (2018-01-01 and 2018-12-31). In terms of proportions, the training, validation, and test data account for approximately 64%, 17%, and 17% of the entire forecasting dataset, respectively.

The lags with a reasonable correlation (≥0.6) [60] with the series itself were incorporated as features for the analysis. 26 lags were included as features, almost all comprising delays within the first 45 days. Notably, two of these lags, namely the 351st and 358th, contribute to capturing a yearly seasonal variation in the series.

The data used in the study consists of daily points; it is therefore broken into parts of the day of the week, day of the month, day of the year, week, and month. The transformation for the timestamp of unit time to the daily, weekly, and monthly counterparts are performed according to Equations 5-14, respectively.

$$DayOfWeek_{X_t} = \sin(2 * \mathrm{pi} * WeekDayOfO_t \,/\, 7)\,\# \tag{5}$$

$$DayOfWeek_{Y_t} = \cos(2 * \mathrm{pi} * WeekDayOfO_t \,/\, 7) \tag{6}$$

$$DayOfMonth_{X_t} = \sin\left(2 * \mathrm{pi} * \frac{MonthDayOfO_t}{\#\ of\ days\ in\ the\ month}\right) \tag{7}$$

$$DayOfMonth_{Y_t} = \cos\left(2 * \mathrm{pi} * \frac{MonthDayOfO_t}{\#\ of\ days\ in\ the\ month}\right) \tag{8}$$

$$DayOfYear_{X_t} = \sin\left(2 * \mathrm{pi} * \frac{YearDayOfO_t}{\#\ of\ days\ in\ the\ year}\right) \tag{9}$$

$$DayOfYear_{Y_t} = \cos\left(2 * \text{pi} * \frac{YearDayOfO_t}{\#\ of\ days\ in\ the\ year}\right) \tag{10}$$

$$Week_{X_t} = \sin(2 * \text{pi} * WeekOfO_t\ /\ 52) \tag{11}$$

$$Week_{Y_t} = \cos(2 * \text{pi} * WeekOfO_t\ /\ 52) \tag{12}$$

$$Month_{X_t} = \sin(2 * \text{pi} * MonthOfO_t\ /\ 12) \tag{13}$$

$$Month_{Y_t} = \cos(2 * \text{pi} * MonthOfO_t\ /\ 12) \tag{14}$$

The Equation 5-14 can be summarized as Equation 15-16:

$$\tau_x^t = \sin\left(2 * pi * \frac{\tau^t}{\rho}\right) \tag{15}$$

$$\tau_y^t = \cos\left(2 * pi * \frac{\tau^t}{\rho}\right) \tag{16}$$

where, τ^t is the corresponding time component to be transformed for the observation O_t at time step t. τ_x^t and τ_y^t are the transformed forms of the time component and ρ is the number of the time component in a specific time period. ρ is 52 and 7, for instance, when time components of observation are week and weekday, respectively.

Another type of feature, the unusualness of any timestamp, are being the weekend and holidays. For any observation belongs to weekends or holidays announced by the official government is tagged in a binomial scheme according to whether the day is the weekend/holiday (1) or not (0).

Average values for any week of the year, month of the year, any day of the week, day of the month, and day of the year are calculated for the period represented by training data considering only the training data. On the contrary, the average values for the period represented by the test data are calculated considering both the training and the validation data that is known anymore.

44 different features were formed through the feature extraction phase as predictors. The predictors consist of 12 time-components, 5 averages, and 27 lags that have at least the correlation coefficient of 0.6 with the time series itself. The predictors are given in Table 1.

Table 1. The predictors for gasoline consumption forecast model.

Time Components	Lags	Average
dayofweek_x	lag_1	weekday_average
dayofweek_y	from 2 to 16	monthday_average
dayofmonth_x	lag_17	yearday_average
dayofmonth_y	lag_21	week_average
dayofyear_x	lag_23	month_average
dayofyear_y	lag_28	
week_x	lag_29	
week_y	lag_30	
month_x	lag_36	
month_y	lag_35	
is_weekend	lag_351	
is_holiday	Lag_358	

When considering the linear relation between all 44 predictors given in Table 1 and the time series, the linear regression, RR, and LR models have an explanation capacity of approximately 84% on the change of the time series for the test data. The linear regression and RR models, which use all available features, result in more complex equations that can be challenging to interpret. LR, however, generates a more interpretable model while preserving explanation capacity. Hence, it is decided as the forecasting model for the consumption of gasoline. By excluding the other 14 features with zero coefficients, LR produces a model with 30 non-zero coefficient features. The features having a non-zero coefficient are shown in Figure 6. The predictors monthX, monthY, dayOfMonthX, dayOfYearX, dayOfYearY, weekAverage, and Lags of 9, 10, 11, 13, 16, 21, 23, 28 and 29 have no effect on the prediction of gasoline consumption.

The R^2 remains constant (approximately 82%) with the LR for the test data despite a reduction in the number of variables having a non-zero coefficient. Table 2 gives the R^2 values and the MAPE level for training, validation, and test data.

While three regression models have similar performance levels for the partitions of data, LR produces slightly better performance for training and test data, according to the values of both MAPE and R^2.

Figure 7 shows the forecasting pattern for the validation data of the LR model. The model predictions deviate from actual validation values as a 4.34% MAPE overall. Upper and lower bounds have been drawn according to limit values imposed by train data sign as an outlier observation in the actual data. This is the reason why the prediction error occurred at this level.

Figure 8 shows the forecasting pattern for the test data of the LR model. The prediction error for the test data is recorded at 4.39%. The test data has more outlier values considering the upper and lower bounds imposed by the training data merged by the test data.

As a result of the proposed forecasting methodology, appropriate strategies for gasoline consumption will be developed with the forecasted daily consumption of gasoline and more accurate planning activities will be developed for the countries. By forecasting the daily consumption of gasoline, the gasoline amount that will be imported will be determined, so that the correct planning and timely implementation of gasoline trading will be realized. By means of this proposed forecasting methodology, which is applied and verified for Türkiye, more planned steps can be taken regarding the gasoline consumption in all countries of the world, and hence a gasoline supply chain system can be established.

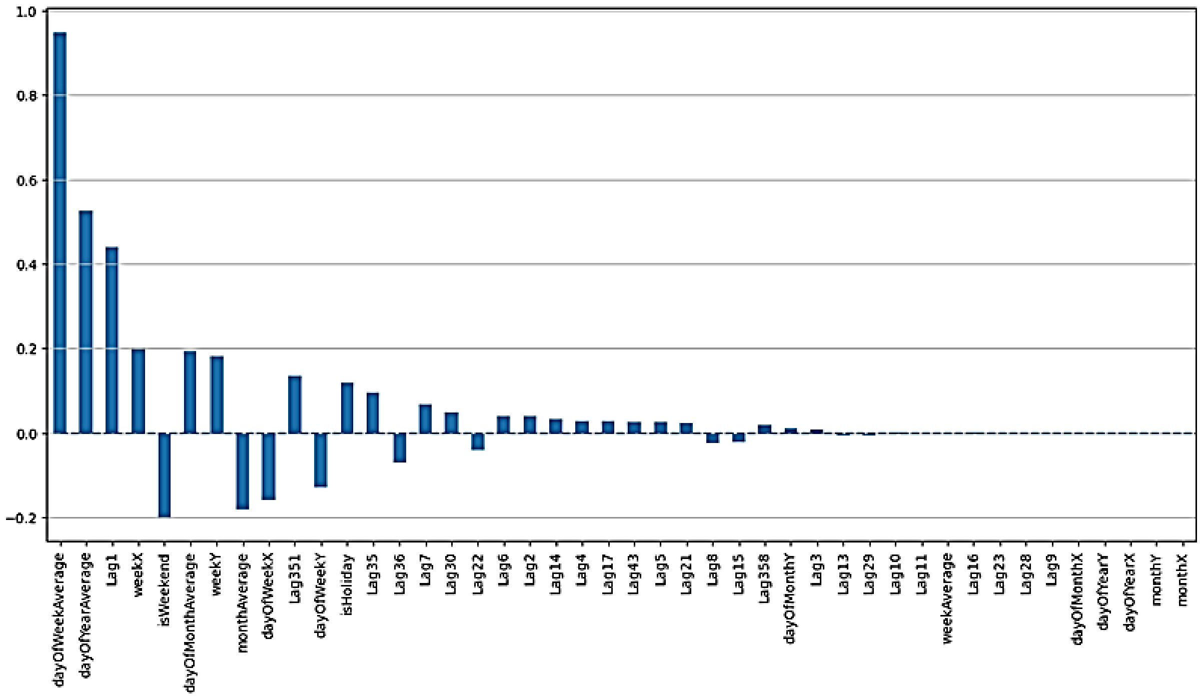


Figure 6. Coefficient of LR model.

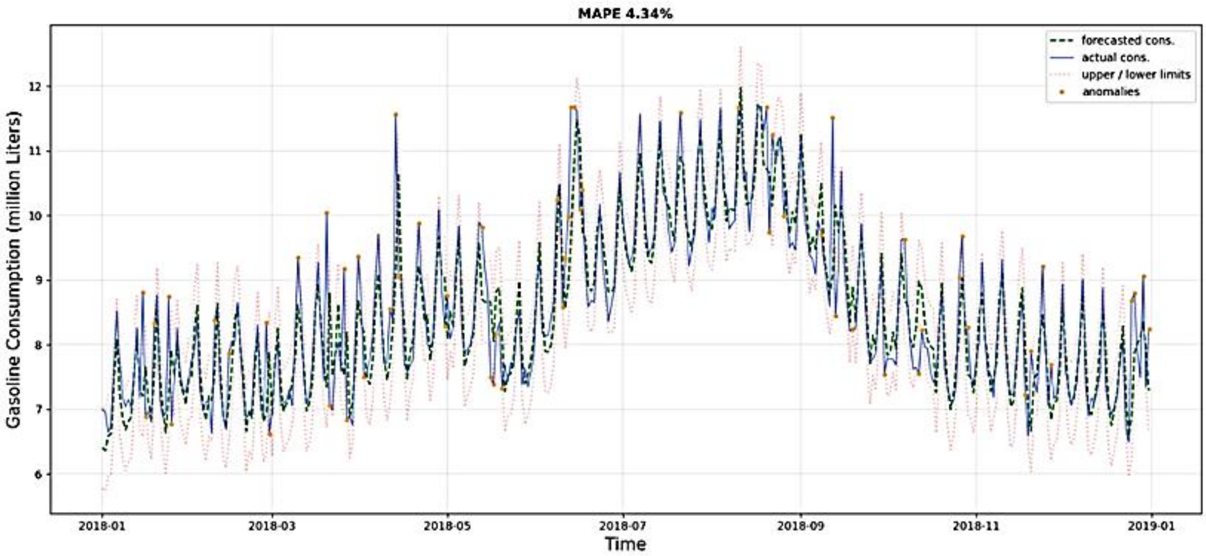


Figure 7. Forecasting pattern for the validation data of the LR model.

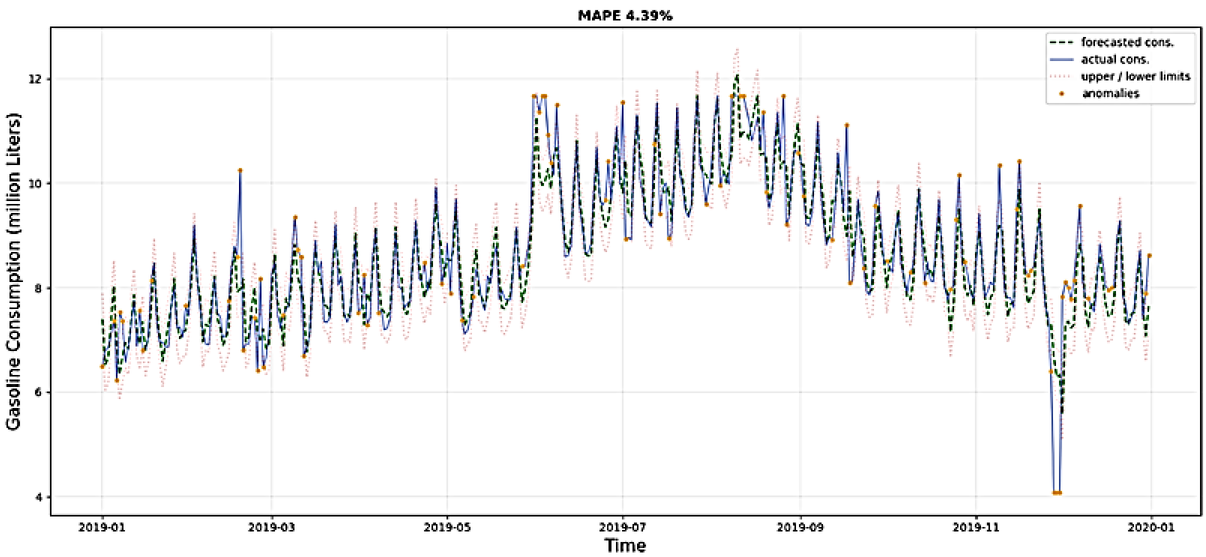


Figure 8. Forecasting pattern for the test data of LR model.

Table 2. R^2 and MAPE values of models.

	Linear regression		Ridge regression		Lasso regression	
	R^2	MAPE	R^2	MAPE	R^2	MAPE
Training	0.837	4.68%	0.835	4.68%	0.835	4.67%
Validation	0.828	4.27%	0.826	4.32%	0.826	4.34%
Test	0.841	4.40%	0.839	4.41%	0.839	4.39%

5. Policy implications

There is a need to implement a more dynamic policy framework that recognizes the variability in daily gasoline consumption. This includes the use of real-time data to improve accuracy and responsiveness, as well as regular updates and adjustments to forecasting models. To improve the accuracy of daily gasoline consumption forecasts, policymakers should consider incorporating advanced forecasting techniques such as lasso regression. This study shows that lasso regression is a reliable tool for planners and policymakers and demonstrates its effectiveness in this context. To address the environmental impacts of gasoline consumption, environmental considerations must be integrated into the policy framework. In this context, supply chain planning should be aligned with sustainability goals. Using alternative energy sources should be promoted, and environmentally friendly practices should be adopted in the gasoline industry. For the successful implementation of the proposed recommendations, cooperation should be established with the petrol industry, research institutions and environmental organizations. To support the development and implementation of changes, adequate funding should be provided. Policy makers should invest in the necessary technological infrastructure to facilitate the seamless integration of improved forecasting methods into the existing management systems of the gasoline supply chain. Integrating environmental considerations into policy frameworks is consistent with global sustainability goals, helps reduce environmental impacts, and promotes a more sustainable gasoline industry. By increasing policy framework adaptability and working with industry stakeholders, policymakers can contribute to the gasoline sector economic stability by ensuring a flexible and responsive supply chain.

The proposed policy implications aim to address the identified gaps in the current policy and provide practical, cost-effective, and socially acceptable solutions. By implementing these recommendations, policymakers can promote positive change, improve the status quo, and contribute to the overall well-being of society.

6. Discussion and Conclusion

Currently, the need for gasoline is increasing day by day. Gasoline is used in almost every industry. Gasoline also has great economic importance. When gasoline consumption continues to have a significant impact on economies worldwide, making accurate forecasts about its consumption is crucial. Such forecasts are essential for implementing necessary actions and developing proper economic strategies for these countries. However, besides these forecasts, it is crucial to predict the gasoline consumption daily which has a huge impact on the environment is currently considered important to make a quick measure to reduce environmental pollution. In this way, accurate planning can be made in the accomplishment of gasoline supply between the customers and suppliers, and it can create a smooth supply chain system for an indispensable product, like gasoline. The daily gasoline consumption that plays a key role in planning gasoline supply and distribution required for the continuum of public and private sector operations should be forecasted. A dynamic forecasting model that can be used daily, especially for short-term operations, is introduced into the literature. The gasoline consumption for Türkiye is forecasted daily with this model using the last six years of data. For this aim, the daily gasoline consumption in Türkiye is forecasted using lasso regression in this paper. The time series for daily gasoline consumption is analyzed based on the feature extraction technique. The data for daily gasoline consumption is created using official data, then the data is cleaned. For this purpose, the abnormal observations in data are updated with the new values calculated by Pchip interpolation. Subsequently, data is analyzed, and features are selected according to the results of the analysis. After the features are determined, values have complied, and the best lasso regression model is structured with these features. As a result of the study, it is shown that the lasso regression is a useful approach to be used to accurately forecast daily gasoline consumption.

The effectiveness of this approach in achieving accurate daily forecasts is demonstrated through the systematic application of feature extraction techniques, data cleaning, and model structuring using lasso regression. Importantly, the study establishes a critical link between daily gasoline consumption, environmental impact, and supply and distribution system planning. By highlighting the need for daily forecasts, the study contributes to a more comprehensive understanding of the complex dynamics that shape gasoline demand. In addition, the proposed forecasting methodology serves as a guide for building a resilient supply/distribution infrastructure that can more effectively respond to the diverse needs of both the public and private sectors. Through these contributions, this research not only provides practical insights for decision makers but also lays the foundation for future efforts to develop and expand forecasting techniques for a more sustainable and responsive gasoline supply chain.

This study can be a guide to establish a smooth and secure collection/distribution infrastructure on supplying gasoline, giving a more effective response to the public and private sectors. The proposed forecasting methodology can be a reference for the uninterrupted flow of gasoline supply and solutions where customers and suppliers can be reached at the maximum level.

6.1 Limitations and future recommendations

There are some limitations despite the findings of this study. The primary limitation is the use of lasso regression as the forecasting method. Lasso regression has proven to be effective in predicting daily gasoline consumption. However, all possible alternative methods can be explored in this study. Future research could benefit from a comparative analysis of different forecasting approaches, such as artificial neural networks or time series models, to improve the robustness of the forecasts. For future directions, the approach outlined in this study could be expanded by incorporating updated data or employing different forecasting methods. Techniques such as artificial neural networks, ARIMA, and Seasonal ARIMA (SARIMA) could compare with the current results. Together with Autometrics, the General to Specific approach shows promise in improving the accuracy of daily gasoline consumption forecasting and can be used for future studies.

Another limitation is that the study focuses only on gasoline consumption in Türkiye. Generalizing the results to other regions or countries may require caution due to specific economic, environmental, and geopolitical factors affecting gasoline demand. A more comprehensive understanding of the factors influencing daily gasoline consumption could be achieved with a broader data set covering different regions. The generalizability of the findings could be enhanced by expanding the study to include international comparisons. Examining the daily gasoline consumption patterns of different countries or regions will allow for a more detailed understanding of the factors that influence demand and enable the development of more adaptable and globally applicable forecasting models.

Future research could further explore the integration of environmental variables into forecasting models, building on the current study. A more sustainable and environmentally friendly approach to supply chain planning could be achieved by examining the impact of environmental policies, climate factors, or alternative energy sources on daily gasoline consumption.

Author contributions

Ertugrul Ayyildiz: Conceptualization, Methodology, Software, Data curation, Original draft preparation, Validation. **Mirac Murat:** Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. World Economic Outlook (2016). World Economic Outlook, April 2016: Too Slow for Too Long.
2. The World Bank (2023). The World Bank in Türkiye. https://www.worldbank.org/en/country/turkey/overview
3. T.C. Ticaret Bakanlığı (2023). Ekonomik Görünüm Mayıs 2023.
4. Ertuğrul, N. A., Bağcı, Z. H., & Ertuğrul, Ö. L. (2018). Aquifer thermal energy storage systems: Basic concepts and general design methods. Turkish Journal of Engineering, 2(2), 38-48. https://doi.org/10.31127/tuje.340334
5. Koyunoğlu, C. (2024). The economic case for blend fuels: A cost-benefit analysis in the European context. Sustainable Technology and Entrepreneurship, 3(2), 100060. https://doi.org/10.1016/j.stae.2023.100060
6. Comert, M., & Yildiz, A. (2021). A novel artificial neural network model for forecasting electricity demand enhanced with population-weighted temperature mean and the unemployment rate. Turkish Journal of Engineering, 6(2), 178-189. https://doi.org/10.31127/tuje.903876
7. Park, S. Y., & Yoo, S. H. (2014). The dynamics of oil consumption and economic growth in Malaysia. Energy Policy, 66, 218-223. https://doi.org/10.1016/j.enpol.2013.10.059
8. Mikayilov, J. I., Mukhtarov, S., Dinçer, H., Yüksel, S., & Aydın, R. (2020). Elasticity analysis of fossil energy sources for sustainable economies: A case of gasoline consumption in Turkey. Energies, 13(3), 731. https://doi.org/10.3390/en13030731
9. PETDER (2019). PETDER Sector Report.
10. Aydın, Ü., Peker, H., & Gönülalan, A. U. (2020). Petrol Sektörü. Türkiye'nin Enerji Görünümü, 161-214
11. Ma, H., & Zhang, Z. (2009). Grey prediction with Markov-Chain for Crude oil production and consumption in China. In The Sixth International Symposium on Neural Networks (ISNN 2009), 551-561. https://doi.org/10.1007/978-3-642-01216-7_58
12. Wang, Q., & Song, X. (2019). Forecasting China's oil consumption: a comparison of novel nonlinear-dynamic grey model (GM), linear GM, nonlinear GM and metabolism GM. Energy, 183, 160-171. https://doi.org/10.1016/j.energy.2019.06.139
13. Yang, Z. Y., Chai, A. H., Yang, Y. F., Li, X. M., Li, P., & Dai, R. Y. (2016). The semiflexible polymer translocation into laterally unbounded region between two parallel flat membranes. Polymers, 8(9), 332. https://doi.org/10.3390/polym8090332
14. Nel, W. P., & Cooper, C. J. (2008). A critical review of IEA's oil demand forecast for China. Energy Policy, 36(3), 1096-1106. https://doi.org/10.1016/j.enpol.2007.11.025
15. Azadeh, A., Moghaddam, M., Khakzad, M., & Ebrahimipour, V. (2012). A flexible neural network-fuzzy mathematical programming algorithm for improvement of oil price estimation and forecasting. Computers & Industrial Engineering, 62(2), 421-430. https://doi.org/10.1016/j.cie.2011.06.019
16. Narayan, P. K., & Wong, P. (2009). A panel data analysis of the determinants of oil consumption: the

case of Australia. Applied Energy, 86(12), 2771-2775. https://doi.org/10.1016/j.apenergy.2009.04.035

17. Yang, Y., Chen, Y., Shi, J., Liu, M., Li, C., & Li, L. (2016). An improved grey neural network forecasting method based on genetic algorithm for oil consumption of China. Journal of Renewable and Sustainable Energy, 8(2), 024104. https://doi.org/10.1063/1.4944977
18. Li, J., Wang, R., Wang, J., & Li, Y. (2018). Analysis and forecasting of the oil consumption in China based on combination models optimized by artificial intelligence algorithms. Energy, 144, 243-264. https://doi.org/10.1016/j.energy.2017.12.042
19. Assareh, E., Behrang, M. A., Assari, M. R., & Ghanbarzadeh, A. (2010). Application of PSO (particle swarm optimization) and GA (genetic algorithm) techniques on demand estimation of oil in Iran. Energy, 35(12), 5223-5229. https://doi.org/10.1016/j.energy.2010.07.043
20. Lin, B., & Xie, C. (2013). Estimation on oil demand and oil saving potential of China's road transport sector. Energy Policy, 61, 472-482. https://doi.org/10.1016/j.enpol.2013.06.017
21. Rao, R. D., & Parikh, J. K. (1996). Forecast and analysis of demand for petroleum products in India. Energy Policy, 24(6), 583-592. https://doi.org/10.1016/0301-4215(96)00019-5
22. Wang, Q., & Song, X. (2019). Forecasting China's oil consumption: a comparison of novel nonlinear-dynamic grey model (GM), linear GM, nonlinear GM and metabolism GM. Energy, 183, 160-171. https://doi.org/10.1016/j.energy.2019.06.139
23. Duan, H., Lei, G. R., & Shao, K. (2018). Forecasting crude oil consumption in China using a grey prediction model with an optimal fractional-order accumulating operator. Complexity, 3869619 https://doi.org/10.1155/2018/3869619
24. Behrang, M. A., Assareh, E., Ghalambaz, M., Assari, M. R., & Noghrehabadi, A. R. (2011). Forecasting future oil demand in Iran using GSA (Gravitational Search Algorithm). Energy, 36(9), 5649-5654. https://doi.org/10.1016/j.energy.2011.07.002
25. Minniear, M. P. (2000). Forecasting the permanent decline in global petroleum production. Journal of Geoscience Education, 48(2), 130-136. https://doi.org/10.5408/1089-9995-48.2.130
26. Al-Qaness, M. A., Abd Elaziz, M., & Ewees, A. A. (2018). Oil consumption forecasting using optimized adaptive neuro-fuzzy inference system based on sine cosine algorithm. IEEE Access, 6, 68394-68402. https://doi.org/10.1109/ACCESS.2018.2879965
27. Fatima, T., Xia, E., & Ahad, M. (2019). Oil demand forecasting for China: a fresh evidence from structural time series analysis. Environment, Development and Sustainability, 21, 1205-1224. https://doi.org/10.1007/s10668-018-0081-7
28. Yu, L., Zhao, Y., Tang, L., & Yang, Z. (2019). Online big data-driven oil consumption forecasting with Google trends. International Journal of Forecasting, 35(1), 213-223. https://doi.org/10.1016/j.ijforecast.2017.11.005
29. Keshavarzian, M., Anaraki, S. K., Zamani, M., & Erfanifard, A. (2012). Projections of oil demand in road transportation sector on the basis of vehicle ownership projections, worldwide: 1972–2020. Economic Modelling, 29(5), 1979-1985. https://doi.org/10.1016/j.econmod.2012.06.009
30. Sadri, A., Ardehali, M. M., & Amirnekooei, K. (2014). General procedure for long-term energy-environmental planning for transportation sector of developing countries with limited data based on LEAP (long-range energy alternative planning) and EnergyPLAN. Energy, 77, 831-843. https://doi.org/10.1016/j.energy.2014.09.067
31. Melikoglu, M. (2014). Demand forecast for road transportation fuels including gasoline, diesel, LPG, bioethanol and biodiesel for Turkey between 2013 and 2023. Renewable Energy, 64, 164-171. https://doi.org/10.1016/j.renene.2013.11.009
32. Azadeh, A., Behmanesh, I., Vafa Arani, H., & Sadeghi, M. H. (2014). An integrated fuzzy mathematical programming-analysis of variance approach for forecasting gasoline consumption with ambiguous inputs: USA, Canada, Japan, Iran and Kuwait. International Journal of Industrial and Systems Engineering, 18(2), 159-184. https://doi.org/10.1504/IJISE.2014.064704
33. Sapnken, E. F., Tamba, J. G., Essiane, S. N., Koffi, F. D., & Njomo, D. (2018). Modeling and forecasting gasoline consumption in Cameroon using linear regression models. International Journal of Energy Economics and Policy, 8(2), 111-120.
34. Anggarani, R., & Watada, J. (2012). A gasoline consumption model based on the harmony search algorithm: Study case of Indonesia. Intelligent Decision Technologies, 6(3), 233-241. https://doi.org/10.3233/IDT-2012-0139
35. Chen, H., Tong, Y., & Wu, L. (2021). Forecast of energy consumption based on FGM (1, 1) model. Mathematical Problems in Engineering, 2021, 1-11. https://doi.org/10.1155/2021/6617200
36. Güngör, B. O., Ertuğrul, H. M., & Soytaş, U. (2021). Impact of Covid-19 outbreak on Turkish gasoline consumption. Technological Forecasting and Social Change, 166, 120637. https://doi.org/10.1016/j.techfore.2021.120637
37. Wang, Q., Li, S., & Li, R. (2018). Forecasting energy demand in China and India: Using single-linear, hybrid-linear, and non-linear time series forecast techniques. Energy, 161, 821-831. https://doi.org/10.1016/j.energy.2018.07.168
38. Wang, Q., Li, S., Li, R., & Ma, M. (2018). Forecasting US shale gas monthly production using a hybrid ARIMA and metabolic nonlinear grey model. Energy, 160, 378-387. https://doi.org/10.1016/j.energy.2018.07.047
39. Wang, Q., Li, S., Zhang, M., & Li, R. (2022). Impact of COVID-19 pandemic on oil consumption in the United States: a new estimation approach. Energy, 239, 122280. https://doi.org/10.1016/j.energy.2021.122280
40. Wang, Q., Li, S., & Jiang, F. (2021). Uncovering the impact of the COVID-19 pandemic on energy consumption: New insight from difference between pandemic-free scenario and actual electricity

consumption in China. Journal of Cleaner Production, 313, 127897. https://doi.org/10.1016/j.jclepro.2021.127897

41. Wang, Q., Li, S., Li, R., & Jiang, F. (2022). Underestimated impact of the COVID-19 on carbon emission reduction in developing countries–a novel assessment based on scenario analysis. Environmental Research, 204, 111990. https://doi.org/10.1016/j.envres.2021.111990
42. Ogutu, J. O., Schulz-Streeck, T., & Piepho, H. P. (2012). Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. BMC proceedings, 6, 1-6. https://doi.org/10.1186/1753-6561-6-S2-S10
43. Farvahari, A., Gozashti, M. H., & Dehesh, T. (2019). The usage of lasso, ridge, and linear regression to explore the most influential metabolic variables that affect fasting blood sugar in type 2 Diabetes patients. Romanian Journal of Diabetes Nutrition and Metabolic Diseases, 26(4), 371-379. https://doi.org/10.2478/rjdnmd-2019-0040
44. Altelbany, S. (2021). Evaluation of ridge, elastic net and lasso regression methods in precedence of multicollinearity problem: A simulation study. Journal of Applied Economics and Business Studies, 5(1), 131-142. https://doi.org/10.34260/jaebs.517
45. McNeish, D. M. (2015). Using lasso for predictor selection and to assuage overfitting: A method long overlooked in behavioral sciences. Multivariate Behavioral Research, 50(5), 471-484. https://doi.org/10.1080/00273171.2015.1036965
46. Trivedi, U. B., Bhatt, M., & Srivastava, P. (2021). Prevent overfitting problem in machine learning: a case focus on linear regression and logistics regression. Innovations in Information and Communication Technologies (IICT-2020) Proceedings of International Conference on ICRIHE-2020, Delhi, India: IICT-2020, 345-349. https://doi.org/10.1007/978-3-030-66218-9_40
47. Jovanovic, M., Radovanovic, S., Vukicevic, M., Van Poucke, S., & Delibasic, B. (2016). Building interpretable predictive models for pediatric hospital readmission using Tree-Lasso logistic regression. Artificial İntelligence in Medicine, 72, 12-21. https://doi.org/10.1016/j.artmed.2016.07.003
48. Jacobson, N. C., Chow, S. M., & Newman, M. G. (2019). The Differential Time-Varying Effect Model (DTVEM): A tool for diagnosing and modeling time lags in intensive longitudinal data. Behavior Research Methods, 51, 295-315. https://doi.org/10.3758/s13428-018-1101-0
49. Chapman, C. S., Gallivan, J. P., Wood, D. K., Milne, J. L., Culham, J. C., & Goodale, M. A. (2010). Reaching for the unknown: multiple target encoding and real-time decision-making in a rapid reach task. Cognition, 116(2), 168-176. https://doi.org/10.1016/j.cognition.2010.04.008
50. T.C. Enerji ve Tabi Kaynaklar Bakanlığı (2020). EIGM Raporları
51. Keogh, E., Chu, S., Hart, D., & Pazzani, M. (2004). Segmenting time series: A survey and novel approach. In Data mining in time series databases, 1-21. https://doi.org/10.1142/9789812565402_0001
52. Pereira, C. V. (2021). Portfolio efficiency with high-dimensional data as conditioning information. International Review of Financial Analysis, 77, 101811. https://doi.org/10.1016/j.irfa.2021.101811
53. Çınaroğlu, S. (2017). Sağlık harcamasının tahmininde makine öğrenmesi regresyon yöntemlerinin karşılaştırılması. Uludağ Üniversitesi Mühendislik Fakültesi Dergisi, 22(2), 179-200. https://doi.org/10.17482/uumfd.338805
54. Teixeira, A. L., Leal, J. P., & Falcao, A. O. (2013). Random forests for feature selection in QSPR Models-an application for predicting standard enthalpy of formation of hydrocarbons. Journal of Cheminformatics, 5, 1-15. https://doi.org/10.1186/1758-2946-5-9
55. Omranian, N., Eloundou-Mbebi, J. M., Mueller-Roeber, B., & Nikoloski, Z. (2016). Gene regulatory network inference using fused LASSO on multiple data sets. Scientific reports, 6(1), 20533. https://doi.org/10.1038/srep20533
56. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
57. Hastie, T., Tibshirani, R., & Wainwright, M. (2015). Statistical learning with sparsity: the lasso and generalizations. CRC Press.
58. Wang, F., Liu, X., Liu, C., Li, H., & Han, Q. (2018). Remaining useful life prediction method of rolling bearings based on Pchip-EEMD-GM (1, 1) model. Shock and vibration, 3013684. https://doi.org/10.1155/2018/3013684
59. Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. Journal of Econometrics, 54(1-3), 159-178. https://doi.org/10.1016/0304-4076(92)90104-Y
60. Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. Malawi Medical Journal, 24(3), 69-71.