**QMM-5520- Forecasting**

**Forecasting U.S. Energy Consumption by Source using Time-Series Models**

**Group 4 members:** Akanksha Rambhad, Subaranjana Giridharan, Suneri Airkar, Vyshali Poola

**Introduction**

Forecasting renewable energy consumption has become increasingly important as the United States expands its commitment to clean and sustain energy sources. Understanding how renewable and biomass energy usage changes over time allows policymakers, energy planners, and sustainability analysts to better prepare for future demand, evaluate policy impact, and support long term resource planning. Accurate forecasts also help improve decision making related to infrastructure, budgeting, and energy reliability.

For this project, we use monthly U.S. energy data from the Energy Information Administration (EIA), focusing on two variables **Total Renewable Energy Consumption** and **Biomass Energy Consumption**. These measures were chosen because they represent major components of the renewable energy sector and show clear trends and seasonal patterns.

**Project Overview & Objectives**

This project focuses on forecasting monthly U.S. renewable and biomass energy consumption using a set of classical time series methods. The analysis is based on monthly data from the Energy Information Administration (EIA), which offers a detailed look at historical energy trends across the country. By studying these patterns, we aim to understand how renewable and biomass energy usage changes over time and how it may evolve in the near future.

### **1. Forecasting Objective**

Our main goal is to generate 12-month forecasts for renewable and biomass energy consumption. By focusing on these two major components of the U.S. renewable energy profile, we aim to capture both seasonal behavior and long term growth patterns.

**2. Modeling Goal**

To build these forecasts, we compare several time series techniques, including Holt’s Winters Exponential Smoothing, Multiple linear and non-linear regression, Seasonal ARIMA, and a Deseasonalized regression model. Each method captures different characteristics of the data, and comparing their performance allows us to identify the most accurate and reliable approach.

**3. Actionable Insights**

The results of this analysis can help guide planning around renewable energy supply and demand. Better forecasts can support decisions related to sustainability, policy development, and resource allocation. These insights can be valuable for government agencies, energy planners, and others working to understand future energy needs and evaluate long term renewable energy strategies.

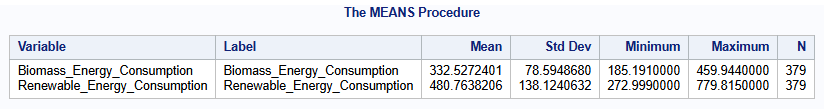
**Data Collection & Data Partition**

We used monthly U.S. energy consumption data from the U.S. Energy Information Administration (EIA), obtained through their open data portal and compiled in the file *Energy\_Consumption\_Production.xlsx*. The dataset includes 379 monthly observations measured in trillion BTUs, providing enough historical depth to analyze long term trends and seasonal behavior.

For model development, we primarily used a **90% training and 10% testing split**, while also experimenting with **80:20** and **70:30** partitions to evaluate the stability and consistency of model performance across different sample sizes.

**Data Source:** U.S. Energy Information Administration (EIA)  
<https://www.eia.gov/totalenergy/data/annual/index.php>

**Graphs and Summary Statistics**

****

Renewable energy consumption is higher on average (481 trillion BTUs) than biomass (333 trillion BTUs), with more variability over time. Both variables show a wide range of values, indicating substantial growth over time and along with meaningful variation from month to month.

| **Renewable Energy Consumption** | **Biomass Energy Consumption** |
| --- | --- |
| Time-series plot | Time-series plot |
| ACF plot | ACF plot |

**Renewable Energy Consumption and Biomass Energy Consumption**

The time-series plot shows a clear upward trend and the presence of noticeable spikes in the pattern. This is confirmed by the ACF plot, where the autocorrelations are not declining quickly towards zero which indicates a trend component. In addition, the autocorrelations are higher at lags 12, 24, and 36 which indicates the presence of a seasonal component. This means the series is **not stationary**.

**Models Developed**

1. **Holt's Winters Exponential Smoothing**
   1. Variable : **renewable\_energy\_consumption**

|  |  |
| --- | --- |
|

**Interpretation of the smoothing parameters**

**Winter’s Additive**

The level smoothing parameter is **0.508**, which indicates that a moderate amount of weight is assigned to the most recent observations. The trend smoothing parameter is **0.0010**, which indicates that the most recent trend has almost no impact on the forecasted value. The seasonal smoothing parameter is **0.377**, which indicates that the most recent seasonality has a less impact on the forecasted values.

**Winter’s Multiplicative**

The level smoothing parameter is **0.515**, which indicates that a moderate amount of weight is assigned to the most recent observations. The trend smoothing parameter is **0.0010**, which indicates that the most recent trend has almost no impact on the forecasted value. The seasonal smoothing parameter is **0.298**, which indicates that the most recent seasonality has a less impact on the forecasted values.

**Actual Vs. Predicted plots**

| Winter’s Additive | Winter’s Multiplicative |
| --- | --- |
|  |  |

**Error Measures of Model Fit and Model Accuracy**

| **Winter’s Additive Model Fit** | **Winter’s Additive Model Accuracy** |
| --- | --- |
|  |  |

| **Winters’s Multiplicative Model Fit** | **Winter’s Multiplicative Model Accuracy** |
| --- | --- |
|  |  |

| **Model** | **Model Fit** | | | **Model Accuracy** | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **MAPE** | **MAE** | **MSE** | **MAPE** | **MAE** | **MSE** |
| **Winter’s Additive** | 2.810% | 11.862 | 256.981 | 2.780% | 19.179 | 543.538 |
| **Winter’s Multiplicative** | 2.795% | 11.807 | 252.257 | 2.629% | 18.205 | 500.567 |

In terms of Model fit, Winter’s Multiplicative model has lower MAPE,MAE and MSE values. In terms of Model accuracy also Winter’s Multiplicative model has lower MAPE,MAE and MSE values. Our ultimate goal is to generate accurate forecasts and since the accuracy of Winter’s Multiplicative model is better, the appropriate model to use is **Winter’s Multiplicative model** to generate forecasts.

* 1. Variable: **biomass\_energy\_consumption**

|  |  |
| --- | --- |
|

**Interpretation of the Smoothing parameters:**

**Winter’s Additive Method**

The level smoothing parameter is **0.472**, which indicates that a moderate amount of weight is assigned to the most recent observations. The trend smoothing parameter is **0.0010**, which indicates that the most recent trend has almost no impact on the forecasted value. The seasonal smoothing parameter is **0.431**, which indicates that the most recent seasonality has a moderate impact on the forecasted values.

**Winter’s Multiplicative Method**

The level smoothing parameter is **0.431**, which indicates that a moderate amount of weight is assigned to the most recent observations. The trend smoothing parameter is **0.0026**, which indicates that the most recent trend has almost no impact on the forecasted values. The seasonal smoothing parameter is **0.478**, which indicates that the most recent seasonality has a moderate impact on the forecasted values.

**Actual Vs. Predicted plots**

| **Winter’s Additive Method** | **Winter’s Multiplicative Method** |
| --- | --- |
|  |  |

**Error Measures of Model Fit and Model Accuracy**

| **Additive’s Model Fit** | **Additive’s Model Accuracy** |
| --- | --- |
|  |  |

| **Winters’s Multiplicative Model Fit** | **Winter’s Multiplicative Model Accuracy** |
| --- | --- |
|  |  |

| **Model** | **Model Fit** | | | **Model Accuracy** | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **MAPE** | **MAE** | **MSE** | **MAPE** | **MAE** | **MSE** |
| **Additive** | 3.333% | 9.562 | 191.310 | 3.393% | 13.922 | 329.897 |
| **Multiplicative** | 3.327% | 9.573 | 189.709 | 3.700% | 15.246 | 352.407 |

In terms of Model fit, the additive and multiplicative models perform very similarly; however, the multiplicative model shows slightly lower MAPE and MSE values, indicating a marginally better fit. In terms of Model accuracy, the additive model has lower MAPE, MAE, and MSE values compared to the multiplicative model. Our ultimate goal is to generate accurate forecasts, and since the additive model provides better forecasting accuracy, the appropriate model to use for generating forecasts is the Winter’s Additive model.

1. **Multiple Linear & Non-linear Regression**
   1. Variable : **renewable\_energy\_consumption**

Multiple Linear Regression

|  |  |
| --- | --- |
|  |

We fitted multiple linear models using time (t) along with 11 monthly dummy variables, where January serves as the reference month.

**Equation of the Fit line**

𝑦̂ = 274.44492 +1.11465(t) −40.19497(m2) +3.92150(m3) −7.57507(m4) +12.15087(m5) −1.93793(m6) +5.52536(m7) −6.62255(m8) −38.70454(m9) −18.22265(m10) −22.33014(m11) −1.51015(m12)​

For January, the estimated renewable energy consumption at t=0 is approximately 274.44 trillion BTUs. On average, renewable energy consumption increases by about 1.11 trillion BTUs per month, regardless of seasonal effects. For February, consumption is 40.19 trillion BTUs lower than in January. For September, consumption is 38.70 trillion BTUs lower than in January.

### **Evaluation**

1. The slope coefficient is positive, indicating a reasonable and upward long-term pattern in the data, which makes the model logical.
2. The p-values for t, m2, and m9 are below 5%, and their confidence intervals do not include zero; therefore, these coefficients are statistically significant. For m3, m4, m5, m6, m7, m8, m10, m11, and m12, the p-values exceed 5%, so those monthly effects are not statistically significant.
3. Since Pr > F is less than 0.05, the overall model is statistically meaningful.
4. The adjusted R² value of 0.8365 is fairly high, suggesting a good fit to the data.
5. All VIFs are well below 10, indicating no multicollinearity issues.

**Assumptions**

|  |  |  |  |
| --- | --- | --- | --- |

1. The relationship between the forecast variable and the predictor variable is linear, which is seen from the time-series plot.
2. The QQ plot and the histogram of residuals show that the residuals are reasonably normally distributed.
3. The residuals vs. predicted values plot displays a pattern, indicating non constant variance in the residuals.
4. From Durbin-Watson D test p-value for positive autocorrelation < 0.05, therefore the errors are not independent. Thus, the independence assumption is violated.

**Non-linear Regression**

|  |  |
| --- | --- |
|  |

We fitted a nonlinear regression model by including both a linear time variable (t) and a squared time term (t2), along with 11 monthly dummy variables, using January as the baseline month.

**Equation of the Fit line**

y^​=342.54485 −0.10111(t) +0.00355(t2) −40.13881(m2) +3.93572(m3) −7.56441(m4) +12.15087(m5 ) −0.47580(m6) +7.68686(m7) −4.18308(m8) −36.25822(m9) −15.78048(m10) −19.86990(m11) +0.91268(m12)

For January, renewable energy consumption at t=0 is estimated to be about 342.54 trillion BTUs.

Because the coefficient of t is negative while the coefficient of t2 is positive, the model suggests that consumption changes over time in a non-linear manner, with the trend gradually shifting rather than remaining constant. For February, consumption is 40.14 trillion BTUs lower than in January. For September, consumption is 36.26 trillion BTUs lower than in January.

### **Evaluation**

1. The slope coefficient is positive, due to which we can say that the model is logical.
2. The p-values indicate that t², m2, and m9 are statistically significant at the 5% level. Their confidence intervals do not contain zero. The remaining monthly dummy variables (m3, m4, m5, m6, m7, m8, m10, m11, m12) have p-values greater than 5%, so their effects are not statistically significant.
3. The value of Pr > F is < 0.0001, confirming that the nonlinear model is statistically significant overall.
4. The adjusted R² value of 0.9035 is higher than that of the linear model, indicating a substantial improvement in fit.
5. All VIF values remain under 10, indicating no multicollinearity issues.

**Assumptions**

|  |  |  |  |
| --- | --- | --- | --- |

1. The QQ Plot and histogram show that the residuals are approximately normally distributed.
2. The residuals vs. predicted values plot again shows a pattern, indicating non constant variance.
3. From Durbin-Watson D test p-value for positive autocorrelation < 0.05, therefore the errors are not independent. Thus, the independence assumption is violated.

**Actual vs Predicted Plots**

| Multiple Linear Regression | Non-Linear Regression |
| --- | --- |
|  |  |

**Error Measures of Model Fit and Model Accuracy**

| **Multiple Linear Regression** | **Non-Linear Regression** |
| --- | --- |
|  |  |

| **Model** | **Model Fit** | | | **Model Accuracy** | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **MAPE** | **MAE** | **MSE** | **MAPE** | **MAE** | **MSE** |
| **Multiple Linear** | 10.359% | 40.57 | 2303.459 | 5.668% | 41.023 | 2151.662 |
| **Non-linear** | 7.407% | 31.347 | 1355.681 | 7.602% | 53.079 | 3280.131 |

In terms of Model fit, the non-linear model has slightly lower values for MAPE, MAE and MSE. But in terms of Model accuracy, the linear model model has lower values for MAPE, MAE and MSE. Since our goal is to generate forecasts, we would say that a linear model is better to generate forecasts.

* 1. Variable : **biomass\_energy\_consumption**

Multiple linear Regression:

|  |  |
| --- | --- |
|  |

We tested multiple linear and non-linear regression models with time (t) and 11 monthly dummy variables, using January as the reference month.  
From the parameter estimates table:  
**Equation of the fit line**   
𝑦̂ = 212.51120 + 0.68574(t) - 31.87850(m2) - 7.10194(m3) - 19.87382(m4) - 7.93784(m5) - 14.47304(m6) + 5.00861(m7) + 5.66379(m8) - 12.81513(m9) + 3.55962(m10) - 4.70580(m11) + 5.20117(m12)

For the month of January, the biomass energy consumed at time t=0 is approximated to be 212.5 trillion BTUs. Every month, on average, energy consumed increases by 0.68 trillion BTUs, independent of the month-specific seasonal effect. For the month of February, there is 31.87 trillion BTUs more energy consumed than the month of January. For the month of August, there is 5.66 trillion BTUs more energy consumed than the month of January.

**Evaluation**

1. The slope of the model appears to be positive, due to which we can say that the model is logical.
2. The p-values for t, m2, m4 are less than 5% and 0 is not a possible value within CI, we can say the slopes for them are statistically significant. Whereas for m3, m5, m6, m7, m8, m9, m11 and m12, the p-value is greater than 5%, hence the slops for them are not statistically significant
3. The value for Pr > F is less than 5%, hence the model is statistically significant.
4. A high adj R² of 0.78 seems to be good.
5. There is no indication of multicollinearity since all the VIF values are less than 10.

**Assumptions**

|  |  |  |  |
| --- | --- | --- | --- |

1. The relationship between the forecast variable and the predictor variable is linear, which we can see from the time-series plot.
2. Quantile plot and histogram indicate that the residuals are normally distributed.
3. The values in the scatter plot appear to have a shape then the residuals do not have constant variance.
4. p-value for positive autocorrelation < 0.05, therefore the errors are not independent.

**Non-linear Regression**

|  |  |
| --- | --- |
|  |

From the parameter estimates table:  
Equation of the fit line :

𝑦̂ = 227.40437 + 0.41986(t) + 0.00077743(t²) - 31.87617(m2) - 7.09883(m3) - 19.87149(m4) - 7.93784(m5) - 13.94517(m6) + 5.54037(m7) + 6.19788(m8) - 12.28026(m9) + 4.09372(m10) - 4.17404(m11) + 5.72904(m12)

For the month of January, the biomass energy consumed at time t=0 is approximated to be 227.4 trillion BTUs. Every month, on average, energy consumed increases by 0.41 trillion BTUs, independent of the month-specific seasonal effect.

For the month of February, there is 31.87 trillion BTUs more energy consumed than the month of January. For the month of August, there is 6.19 trillion BTUs more energy consumed than the month of January.

**Evaluations**:

1. The slope of the model appears to be positive, due to which we can say that the model is logical.
2. The p-values for t, t2, m2, m4 are less than 5% and 0 is not a possible value within CI, we can say the slopes for them are statistically significant. Whereas for m3, m5, m6, m7, m8, m9, m11 and m12, the p-value is greater than 5%, hence the slops for them are not statistically significant
3. The value for Pr > F is less than 5%, hence the model is statistically significant.
4. A high adj R² of 0.79 seems to be good.
5. There is multicollinearity since the VIF values for t and t2 are greater than 10.

**Assumptions**:

|  |  |  |  |
| --- | --- | --- | --- |

1. Quantile plot and histogram indicate that the residuals are normally distributed.
2. The values in the scatter plot appear to have a shape then the residuals do not have constant variance.
3. p-value for positive autocorrelation < 0.05, therefore the errors are not independent.

**Actual Vs. Predicted plots**

| Multiple Linear Regression | Non-Linear Regression |
| --- | --- |
|  |  |

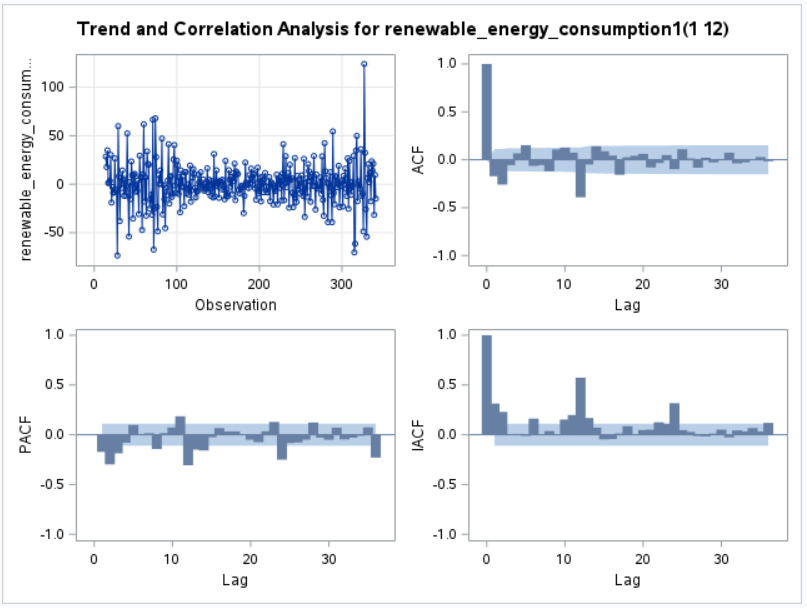
**Error Measures of Model Fit and Model Accuracy:**

| **Multiple Linear Regression** | **Non-Linear Regression** |
| --- | --- |
|  |  |

| **Model** | **Model Fit** | | | **Model Accuracy** | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **MAPE** | **MAE** | **MSE** | **MAPE** | **MAE** | **MSE** |
| **Linear** | 10.244% | 30.506 | 1239.714 | 8.657% | 35.742 | 1466.161 |
| **Non-linear** | 9.696% | 29.612 | 1194.383 | 13.578% | 56.235 | 3417.501 |

In terms of Model fit, the non-linear model has slightly lower values for MAPE, MAE and MSE. But in terms of Model accuracy, the linear model model has lower values for MAPE, MAE and MSE. Since our goal is to generate forecasts, we would say that a linear model is better.

1. **Seasonal ARIMA**
2. Variable : **renewable\_energy\_consumption**

****

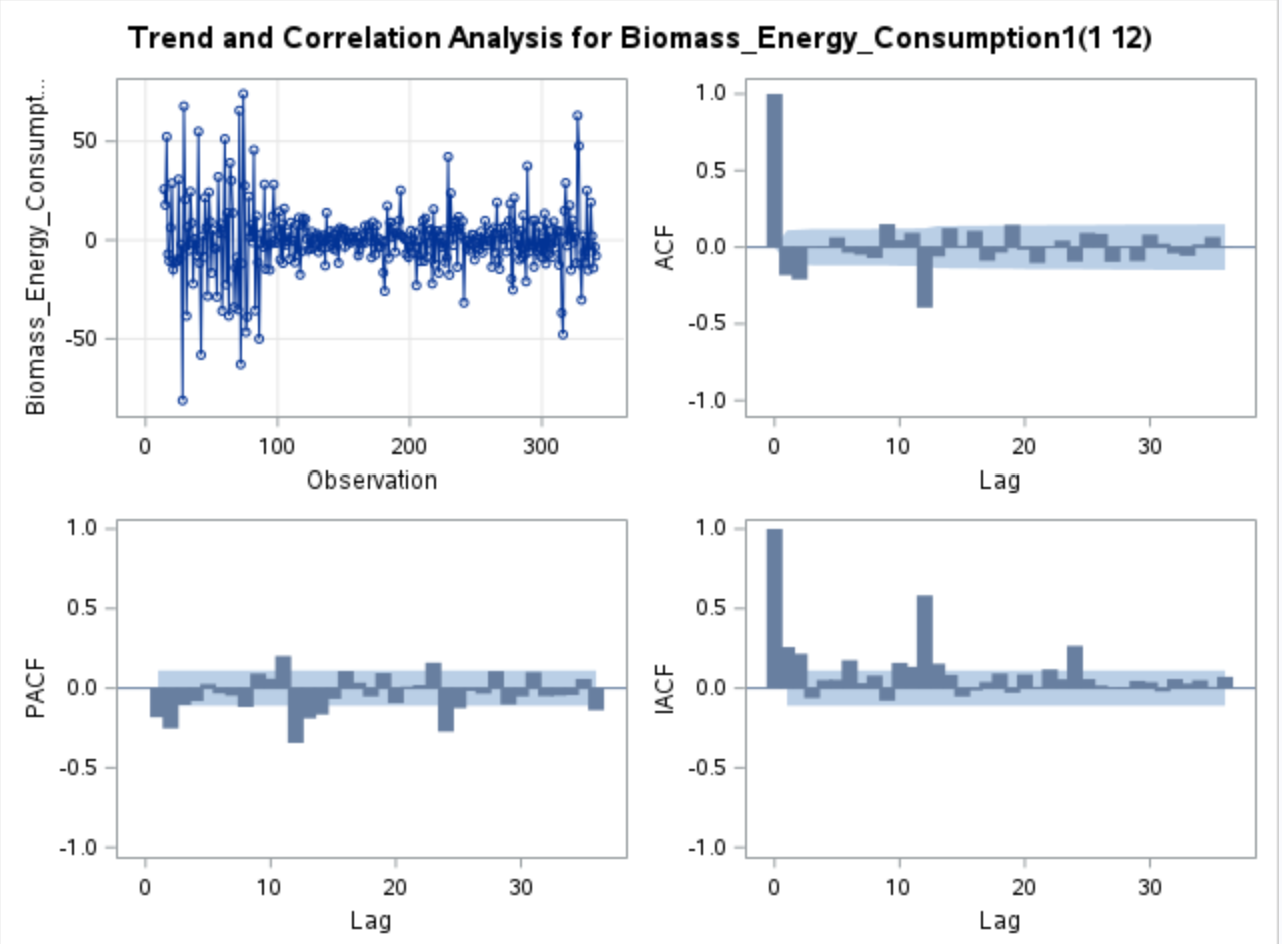
In this case, second-order differencing is done to make ACF plot stationary, there are 3 significant spikes in PACF plot i.e at 12th lag, 24th lag and 36th lag (seasonal AR term) and 1 significant spike in ACF plot at 12th lag (seasonal MA term), so therefore (P,D,Q) would be (3,1,1). Next I considered the ARMA model here as I could see a significant spike in PACF plot at 11th lag and in ACF plot at 5th lag so (p,d,q) would be (11,1,5). Therefore the ARIMA model which I choose is **ARIMA(11,1,5)(3,1,1).** The other ARIMA model I tried is **ARIMA(14,1,17)(3,1,1)** , where I could see a significant spike in PACF plot at 14th lag and in ACF plot at 17th lag, so (p,d,q) would be (14,1,17).

| **ARIMA(11,1,5)(3,1,1)** | **ARIMA(14,1,17)(3,1,1)** |
| --- | --- |
|  |  |

The residuals of the both selected models are not white noise residuals which indicates that the selected models are not appropriate. This means the model did not fully remove autocorrelation, so it wasn’t able to capture the seasonal structure of the data.

Even though multiple combinations were tried here i.e ARIMA(3,1,0)(3,1,1) and ARIMA(3,1,2)(3,1,1) with low order, none of the ARIMA models passed the residuals test, which means ARIMA is not a reliable model for forecasting this dataset.

1. Variable: **biomass\_energy\_consumption**

****

In this case, first-order and seasonal differencing were applied to make the series stationary. The seasonal PACF shows significant spikes at lags 12, 24, and 36, indicating a seasonal AR order of 3, while the ACF shows a strong spike at lag 12, suggesting a seasonal MA order of 1. Thus, the seasonal component is (P, D, Q) = (3, 1, 1).For the non-seasonal part, the PACF has a clear spike at lag 11 (AR term), and the ACF shows a significant spike at lag 9 (MA term), giving (p, d, q) (11, 1, 9). Therefore, one model is **ARIMA(11,1,9)(3,1,1).**

The second model is **ARIMA(14,1,2)(3,1,1),** based on a significant spike at lag 14 in the PACF (AR term) and a spike at lag 2 in the ACF (MA term), giving the non-seasonal orders (14, 1, 2).

| **ARIMA(11,1,9)(3,1,1)** | **ARIMA(14,1,2)(3,1,1)** |
| --- | --- |
|  |  |

The residuals of both selected models are not white noise, indicating that neither model is appropriate. The Ljung–Box p-values are all significant, showing that the residuals still contain autocorrelation. This means the ARIMA models were not able to fully capture the seasonal structure of the data. Even though several ARIMA model combinations were tested, none of them passed the white-noise residual test. Therefore, ARIMA is not a reliable modeling approach for forecasting this dataset. Several ARIMA combinations were explored, including ARIMA(2,1,2)(3,1,1) and ARIMA(2,1,9)(3,1,1).

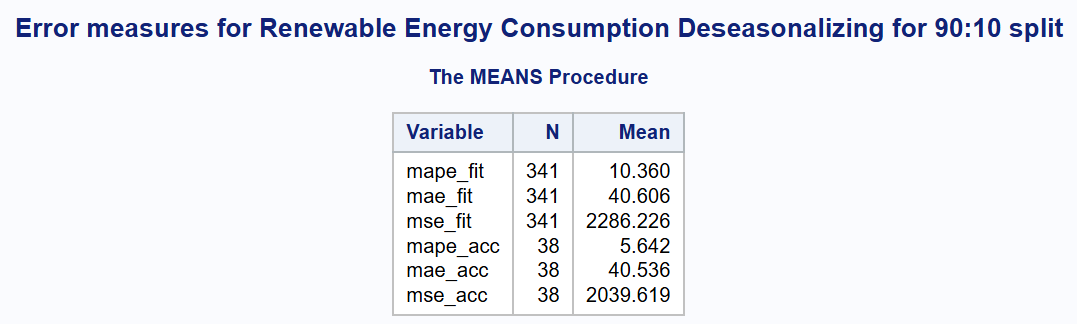
1. **Deseasonalization**
2. Variable : **renewable\_energy\_consumption**

|  | Equation:  𝑦̂ = 264.891 + 1.114t  This equation tells us that renewable energy consumption is increasing by about 1.114 trillion BTUs per month on average. When t=0, the model predicts a baseline level of about 264.89 trillion BTUs of renewable energy consumption. |
| --- | --- |

| **Deseasonalized** | **Reseasonalized** |
| --- | --- |
|  |  |

Visually, we can see that the model captures the trend well, but it doesn’t fully adapt to monthly fluctuations.

**Error Measures of Model Fit and Model Accuracy:**

****

| **Model** | **Model Fit** | | | **Model Accuracy** | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **MAPE** | **MAE** | **MSE** | **MAPE** | **MAE** | **MSE** |
| **Deseasonalization** | 10.360% | 40.606 | 2286.226 | 5.642% | 40.536 | 2039.619 |

The deseasonalized regression model shows lower MAPE, MAE, and MSE in the accuracy set compared to the fit set. This indicates that the model generalizes reasonably well to unseen data and maintains consistent forecasting performance.

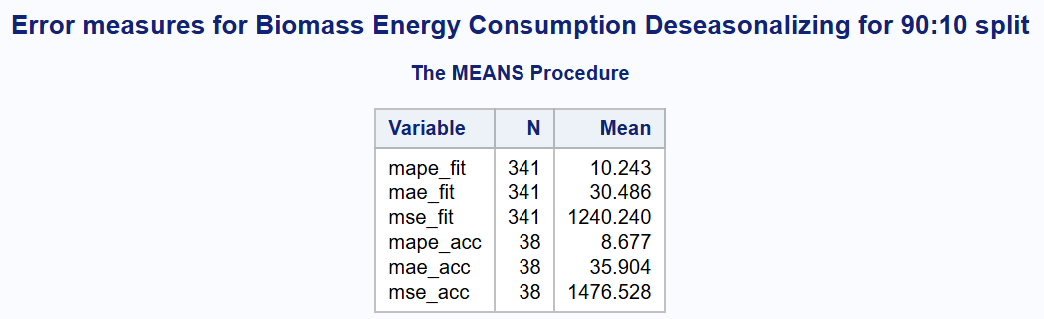
1. Variable: **biomass\_energy\_consumption**

|  | Equation:  𝑦̂ = 205.701 + 0.687t  This equation tells us that biomass energy consumption is increasing by about 0.687 trillion BTUs per month on average. When t=0, the model predicts a baseline level of about 205.7 trillion BTUs of biomass energy consumption. |
| --- | --- |

| **Deseasonalized** | **Reseasonalized** |
| --- | --- |
|  |  |

Visually, we can see that the model captures the trend well, but it doesn’t fully adapt to monthly fluctuations.

**Error Measures of Model Fit and Model Accuracy:**

****

| **Model** | **Model Fit** | | | **Model Accuracy** | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **MAPE** | **MAE** | **MSE** | **MAPE** | **MAE** | **MSE** |
| **Deseasonalization** | 10.243% | 30.486 | 1240.240 | 8.677% | 35.904 | 1476.528 |

The deseasonalization model shows strong performance, with relatively low error values for both model fit and model accuracy. Its lower MAPE, MAE, and MSE values for model accuracy indicate that the model effectively captures underlying trends while keeping prediction errors within a reasonable range.

**Comparison of all models**

1. Variable : **renewable\_energy\_consumption**

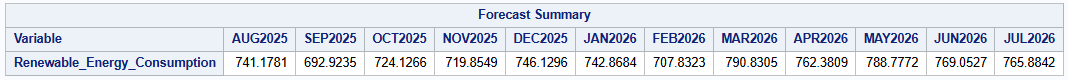
| **Model** | **Model Fit** | | | **Model Accuracy** | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **MAPE** | **MAE** | **MSE** | **MAPE** | **MAE** | **MSE** |
| **Holt’s Winters Exponential Smoothing** | 2.795% | 11.807 | 252.257 | 2.629% | 18.205 | 500.567 |
| **Multiple Linear Regression** | 10.359% | 40.57 | 2303.459 | 5.668% | 41.023 | 2151.662 |
| **Non-linear Regression** | 7.407% | 31.347 | 1355.681 | 7.602% | 53.079 | 3280.131 |
| **Deseasonalization** | 10.360% | 40.606 | 2286.226 | 5.642% | 40.536 | 2039.619 |

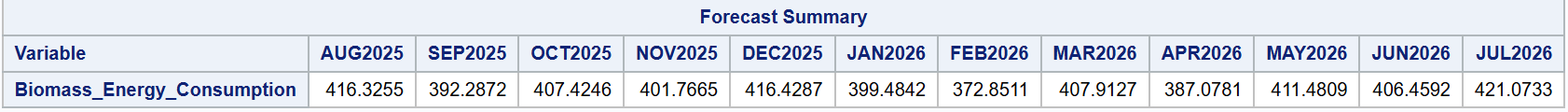
b.Variable: **biomass\_energy\_consumption**

| **Model** | **Model Fit** | | | **Model Accuracy** | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **MAPE** | **MAE** | **MSE** | **MAPE** | **MAE** | **MSE** |
| **Holt’s Winters Exponential Smoothing** | 3.333% | 9.562 | 191.310 | 3.393% | 13.922 | 329.897 |
| **Multiple Linear Regression** | 10.244% | 30.506 | 1239.714 | 8.657% | 35.742 | 1466.161 |
| **Non-linear Regression** | 9.696% | 29.612 | 1194.383 | 13.578% | 56.235 | 3417.501 |
| **Deseasonalization** | 10.243% | 30.486 | 1240.240 | 8.677% | 35.904 | 1476.528 |

For both variables Renewable and Biomass energy consumption, Holt’s Winters exponential model gives the lowest errors (MAPE,MAE and MSE) among all models in both model fit as well as model accuracy, making it the most accurate. Holt’s winter’s model is **parsimonious** as it uses fewer parameters but still performs better. Its simplicity and strong performance make it the best choice for forecasting this series.

**12-Month Forecast Summary**

****

****

Since Holt’s Winters exponential smoothing model performed the best, it was used to generate the 12-month forecast for both Renewable and Biomass Energy Consumption. The forecasts show steady seasonal patterns and gradual growth, providing a clear outlook for the upcoming year.

**Conclusion**

This analysis indicates that Holt's Winters exponential smoothing models provide the most reliable short‑term forecasts of U.S. renewable and biomass energy consumption. A multiplicative version would be preferred for renewable energy and an additive one for biomass, as they have better accuracy on test sets.

Both series continue to rise with clear seasonal peaks and troughs in the 12‑month forecasts, which provide useful guidance for capacity planning, budgeting, and near‑term policy decisions related to renewable energy.

However, these models use only historical consumption and do not take into account factors such as economic conditions, prices, technology, or new policies; alternative models have issues such as residual autocorrelation, and thus future work must consider models using external predictors and more extensive forecast validation.

**SAS Code**

/\* Importing Excel file into SAS \*/

proc import out=energy datafile="/home/u64128357/sasuser.v94/Energy\_Consumption\_Production.xlsx"

dbms=xlsx replace;

run;

/\* Descriptive statistics \*/

proc means data=energy chartype mean std min max n vardef=df;

var Biomass\_Energy\_Consumption Renewable\_Energy\_Consumption;

run;

/\* Biomass Energy Consumption \*/

/\* Creating a time series plot \*/

proc sgplot data=energy;

series x= month y=Biomass\_Energy\_Consumption;

title "Monthly Consumption for Biomass Energy Consumption (1994-2025)";

run;

/\*Creating an ACF plot\*/

proc timeseries data=energy plots=acf out=\_null\_;

var Biomass\_Energy\_Consumption ;

corr acf/nlag=36;

run;

/\* Renewable Energy Consumption \*/

/\* Creating a time series plot \*/

proc sgplot data=energy;

series x= month y=Renewable\_Energy\_Consumption;

title "Monthly Consumption for Renewable Energy Consumption (1994-2025)";

run;

/\*Creating an ACF plot\*/

proc timeseries data=energy plots=acf out=\_null\_;

var Renewable\_Energy\_Consumption ;

corr acf/nlag=36;

run;

/\* 90:10 split \*/

data energy;

set energy;

t=\_n\_;

Biomass\_Energy\_Consumption\_new=Biomass\_Energy\_Consumption;

Renewable\_Energy\_Consumption\_new=Renewable\_Energy\_Consumption;

if t>341 then Biomass\_Energy\_Consumption\_new=. ;

if t>341 then Renewable\_Energy\_Consumption\_new=.;

mth=month(month);

if mth=2 then m2=1; else m2=0;

if mth=3 then m3=1; else m3=0;

if mth=4 then m4=1; else m4=0;

if mth=5 then m5=1; else m5=0;

if mth=6 then m6=1; else m6=0;

if mth=7 then m7=1; else m7=0;

if mth=8 then m8=1; else m8=0;

if mth=9 then m9=1; else m9=0;

if mth=10 then m10=1; else m10=0;

if mth=11 then m11=1; else m11=0;

if mth=12 then m12=1; else m12=0;

t2=t\*t;

/\* t3=t\*t\*t; \*/

/\* t4=t\*t\*t\*t; \*/

/\* t5=t\*t\*t\*t\*t; \*/

run;

proc freq data=energy;

table mth m2 m3 m4 m5 m6 m7 m8 m9 m10 m11 m12;

run;

/\* Biomass Energy Consumption \*/

/\* Deseasonalizing (90:10)\*/

proc timeseries data=energy outdecomp=sa\_energy1 out=null;

decomp sa;

id month interval=month;

var Biomass\_Energy\_Consumption;

run;

data sa\_energy1a;

merge energy sa\_energy1(keep=sa);

si = Biomass\_Energy\_Consumption/sa;

t=\_n\_;

sa1=sa;

if t>341 then sa1=.;

run;

proc sgplot data=sa\_energy1a;

series x=month y=Biomass\_Energy\_Consumption;

series x=month y=sa;

title "Monthly Biomass Energy Consumption (Desasonalized) (90:10)";

run;

proc reg data=sa\_energy1a;

model sa1=t;

output out=sa\_energyout1 r=sa\_resid p=sa\_predict;

run;

data sa\_energyout1;

set sa\_energyout1;

energy\_reseason=sa\_predict\*si;

if t<=341 then

do;

mape\_fit=abs(sa\_resid/sa1)\*100;

mae\_fit=abs(sa\_resid);

mse\_fit=sa\_resid\*\*2;

end;

else if t>341 then

do;

mape\_acc=abs((sa-sa\_predict)/sa)\*100;

mae\_acc=abs(sa-sa\_predict);

mse\_acc=(sa-sa\_predict)\*\*2;

end;

run;

proc means data=sa\_energyout1 n mean maxdec=3;

var mape\_fit mae\_fit mse\_fit mape\_acc mae\_acc mse\_acc;

title "Error measures for Biomass Energy Consumption Deseasonalizing for 90:10 split ";

run;

proc sgplot data=sa\_energyout1;

series x=month y=Biomass\_Energy\_Consumption;

series x=month y=energy\_reseason;

title "Monthly Biomass Energy Consumption (Resasonalized)(90:10)";

run;

/\* Linear Regression using dummy variables for Biomass\_Energy\_Consumption \*/

proc reg data=energy;

model Biomass\_Energy\_Consumption\_new=t m2 m3 m4 m5 m6 m7 m8 m9 m10 m11 m12/clb vif dwprob;

output out=energyout1a p=biomass\_predict r=biomass\_resid;

run;

proc sgplot data=energyout1a;

series x=month y=Biomass\_Energy\_Consumption;

series x=month y=biomass\_predict;

title "Linear Trend for Biomass Energy Consumption for 90:10 split ";

run;

data energyout1a;

set energyout1a;

if t<=341 then

do;

mape\_fit=abs(biomass\_resid/Biomass\_Energy\_Consumption\_new)\*100;

mae\_fit=abs(biomass\_resid);

mse\_fit=(biomass\_resid)\*\*2;

end;

else if t>341 then

do;

mape\_acc=abs((Biomass\_Energy\_Consumption-biomass\_predict)/Biomass\_Energy\_Consumption)\*100;

mae\_acc=abs(Biomass\_Energy\_Consumption-biomass\_predict);

mse\_acc=(Biomass\_Energy\_Consumption-biomass\_predict)\*\*2;

end;

run;

proc means data=energyout1a n mean maxdec=3;

var mape\_fit mae\_fit mse\_fit mape\_acc mae\_acc mse\_acc;

title "Error measures for Biomass Energy Consumption Linear Trend for 90:10 split ";

run;

/\* Non-linear Trend for Biomass\_Energy\_Consumption \*/

proc reg data=energy;

model Biomass\_Energy\_Consumption\_new=t t2 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11 m12/clb vif dwprob;

output out=energyout1b p=biomass\_predict r=biomass\_resid;

run;

proc sgplot data=energyout1b;

series x=month y=Biomass\_Energy\_Consumption;

series x=month y=biomass\_predict;

title "Non-Linear Trend for Biomass Energy Consumption for 90:10 split ";

run;

data energyout1b;

set energyout1b;

if t<=341 then

do;

mape\_fit=abs(biomass\_resid/Biomass\_Energy\_Consumption\_new)\*100;

mae\_fit=abs(biomass\_resid);

mse\_fit=(biomass\_resid)\*\*2;

end;

else if t>341 then

do;

mape\_acc=abs((Biomass\_Energy\_Consumption-biomass\_predict)/Biomass\_Energy\_Consumption)\*100;

mae\_acc=abs(Biomass\_Energy\_Consumption-biomass\_predict);

mse\_acc=(Biomass\_Energy\_Consumption-biomass\_predict)\*\*2;

end;

run;

proc means data=energyout1b n mean maxdec=3;

var mape\_fit mae\_fit mse\_fit mape\_acc mae\_acc mse\_acc;

title "Error measures for Biomass Energy Consumption Non Linear Trend for 90:10 split ";

run;

/\* Holt's Exponential Smoothing Method (90:10 split) \*/

/\* Additive Winter's \*/

proc esm data=energy print=all lead=38 back=38 outfor=energyout1c out=\_null\_;

id month interval=month;

forecast Biomass\_Energy\_Consumption/model=addwinters;

run;

proc sgplot data=energyout1c;

series x=month y=actual;

series x=month y=predict;

title " Actual vs Predicted values for Biomass Energy Consumption ";

run;

/\* Multiplicative Winter's \*/

proc esm data=energy print=all lead=38 back=38 outfor=energyout1d out=\_null\_;

id month interval=month;

forecast Biomass\_Energy\_Consumption/model=winters;

run;

proc sgplot data=energyout1d;

series x=month y=actual;

series x=month y=predict;

title " Actual vs Predicted values for Biomass Energy Consumption ";

run;

/\* Seasonal ARIMA (90:10 split)\*/

/\*Model 1\*/

data energy;

set energy;

t=\_n\_;

Biomass\_Energy\_Consumption1=Biomass\_Energy\_Consumption;

if t>341 then Biomass\_Energy\_Consumption1=.;

run;

proc arima data=energy;

identify var=Biomass\_Energy\_Consumption1(12,1) whitenoise=ignoremiss nlag=36;

estimate p=(11)(12)(24)(36) q=(9)(12) whitenoise=ignoremiss; /\*(11,1,9)(3,1,1)\*/

\*estimate p=(2)(12)(24)(36) q=(2)(12) whitenoise=ignoremiss; /\*(2,1,2)(3,1,1)\*/

\*estimate p=(2)(12)(24)(36) q=(9)(12) whitenoise=ignoremiss; /\*(2,1,9)(3,1,1)\*/

forecast id=month interval=month lead=38 out=energyout1e;

run;

data energyout1e;

merge energy energyout1e;

if t<=341 then

do;

mape\_fit=abs(residual/Biomass\_Energy\_Consumption1)\*100;

mae\_fit=abs(residual);

mse\_fit=residual\*\*2;

end;

else if t>341 then

do;

mape\_acc=abs((Biomass\_Energy\_Consumption-forecast)/Biomass\_Energy\_Consumption)\*100;

mae\_acc=abs(Biomass\_Energy\_Consumption-forecast);

mse\_acc=abs(Biomass\_Energy\_Consumption-forecast)\*\*2;

end;

run;

proc means data=energyout1e n mean maxdec=3;

var mape\_fit mae\_fit mse\_fit mape\_acc mae\_acc mse\_acc;

run;

data energyout1e;

merge energy energyout1e;

run;

proc sgplot data= energyout1e;

series x=month y=Biomass\_Energy\_Consumption;

series x=month y=forecast;

title "Actual Vs Predicted plot for Biomass Energy Consumption";

run;

/\* Model 2\*/

data energy;

set energy;

t=\_n\_;

Biomass\_Energy\_Consumption1=Biomass\_Energy\_Consumption;

if t>341 then Biomass\_Energy\_Consumption1=.;

run;

proc arima data=energy;

identify var=Biomass\_Energy\_Consumption1(12,1) whitenoise=ignoremiss nlag=36;

estimate p= (14)(12)(24)(36) q=(2)(12) whitenoise=ignoremiss; /\*(14,1,2)(3,1,1)\*/

forecast id=month interval=month lead=38 out=energyout1f;

run;

data energyout1f;

merge energy energyout1f;

if t<=341 then

do;

mape\_fit=abs(residual/Biomass\_Energy\_Consumption1)\*100;

mae\_fit=abs(residual);

mse\_fit=residual\*\*2;

end;

else if t>341 then

do;

mape\_acc=abs((Biomass\_Energy\_Consumption-forecast)/Biomass\_Energy\_Consumption)\*100;

mae\_acc=abs(Biomass\_Energy\_Consumption-forecast);

mse\_acc=abs(Biomass\_Energy\_Consumption-forecast)\*\*2;

end;

run;

proc means data=energyout1f n mean maxdec=3;

var mape\_fit mae\_fit mse\_fit mape\_acc mae\_acc mse\_acc;

run;

data energyout1f;

merge energy energyout1f;

run;

proc sgplot data= energyout1f;

series x=month y=Biomass\_Energy\_Consumption;

series x=month y=forecast;

title "Actual Vs Predicted plot for Biomass Energy Consumption";

run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* Renewable Energy Consumption \*/

/\* Deseasonalizing (90:10)\*/

proc timeseries data=energy outdecomp=sa\_energy2 out=null;

decomp sa;

id month interval=month;

var renewable\_energy\_consumption;

run;

data sa\_energy2a;

merge energy sa\_energy2(keep=sa);

si = renewable\_energy\_consumption/sa;

t=\_n\_;

sa1=sa;

if t>341 then sa1=.;

run;

proc sgplot data=sa\_energy2a;

series x=month y=renewable\_energy\_consumption;

series x=month y=sa;

title "Monthly Renewable Energy Consumption (Desasonalized) (90:10)";

run;

proc reg data=sa\_energy2a;

model sa1=t;

output out=sa\_energyout2 r=sa\_resid p=sa\_predict;

run;

data sa\_energyout2;

set sa\_energyout2;

energy\_reseason=sa\_predict\*si;

if t<=341 then

do;

mape\_fit=abs(sa\_resid/sa1)\*100;

mae\_fit=abs(sa\_resid);

mse\_fit=sa\_resid\*\*2;

end;

else if t>341 then

do;

mape\_acc=abs((sa-sa\_predict)/sa)\*100;

mae\_acc=abs(sa-sa\_predict);

mse\_acc=(sa-sa\_predict)\*\*2;

end;

run;

proc means data=sa\_energyout2 n mean maxdec=3;

var mape\_fit mae\_fit mse\_fit mape\_acc mae\_acc mse\_acc;

title "Error measures for Renewable Energy Consumption Deseasonalizing for 90:10 split ";

run;

proc sgplot data=sa\_energyout2;

series x=month y=renewable\_energy\_consumption;

series x=month y=energy\_reseason;

title "Monthly Energy Consumption (Resasonalized)(90:10)";

run;

/\* Linear Regression using dummy variables for Renewable\_Energy\_Consumption \*/

proc reg data=energy;

model Renewable\_Energy\_Consumption\_new=t m2 m3 m4 m5 m6 m7 m8 m9 m10 m11 m12/clb vif dwprob;

output out=energyout2a p=renew\_predict r=renew\_resid;

run;

proc sgplot data=energyout2a;

series x=month y=Renewable\_Energy\_Consumption;

series x=month y=renew\_predict;

title "Linear Trend for Renewable Energy Consumption for 90:10 split ";

run;

data energyout2a;

set energyout2a;

if t<=341 then

do;

mape\_fit=abs(renew\_resid/Renewable\_Energy\_Consumption\_new)\*100;

mae\_fit=abs(renew\_resid);

mse\_fit=(renew\_resid)\*\*2;

end;

else if t>341 then

do;

mape\_acc=abs((Renewable\_Energy\_Consumption-renew\_predict)/Renewable\_Energy\_Consumption)\*100;

mae\_acc=abs(Renewable\_Energy\_Consumption-renew\_predict);

mse\_acc=(Renewable\_Energy\_Consumption-renew\_predict)\*\*2;

end;

run;

proc means data=energyout2a n mean maxdec=3;

var mape\_fit mae\_fit mse\_fit mape\_acc mae\_acc mse\_acc;

title "Error measures for Renewable Energy Consumption Linear Trend for 90:10 split ";

run;

/\* Non-linear Trend for Renewable\_Energy\_Consumption \*/

proc reg data=energy;

model Renewable\_Energy\_Consumption\_new=t t2 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11 m12/clb vif dwprob;

output out=energyout2b p=renew\_predict r=renew\_resid;

run;

proc sgplot data=energyout2b;

series x=month y=Renewable\_Energy\_Consumption;

series x=month y=renew\_predict;

title "Non-Linear Trend for Renewable Energy Consumption for 90:10 split ";

run;

data energyout2b;

set energyout2b;

if t<=341 then

do;

mape\_fit=abs(renew\_resid/Renewable\_Energy\_Consumption\_new)\*100;

mae\_fit=abs(renew\_resid);

mse\_fit=(renew\_resid)\*\*2;

end;

else if t>341 then

do;

mape\_acc=abs((Renewable\_Energy\_Consumption-renew\_predict)/Renewable\_Energy\_Consumption)\*100;

mae\_acc=abs(Renewable\_Energy\_Consumption-renew\_predict);

mse\_acc=(Renewable\_Energy\_Consumption-renew\_predict)\*\*2;

end;

run;

proc means data=energyout2b n mean maxdec=3;

var mape\_fit mae\_fit mse\_fit mape\_acc mae\_acc mse\_acc;

title "Error measures for Renewable Energy Consumption Non Linear Trend for 90:10 split ";

run;

/\* Holt's Exponential Smoothing Method (90:10 split) \*/

/\* Additive Winter's \*/

proc esm data=energy print=all lead=38 back=38 outfor=energyout2c out=\_null\_;

id month interval=month;

forecast renewable\_energy\_consumption/model=addwinters;

run;

proc sgplot data=energyout2c;

series x= month y= actual;

series x= month y =predict;

title "Actual vs Predicted values for Renewable Energy Consumption";

run;

/\* Multiplicative Winter's \*/

proc esm data=energy print=all lead=38 back=38 outfor=energyout2d out=\_null\_;

id month interval=month;

forecast renewable\_energy\_consumption/model=winters;

run;

proc sgplot data=energyout2d;

series x= month y= actual;

series x= month y =predict;

title "Actual Vs Predicted values for Renwable Energy Consumption";

run;

/\* Seasonal ARIMA (90:10 split)\*/

/\* Model 1 \*/

data energy;

set energy;

t=\_n\_;

renewable\_energy\_consumption1=renewable\_energy\_consumption;

if t>341 then renewable\_energy\_consumption1=.;

run;

proc arima data=energy;

identify var=renewable\_energy\_consumption1(12,1) whitenoise=ignoremiss nlag=36;

estimate p=(11)(12)(24)(36) q=(5)(12) whitenoise=ignoremiss; /\*ARIMA(11,1,5)(3,1,1) \*/

\*estimate p=(3)(12)(24)(36) q=(12) whitenoise=ignoremiss; /\*ARIMA(3,1,0)(3,1,1) \*/

\*estimate p=(3)(12)(24)(36) q=(2)(12) whitenoise=ignoremiss; /\*ARIMA(3,1,2)(3,1,1) \*/

forecast id=month interval=month lead=38 out=energyout2e;

run;

data energyout2e;

merge energy energyout2e;

if t<=341 then

do;

mape\_fit=(abs((residual)/renewable\_energy\_consumption1))\*100;

mae\_fit=abs(residual);

mse\_fit=residual\*\*2;

end;

else if t>341 then

do;

mape\_acc=abs((renewable\_energy\_consumption-forecast)/renewable\_energy\_consumption)\*100;

mae\_acc=abs(renewable\_energy\_consumption-forecast);

mse\_acc=(renewable\_energy\_consumption-forecast) \*\*2;

end;

run;

proc means data=energyout2e n mean maxdec=3;

var mape\_fit mae\_fit mse\_fit mape\_acc mae\_acc mse\_acc;

run;

data energyout2e;

merge energy energyout2e;

run;

proc sgplot data= energyout2e;

series x=month y=renewable\_energy\_consumption;

series x=month y=forecast;

title "Actual Vs Predicted values for Renewable Energy Consumption";

run;

/\* Model 2 \*/

data energy;

set energy;

t=\_n\_;

renewable\_energy\_consumption1=renewable\_energy\_consumption;

if t>341 then renewable\_energy\_consumption1=.;

run;

proc arima data=energy;

identify var=renewable\_energy\_consumption1(12,1) whitenoise=ignoremiss nlag=36;

estimate p=(14)(12)(24)(36) q=(17)(12) whitenoise=ignoremiss; /\*ARIMA(14,1,17)(3,1,1) \*/

forecast id=month interval=month lead=38 out=energyout2f;

run;

data energyout2f;

merge energy energyout2f;

if t<=341 then

do;

mape\_fit=(abs((residual)/renewable\_energy\_consumption1))\*100;

mae\_fit=abs(residual);

mse\_fit=residual\*\*2;

end;

else if t>341 then

do;

mape\_acc=abs((renewable\_energy\_consumption-forecast)/renewable\_energy\_consumption)\*100;

mae\_acc=abs(renewable\_energy\_consumption-forecast);

mse\_acc=(renewable\_energy\_consumption-forecast) \*\*2;

end;

run;

proc means data=energyout2f n mean maxdec=3;

var mape\_fit mae\_fit mse\_fit mape\_acc mae\_acc mse\_acc;

run;

data energyout2f;

merge energy energyout2f;

run;

proc sgplot data= energyout2f;

series x=month y=renewable\_energy\_consumption;

series x=month y=forecast;

title "Actual Vs Predicted values for Renewable Energy Consumption";

run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* Forecast for next 12-month period using Holt's Winter Exponential Smoothing model\*/

/\* Renewable Energy Consumption \*/

proc esm data=energy lead=12 print=all outfor=energyout3 out=null;

id month interval=month;

forecast renewable\_energy\_consumption/model=winters;

run;

proc sgplot data=energyout3;

series x= month y= actual;

series x= month y =predict;

title " Actual vs Predicted values for Renewable Energy Consumption ";

run;

/\* Biomass Energy Consumption \*/

proc esm data=energy print=all lead=12 outfor=energyout4 out=\_null\_;

id month interval=month;

forecast Biomass\_Energy\_Consumption/model=addwinters;

run;

proc sgplot data=energyout4;

series x=month y=actual;

series x=month y=predict;

title " Actual vs Predicted values for Biomass Energy Consumption ";

run;