Time Series Analysis On Fuel Prices

KANISHK DAD (202103005),* ARSH JINDAL (202103021),† and AARZOO KHAMBHOO (202103026)‡
Dhirubhai Ambani Institute of Information & Communication Technology,
Gandhinagar, Gujarat 382007, India
SC-475, Time Series Analysis

In this project, we have analysed the time series data of the weekly data set comprising fuel prices in Italy over 18 years taken from the government site which is publicly available to use. We went through many different time series models using different methods to check the stationarity and patterns in our data and finally achieved the best fit using an ARIMA model.

I. INTRODUCTION

In this report, we explore the history of fuel prices in Italy over 18 years. Fuel prices are a crucial economic indicator. It reflects not only market trends but also geopolitical shifts, environmental policies, and consumer behaviour. In this report, we do a detailed analysis of eight distinct fuel types, with our aim being, to discover the underlying patterns, trends, and anomalies within the dataset.

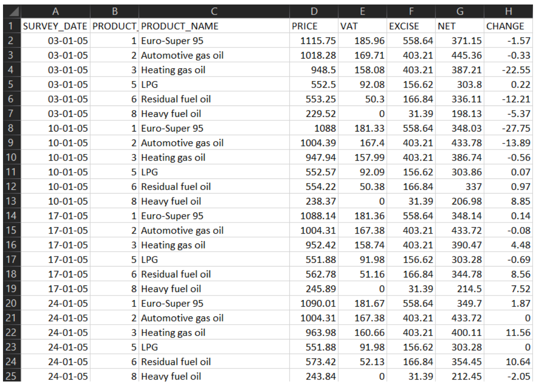
Our primary focus in this report is on the time series analysis of one fuel product: Euro-Super 95, given by product ID 1 in our dataset. We try to extract meaningful information through the time series methods and statistical techniques taught to us in class, we try to get a deeper understanding of the fuel pricing in the Italian market.

II. THE DATASET

In our project, we have used the data available openly on site. The data shows Italy's weekly average fuel price for 6 different types of fuels namely, Euro-Super 95, Automotive gas oil, Heating gas oil, LPG, Residual fuel oil, and Heavy fuel oil for which the product ID that were given are 1, 2, 3, 5, 6, 8 (the same have been used in code and report), over a span of 19 years approximately for the year 2005 to 2024. We have performed the necessary Exploratory data analysis on the dataset and refined it to our needs.

However, we have used only three columns for this project, SURVEY DATE, PRODUCT ID, and PRICE, for understanding. For modelling purposes, we have used the fuel with product ID 1.

*Electronic address: 202103005@daiict.ac.in
†Electronic address: 202103021@daiict.ac.in
‡Electronic address: 202103026@daiict.ac.in



	A	B	C	D	E	F	G	H
1	SURVEY_DATE	PRODUCT_	PRODUCT_NAME	PRICE	VAT	EXCISE	NET	CHANGE
2	03-01-05	1	Euro-Super 95	1115.75	185.96	558.64	371.15	-1.57
3	03-01-05	2	Automotive gas oil	1018.28	169.71	403.21	445.36	-0.33
4	03-01-05	3	Heating gas oil	948.5	158.08	403.21	387.21	-22.55
5	03-01-05	5	LPG	552.5	92.08	156.62	303.8	0.22
6	03-01-05	6	Residual fuel oil	553.25	50.3	166.84	336.11	-12.21
7	03-01-05	8	Heavy fuel oil	229.52	0	31.39	198.13	-5.37
8	10-01-05	1	Euro-Super 95	1088	181.33	558.64	348.03	-27.75
9	10-01-05	2	Automotive gas oil	1004.39	167.4	403.21	433.78	-13.89
10	10-01-05	3	Heating gas oil	947.94	157.99	403.21	386.74	-0.56
11	10-01-05	5	LPG	552.57	92.09	156.62	303.86	0.07
12	10-01-05	6	Residual fuel oil	554.22	50.38	166.84	337	0.97
13	10-01-05	8	Heavy fuel oil	238.37	0	31.39	206.98	8.85
14	17-01-05	1	Euro-Super 95	1088.14	181.36	558.64	348.14	0.14
15	17-01-05	2	Automotive gas oil	1004.31	167.38	403.21	433.72	-0.08
16	17-01-05	3	Heating gas oil	952.42	158.74	403.21	390.47	4.48
17	17-01-05	5	LPG	551.88	91.98	156.62	303.28	-0.69
18	17-01-05	6	Residual fuel oil	562.78	51.16	166.84	344.78	8.56
19	17-01-05	8	Heavy fuel oil	245.89	0	31.39	214.5	7.52
20	24-01-05	1	Euro-Super 95	1090.01	181.67	558.64	349.7	1.87
21	24-01-05	2	Automotive gas oil	1004.31	167.38	403.21	433.72	0
22	24-01-05	3	Heating gas oil	963.98	160.66	403.21	400.11	11.56
23	24-01-05	5	LPG	551.88	91.98	156.62	303.28	0
24	24-01-05	6	Residual fuel oil	573.42	52.13	166.84	354.45	10.64
25	24-01-05	8	Heavy fuel oil	243.84	0	31.39	212.45	-2.05

FIG. 1: DATASET

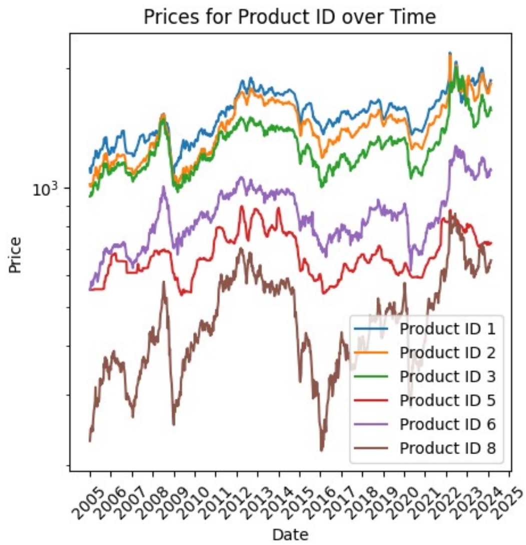


FIG. 2: Prices of Fuels for every product ID over time

III. DATA ANALYSIS

Moving Average: Initially, after observing our dataset we began our analysis by using the method of moving average. The below diagram shows the fuel price over time along with its moving average counterpart. The process of moving the average smoothens the curve and helps in observing the trend. The yellow curve in the figure shows the approximate trend in our fuel prices over the years.

FIG. 3: Moving average of prices over time

Detrending using moving average: Next, to make our data stationary, we tried to detrend our data. The below figure shows the detrended data using two different methods, the additive model and the multiplicative model. In a time series model, the trend can be in added form to the stationary part or could be in multiplicative form. To make sure we cover both ways, we showed the detrended data using both methods.

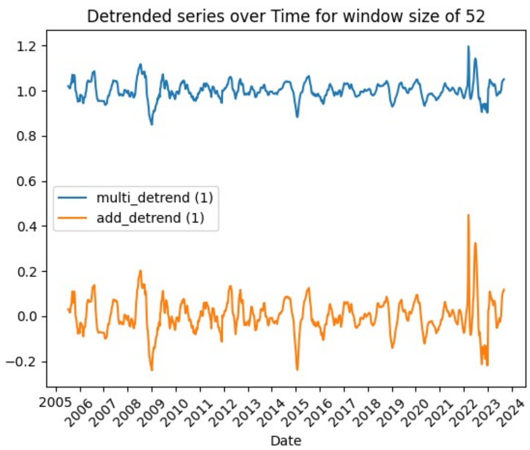


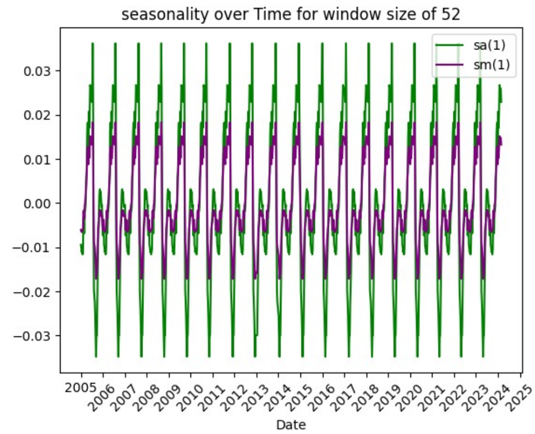
FIG. 4: Detrended series over time

To show the graph on a similar scale we scaled down the additive model by 800.

Seasonality

Next, we checked for seasonality in our data. In the seasonality too we checked for both multiplicative and additive models.

In the figure, we can see that there is a difference in both the model's output for seasonality. To make sure there is not any scaling difference. We scaled down the additive model by 800 and also subtracted the multiplicative model by 1 to get a similar scale. However, this scaling works well only for product 1. So, for different products, we will be using different scaling.



[h]

FIG. 5: Seasonality over time

IV. RESIDUAL

Once, we have the seasonality and trend of the time series, we can get the stationary part to fit the model by removing the seasonality and trend and getting the residual

The below figure shows the residual after removing the seasonality and trend. Here, too we are checking for both the multiplicative model and the additive model.

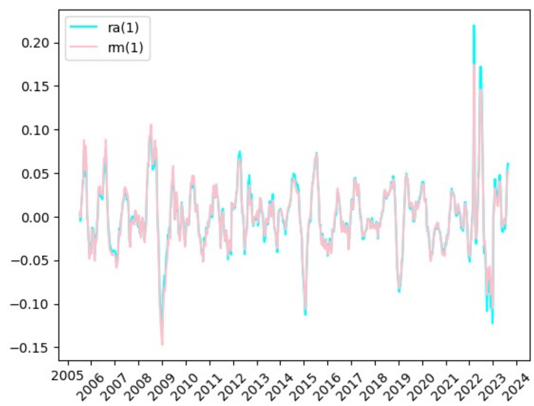


FIG. 6: Residual after removing the seasonality and trend

Once we have the residual data, we can try to find the autocorrelation of this data, to check which model will be best fit for our data.

Once we have the residual, we check for the stationarity of the residual data. We can check for the stationarity of our data using the ADF(Augmented Dickey-Fuller).

When we plot the autocorrelation function of the residual, we see that the autocorrelation of the residual depends upon time and not just on lag. This shows that this data is not stationary.

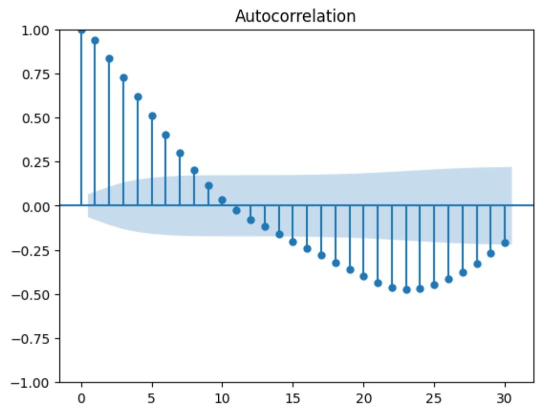


FIG. 7: Autocorrelation of the residual data showing non-stationarity

V. USING LOARITHM

Since the residual data is not stationary, we then tried to check the stationarity of data by applying log on the data.

Log helps in getting a better representation of data and also helps in getting a better scale. However, we can see in the figure that the autocorrelation of data after applying log is not independent of time, i.e. that the autocorrelation is dependent upon time and not just lag. So we can conclude that this data is not stationary and try another approach.

VI. ANALYZING DIFFERENCED DATA

Since the residual after detreding and removing seasonality is not stationary, we will now check for the stationarity of the data after the difference of two consecutive values.

On applying the ADF test on the differenced data, we observe that the data is stationary, thus we can now try to fit a time series model on this data. This statement is also supported by the below figure, which shows that the autocorrelation and partial-autocorrelation of the differenced data are independent of time.

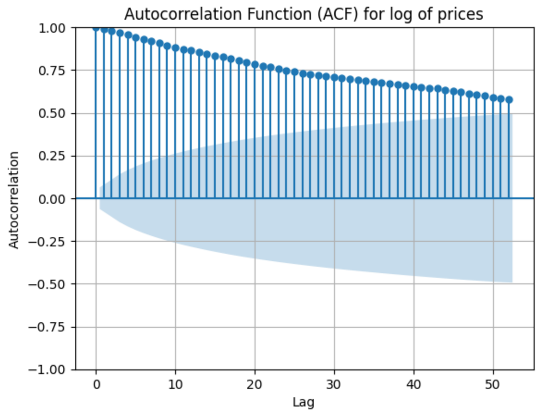


FIG. 8: Autocorrelation of the log data showing non-stationarity

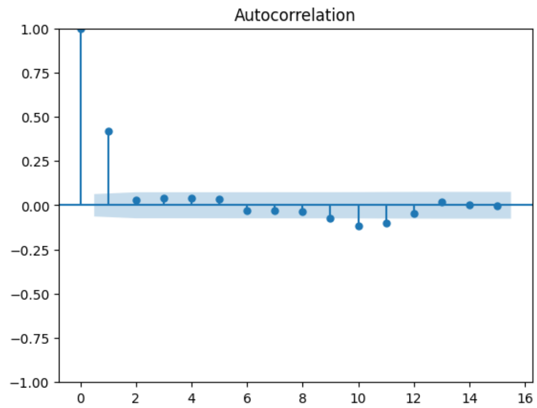


FIG. 9: Autocorrelation of the differenced data after removing null values

The autocorrelation of the differenced data is 0.5 at lag 1 and 0 at all other values, and the partial-autocorrelation function is 1 at lag 1 and -0.25 at lag 2.

The autocorrelation at lag 1 being 0.5 indicates a moderate positive correlation, while the autocorrelation at all other lags being 0 suggests that the data has no significant autocorrelation beyond lag 1. The partial autocorrelation at lag 1 being 1 and at lag 2 being -0.25, with 0 at all other lags, suggests a possible ARIMA(1,0,2) model on the differnced data.

Since this is on 1 differed data, we come to the conclusion of using an ARIMA(1, 1, 2) model on our original data.

VII. ARIMA(1, 1, 2)

We used an ARIMA(1, 1, 2) model to analyze our time series data. This model comprises an autoregressive component of order 1 (AR(1)), accounting for the influence of

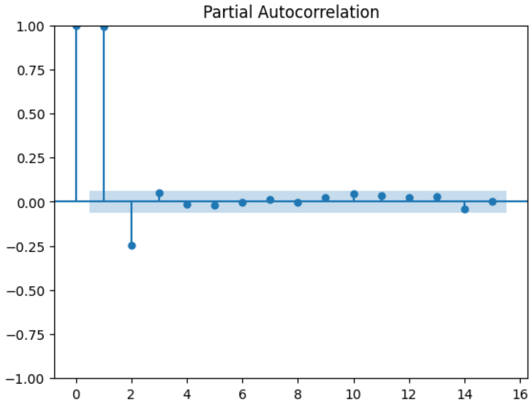


FIG. 10: Partial-autocorrelation of the differenced data

the previous observation, an integrated component of order 1 (I(1)), indicating differencing to achieve stationarity, and a moving average component of order 2 (MA(2)), capturing the effects of past forecast errors.

VIII. PREDICTION

After obtaining ARIMA(1, 1, 2) as the best fit for our data, we try predicting the values of 5 future values. The forecasted values for the next 5 days are:
961 1850.216207
962 1850.949371
963 1850.242452
964 1850.924066
965 1850.266852

The predictions are not perfect, but we can further improve these predictions by using machine learning techniques. Since fuel prices are not just dependent on the previous pattern, but also on other external factors, a simple time series model can not give us accurate predictions, but it can still provide us with a general overview and baseline results to move forward with.

IX. CONCLUSIONS

On applying time series analysis methods to our data on fuel prices, we can get an ARIMA(1, 1, 2) model to fit our data. This model can then be used to predict future fuel prices. However, since fuel prices are also affected by other external factors, this model further needs external information and machine learning techniques to get a better prediction. Also as we haven't used all the columns which include the effect of taxes, thus the data may be nonlinear with previous values. Future work on this data can include the study of the influence of external factors as well as including other factors from the data.

X. VIDEO PRESENTAION

Here is the link to our video presentation of our project.
link

[1]
[2]