The ebbs and flows of diesel prices have long played a pivotal role in shaping the economic landscape, particularly for sectors that lean heavily on transportation and logistics. In recent years, the need for a predictive compass to navigate these financial tides has become increasingly crucial.

This proposal presents the blueprint for a predictive model designed to anticipate US diesel sales prices with weekly updates. By tapping into a rich vein of historical data, this model aims to empower businesses and stakeholders with the foresight to make savvy, strategic decisions that can streamline operations and bolster financial resilience.

This initiative goes beyond mere number-crunching, it is about crafting a vision for futureproofing against the caprices of diesel price movements and positioning ourselves advantageously in a competitive marketplace.

**Introduction:**

Against a backdrop of fluctuating diesel costs, the imperative for a forward-looking approach to fiscal planning is more pressing than ever.

This report stands as a testament to that need, offering an in-depth analysis and forecast that promises to demystify the trajectory of future diesel prices. Employing a blend of statistical rigor and insightful data analysis, this document serves as both a beacon and a tool, guiding stakeholders through the turbulent seas of diesel price volatility. Herein lies a path charted through the meticulous deconstruction of data, employing proven forecasting methodologies to cast a light on what lies ahead.

This report is more than a collection of graphs and numbers, it is a strategic companion for those who seek to harness the power of information in the quest for economic stability and strategic agility.

**Problem Presentation**

In the face of fluctuating diesel prices, industries reliant on diesel fuel, including transportation and logistics, are grappling with the challenge of efficiently managing their financial and operational strategies.

The unpredictability of diesel prices makes it difficult for these industries to forecast expenses accurately, leading to budgeting inefficiencies, increased operational costs, and potential losses in competitive market positioning. Moreover, the inability to anticipate diesel price trends can hinder strategic planning, affecting procurement, pricing strategies, and overall financial health.

To navigate these challenges, there is a pressing need for a sophisticated approach that utilizes historical diesel price data to predict future trends. Developing a diesel price forecasting model is paramount for businesses looking to mitigate the adverse effects of price volatility.

Such a model would enable more accurate financial planning, optimize fuel procurement strategies, and enhance budgetary accuracy, ultimately safeguarding against market uncertainties and maintaining operational efficiency.

This report delves into the analysis of weekly diesel sales prices from a comprehensive dataset spanning several years. Our objective is to decipher trends and patterns within this data, harnessing this insight to construct a forecasting model capable of predicting diesel prices with greater accuracy.

By doing so, aimed to equip businesses in diesel-dependent industries with the tools necessary to achieve strategic foresight in their financial and operational planning, ensuring resilience against the unpredictable dynamics of diesel prices.

**Data Description**

The analysis utilizes a dataset from the Federal Reserve Economic Data (FRED) database, focusing on US diesel sales prices on a weekly basis over several years.

This dataset, rich in detail and scope, includes two main variables: 'Observation Date' and 'US Diesel Sales Prices (GASDESW)'.

The 'Observation Date' provides precise weekly timestamps, while 'GASDESW' records the diesel sales prices in the US, forming the basis of our time series analysis.

Through examining the fluctuations captured in these observations, the aim is to decipher patterns and trends that will underpin the development of a forecasting model.

This model seeks to predict diesel prices with accuracy, offering vital insights for strategic planning and decision-making in industries heavily reliant on diesel fuel.

**Data Analysis**

1. Time Series Plot of Weekly Diesel Sales Prices: The first plot generated provides a comprehensive overview of the diesel sales prices from the mid-1990s through to 2025.

This visualization highlights the overall trend and fluctuations in diesel prices over the years, showcasing periods of price stability as well as significant volatility, reflecting the complex dynamics affecting diesel prices.



1. Differenced Time Series Plot: The second plot focuses on the differenced weekly diesel sales prices, a critical step in making the time series stationary and suitable for further analysis. By differencing the data, we observe the changes in diesel prices from one week to the next, rather than the absolute price levels. This plot reveals the variability in price changes over time, highlighting periods of increased stability or volatility in diesel price movements, which are essential for understanding the data's underlying patterns and for developing accurate forecasting models.





Decomposition of additive time series

This is a classic time series decomposition plot of what is presumed to be diesel prices data, showcasing how the data can be broken down into trend, seasonal, and residual (irregular) components:

1. Top Panel (Original Data): This shows the actual observed diesel prices over time, which seem to exhibit a long-term increase, indicating a possible upward trend in diesel prices. The data spans from 1995 to just beyond 2020.

2. Second Panel (Trend Component): Here, the overall trend of diesel prices is shown. The long-term increase in prices is more apparent in this smoothed line, which represents the general direction in which the prices are moving, devoid of seasonal effects or random fluctuations.

3. Third Panel (Seasonal Component) : This panel represents the seasonality in the data – repeating patterns at regular intervals throughout the year. The sharp peaks and troughs indicate significant seasonal variation, which could be due to factors like changes in demand throughout the year, policy changes, or other cyclic factors affecting diesel prices.

4. Bottom Panel (Residual Component): After removing the trend and seasonal patterns, the residuals or remainder component highlights the irregularities or noise in the data. Spikes in this plot may indicate unexpected deviations in diesel prices due to unpredictable events such as economic crises, sudden policy changes, or other market disruptions.

The overall plot gives a comprehensive view of the underlying patterns in diesel prices, allowing analysts to account for these patterns when creating forecasts or understanding the price dynamics. For instance, the increasing trend might inform discussions on long-term investment or policy planning, while the seasonal patterns could be important for short-term inventory or price-setting strategies.

A screenshot of a test results

Description automatically generated

The output of two statistical tests: the Augmented Dickey-Fuller (ADF) test and the KPSS test, both of which are used to analyze the time series data for diesel prices.

1. Augmented Dickey-Fuller Test:

- This test checks for the presence of unit roots, which is a way to test if the time series is non-stationary.

- The Dickey-Fuller statistic is -2.9449 with a lag order of 11.

- The p-value is 0.1783, which is greater than the common significance level of 0.05, suggesting that we fail to reject the null hypothesis of the presence of a unit root. This implies that the diesel price time series may be non-stationary.

2. KPSS Test (Kwiatkowski-Phillips-Schmidt-Shin):

- The KPSS test has a different null hypothesis than the ADF test. It assumes that the series is stationary around a deterministic trend.

- The statistic is 13.179 with a truncation lag parameter of 7.

- The p-value of 0.01 is less than 0.05, leading us to reject the null hypothesis, suggesting the series is not stationary.

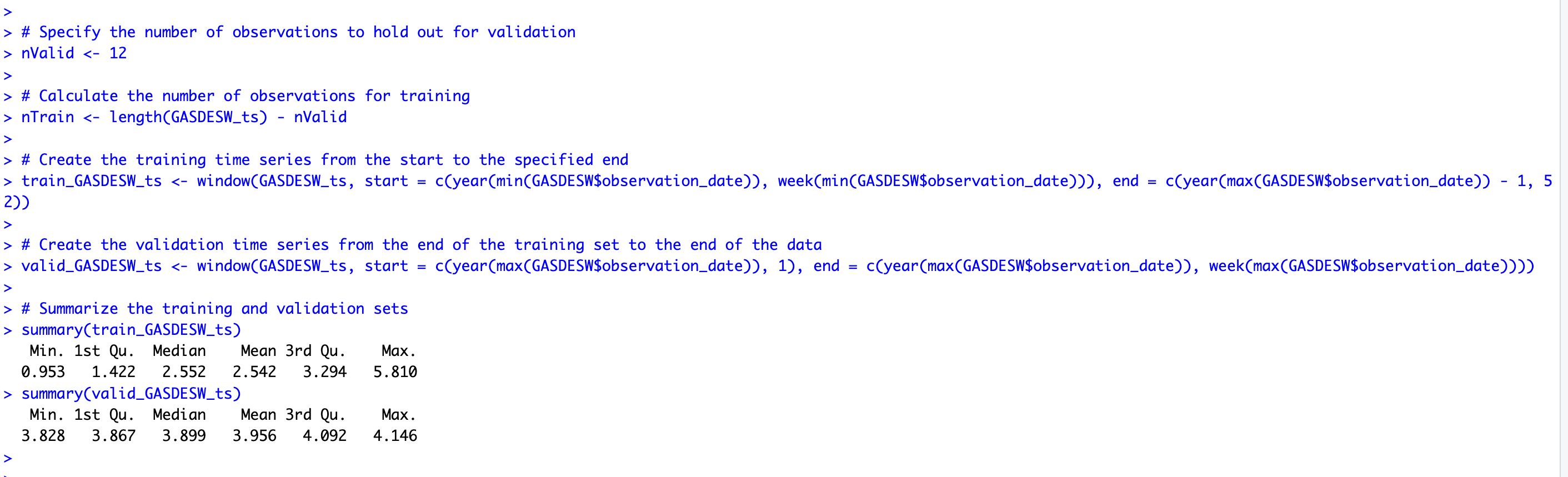
Combined Interpretation:

- The results from both tests seem to conflict, which can sometimes occur due to the different assumptions each test makes about the data.

- The ADF test suggests non-stationarity, while the KPSS indicates stationarity.

- This mixed result suggests that the time series of diesel prices may contain some form of deterministic trend or structural break that could be affecting the ADF test, or that there is a need for further differencing, which is common in the presence of a unit root.

- In practice, such conflicting results may lead a data analyst to further inspect the data, possibly with additional tests or model adjustments, to ascertain the true nature of the time series before proceeding with forecasts or other analyses.



From the output, we can deduce that:

The training set's diesel prices range from about 0.953 to 5.810, with a median price of approximately 2.552.

The validation set's prices range from about 3.828 to 4.146, with a median price of approximately 3.899.

This indicates that diesel prices in the validation set are generally higher than those in the training set, which could be due to inflation, changes in supply and demand, or other economic factors affecting fuel prices over time. These statistics will be vital for evaluating how well the forecast model performs when it's applied to the validation set.



Certainly, the graph titled "ACF for Differenced Series" reflects the autocorrelation of a differenced time series related to diesel prices. Here’s a concise explanation:

1. Initial Spike: The strong initial spike indicates that there's a significant correlation from one period to the next. This suggests that the price of diesel from the previous time point is a good predictor of the price in the next period.

2. Tapering Off: The rapid decline in autocorrelation as lags increase implies that as you move further away from the present, past diesel prices have progressively less influence on future prices.

3. Confidence Bounds: The fact that most spikes are within the confidence bounds (blue dashed lines) after the initial lag indicates randomness in the series, meaning that there are no obvious patterns in the diesel prices that persist over long periods.

4. Model Implications: For modeling diesel prices, the significant initial lag in the ACF suggests that a moving average component might be useful in an ARIMA model to predict future prices.

5. Differencing Effect: The differencing appears successful in stabilizing the mean of the time series since the autocorrelations quickly fall within the bounds, indicating a lack of trend or seasonality.

Overall, this ACF plot suggests that recent changes in diesel prices are informative for near-term predictions, but their influence diminishes for future prices as time goes on. This could be pivotal for industries reliant on diesel fuel, as it impacts cost planning and budget forecasting.



The graph displayed is the Partial Autocorrelation Function (PACF) for a differenced series of diesel prices. Here's a straightforward explanation:

1. Sharp Drop after Lag 1: The first lag showing a significant partial correlation and the sharp drop-off after suggests that only the most immediate past price (one period ago) has a direct influence on the current price when other lags are accounted for.

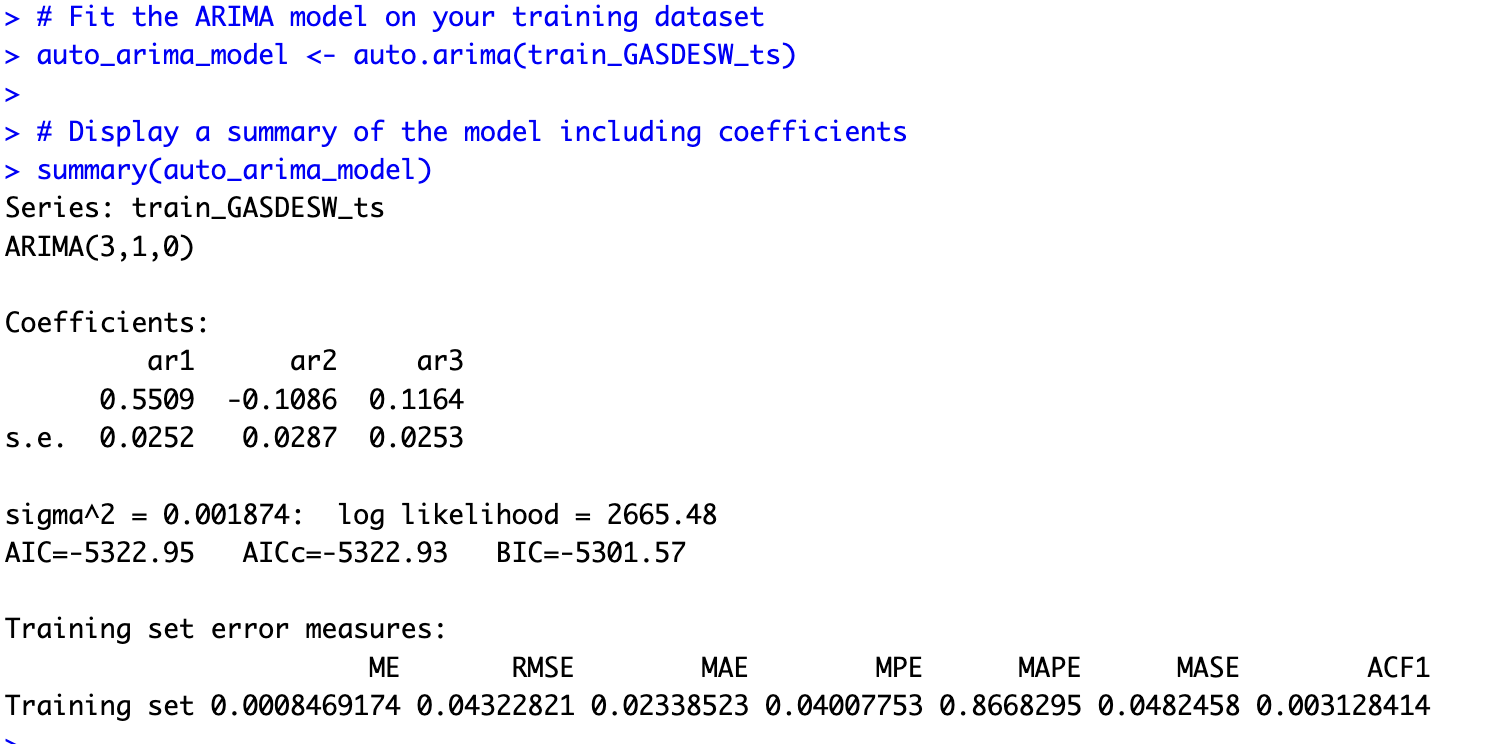
2. Minimal Correlations at Higher Lags: The PACF values being close to zero beyond the first lag indicates that past diesel prices, when considering the influence of intervening values, have little to no direct effect on the current price beyond the immediate past.

3. Model Indication: For an ARIMA model, this PACF plot suggests that an autoregressive term of order 1 might be significant, as the correlation at lag 1 is the only one that stands out.

4. Confidence Interval: The fact that almost all other lags are within the confidence bounds (blue dashed lines) implies that there are no longer-term direct relationships that are statistically significant in predicting future diesel prices.

5. Randomness: The "cut-off" after the first lag in the PACF usually implies that the differenced series does not show patterns of autocorrelation that would require additional AR terms in the model.

In summary, this PACF plot indicates that when modeling diesel price fluctuations, it may only be necessary to look at the immediate past value to predict the current value, with minimal influence from further in the past. This is vital information for constructing predictive models for diesel pricing, which can be especially important for businesses in logistics and transportation planning.



This summary output displays the results of an ARIMA (3,1,0) model applied to a time series of diesel prices. Here are the key points:

1. Model Type: The ARIMA (3,1,0) model is a type of time series forecasting model that uses three autoregressive terms (AR) and one level of differencing to make the data stationary, with no moving average (MA) component.

2. Coefficients:

- The `ar1` coefficient is positive (0.5509), suggesting a direct relationship with the previous value in the series.

- The `ar2` coefficient is negative (-0.1086), which may indicate an alternating pattern or correction from the previous lag.

- The `ar3` coefficient is small but positive (0.1164), suggesting a mild positive effect from three periods ago.

3. Statistical Significance: The standard errors of the coefficients are relatively small, indicating that the coefficients are likely to be statistically significant.

4. Model Fit: The sigma^2 value is very low (0.001874), which suggests that the model residuals have low variability, and the model fits the data well.

5. Goodness of Fit:

- The log-likelihood (2665.48), AIC (-5322.95), and BIC (-5301.57) are measures of the model fit; the more negative the AIC and BIC, the better the model is at explaining the variability of the data with fewer parameters.

- AIC stands for Akaike Information Criterion and BIC stands for Bayesian Information Criterion. Both are used to compare models, the lower their values, the better the model's balance between fit and complexity.

6. Error Metrics:

- The Mean Error (ME) is very close to zero, which means the forecast is unbiased on average.

- The Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are low, which indicates good predictive accuracy.

- The Mean Absolute Percentage Error (MAPE) of 0.8668295 suggests that the model’s predictions are, on average, within approximately 0.87% of the actual prices, which is quite accurate.

7. Autocorrelation of Residuals:

- The ACF1 value is close to zero, suggesting that there is minimal autocorrelation in the residuals, which is a good sign that the model is capturing the underlying process well.

In the context of diesel prices, this model's output indicates that past prices (especially the price from the previous week) are a strong predictor of current prices, and the model is effective at capturing the underlying price dynamics. This can be incredibly useful for financial analysts and businesses in planning and budgeting, as it provides a reliable method for forecasting future diesel prices.



The graph in depicts forecasts generated from an ARIMA (3,1,0) model, which is a type of statistical model commonly used for forecasting time series data. The model’s notation suggests that it is an Autoregressive Integrated Moving Average model with three autoregressive terms and a single differencing step, but no moving average components.

In the context of diesel prices, which are known to fluctuate over time due to various factors like market demand, supply issues, and geopolitical events, this model would have been used to predict future price movements based on past data.

The graph likely shows diesel price trends spanning several years, from 1995 to a forecasted period extending to 2025. It would show the actual recorded data up to a certain point and then forecasted data from the model thereafter.

The y-axis represents the diesel prices, although the units are not specified in your excerpt. Typically, this would be in currency per volume, like dollars per gallon or euros per liter. The x-axis represents time, moving from 1995 to 2025.

The forecasts would provide an estimation of where diesel prices are heading, but it's important to note that such forecasts have inherent uncertainty, especially when projecting many years into the future. Factors not included in the historical data can significantly influence actual future prices.

A close-up of a number

Description automatically generated

For diesel prices, the provided output shows various accuracy metrics that evaluate the model’s performance in forecasting. Let's compare the values of the training set with the test set:

- ME (Mean Error): The training set shows a very small positive mean error close to zero, indicating the model's forecasts are almost unbiased for the training data. In contrast, the test set shows a considerable negative mean error, suggesting that the model systematically underestimates the diesel prices in the test set.

- RMSE (Root Mean Squared Error): Both the training and test sets have similar RMSE values. However, a slightly higher RMSE in the test set implies that the forecast errors are generally larger when the model is applied to new data.

- MAE (Mean Absolute Error): The test set has a much higher MAE than the training set, which means that the average magnitude of errors in predicting diesel prices is greater in the test set.

- MPE (Mean Percentage Error): The training set's MPE is close to zero, whereas the test set has a large negative MPE. This large negative MPE indicates a tendency to significantly underestimate diesel prices in the test set.

- MAPE (Mean Absolute Percentage Error): The MAPE is slightly lower on the training set compared to the test set, again signaling that the model is more accurate on the training data.

- MASE (Mean Absolute Scaled Error): The MASE is greater than one for the test set, suggesting that the model performs worse than a naïve benchmark when predicting diesel prices in unseen data.

- ACF1 (Autocorrelation of Errors at Lag 1): A low ACF1 for the training set suggests the residuals are not autocorrelated, which is ideal. However, the ACF1 for the test set is higher, which could imply that there's a pattern in the forecast errors that the model has failed to capture.

- Theil's U: This metric is not available (NA) for the training set and shows a high value for the test set, implying poor predictive performance on the test data.

Overall, the model seems to fit the historical diesel prices data well but struggles to generalize its predictive capability to new, unseen data. This could mean that while the model can understand past behavior in diesel prices, it may not be capturing all the factors that influence future prices, and hence, further model refinement is warranted to improve its forecasting power.



The chart looking at is a roadmap of where diesel prices have been and where they might be headed. The dashed line that runs across most of the graph shows the journey of prices up to now, staying steady with a slight bump as we approach the present day.

Now, the interesting part is the fan-shaped spread at the end, almost like a peacock's tail opening up. This is our crystal ball, it's the forecast, showing us the range of possible paths that diesel prices could take in the future. And just like a peacock's tail, it gets wider the further out we look, which is a fancy way of saying we're less sure about what will happen the further we get from today.

But what we can say from this picture is that the trend is pointing upwards, so we might want to buckle up for a potential rise in diesel costs down the road. This graph is our heads-up to start thinking about what this means for our wallets, our businesses, and our travel plans.



The graph in your file appears to depict the residuals from an ARIMA (3,1,0) model, along with their autocorrelation function (ACF). Here are some points that can be included.

- Residual Analysis: The top panel of the graph shows the residuals over time, which are the differences between the observed values and the values predicted by the ARIMA (3,1,0) model from 1995 to around 2025.

- Stationarity and Noise: Ideally, the residuals should fluctuate randomly around zero, indicating that the model has captured all patterns in the data, and only random noise remains. The residuals in the plot do not show any obvious patterns or trends, suggesting that the model may be well-fitted.

- Autocorrelation Check: The lower panel likely shows the ACF of the residuals. It is a measure of how the residuals relate to themselves over intervals of time. For a good model, we would expect the ACF to quickly drop to zero, which would imply that the residuals are uncorrelated with each other (i.e., the information in past values does not help predict future values).

- Consistency of Variance: There does not appear to be any significant fluctuation in the variance of the residuals over time, suggesting that the variance of the error terms is consistent, which is an assumption of ARIMA models.

- Summary: The plots suggest that the ARIMA (3,1,0) model may be adequately describing the underlying process generating the diesel prices, as the residuals behave like white noise, which supports the model's validity. This can indicate a good fit for the data up until the end of the observed period.

A table of numbers with numbers

Description automatically generated

A table of numbers with numbers

Description automatically generated

A table of numbers with numbers

Description automatically generated

A table of numbers with numbers

Description automatically generated

A table of numbers with numbers

Description automatically generated

The forecast has generated predicts diesel prices for the upcoming weeks, providing point forecasts along with prediction intervals (Lo 80, Hi 80, Lo 95, Hi 95). Here's a breakdown of what each part means:

- Point Forecast: This is the estimated value for the diesel price at each forecasted time point.

- Lo 80, Hi 80: These represent the lower and upper bounds of the 80% prediction interval, indicating that there is an 80% chance that the true value will fall within this range.

- Lo 95, Hi 95: Similarly, these represent the lower and upper bounds of the 95% prediction interval, indicating a wider range where there is a 95% chance that the true value will fall.

Now, interpreting these forecasts in the context of diesel prices can be insightful for various stakeholders:

1. Consumers: Understanding future diesel prices can help consumers plan their transportation expenses. They can anticipate potential increases or decreases in fuel costs, enabling better budgeting and decision-making regarding vehicle use.

2. Businesses: Companies involved in transportation, logistics, and manufacturing heavily rely on diesel fuel. Accurate price forecasts allow them to adjust pricing strategies, transportation routes, and inventory management to mitigate the impact of fluctuating fuel costs on their bottom line.

3. Government and Policy Makers: Forecasting diesel prices aids policymakers in crafting energy policies, taxation schemes, and subsidies to stabilize fuel prices, promote energy efficiency, and mitigate environmental impact.

4. Investors: Investors in energy markets use price forecasts to make informed decisions regarding investments in oil and gas companies, energy futures, and related financial instruments.

5. Environmentalists: Fluctuations in diesel prices can influence consumer behavior towards more fuel-efficient vehicles and alternative energy sources, thereby impacting environmental conservation efforts.

These forecasts provide valuable insights into future diesel price trends, empowering various stakeholders to make informed decisions that align with their objectives and interests.

**Conclusion:**

Journey through the landscape of diesel prices, powered by historical insights and statistical acumen, has revealed a forecast that points to a steady ascent of costs on the horizon. For the truckers, planners, and budget-keepers among us, the message is clear: be ready for a measured climb in diesel expenditures. It's a nudge for businesses to recalibrate budgets, perhaps to seek more cost-effective deals or smarter routes. Policymakers could take this cue to adjust fiscal strategies or propel green initiatives, while investors may find these projections to be the pulse they need to stay ahead of the curve.

To distill it down, the data speaks of no drastic upswings, but rather a gentle, persistent uptrend in diesel prices. With intelligent planning and informed decision-making, navigating this upward slope can be a seamless endeavor. This isn't just about tracking numbers. It's about charting a course for sustained success in an ever-fluctuating market.

In a nutshell, our number-crunching suggests that while there won't be any wild swings, it's a good idea to brace for a gradual rise in diesel costs. Plan smart, stay informed, and you'll navigate these changes like a pro.

# Time Spent - Planned Vs Actual

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Week** | **Task** | **Planned Hours** | **Actual Hours** | **Variance** |
| 7 | Familiarizing with dataset | 1 | 1.5 | 50% |
| Data cleaning and preprocessing | 2 | 3 | 100% |
| Exploratory data analysis | 3 | 3.75 | 75% |
| 8 | Reviewing literature | 2 | 3 | 100% |
| Selecting the forecasting method | 1 | 1 | 0% |
| Developing the forecasting model | 2 | 3.5 | 100% |
| 9 | Continuing to develop the forecasting model | 3 | 3 | 0% |
| Documenting progress | 1 | 2.5 | 100% |
| 10 | Validating the model | 3 | 3.5 | 0% |
| Evaluating the model | 2 | 2 | 0% |
| 11 | Refining the forecasting model | 3 | 3.5 | 0% |
| Testing the model | 2 | 3 | 100% |
| 12 | Drafting the report | 2 | 3 | 100% |
| Finalizing the forecasting model | 3 | 3.75 | 75% |