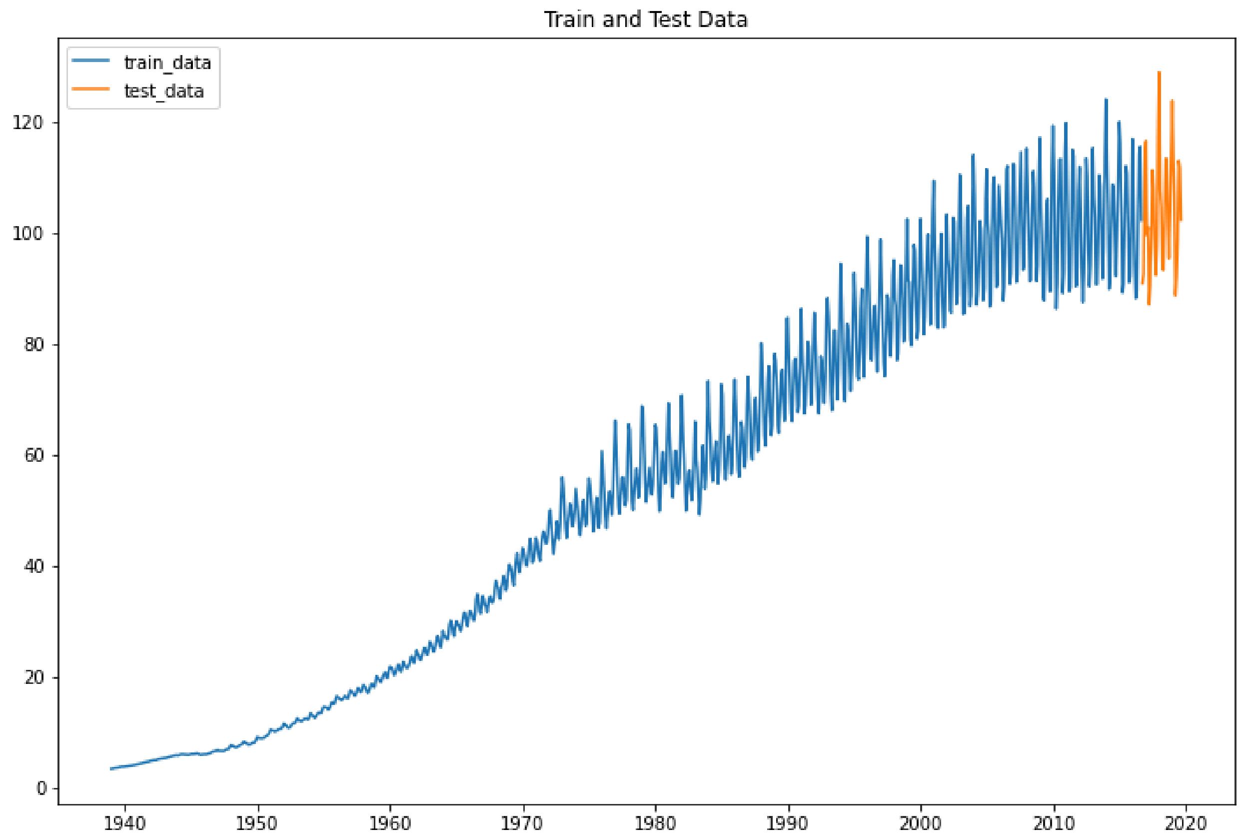
Forecasting Energy Consumption using SARIMA Models

This report presents an analysis of historical energy consumption data as seen earlier while using exponential smoothing models for forecasting. This currently uses and compares two SARIMA models (i.e. **SARIMA(2,1,1)(1,1,1,12)** and **SARIMA(1,1,1)(1,1,1,12)**) for fitting and testing the model fit and prediction accuracy on the test data.

The dataset as seen earlier consists of historical data of monthly energy consumption from Jan 1939 till Sep 2019 and we try to forecast the demand for the next 3 years. We will use RMSE and MAE in order to assess model performance.



The summary statistics for the dataset is shown below:

	ENERGY_INDEX
count	969.000000
mean	54.657608
std	35.455379
min	3.384200
25%	19.581700
50%	54.763700
75%	87.729300
max	128.907100

Analysis

I have first tested the data for stationarity using both Augmented Dickey Fuller (ADF) and KPSS tests. The null hypothesis in ADF states that the series is not stationary. So if the test statistic is less than the critical value, we can reject the null hypothesis (aka the series is stationary). When the test statistic is greater than the critical value, we fail to reject the null hypothesis (which means the series is not stationary).

```
In [49]: adf_test(train['ENERGY_INDEX'])
Results of Dickey-Fuller Test:
Test Statistic                  -0.374861
p-value                          0.914212
#Lags Used                      21.000000
Number of Observations Used    911.000000
Critical Value (1%)             -3.437548
Critical Value (5%)             -2.864718
Critical Value (10%)            -2.568462
```

Here as can be seen from ADF test results above, test statistic is > than critical. Hence the series is not stationary.

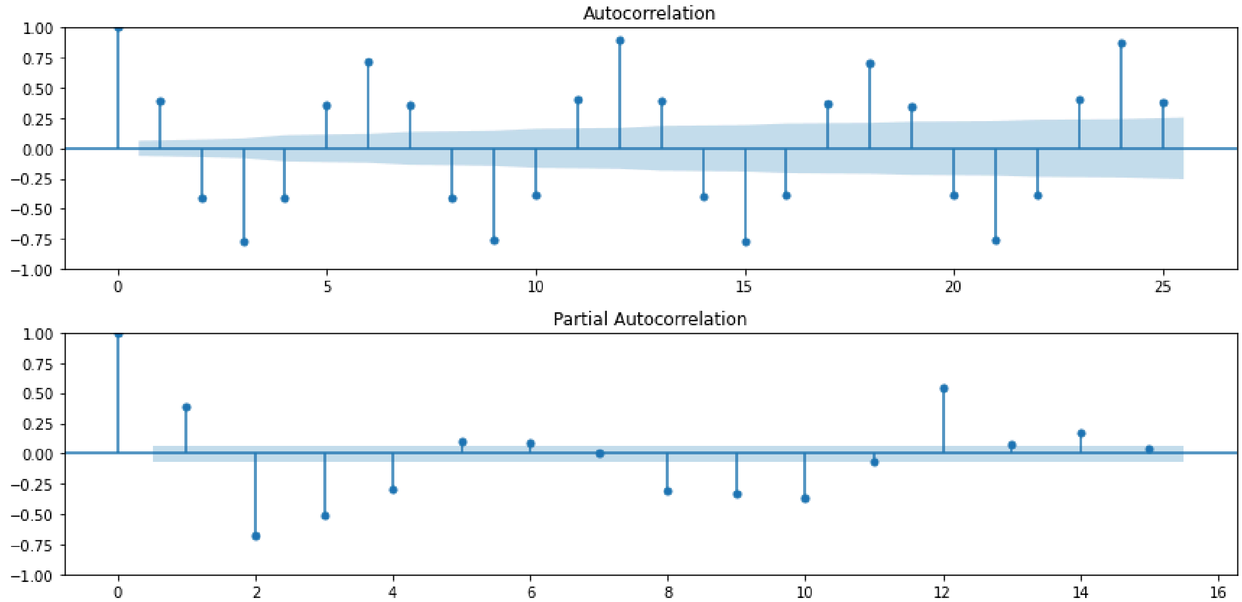
On the other hand, for KPSS test the null hypothesis is that the series is trend stationary. If the test statistic is greater than the critical value, we reject the null hypothesis (series is not stationary). If the test statistic is less than the critical value, we fail to reject the null hypothesis (series is stationary).

```
In [51]: kpss_test(train['ENERGY_INDEX'])
Results of KPSS Test:
Test Statistic            5.024274
p-value                   0.010000
Lags Used                18.000000
Critical Value (10%)      0.347000
Critical Value (5%)       0.463000
Critical Value (2.5%)     0.574000
Critical Value (1%)       0.739000
```

Here as seen above for KPSS test results, test statistic is > than critical. Hence the series is not stationary.

I then checked for the ACF and PACF plots in order to infer the p, q values. Based on the ACF plot showing significant spikes at lag 1 and seasonal lags (12, 24), while the PACF showing significant spike at lag 1 and cuts off after lag 2, I choose the SARIMA (2,1,1) (1,1,1,12) model and also experimented with the SARIMA (1,1,1) (1,1,1,12) model for a comparison.

I have also applied a log transformation to stabilise variance and used first differencing (d=1) to make series stationary.



Results

The SARIMA (2,1,1) (1,1,1,12) model and SARIMA (1,1,1) (1,1,1,12) results are shown below:

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                        energyindex_log   No. Observations:              933
Model:             SARIMAX(2, 1, 1)x(1, 1, 1, 12)   Log Likelihood            2217.325
Date:                            Sat, 05 Apr 2025   AIC                      -4422.650
Time:                                    17:22:07   BIC                      -4393.796
Sample:                                01-01-1939   HQIC                     -4411.631
                                     - 09-01-2016
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.6684      0.043     15.559      0.000       0.584       0.753
ar.L2         -0.0953      0.032     -2.961      0.003      -0.158      -0.032
ma.L1         -0.8495      0.036    -23.772      0.000      -0.919      -0.779
ar.S.L12       0.0322      0.042      0.775      0.439      -0.049       0.114
ma.S.L12      -0.7050      0.031    -22.762      0.000      -0.766      -0.644
sigma2         0.0004   1.64e-05     26.269      0.000       0.000       0.000
===================================================================================
Ljung-Box (L1) (Q):                   0.14   Jarque-Bera (JB):                67.07
Prob(Q):                              0.71   Prob(JB):                         0.00
Heteroskedasticity (H):               4.05   Skew:                             0.03
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.33
===================================================================================
```

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                      energyindex_log   No. Observations:                  933
Model:             SARIMAX(1, 1, 1)x(1, 1, 1, 12)   Log Likelihood                2213.104
Date:                            Sat, 05 Apr 2025   AIC                          -4416.208
Time:                                    17:46:05   BIC                          -4392.163
Sample:                                01-01-1939   HQIC                         -4407.025
                                     - 09-01-2016
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.6672      0.040     16.663      0.000       0.589       0.746
ma.L1         -0.8964      0.026    -34.130      0.000      -0.948      -0.845
ar.S.L12       0.0755      0.043      1.747      0.081      -0.009       0.160
ma.S.L12      -0.7085      0.032    -22.106      0.000      -0.771      -0.646
sigma2         0.0004   1.69e-05     25.958      0.000       0.000       0.000
===================================================================================
Ljung-Box (L1) (Q):                   2.53   Jarque-Bera (JB):                56.21
Prob(Q):                              0.11   Prob(JB):                         0.00
Heteroskedasticity (H):               4.06   Skew:                             0.06
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.21
===================================================================================
```

The key residual diagnostics from the 2 SARIMA models include:

- Ljung-Box Test (autocorrelation)
- Jarque-Bera Test (normality)
- Heteroskedasticity Test

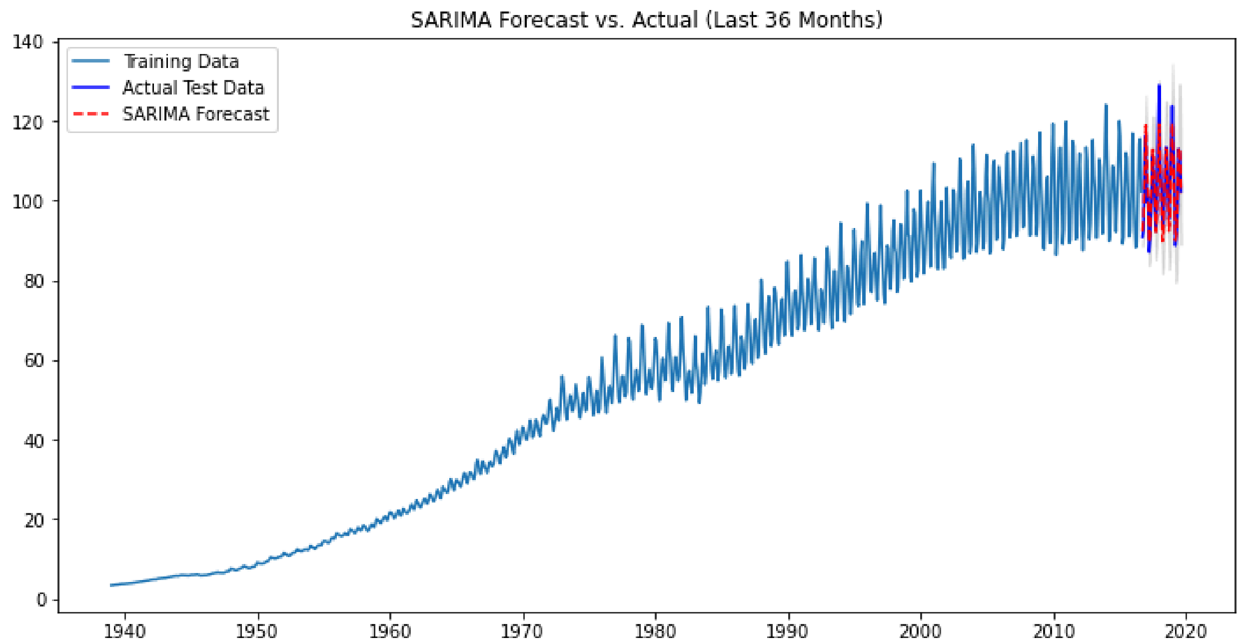
From each of the Ljung-Box test statistics, we can infer there is no autocorrelation ($p > 0.05$). While the Jarque-Bera and heteroskedasticity tests statistics show that residuals are not normal and variance does change over time. And SARIMA (2,1,1) has a lower AIC (-4422) than SARIMA (1,1,1) (-4416), thus suggesting a slightly better fit.

To assess model performance, Root Mean Squared Error (RMSE) was calculated using the last 36 months of data as a test set:

Model	**RMSE value**	**MAE value**
SARIMA (2,1,1) (1,1,1,12)	4.163	3.117
SARIMA (1,1,1) (1,1,1,12)	4.169	3.139

From the RMSE values, we can see SARIMA (2,1,1) performs slightly better than SARIMA (1,1,1), although the difference is marginal, thus suggesting similar predictive accuracy between the models. The MAE evaluation metric also shows slightly better result for SARIMA (2,1,1) , which is consistent with RMSE metric.

SARIMA(2,1,1)



SARIMA(1,1,1)