Time Series Fitting and Forecasting: Residential Electricity Prices in the United States

FRE-GY 6351 – Econometrics and Time Series Analysis

Submitted By

Haritha GB (hg2590)
Naman Rathi (nr2775)
Syed Ahzam Tariq (sat10045)

May 09, 2024

Table of Contents

1. ABSTRACT

In this project, we intend to undertake a time series analysis of electricity pricing data. Our objective is to decompose the series into its constituent trend and seasonality components and to scrutinize the residuals for any overlooked patterns. The study escalates from elementary to intricate models to refine forecast precision and enhances understanding of price dependencies. Detailed statistical analysis reveals significant seasonal trends and persistent non-stationarities, prompting the application of modeling techniques such as ARIMA, SARIMA, and Fourier transforms. The aim is to offer robust predictions and insights into the factors driving price fluctuations.

2. DATASET

The dataset utilized in this study consists of the monthly average residential electricity prices in the United States. The data, sourced from the U.S. Energy Information Administration (EIA), exhibits clear monthly seasonality, which enhances the complexity and interest of our analysis by providing a predictable cyclical pattern that can be exploited. This characteristic of the dataset is crucial, as it allows for a detailed exploration of seasonal trends in electricity pricing, a key factor in our time series analysis. The dataset is publicly accessible via the EIA's official website, ensuring transparency and reproducibility of our research findings. The dataset can be found at [1].

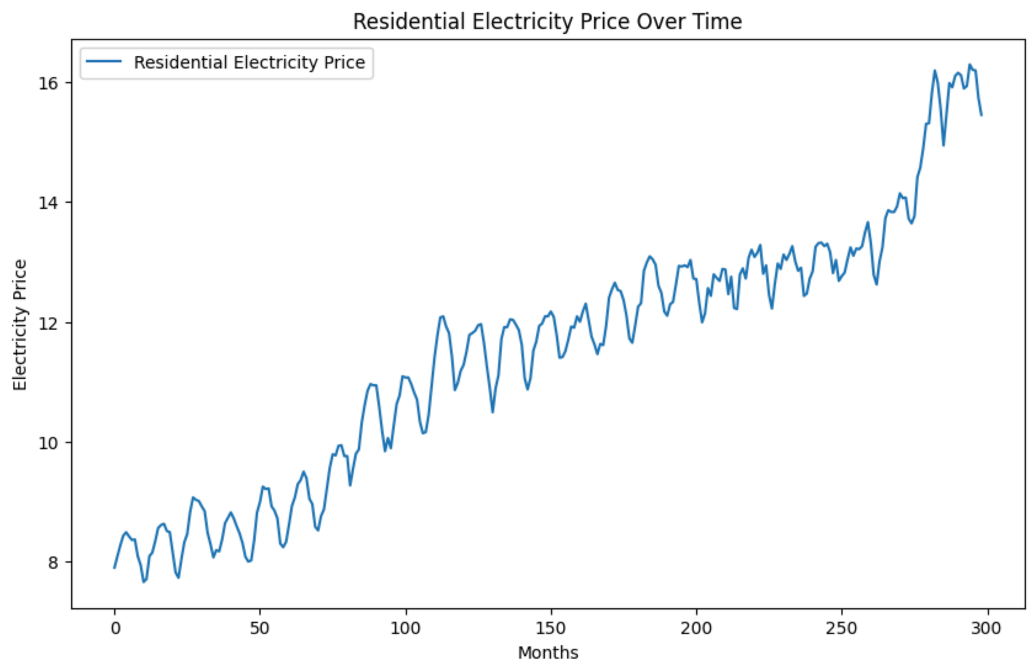


Figure 1

3. METHODOLOGY & MODEL EVALUATION

Original Data

First, we perform stationarity tests on the original dataset, to check the applicability of time series models. We use the Augmented Dickey-Fuller test to find the stationarity of the series [2].

```
Results of Dickey-Fuller Test:
Test Statistic                   0.512450
p-value                          0.985247
#Lags Used                      13.000000
Number of Observations Used    287.000000
Critical Value (1%)             -3.453342
Critical Value (5%)             -2.871664
Critical Value (10%)            -2.572164
dtype: float64
```

Figure 2: ADF results, Original Data

Fig 2 shows the ADF test results of the original data series. For all confidence intervals, the p-value is greater than the critical value, and hence we reject the null and hence conclude the original data has stationary components to it.

Regardless of the present stationarity, we try to decompose the original data into trend seasonality in order to study the residuals and their structure.

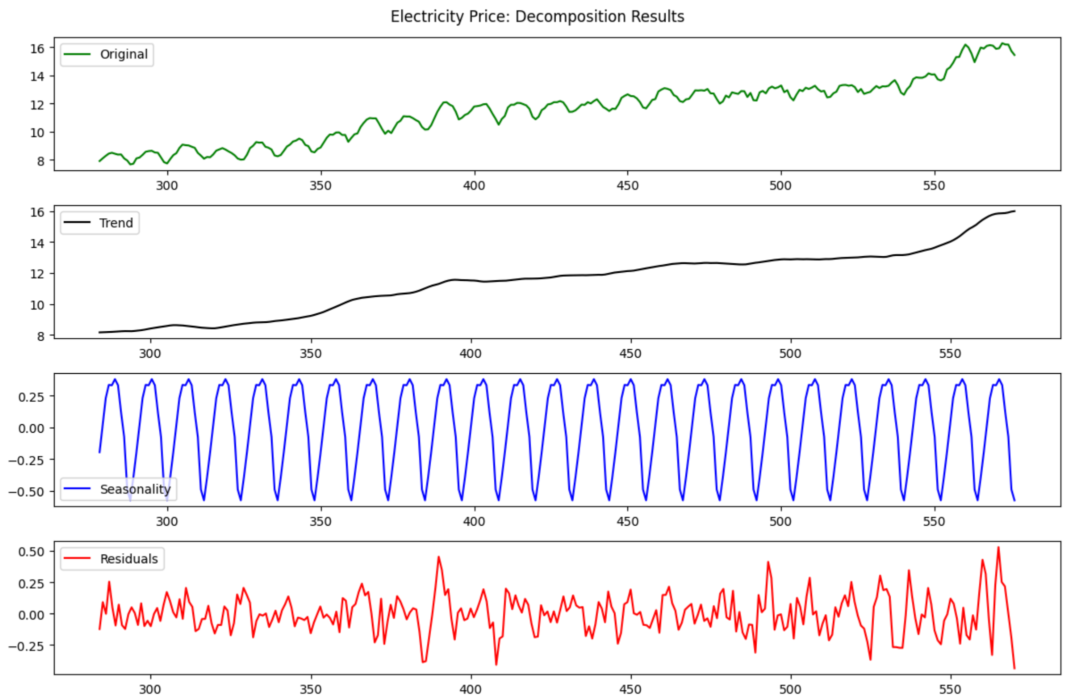
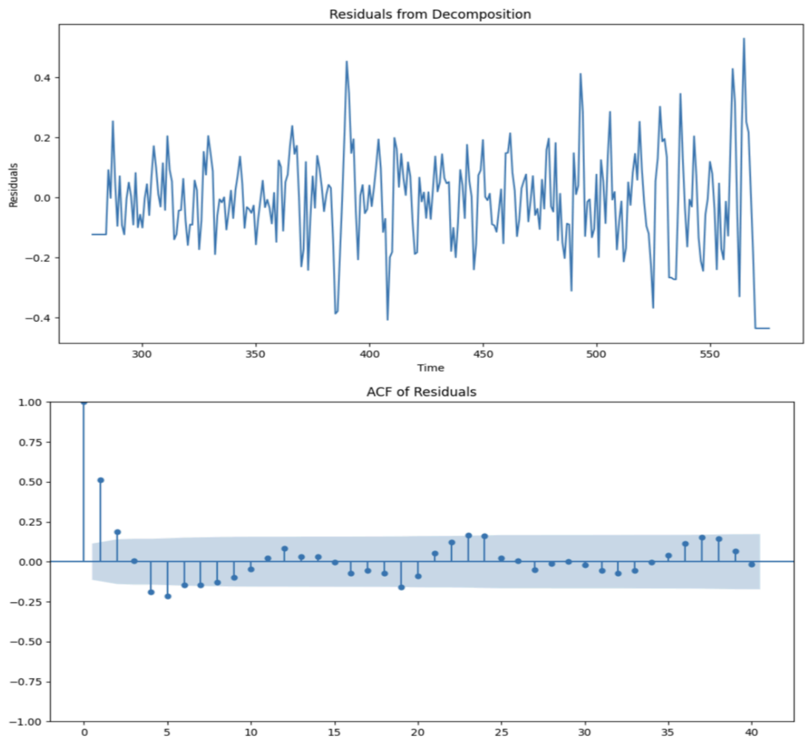


Figure 3: Decomposition of Original Dataset

Fig 3 shows the decomposition results of the original dataset. The seasonal_decompose function from the statsmodels library implements classical time series decomposition. It decomposes a time series into three components: trend, seasonality, and residuals. The trend component represents the long-term progression of the series, showing how the central tendency changes over time. The trend component is typically calculated using a moving average, smoothing out short-term fluctuations and highlighting longer-term trends in the data. The seasonal component captures systematic, calendar-related movements and is repeated at regular intervals over the series. Seasonality is determined by identifying and averaging the repeating short-term cycles within the specified period. Lastly, the residuals comprise the remainder of the time series after the trend and seasonal components have been removed, capturing irregular fluctuations not explained by the other two components. This method assumes an additive or multiplicative model depending on the nature of the seasonal effect. This decomposition can be applied under an additive or multiplicative model, depending on the nature and amplitude of the seasonal variation relative to the trend.

On applying this decomposition, we perform the Ljung-Box test [4] and plot the ACFs [5] to study the issues of this decomposition. The results are found at Fig 4. The results in Fig 4 show us that the residuals are not random, solidifying the results of the original data not being appropriate for direct application to a time series model.

The most common next step is to differentiate the data series and repeat analyses to find the change in resulting statistics.




```
Ljung-Box Test Results:
      lb_stat     lb_pvalue
10  136.904919  1.813695e-24
The residuals are not random (i.e., there is autocorrelation at lag 10).
```

Figure 4: Results of the residual analysis: Original Data

We differentiate the original price series and redo the analyses outlined.

```
Results of Dickey-Fuller Test:
Test Statistic                  -3.887905
p-value                          0.002126
#Lags Used                      12.000000
Number of Observations Used    287.000000
Critical Value (1%)             -3.453342
Critical Value (5%)             -2.871664
Critical Value (10%)            -2.572164
dtype: float64
```

Figure 5: ADF test on differenced series

Fig 5 shows the differenced data's ADF test results [2]. The test statistic meets the criteria for all three confidence intervals, and we hence conclude that the differenced data is more stationary. We now decompose this differenced series into trend, and seasonality and study the implied residuals [3].

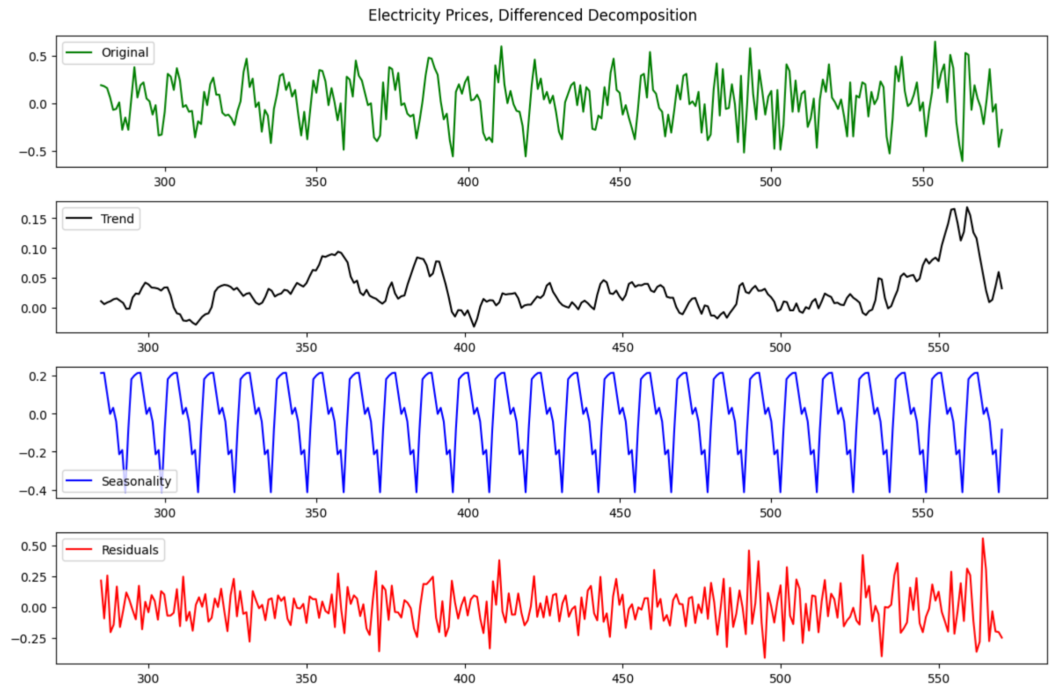


Figure 6: Decomposition of the differenced data series

Fig 6 shows the decomposition results of the differenced data series. We immediately notice a more granular seasonality series. We notice that the trend is not linearly increasing like before, but this is acceptable and required since we are decomposing the differenced series and not the original series. Now we proceed to perform the same statistical analysis on the residuals as previously.

We then perform ACF plotting [5] and Ljung-Box Test [4] on the differenced data residuals. Fig 7 shows us these results. We immediately see better results in the ACF plot and the Ljung-Box Test shows that the residuals appear random.

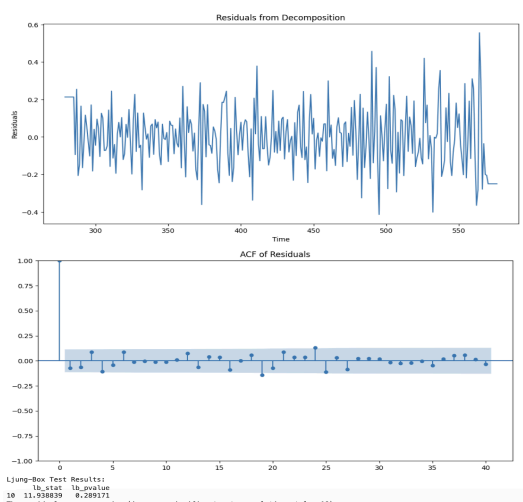


Figure 7

4. MODELLING

We now move on the modelling of the data as a time series. Our aim is to investigate the feasibility of implementing the AR, MA, ARIMA and Fourier Model. At the end of each model, we plot the fit, analyse the statistics of fit and try to forecast 5 steps into the future.

4.1 Autoregressive model (AR)

An autoregressive (AR) model is a popular type of time series model used to describe certain time-dependent data behaviors [6]. The AR model is defined by the relationship between a variable and its own lagged (or previous) values. The term AR(p) represents an autoregressive model of order p, where p indicates the number of lag terms used by the model.

The AR(p) model is expressed as:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t$$

where:

- X_t is the value of the series at time t,
- c is a constant (also called the intercept),
- $\phi_1, \phi_2, \ldots, \phi_p$ are parameters of the model,
- ϵ_t is white noise error at time t.

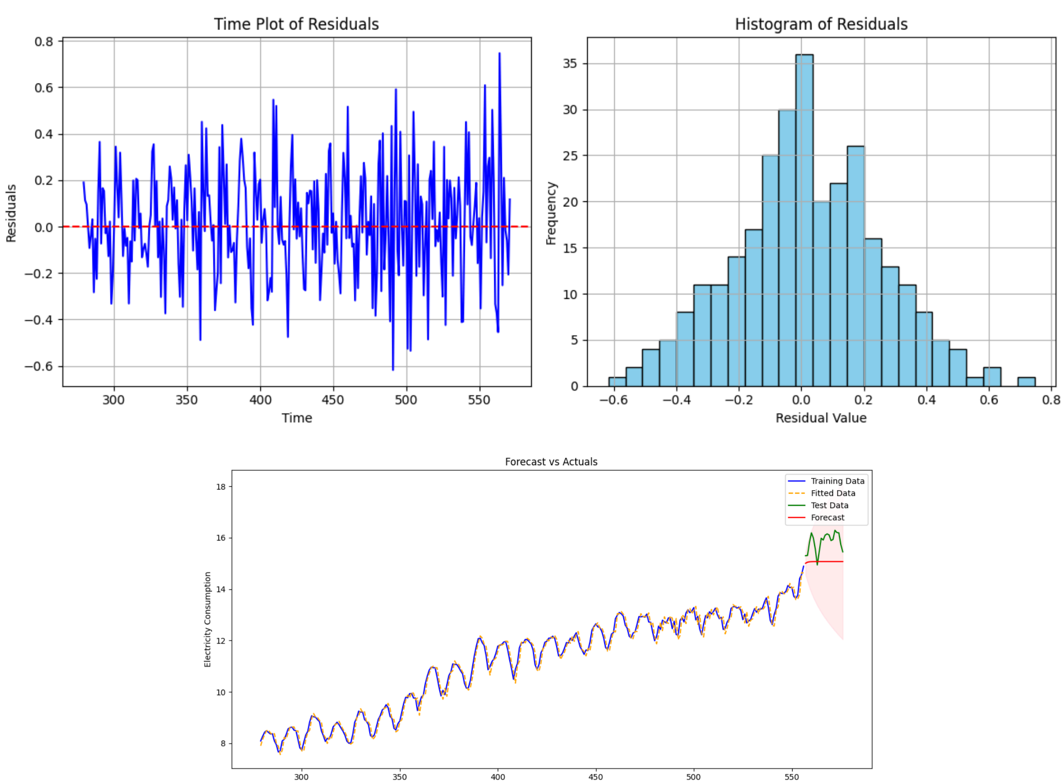


Figure 8

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  279
Model:               SARIMAX(1, 1, 0)   Log Likelihood                  18.609
Date:                Wed, 08 May 2024   AIC                            -33.217
Time:                        22:27:35   BIC                            -25.962
Sample:                             0   HQIC                           -30.307
                                - 279
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.3620      0.060      6.013      0.000       0.244       0.480
sigma2         0.0512      0.004     11.413      0.000       0.042       0.060
===================================================================================
Ljung-Box (L1) (Q):                   0.14   Jarque-Bera (JB):                 0.17
Prob(Q):                              0.71   Prob(JB):                         0.92
Heteroskedasticity (H):               1.89   Skew:                            -0.04
Prob(H) (two-sided):                  0.00   Kurtosis:                         2.92
===================================================================================
```

Figure 9

Evaluation of the AR(p) model:

- **Model Hypothesis:** The analysis begins with the hypothesis that the dataset, representing a time series, can be aptly modeled using an Autoregressive model of order p (AR(p)). This approach assumes that the current values of the series can be predicted from its own previous values**.**

- **Model Order Determination:** To determine the optimal order (p) of the AR model, a grid search was performed ranging from 0 to 8. This method aims to identify the order that best captures the dependencies in the data while balancing model complexity and fit.

- **Diagnostic Checking**
  - **Ljung-Box Test:** The Ljung-Box test was applied to the residuals of the fitted AR(p) model to check for autocorrelation. The absence of significant autocorrelation among the residuals suggests that the AR(p) model captures the underlying data structure adequately, confirming the appropriateness of the model.
  - **Residual Analysis:** Despite the residuals centering around zero, indicating unbiased predictions, there remains some observable patterns in the residuals. This could point to model misspecifications or potential dynamic patterns in the series not captured by the model.
  - **Residual Normality:** The histogram of residuals show that the residuals satisfactorily show a normal distribution. The JB test shows the goodness of fit of residuals into a normal distribution, and we see conducive results with a probability of 0.92.

- **Model Performance:** The AR(p) model demonstrated commendable performance in tracking the historical data for the majority of the series. However, towards the end of the series, the model failed to predict a significant rise in electricity consumption, diverging noticeably. This divergence might suggest potential misfit, or it may indicate the model's inability to adapt to structural breaks or abrupt changes in the series behaviour.

4.2 Moving Average Model (MA)

An MA (Moving Average) model is another classical approach for modelling time series data, particularly useful for describing time series that show short-term dependencies between successive observations [6].

The MA(q) model is defined as:

$$X_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

where:

- X_t is the value of the series at time t,
- μ is the mean of the series,
- ϵ_t represents the white noise error terms at time t,
- $\theta_1, \theta_2, \ldots, \theta_q$ are the parameters of the model,
- q is the order of the moving average model.

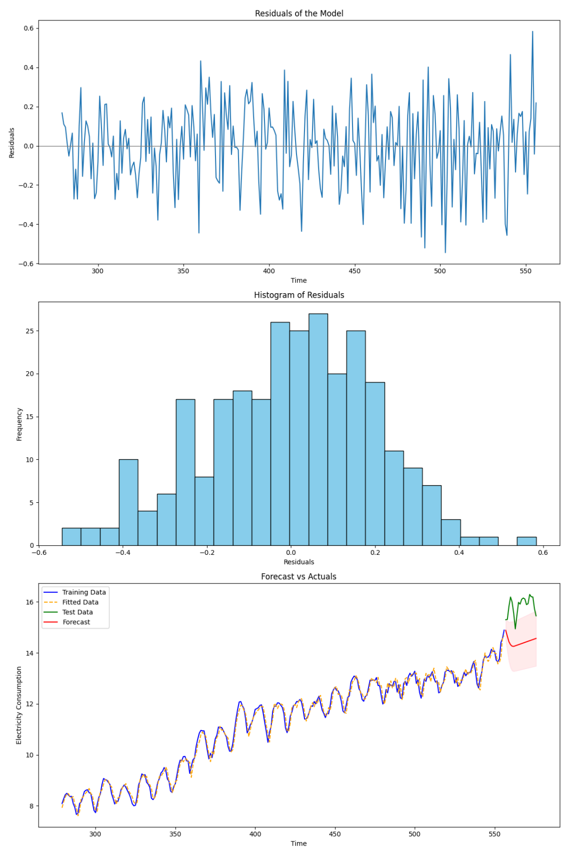


Figure 10

```
                                   SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  279
Model:               SARIMAX(0, 1, 6)   Log Likelihood                  51.073
Date:                Wed, 08 May 2024   AIC                            -86.146
Time:                        22:27:37   BIC                            -57.125
Sample:                             0   HQIC                           -74.503
                                - 279
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0222      0.004      5.499      0.000       0.014       0.030
ma.L1          0.1685      0.059      2.877      0.004       0.054       0.283
ma.L2          0.0512      0.052      0.987      0.324      -0.051       0.153
ma.L3         -0.0361      0.053     -0.681      0.496      -0.140       0.068
ma.L4         -0.3458      0.056     -6.204      0.000      -0.455      -0.237
ma.L5         -0.3898      0.054     -7.226      0.000      -0.496      -0.284
ma.L6         -0.1370      0.060     -2.298      0.022      -0.254      -0.020
sigma2         0.0403      0.004     10.870      0.000       0.033       0.048
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                 3.12
Prob(Q):                              0.91   Prob(JB):                         0.21
Heteroskedasticity (H):               1.82   Skew:                            -0.24
Prob(H) (two-sided):                  0.00   Kurtosis:                         2.80
===================================================================================
```

Figure 11

Evaluation of the MA(q) model:

- **Model Extension:** Following the assessment of the Autoregressive (AR) model, an exploration into the Moving Average (MA) component was conducted to determine if it could enhance the model's ability to capture additional dynamics in the time series data.

- **Residual Analysis**
  - **Stabilisation of Residuals:** The analysis indicates that the residuals generally stabilise around zero, reinforcing the model's effectiveness in capturing the primary trends and cyclicality of the data.
  - **Residual Distribution:** A histogram of the residuals revealed asymmetry and slight skewness. These characteristics suggest a deviation from the ideal normal distribution, indicating that the assumption of normality may not be fully satisfied in the current model setup, and the JB test results support the same.

- **Forecast Accuracy:** Comparison of Forecasts to Actual Values: A plot comparing the model's forecasts against actual values shows that while the model follows the historical trend with reasonable accuracy, it struggles with accurately predicting sharper, more abrupt changes observed in the test set. This discrepancy between forecasted and actual values underscores a potential limitation of the model's adaptability to sudden shifts in the series or its ability to generalise effectively under varying conditions.

- **Implications for Model Enhancement:** The preliminary findings from incorporating an MA component suggest that while the model can delineate basic patterns in the dataset, improvements are necessary for handling non-linear dynamics and outliers. The presence of skewness and outliers in the residuals may warrant a revision of the model's error distribution assumptions or the inclusion of additional data preprocessing steps.

4.3 Autoregressive Integrated Moving Average Model (ARIMA)

The ARIMA (Autoregressive Integrated Moving Average) model combines elements of AR (Autoregressive) and MA (Moving Average) models and includes differencing to make the time series data stationary. It is one of the most popular and broadly used statistical methods for time series forecasting [6].

The ARIMA model is typically denoted as ARIMA(p, d, q), where:

- p is the number of autoregressive terms,
- d is the number of differences needed to make the series stationary,
- q is the number of moving average terms.

The mathematical representation of the ARIMA model is:

$$X'_t = c + \phi_1 X'_{t-1} + \phi_2 X'_{t-2} + \cdots + \phi_p X'_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

where:

- X'_t denotes the differenced series (i.e., the series after applying differencing d times to achieve stationarity),
- c is a constant,
- ϕ_i are the coefficients for the AR terms,
- θ_i are the coefficients for the MA terms,
- ϵ_t represents the white noise error terms at time t.

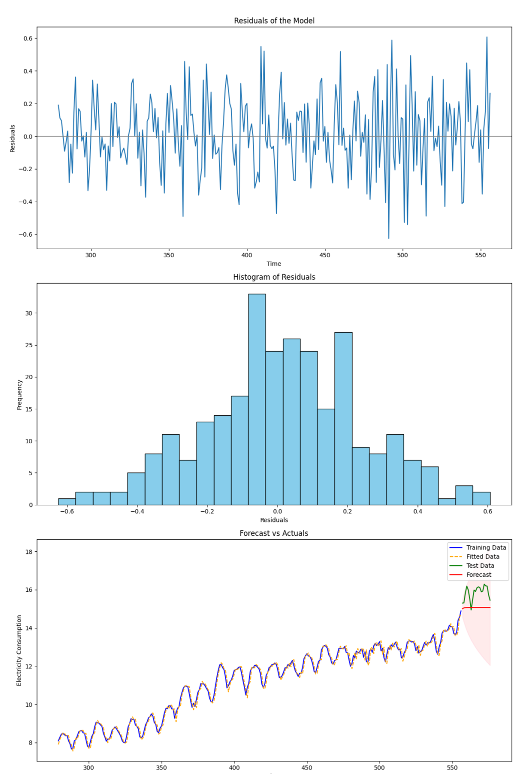


Figure 12
ARIMA Residuals and Forecast

```
==============================================================================
Dep. Variable:                      y   No. Observations:                  279
Model:               SARIMAX(1, 1, 0)   Log Likelihood                  18.609
Date:                Sat, 04 May 2024   AIC                            -33.217
Time:                        02:10:48   BIC                            -25.962
Sample:                             0   HQIC                           -30.307
                                - 279
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.3620      0.060      6.013      0.000       0.244       0.480
sigma2         0.0512      0.004     11.413      0.000       0.042       0.060
===================================================================================
Ljung-Box (L1) (Q):                   0.14   Jarque-Bera (JB):                 0.17
Prob(Q):                              0.71   Prob(JB):                         0.92
Heteroskedasticity (H):               1.89   Skew:                            -0.04
Prob(H) (two-sided):                  0.00   Kurtosis:                         2.92
===================================================================================
```

Figure 13: ARIMA Fitting Statistics

Evaluation of the ARIMA(p, d, q) Model:

- **Model Hypothesis:** The investigation commences with the hypothesis that the dataset, which exhibits complex dynamic patterns, can be effectively modeled using an ARIMA(p, d, q) framework. This model integrates autoregressive (AR) elements, differencing (d) for stationarity, and moving average (MA) components, aiming to capture a comprehensive range of autocorrelations within the time series data.

- **Diagnostic Checks:**
  - **Ljung-Box Test:** The Ljung-Box test was applied to assess the presence of autocorrelation in the residuals of the ARIMA model (Fig 12). The test results indicated no significant autocorrelation, suggesting that the model successfully captures the underlying data structure and the dependencies among the observations.
  - **Residual Analysis:** Despite the residuals centering approximately around zero, indicative of an unbiased model, patterns within the residuals persisted. This observation hints at potential model misspecifications or unaccounted dynamics within the time series that the current ARIMA model configuration may not be capturing.

- **Residual Histogram Analysis:** The histogram of residuals (Fig 12) show that the residuals satisfactorily show a normal distribution. The JB test (Fig 13) shows the goodness of fit of residuals into a normal distribution, and we see conducive results with a probability of 0.92.

- **Implications for Model Refinement:** These findings underscore the need for further refinement of the ARIMA model parameters or perhaps the inclusion of additional data preprocessing steps to better adhere to the assumptions of normality. The persistence of patterns in the residuals suggests exploring alternative model specifications, including higher-order differencing or the incorporation of seasonal components if justified by the data.

4.4 SARIMA

The SARIMA (Seasonal Autoregressive Integrated Moving Average) model is an extension of the ARIMA model designed to more effectively handle seasonal variations in time series data. It is particularly useful for data that show patterns and behaviors that repeat over a fixed period, such as daily temperatures, monthly sales, or annual trends [7].

SARIMA models are denoted as SARIMA(p, d, q)(P, D, Q)_s, where:

- p, d, q are the non-seasonal components (as in ARIMA) representing the autoregressive order, degree of differencing, and moving average order, respectively.
- P, D, Q represent the seasonal aspects of the model, corresponding to the seasonal autoregressive order, seasonal differencing degree, and seasonal moving average order.
- s is the length of the seasonal cycle. For example, s = 12 for monthly data with an annual cycle, or s = 4 for quarterly data.

The SARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative manner.

The model can be expressed as:

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D X_t = \Theta_Q(B^s)\theta_q(B)\epsilon_t$$

where:

- B is the backshift operator, $B^k X_t = X_{t-k}$.
- Φ_P and ϕ_p are the seasonal and non-seasonal autoregressive operators.
- Θ_Q and θ_q are the seasonal and non-seasonal moving average operators.
- $(1-B)^d$ and $(1-B^s)^D$ are the differencing operators to achieve stationarity on non-seasonal and seasonal levels.

Model Hypothesis

The SARIMAX(1, 1, 0) model is employed to forecast a time series, positing that the data's inherent dynamics can be modeled effectively by integrating an autoregressive term and a differencing step, while also accounting for external influences. This model configuration suggests an emphasis on capturing the linear dependencies and addressing non-stationarity, with the expectation that external factors significantly influence the series' behavior.

- **Diagnostic Checks**

  Ljung-Box Test: The results from the Ljung-Box test indicate no significant autocorrelation in the residuals (Q = 0.14, p = 0.71), suggesting that the model captures the temporal dependencies adequately.

- **Residual Analysis**: The residuals are centered around zero, indicating an unbiased model forecast. However, the dispersion and frequency of spikes suggest potential volatility or model inadequacies in fully capturing the data dynamics.
- **Residual Histogram Analysis:** The histogram of the residuals displays a fairly symmetric distribution around zero but the JB test shows a surprising statistic: the residuals are less likely to be normally distributed. An additional difference on data aided this situation, but the variance does not seem to be constant and it looks like it would need an additional function to fit on the volatility.
- **Implications for Model Refinement:** The presence of heteroskedasticity and the pattern in residuals suggest exploring models that can accommodate changing variance, such as incorporating a GARCH component. The leptokurtic nature of the residual distribution prompts a review of outlier effects or the need for transformations to stabilise variance across the series. Given the model's performance and external variable inclusion, further refinement could involve tuning the parameters or extending the model to include additional lags or moving average components to better adapt to complex dynamics.

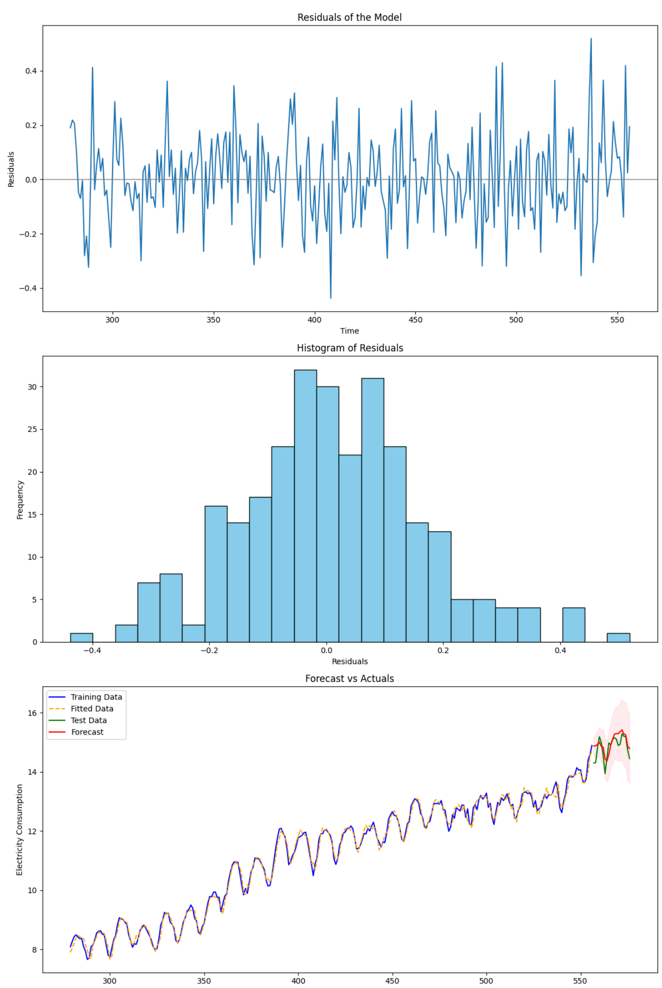


Figure 14: SARIMA Residuals and Forecast

```
==============================================================================
Dep. Variable:                                  y   No. Observations:                  279
Model:             SARIMAX(0, 1, 1)x(1, 0, 1, 12)   Log Likelihood                 120.463
Date:                            Sat, 04 May 2024   AIC                           -232.925
Time:                                    02:11:57   BIC                           -218.415
Sample:                                         0   HQIC                          -227.104
                                            - 279
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.2094      0.057     -3.686      0.000      -0.321      -0.098
ar.S.L12       0.9835      0.009    110.940      0.000       0.966       1.001
ma.S.L12      -0.7330      0.044    -16.547      0.000      -0.820      -0.646
sigma2         0.0229      0.002     12.963      0.000       0.019       0.026
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                 4.36
Prob(Q):                              0.94   Prob(JB):                         0.11
Heteroskedasticity (H):               1.86   Skew:                             0.21
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.46
===================================================================================
```

Figure 15: Fitting Statistics, SARIMA

4.5 Fast Fourier Transform

The Fast Fourier Transform (FFT) is a powerful tool in time series analysis, particularly useful for identifying periodicities, trends, and filtering noise in frequency domain data [8]. At the heart of FFT-based time series analysis is the Fourier Transform, a mathematical transformation used to decompose a function (in this case, a time series) into its constituent frequencies. The Fourier Transform of a time series tells us how much of the signal is made up of different frequency components. Time series data is naturally in the time domain, where each data point represents a value at a specific time. The Fourier Transform converts this data into the frequency domain, where each point represents a particular frequency component's contribution to the overall time series. By analysing the frequency domain representation of the data, one can identify the main periodic components of the time series. For instance, dominant frequencies can indicate regular cycles in the data, which are crucial for tasks like trend analysis and forecasting.

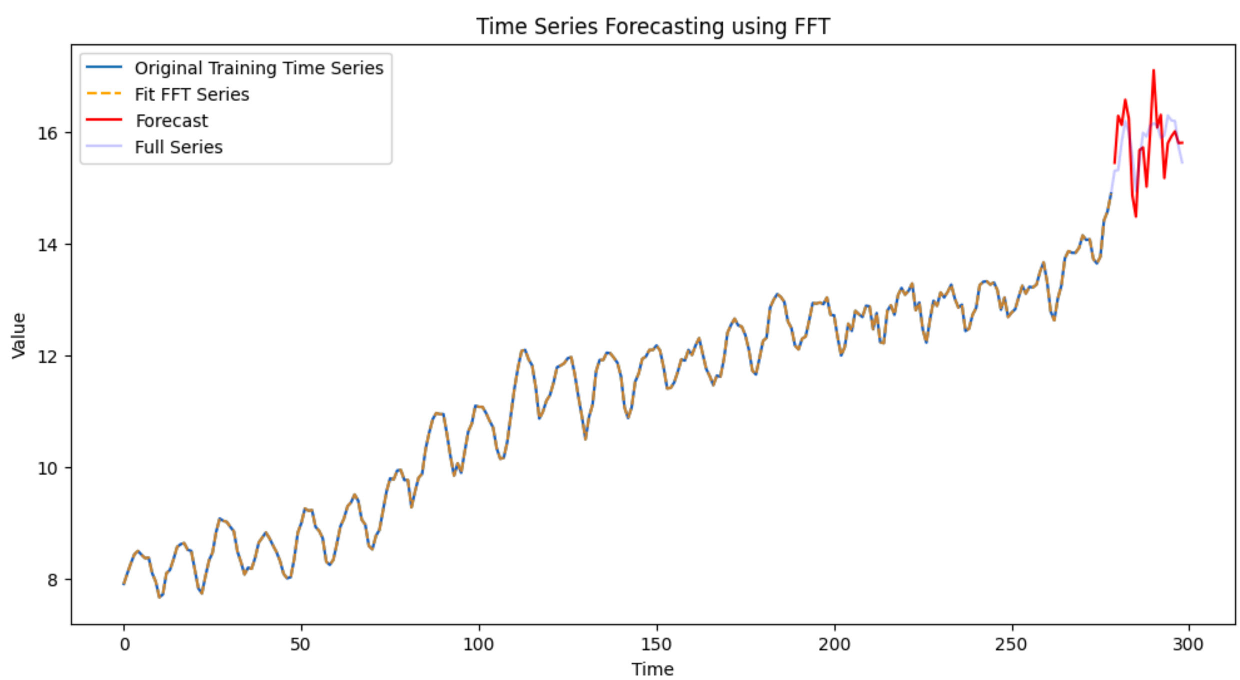


Figure 16: Forecasting using FFT

Applying FFT to Time Series Modeling

Transformation to Frequency Domain

- The first step is to apply FFT to the time series data, transforming it from the time domain to the frequency domain. This process converts the temporal sequence of data into a function of frequencies that shows how different frequencies contribute to the time series signal.

Frequency Domain Analysis

- Identifying Significant Frequencies: Once the time series is transformed, the next step is to analyze the spectrum to identify significant frequencies. Peaks in the Fourier Transform indicate strong periodic components at those frequencies.
- Filtering: FFT can also be used to filter out noise or insignificant frequencies. By setting a cutoff threshold, you can zero out frequencies that contribute noise or are of no interest, thereby smoothing or denoising the data.

Inverse FFT

- After modifying the frequency domain data (like filtering out certain frequencies), the inverse FFT is applied to convert the data back to the time domain. This step reconstructs the time series data, now altered based on the manipulations in the frequency domain (e.g., removal of noise, emphasis on significant frequencies).

Forecasting

- Forecast Construction: For forecasting, the model can be constructed either directly in the frequency domain by modelling how frequency components evolve or by reconstructing the time series in the time domain after making adjustments in the frequency domain and using this reconstructed series to forecast future values.
- Iterative Forecasts: In a one-step-ahead forecasting model using FFT, each new forecast can be iteratively added to the time series, and the FFT process can be reapplied to include this new data point, thus dynamically updating the forecast model.

5. FORECASTING RESULTS

Model	RMSE
AR(1)	0.83837
MA(6)	1.45275
ARIMA(1,1,0)	0.83837
SARIMA (0,1,1)(1,0,1)[12]	0.88234
FFT	**0.77330**

6. CONCLUSION

In the initial phase of our study, we employed a variety of statistical models to predict a time series dataset, beginning with simpler models such as the AR and progressing to more complex ones like SARIMA and FFT. Our evaluation, based on the RMSE metric, reveals that while simpler models like the AR and ARIMA have comparable performance, more complex approaches such as the FFT provided the lowest RMSE, indicating superior accuracy in handling the dataset's inherent fluctuations.

The SARIMA model, designed to account for seasonal patterns, delivered moderate success. It managed to capture the periodicity of the dataset, which is crucial for accurate forecasting in seasonal time series. It however did not render normally distributed residuals, which indicates room for better fit. However, its performance was slightly outperformed by the FFT model, suggesting that the transformation-based approach of FFT might be more effective in capturing complex patterns and noise in the data.

Incorporating external variables using the SARIMAX model could substantially improve forecast accuracy. An additional trial may involve modelling volatility using GARCH models. This model extension is particularly beneficial in scenarios where the series is influenced by identifiable external factors, enabling the model to account for additional variability and potentially unforeseen events. Machine learning models, especially neural networks such as LSTM and GRU, are well-suited for modelling sequences and can learn complex patterns from data. Their ability to remember information over long periods makes them ideal for time series forecasting. Experimentation with different architectures could unveil significant improvements in predictive performance.

Combining the strengths of traditional statistical models with machine learning could yield a hybrid approach that leverages the robustness and efficiency of statistical methods alongside the adaptability and learning capabilities of machine learning models. Enriching our model inputs with data from related domains (e.g., Economic indicators for stock price prediction or weather data for energy demand forecasting) could provide deeper insights and enhance the model's predictive power. Implementing rigorous cross-validation methods will be essential in assessing the models' performance and ensuring their reliability and consistency across different temporal segments.

7. REFERENCES

[1] U.S. Energy Information Administration. (2024). Monthly Energy Review. Retrieved from https://www.eia.gov/totalenergy/data/monthly/

[2] Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. Journal of the American Statistical Association, 74(366), 427-431.

[3] Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In *Proceedings of the 9th Python in Science Conference*.

[4] Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297-303.

[5] Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. 3rd Edition. Prentice-Hall.

[6] Zhang, Mingda. "Time series: Autoregressive models AR, MA, ARMA, ARIMA." University of Pittsburgh (2018).

[7] Chikobvu, Delson, and Caston Sigauke. "Regression-SARIMA modelling of daily peak electricity demand in South Africa." Journal of Energy in Southern Africa 23.3 (2012): 23-30.

[8] Musbah, H., El-Hawary, M., & Aly, H. (2019, October). Identifying seasonality in time series by applying fast fourier transform. In 2019 IEEE Electrical Power and Energy Conference (EPEC) (pp. 1-4). IEEE.

[9] GB, Tariq, Rathi (2024). [https://github.com/harithagb/timeseries]