# ELECTRICITY DEMAND FORECASTING IN VICTORIA , AUSTRALIA

Mudita Garg

# 1. INTRODUCTION

Electricity demand is a measure of the average rate at which your home or business consumes electricity in a defined time interval. In general, the more electrical devices you operate at one time, the higher is your demand.

Forecasting of electricity demand plays an essential role in the electrical industry, as it provides the basis for making decisions in power system planning and operation.

A great variety of methods for predicting electricity demand are being used by electrical companies, which apply to short-term, medium-term, or long-term forecasting.

But the use of electricity arises from complex interactions between meteorological and socio-economic factors. In such a dynamic environment, ordinary forecasting techniques are not sufficient, and more sophisticated methods are needed.

Lockdown measures have significantly reduced electricity demand, affecting in turn the power mix. Increases in residential demand were far outweighed by reductions in commercial and industrial operations.

Daily data collected for more than 30 countries, representing over one-third of global electricity demand, show that the extent of demand declines depends on the duration and stringency of lockdowns. On average we find that every month of full lockdown reduces demand by 20% on average, or over 1.5% on an annual basis.

During COVID, after a significant rise in new daily cases, Victoria, the second-largest state in Australia by population, introduced some of the strictest movement control measures in the world. Travel restrictions, widely credited for helping to reduce the virus transmission rate, have harmed businesses. Among other indicators, electricity demand is argued to reflect a level of economic activity. The deeper economic activity falls during the pandemic, the further, and likely longer, the climb back to pre-pandemic level may take.

## 2. LITERATURE REVIEW

**Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China Yuanyuan Wang a, Jianzhou Wang b,n, Ge Zhao b, Yao Dong b**

Although a seasonal ARIMA model is widely used in electricity demand analysis and is a high-precision approach for seasonal data forecasting, errors are unavoidable in the forecasting process. Consequently, a significant research goal is to further improve forecasting precision. In this study, the PSO optimal Fourier method, seasonal ARIMA model, and combined models of PSO optimal Fourier method with seasonal ARIMA are applied in the Northwest electricity grid of China to correct the forecasting results of seasonal ARIMA. The results indicate that the prediction accuracy of the three residual modification models is higher than the single seasonal ARIMA model and that the combined model is the most satisfactory of the three models.[1]

**Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia Mohanad S. Al-Musaylha,b,⁎, Ravinesh C. Deoa,d,⁎, Jan F. Adamowskic, Yan Lia**

Data-driven techniques for short-term G-data forecasting are used for 0.5 h, 1.0 h, and 24 h forecasting horizons in this analysis. The research focuses on the second largest state of Australia, Queensland, with a growing demand for electricity for end-users. The partial autocorrelation function was applied during the training period to evaluate MARS and SVR model inputs for the historical G data, to distinguish significant inputs.

About measurable data G, the accuracy of G forecasting is evaluated using statistical metrics such as Pearson Product Moment, Root Square Error, and Mean Absolute Error correction coefficients. The SVR model, on the other hand, is superior to the MARS and ARIMA for the regular horizon which shows a wider WI and MAE. Consequently, the MARS and SVR models in Queensland, Australia can be found to be more appropriate than the ARIMA model for short-term G forecasts. They are also useful scientific instruments for further exploring the data forecasting of demand for energy in real-time.[2]

**Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods**
**Erick Meira de Oliveira a, b, * , Fernando Luiz Cyrino Oliveira a**

In the last decades, the world's energy consumption has increased rapidly due to fundamental changes in the industry and economy. This paper expands the fields of application of combined Bootstrap aggregating (Bagging) and forecasting methods to the electric energy sector, a novelty in literature, to obtain more accurate demand forecasts. The results show that the proposed methodologies substantially improve the forecast accuracy of the energy demand endures services in both developed and developing countries.

The electricity demand is forecasted 24 months in advance in various countries. The prospects of bagging technology are discussed to improve forecasting. The bagging process is proposed to be changed again. In certain cases, predictions were consistent with the methods suggested.[3]

**Review analysis of COVID-19 impact on electricity demand for residential buildings Author links open overlay panelMoncefKrarti[ab]MohammadAldubyan[b]**

Weather adjusted time series data of electricity demand before and after COVID-19 lockdowns are used to determine the magnitude of changes in electricity demand and residential energy use patterns. The analysis results indicate that while overall electricity demand is lower because of lockdowns that impact commercial buildings and manufacturing sectors, the energy consumption for the housing sector has increased by as much as 30% during the full 2020 lockdown period. Analysis of reported end-user data indicates that most of the increase in household energy demand is due to higher occupancy patterns during daytime hours, resulting in increased use of energy-intensive systems such as heating, air conditioning, lighting, and appliances. Several energy efficiencies and renewable energy solutions are presented to cost-effectively mitigate the increase in energy demands due to extended stay home living patterns.

The impacts of lockdowns and stay home orders due to COVID-19 on global electricity demand are discussed. Data for specific US regions and residential building stocks are obtained and analyzed. Specific sets of recommendations to improve the indoor air quality and energy efficiency for existing homes are outlined.[4]

**The application of seasonal latent variable in forecasting electricity demand as an alternative method**
**Kutluk KaganSumer, Ozlem Goktas, AycanHepsag**

In this study, we used ARIMA, seasonal ARIMA (SARIMA), and alternatively the regression model with a seasonal latent variable in forecasting electricity demand by using data that belongs to ‘‘Kayseri and Vicinity Electricity Joint-Stock Company’’ over the 1997:1–2005:12 periods. This study tries to examine the advantages of forecasting with ARIMA, SARIMA methods and with the model has seasonal latent variable to each other. The results support that ARIMA and SARIMA models are unsuccessful in forecasting electricity demand.

Seasonal ARIMA (SARIMA) is an extension of autoregressive integrated moving average (ARIMA), where seasonality in the data is accommodated using seasonal difference (Goh and Law, 2002). Also in applied studies, it is possible to model the seasonal and non-seasonal processes together through seasonal Box-Jenkins (SARIMA) models.[5]

## 2. HYPOTHESIS

This analysis will provide Victorians an insight into how much electricity is consumed in comparison with previous years to the current year.
The data set contains 14 variables where 13 are independent variables and 1 is the dependent variable.

The dependent variables are Date, RRP, demand_pos_RRP, RRP_positive, demand_neg_RRP, RRP_negative, frac_at_neg_RRP, min_temperature, max_temperature, solar exposure, rainfall,school_day, holiday.
The dependent variable is Demand.

**H0 (Null hypothesis) :**
The model does not show a lack of fit (or in simple terms—the model is just fine)
The difference between the demand and rrp (price) will decrease or remain constant.
**H1(Alternate hypothesis):**
The model does show a lack of fit.
The difference between the demand and rrp (price) will increase.

## 3. DATASET

Source of the Dataset-
https://www.kaggle.com/aramacus/electricity-demand-in-victoria-australia

The dataset contains various parameters like demand, rainfall, solor_exposure, RRP, etc

We have chosen Date with Demand and RRP to plot the time series to do the analysis.

The dataset is from 2015-2020, so we will first convert the daily data to monthly and then segregate the data into different parts.

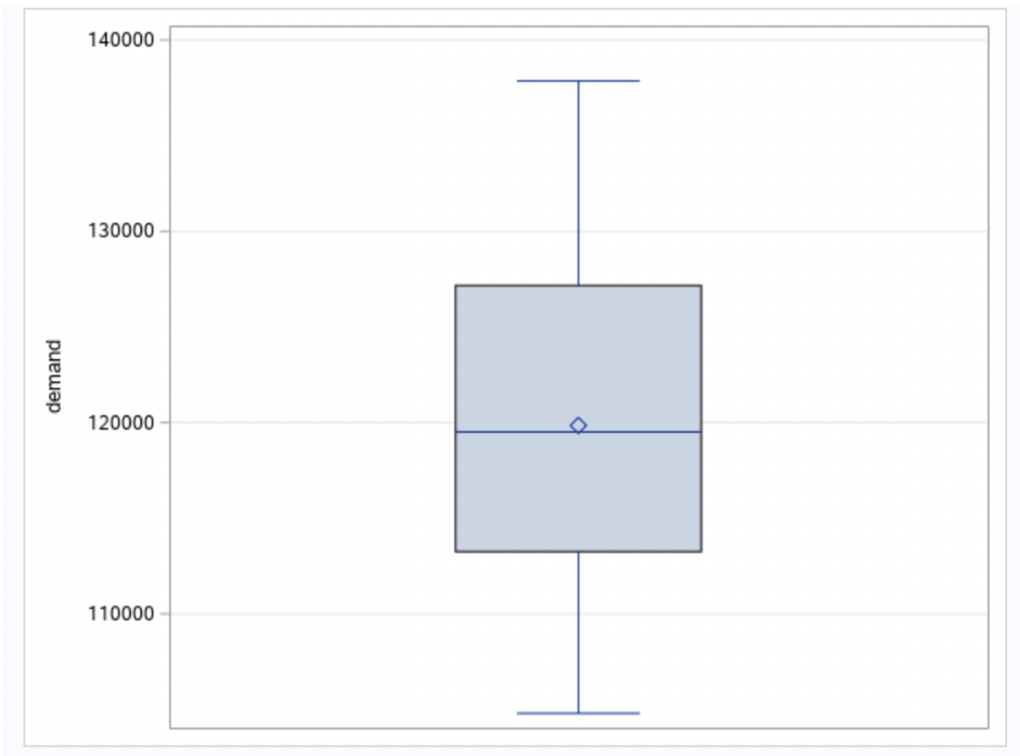
PART 1: Monthly data from 2015-2019

PART 2: 2020 Data

There are no missing values in the dataset.

## OUTLIER DETECTION:

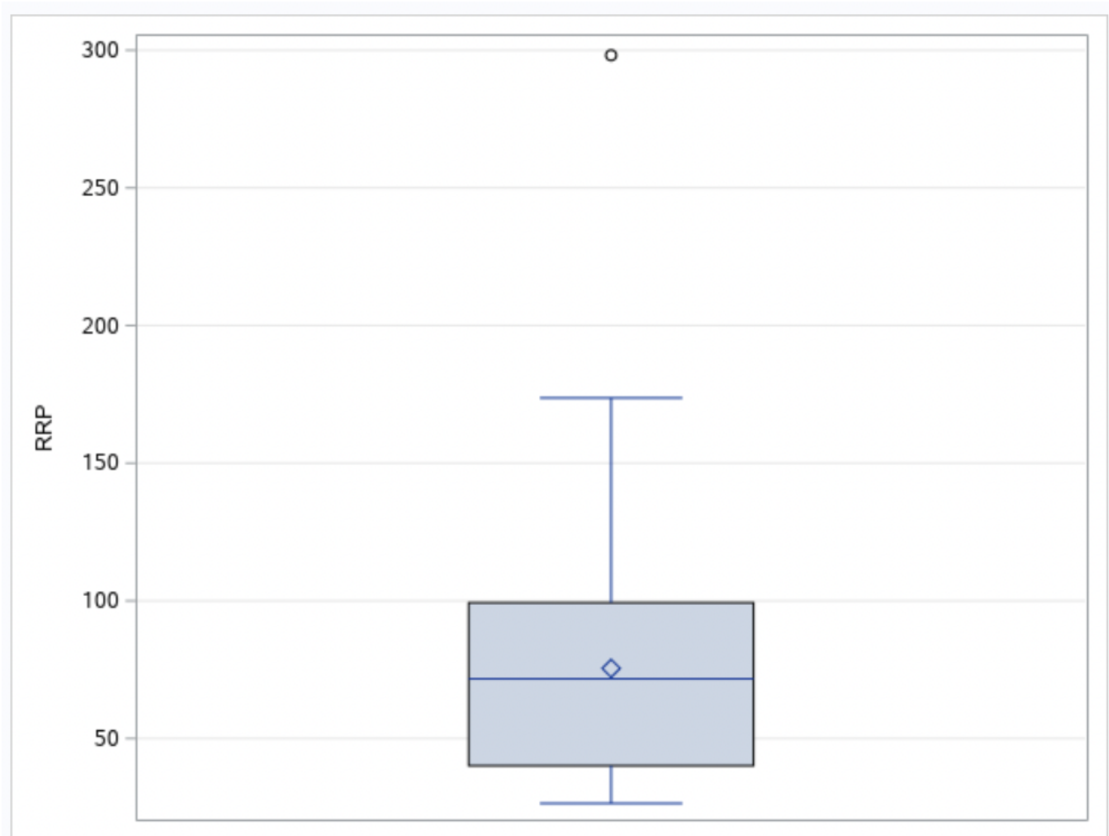
### 1. DEMAND




There are no outliers in the demand variable.

### 2. RRP:

There is one outlier in the RRP variable.

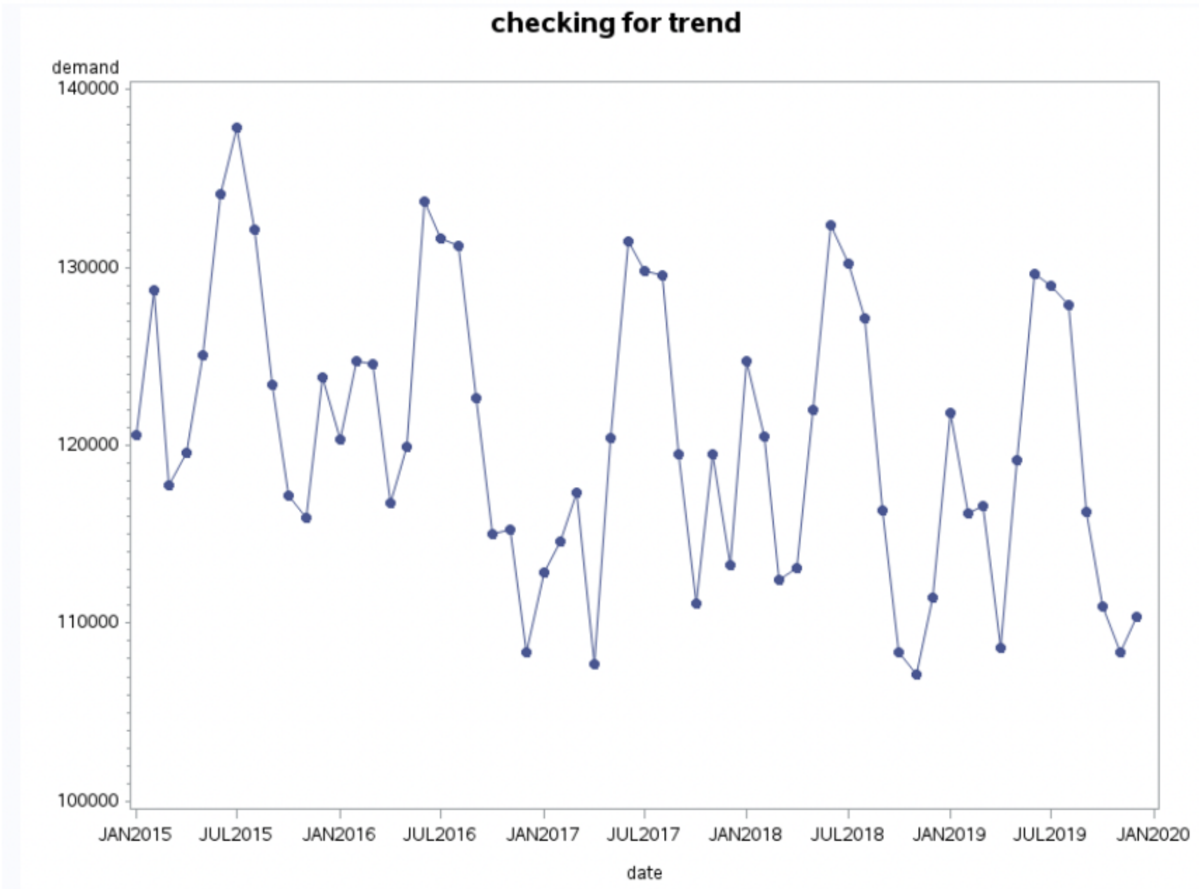


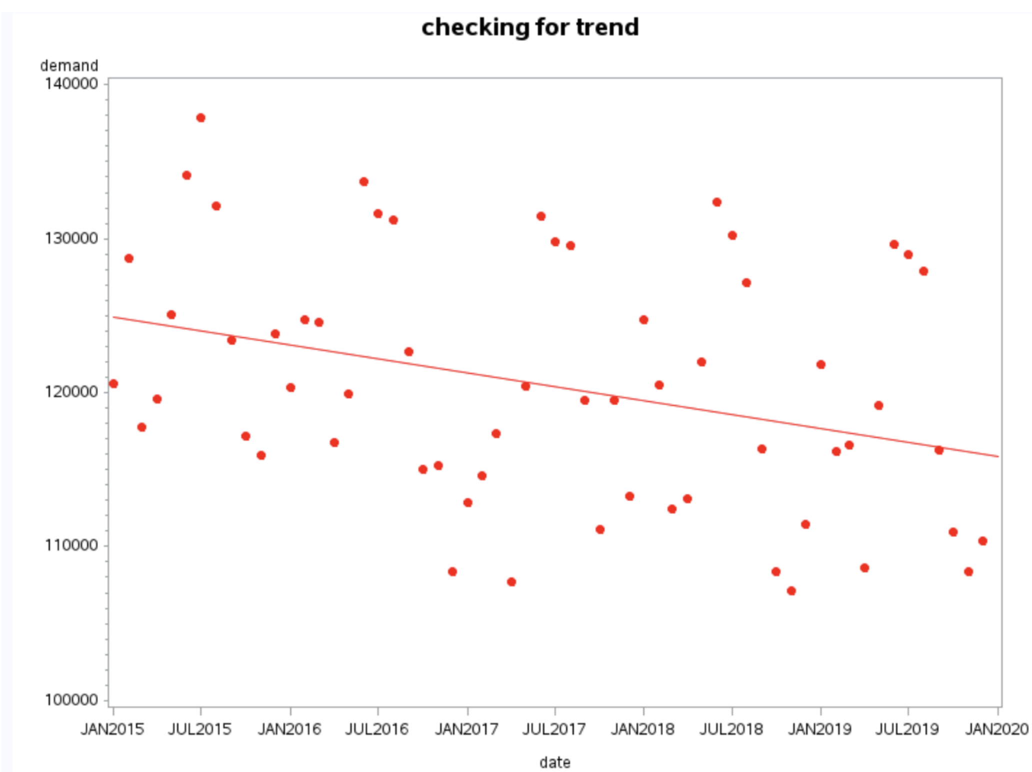

## 4. METHODOLOGY

### TREND OF THE TIME SERIES:

### Date with Demand

There is a decrease in the trend this means that the demand has reduced over time. As you can see in the plot, the peaks are in the months of June-July which makes sense as the weather is hot thus due to Air conditioning the electricity demand is high.

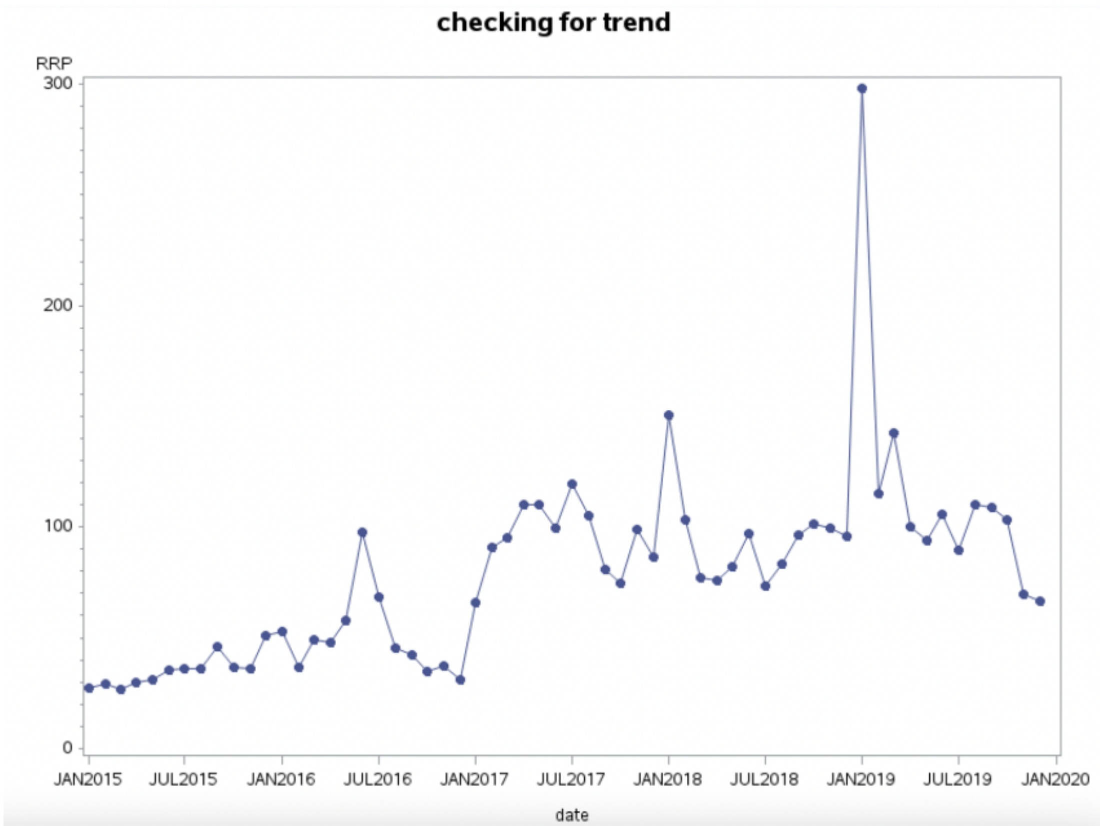

checking for trend
demand
140000
130000
120000
110000
100000
JAN2015
JUL2015
JAN2016
JUL2016
JAN2017
JUL2017
JAN2018
JUL2018
JAN2019
JUL2019
JAN2020
date

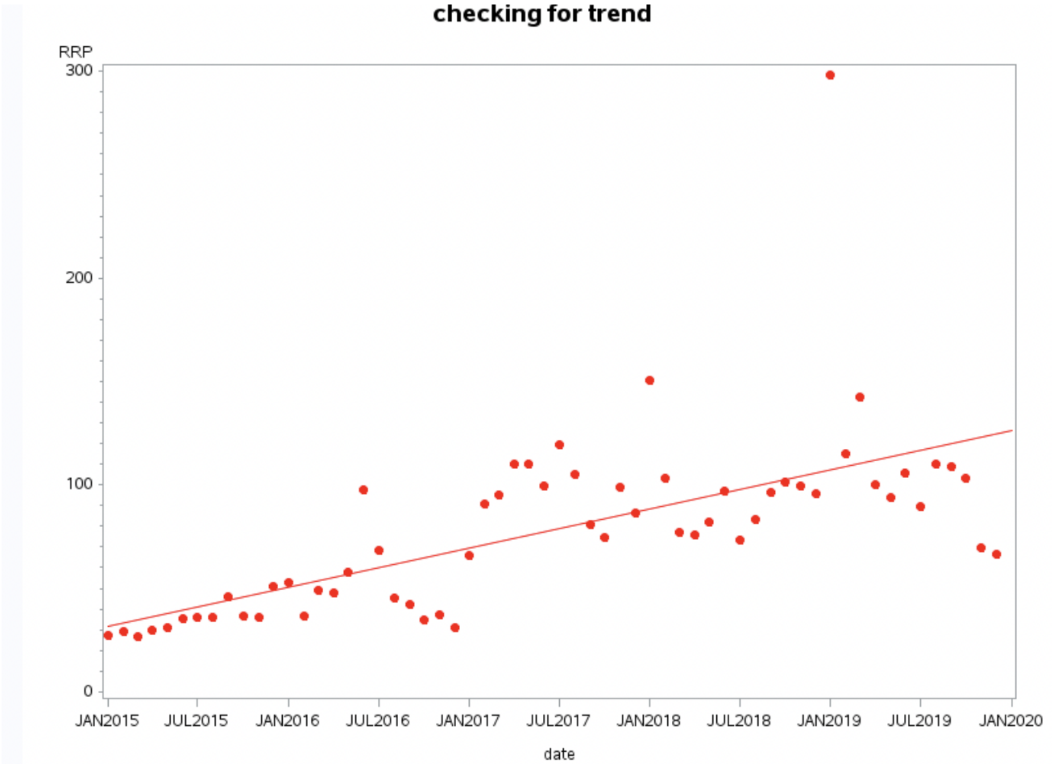

checking for trend
demand
140000
130000
120000
110000
100000
JAN2015
JUL2015
JAN2016
JUL2016
JAN2017
JUL2017
JAN2018
JUL2018
JAN2019
JUL2019
JAN2020
date

**Date with RRP (Recommended Retail Price)**

There is an increase in the trend this means that the retail price has increased over time. As you can see in the graph, the peaks are very random in order, this means there is no fixed period of increasing retail price. Jan 2019 seems to be an outlier in the trend series.

In this plot, there has been an increase in the retail price of electricity during the year 2019-2020.






## AUTOCORRELATION TEST:

The Durbin-Watson test is a widely used method of testing for autocorrelation. The first-order Durbin-Watson statistic is printed by default.

The Durbin-Watson Test is a test for serial correlation in the errors $\varepsilon 1$, $\varepsilon 2$,$\varepsilon n$.

This is a type of dependence or relationship among the errors $\varepsilon t$, which is a violation of the assumption of independence.

Positive serial correlation occurs when neighboring errors (in time) $\varepsilon t$ and $\varepsilon t+1$ tend to be similar. With positive serial correlation, positive errors tend to be followed by positive errors, and negative errors followed by negative errors.

Negative serial correlation occurs when neighboring errors (in time) tend to be dissimilar. With negative serial correlation, positive errors tend to be followed by negative errors, and negative errors are followed by positive errors.

**Positive serial correlation leads to DW < 2.**

**Negative serial correlation leads to DW >2**

## WHITE NOISE TEST:

White noise is an approximate statistical test of the hypothesis that none of the autocorrelations of the series up to a given lag are significantly different from 0. If this is true for all lags, then there is no information in the series to model, and no ARIMA model is needed for the series.

The Correlation for demand with 1st Lag indicates the presence of trend and that with 12th lag indicates an annual seasonality. Hence, we can consider differencing at first and Twelfths orders.

The Correlation for RRP with 1$^{st}$ lag indicates the presence of a trend, Hence we will consider differencing in the first order.

Since the variables are not stationary yet, thus the p-values are less than alpha.

| Durbin-Watson Statistics | | | |
|---|---|---|---|
| Order | DW | Pr < DW | Pr > DW |
| 1 | 0.8762 | <.0001 | 1.0000 |

| Durbin-Watson Statistics | | | |
|---|---|---|---|
| Order | DW | Pr < DW | Pr > DW |
| 1 | 1.4935 | 0.0161 | 0.9839 |

| Autocorrelation Check for White Noise | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 32.82 | 6 | <.0001 | 0.596 | 0.135 | -0.218 | -0.290 | -0.089 | 0.007 |
| 12 | 81.81 | 12 | <.0001 | -0.037 | -0.189 | -0.223 | 0.025 | 0.400 | 0.628 |

Demand Autocorrelation

| Autocorrelation Check for White Noise | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 67.02 | 6 | <.0001 | $\boxed{0.547}$ | 0.528 | 0.403 | 0.343 | 0.329 | 0.255 |
| 12 | 93.17 | 12 | <.0001 | 0.310 | 0.262 | 0.249 | 0.172 | 0.180 | 0.262 |

RRP autocorrelation

## STATIONARITY TEST:

**Null Hypothesis**: Non-Stationary

**Alternative Hypothesis**: Stationary

There are three types by which you can calculate test statistics of dickey-fuller test-

- **Zero Mean** - No Intercept. Series is a random walk without drift.
- **Single Mean** - Includes Intercept. Series is a random walk with drift.
- **Trend** - Includes Intercept and Trend. Series is a random walk with a linear trend.

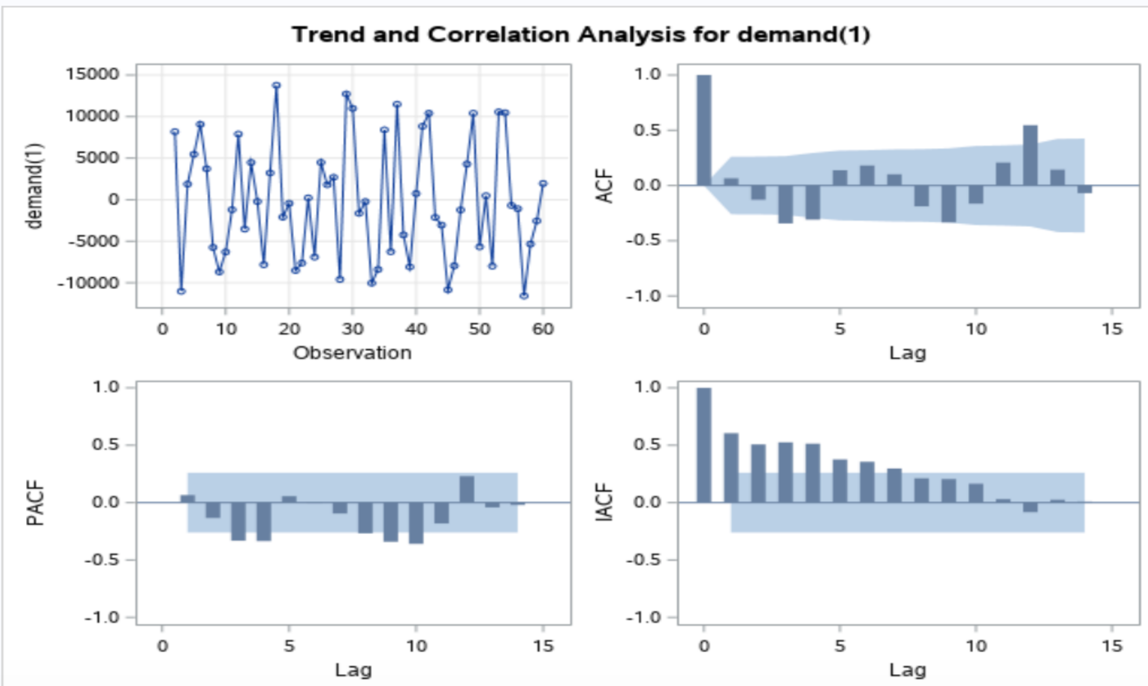
We can check the stationarity test by :

```
title "to check stationarity using unit root test";
proc arima data = work.filter1;
identify var = demand stationarity = (adf) ;
*can see the demand is not stationary so we have to do the differencing;
identify var = demand(1) stationarity = (adf);
*IACF is dying down very slowly so its overdifferenced;
identify var = demand(1,12) stationarity = (adf) ;
 identify var = demand(12) stationarity = (adf) ;
 *trend is not significant, hence we have to add 1 to differencing;
identify var = demand (6,12) stationarity = (adf) ;
*NO SIGNIFCANT P-VALUES;
run;
```

**demand(1):**

The p-values in the output are significant but the IACF is dying slowly.

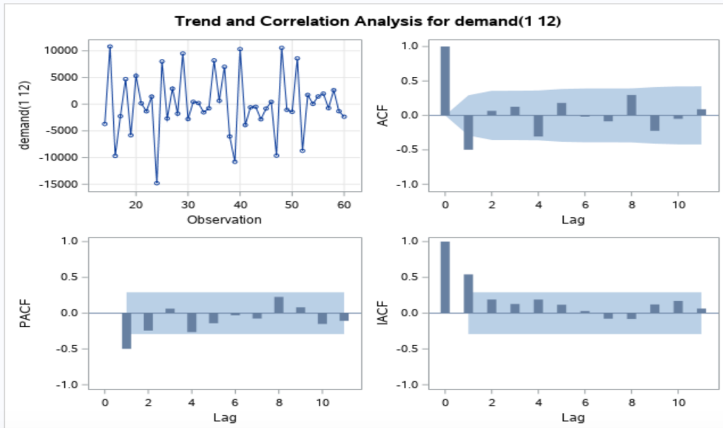
| Augmented Dickey-Fuller Unit Root Tests | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type | Lags | Rho | Pr < Rho | Tau | Pr < Tau | F | Pr > F |
| Zero Mean | 0 | -54.1349 | <.0001 | -7.14 | <.0001 | | |
| | 1 | -67.8212 | <.0001 | -5.81 | <.0001 | | |
| | 2 | -354.761 | 0.0001 | -6.39 | <.0001 | | |
| Single Mean | 0 | -54.2087 | 0.0005 | -7.09 | 0.0001 | 25.15 | 0.0010 |
| | 1 | -67.9205 | 0.0005 | -5.75 | 0.0001 | 16.57 | 0.0010 |
| | 2 | -358.535 | 0.0001 | -6.34 | 0.0001 | 20.09 | 0.0010 |
| Trend | 0 | -54.2222 | 0.0001 | -7.02 | <.0001 | 24.69 | 0.0010 |
| | 1 | -68.3471 | 0.0001 | -5.73 | 0.0001 | 16.41 | 0.0010 |
| | 2 | -365.466 | 0.0001 | -6.30 | <.0001 | 19.85 | 0.0010 |




## demand(1,12):

As per the differencing for demand(1,12). We found the p-values are significant

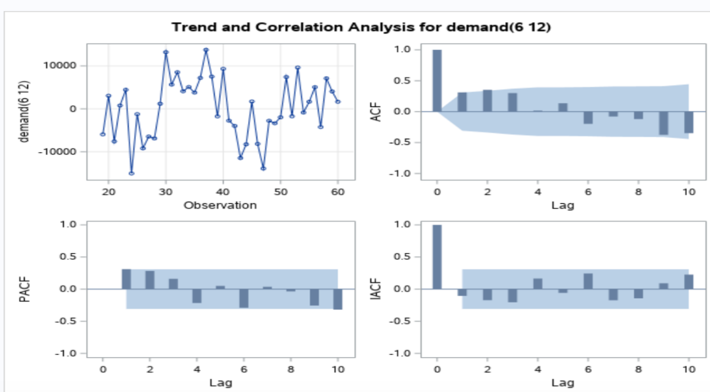
| Augmented Dickey-Fuller Unit Root Tests | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type | Lags | Rho | Pr < Rho | Tau | Pr < Tau | F | Pr > F |
| Zero Mean | 0 | -69.0079 | <.0001 | -11.66 | <.0001 | | |
| | 1 | -107.473 | 0.0001 | -7.39 | <.0001 | | |
| | 2 | -75.0332 | <.0001 | -4.37 | <.0001 | | |
| Single Mean | 0 | -69.0118 | 0.0004 | -11.54 | 0.0001 | 66.54 | 0.0010 |
| | 1 | -107.351 | 0.0001 | -7.30 | 0.0001 | 26.68 | 0.0010 |
| | 2 | -75.0324 | 0.0003 | -4.31 | 0.0013 | 9.30 | 0.0010 |
| Trend | 0 | -69.0173 | <.0001 | -11.40 | <.0001 | 65.03 | 0.0010 |
| | 1 | -108.358 | 0.0001 | -7.26 | <.0001 | 26.39 | 0.0010 |
| | 2 | -76.6467 | <.0001 | -4.27 | 0.0078 | 9.15 | 0.0036 |




**demand(6,12):**

As per the differencing (6,12) the values are higher than the p-value. Hence it is not significant.

| Augmented Dickey-Fuller Unit Root Tests | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type | Lags | Rho | Pr < Rho | Tau | Pr < Tau | F | Pr > F |
| Zero Mean | 0 | -28.1889 | <.0001 | -4.62 | <.0001 | | |
| | 1 | -14.9563 | 0.0046 | -2.66 | 0.0091 | | |
| | 2 | -10.0522 | 0.0226 | -2.03 | 0.0417 | | |
| Single Mean | 0 | -28.2270 | 0.0003 | -4.57 | 0.0007 | 10.46 | 0.0010 |
| | 1 | -14.9869 | 0.0258 | -2.63 | 0.0960 | 3.45 | 0.2192 |
| | 2 | -10.0731 | 0.1080 | -2.01 | 0.2800 | 2.08 | 0.5524 |
| Trend | 0 | -28.4465 | 0.0027 | -4.52 | 0.0043 | 10.25 | 0.0010 |
| | 1 | -15.2774 | 0.1229 | -2.62 | 0.2731 | 3.45 | 0.5084 |
| | 2 | -10.1269 | 0.3731 | -1.99 | 0.5907 | 1.98 | 0.7856 |

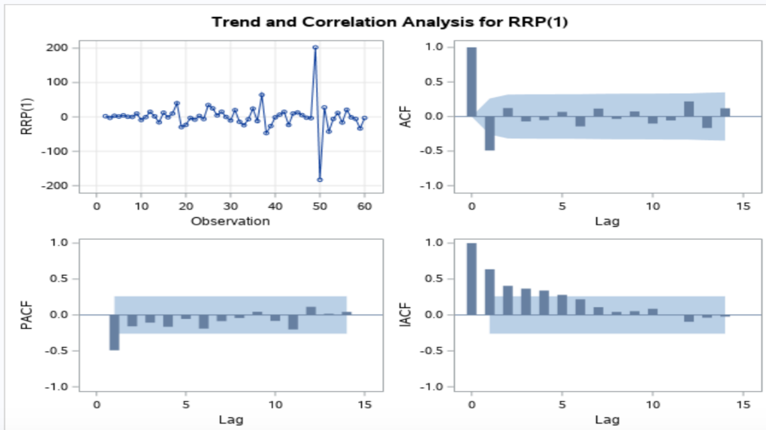



**RRP (1):**

As we see the output of differencing 1 the p-value is stationary, but the IACF dies down very slowly.

Augmented Dickey-Fuller Unit Root Tests

| Type | Lags | Rho | Pr < Rho | Tau | Pr < Tau | F | Pr > F |
|---|---|---|---|---|---|---|---|
| Zero Mean | 0 | -86.4919 | <.0001 | -12.93 | <.0001 | | |
| | 1 | -117.398 | 0.0001 | -7.47 | <.0001 | | |
| | 2 | -194.517 | 0.0001 | -5.71 | <.0001 | | |
| Single Mean | 0 | -86.5182 | 0.0005 | -12.82 | 0.0001 | 82.16 | 0.0010 |
| | 1 | -117.872 | 0.0001 | -7.41 | 0.0001 | 27.49 | 0.0010 |
| | 2 | -199.054 | 0.0001 | -5.67 | 0.0001 | 16.10 | 0.0010 |
| Trend | 0 | -86.6992 | 0.0001 | -12.76 | <.0001 | 81.37 | 0.0010 |
| | 1 | -119.240 | 0.0001 | -7.41 | <.0001 | 27.44 | 0.0010 |
| | 2 | -207.230 | 0.0001 | -5.69 | 0.0001 | 16.20 | 0.0010 |




When a time series has a unit root, the series is nonstationary and the ordinary least squares (OLS) estimator is not normally distributed.

As we see the output after differencing for Demand and RRP, We also check the pr<tau values and all the values are less than alpha thus the differencing is stationary.

## SCAN ESCAF MINIC

The smallest canonical **(SCAN)** correlation method can tentatively identify the orders of a *stationary or nonstationary* ARMA process. Given a stationary or nonstationary time series {zt:1≤t≤n} with mean corrected form z~t=zt−μz with a true autoregressive order of p+d and with a true moving-average order of, you can use the SCAN method to analyze eigenvalues of the correlation matrix of the ARMA process.
The extended sample autocorrelation function **(ESACF)** method can tentatively identify the orders of a stationary or nonstationary

ARMA process based on iterated least squares estimates of the autoregressive parameter
The minimum information criterion **(MINIC)** method can tentatively identify the order of a stationary and invertible ARMA process. with a true autoregressive order of p and with a true moving-average order of q, you can use the MINIC method to compute information criteria (or penalty functions) for various autoregressive and moving average orders.

**Minimum Table Value: BIC(1,0) = 16.89245**

| ARMA(p+d,q) Tentative Order Selection Tests | | | | | |
|---|---|---|---|---|---|
| SCAN | | | ESACF | | |
| p+d | q | BIC | p+d | q | BIC |
| 3 | 0 | 17.00705 | 2 | 2 | 17.08912 |
| 2 | 3 | 17.15496 | 0 | 3 | 17.0423 |
| 0 | 4 | 17.10127 | 4 | 3 | 17.24738 |

**(5% Significance Level)**

**SCAN:**

As you can see in the table, the scan method suggests us that we should use p+d=3 q=0. Here p means AR model, q means MA model and d is differencing.

So, we can use differencing (2) + p(1) for estimating the model.

Similarly we can check for other options as well.

**MIMIC:**

We check the lowest value in the table,here our lowest value comes out to be 16.8 which suggests that BIC(1,0). It means that we can use p=1 maximum order and q=0.

**The PROC ARIMA**

•ARIMA stands for the autoregressive integrated moving average. It is also known as the Box-Jenkins model, as the ARIMA has been a technique popularized by Box and Jenkins. For ARIMA forecasting, data needs to be stationary.

•The ARIMA procedure analyzes and forecasts equally spaced univariate time series data, transfer function data, and intervention data by using autoregressive integrated moving averages.

**Identification Stage**

•The Identification Stage computes auto-correlation, inverse autocorrelations, partial autocorrelations, and cross-correlations. Stationarity tests can be performed to identify whether differencing is necessary. It also provides descriptive statistics.

•nlag controls the number of lags for which autocorrelation is shown. It should be always less than the number of observations in your dataset.

•var is used to specify the name of the variable that needs to forecast.

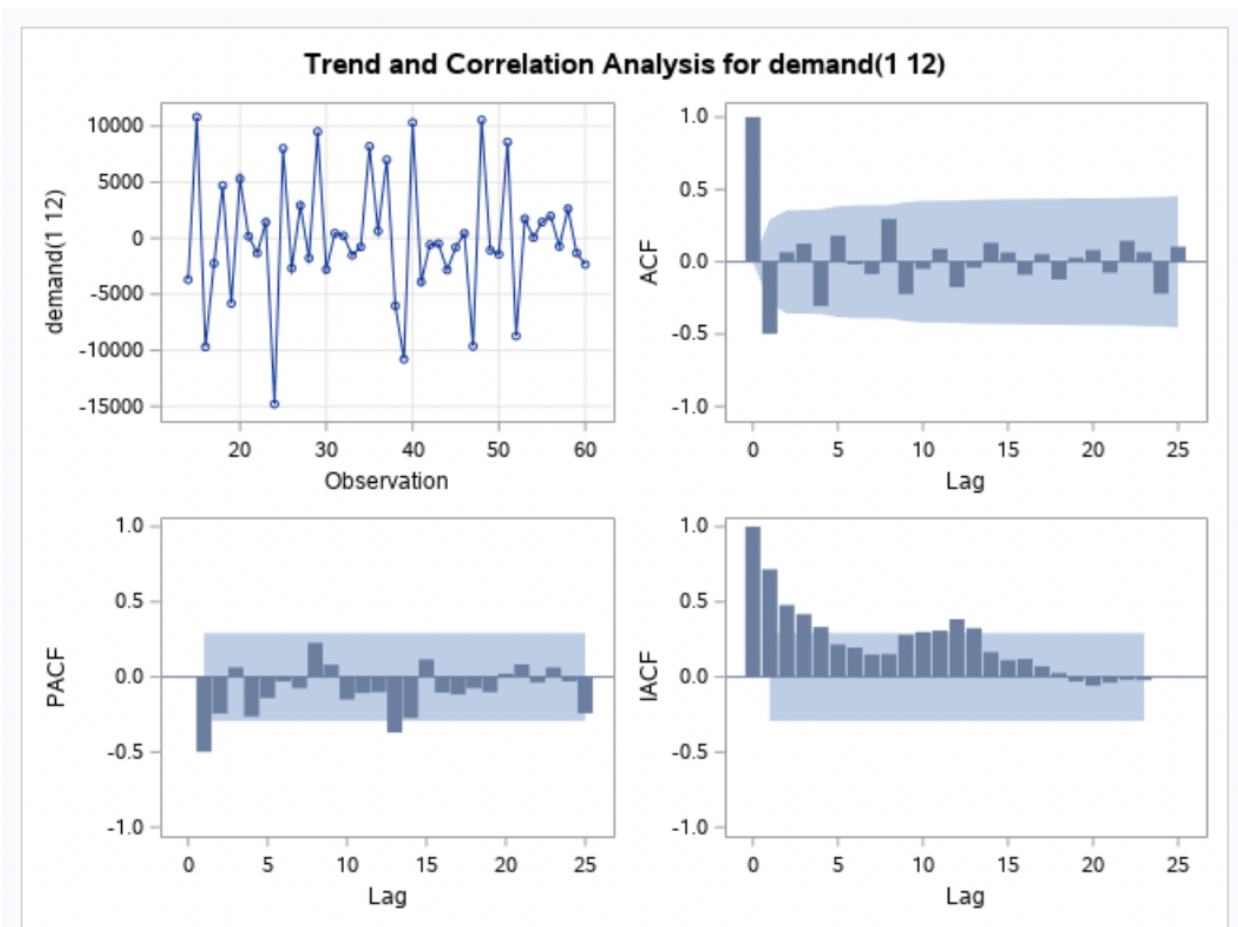
•The identify statement produces panels of plots for auto-correlation and trend analysis.

- Time series plot of the series.
- Auto-correlation function plot (ACF).
- Inverse autocorrelation function plot (IACF).
- Partial autocorrelation function plot (PACF).

```
proc arima data = work.filter1;
identify var = demand(1,12) nlag=25 ;
```

$$z_t = (1 - \theta_6 B^6)(1 - \theta_{1,12} B^{12}) a_t$$

**Equation for differencing used in the model : demand (1,12)**



When we identify the demand variable in the time series:

**ACF PLOT (SAC)-** Spike at lag 1

**PACF PLOT (SPAC)-** Spike at lag 1,13

## Estimation and Diagnostic Stage

•The estimate statement is used to specify the ARIMA model to fit the variable specified in the previous identity statement and to estimate the parameters of that model.

•The estimate statement also produces diagnostic statistics to help you judge the adequacy of the model.

## Ljung-Box test

• Ljung-Box test is an important diagnostic to check if residuals from the time series model are independently distributed. In SAS, it is easy to perform Ljung-Box with the ARIMA procedure.

When modeling an autoregressive-moving average time series we typically use the Ljung-Box statistic on the residuals to see if our fitted model is adequate. However, much like checking the adequacy of the regression through an F-test, checking the adequacy of the fitted time-series model is of the utmost importance.

## Ljung Box Test Hypotheses

H0: null hypothesis :

The model does not show lack of fit (or in simple terms—the model is just fine).

Ha: The alternate hypothesis :

The model does show a lack of fit. A significant p-value in this test rejects the null hypothesis that the time series isn’t autocorrelated.

**Estimating different models-**

**Model 1-**

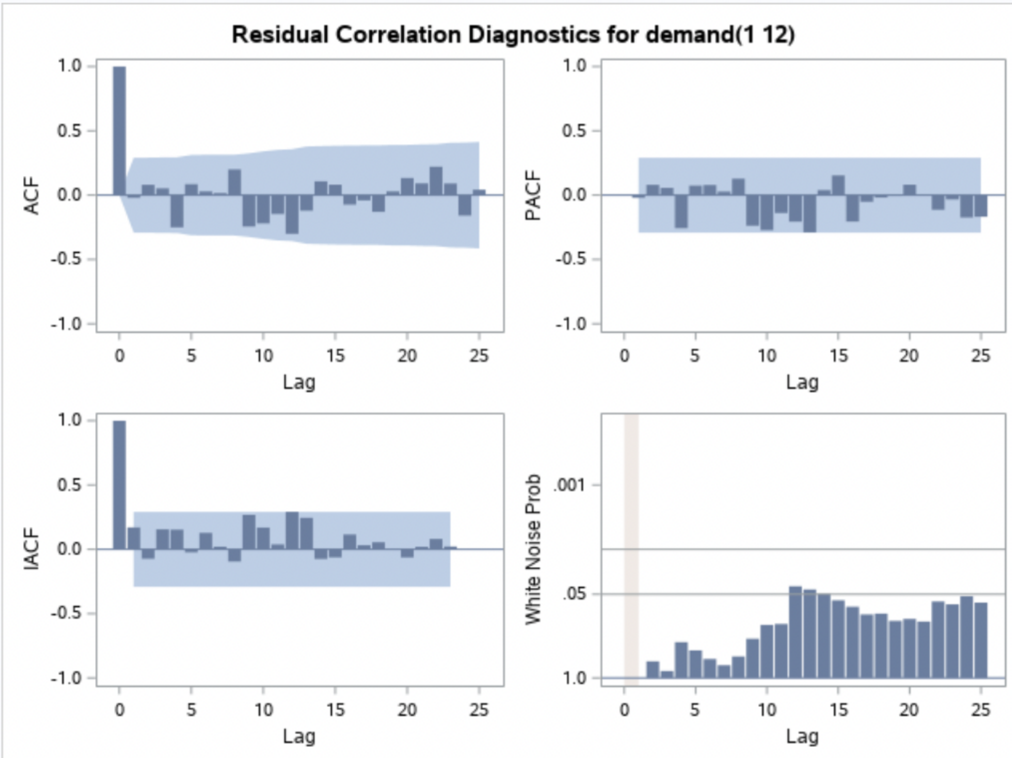
```
title "model 1";
proc arima data = work.filter1;
identify var = demand(1,12) nlag=25 ;
estimate q= (1) noint printall plot;
forecast lead=5 ;
run;
```

| Conditional Least Squares Estimation | | | | | |
|---|---|---|---|---|---|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > \|t\| | Lag |
| MA1,1 | 0.63381 | 0.11390 | 5.56 | <.0001 | 1 |

| | |
|---|---|
| Variance Estimate | 22888438 |
| Std Error Estimate | 4784.186 |
| AIC | 930.8381 |
| SBC | 932.6883 |
| Number of Residuals | 47 |

The model 1 for lag=12, the p-value is less than alpha so it’s not significant. Thus, this lag is not adequate.

| Autocorrelation Check of Residuals | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 4.30 | 5 | 0.5076 | -0.020 | 0.080 | 0.053 | -0.250 | 0.085 | 0.029 |
| 12 | 20.61 | 11 | 0.0377 | 0.016 | 0.200 | -0.244 | -0.220 | -0.146 | -0.301 |
| 18 | 24.75 | 17 | 0.1005 | -0.121 | 0.108 | 0.082 | -0.073 | -0.040 | -0.130 |
| 24 | 34.90 | 23 | 0.0532 | 0.029 | 0.133 | 0.092 | 0.221 | 0.091 | -0.159 |



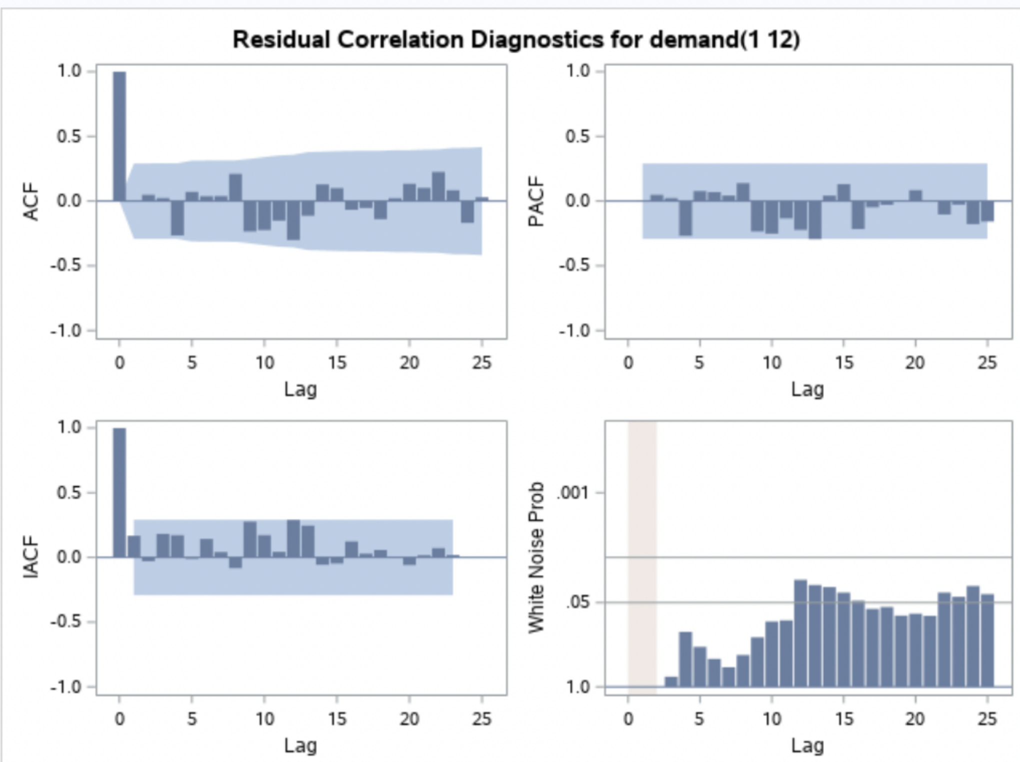

## Model 2-

```
title "model 2";
proc arima data = work.filter1;
identify var = demand(1,12) nlag=25 ;
estimate p=(1) q=(1) printall plot; *
forecast lead=5;
run;
```

| Conditional Least Squares Estimation | | | | | |
|---|---|---|---|---|---|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > \|t\| | Lag |
| MU | 17.45803 | 289.21174 | 0.06 | 0.9521 | 0 |
| MA1,1 | 0.57172 | 0.19816 | 2.89 | 0.0060 | 1 |
| AR1,1 | -0.08240 | 0.24141 | -0.34 | 0.7345 | 1 |

| | |
|---|---|
| Constant Estimate | 18.89653 |
| Variance Estimate | 23885330 |
| Std Error Estimate | 4887.262 |
| AIC | 934.7526 |
| SBC | 940.3031 |
| Number of Residuals | 47 |

| Autocorrelation Check of Residuals | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 4.27 | 4 | 0.3701 | 0.004 | 0.049 | 0.024 | -0.265 | 0.071 | 0.038 |
| 12 | 20.85 | 10 | 0.0222 | 0.039 | 0.211 | -0.235 | -0.224 | -0.152 | -0.301 |
| 18 | 25.71 | 16 | 0.0583 | -0.113 | 0.129 | 0.100 | -0.068 | -0.053 | -0.139 |
| 24 | 36.35 | 22 | 0.0278 | 0.023 | 0.133 | 0.102 | 0.225 | 0.084 | -0.168 |



The intercept and AR (1) are not significant in this model, also the p values are not significant for lags.

**Model 3-**

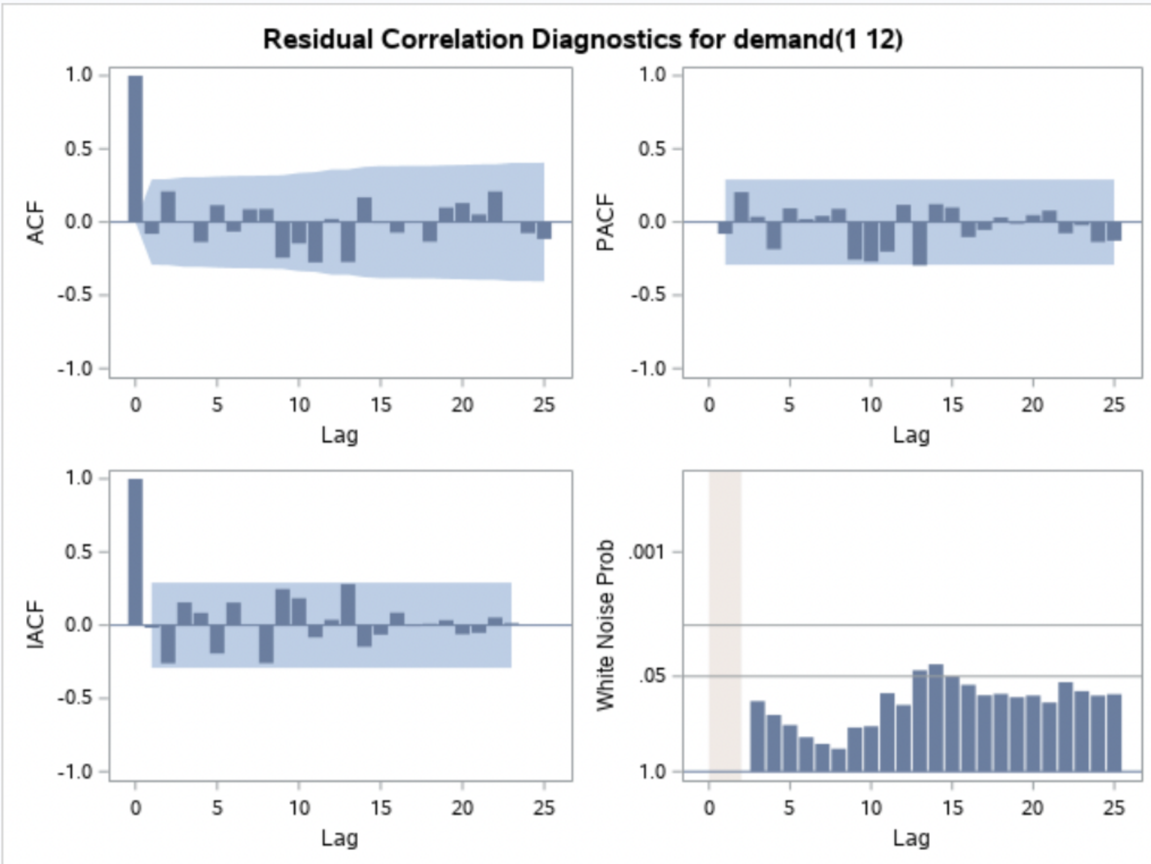
```
proc arima data = work.filter1;
identify var = demand(1,12) nlag=25 ;
estimate q= (1)(12) noint printall plot;
forecast lead=10 ;
run;
```

A factored model (also referred to as a multiplicative model) represents the ARIMA model as a product of simpler ARIMA models. For example, you might model SALES as a combination of an AR(1) process that reflects short-term dependencies and an AR(12) model that reflects the seasonal pattern.

| Conditional Least Squares Estimation | | | | | |
|---|---|---|---|---|---|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > \|t\| | Lag |
| MA1,1 | 0.80522 | 0.08927 | 9.02 | <.0001 | 1 |
| MA2,1 | 0.58789 | 0.14146 | 4.16 | 0.0001 | 12 |

| | |
|---|---|
| Variance Estimate | 18967851 |
| Std Error Estimate | 4355.21 |
| AIC | 922.9745 |
| SBC | 926.6748 |
| Number of Residuals | 47 |

| Autocorrelation Check of Residuals | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 4.57 | 4 | 0.3342 | -0.077 | 0.213 | 0.009 | -0.133 | 0.118 | -0.063 |
| 12 | 15.16 | 10 | 0.1264 | 0.090 | 0.092 | -0.240 | -0.141 | -0.275 | 0.022 |
| 18 | 24.13 | 16 | 0.0866 | -0.274 | 0.166 | -0.012 | -0.077 | -0.006 | -0.138 |
| 24 | 30.94 | 22 | 0.0974 | 0.095 | 0.127 | 0.050 | 0.207 | -0.007 | -0.070 |




Everything seems fine in model 3, so we will choose model 3 for forecasting.

**•Forecasting Stage**

•The FORECAST statement is used to forecast future values of the time series and to generate confidence intervals for these forecasts from the ARIMA model produced by the preceding ESTIMATE statement.

•lead specifies how many periods ahead to forecast. (12 months is our example).

•id specifies the ID variable (which is generally SAS date, time, and DateTime).

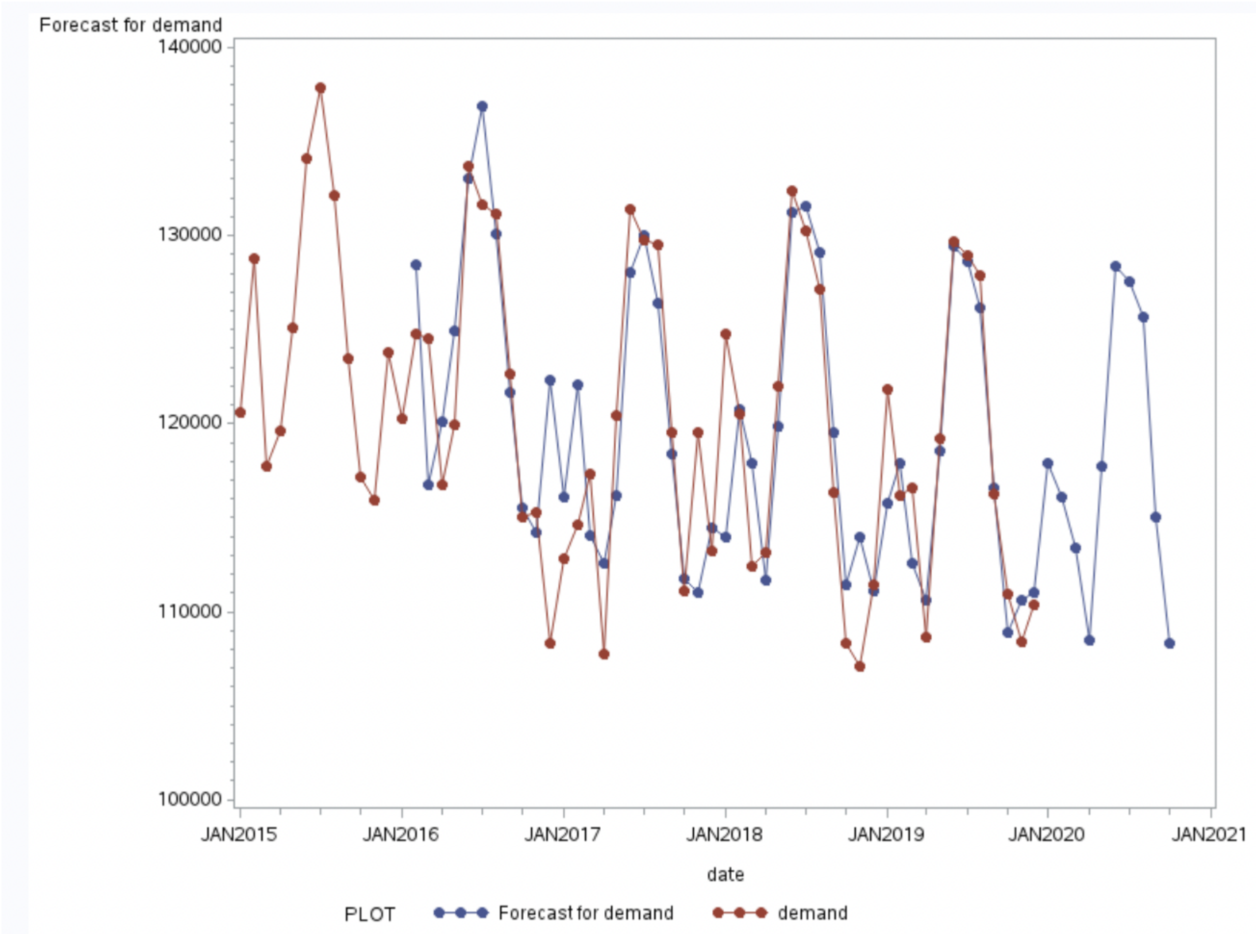
•interval indicates the data are monthly.

•out allows us to write the forecast data to the dataset's results.

| Forecasts for variable demand | | | | |
|---|---|---|---|---|
| Obs | Forecast | Std Error | 95% Confidence Limits | |
| 61 | 117871.3 | 4355.21 | 109335.3 | 126407.4 |
| 62 | 116125.8 | 4435.93 | 107431.5 | 124820.0 |
| 63 | 113405.4 | 4515.19 | 104555.8 | 122255.0 |
| 64 | 108497.2 | 4593.10 | 99494.9 | 117499.5 |
| 65 | 117718.6 | 4669.70 | 108566.2 | 126871.0 |
| 66 | 128348.8 | 4745.06 | 119048.7 | 137649.0 |
| 67 | 127583.9 | 4819.25 | 118138.4 | 137029.5 |
| 68 | 125635.8 | 4892.31 | 116047.0 | 135224.5 |
| 69 | 115053.2 | 4964.30 | 105323.4 | 124783.0 |
| 70 | 108337.9 | 5035.25 | 98469.0 | 118206.8 |

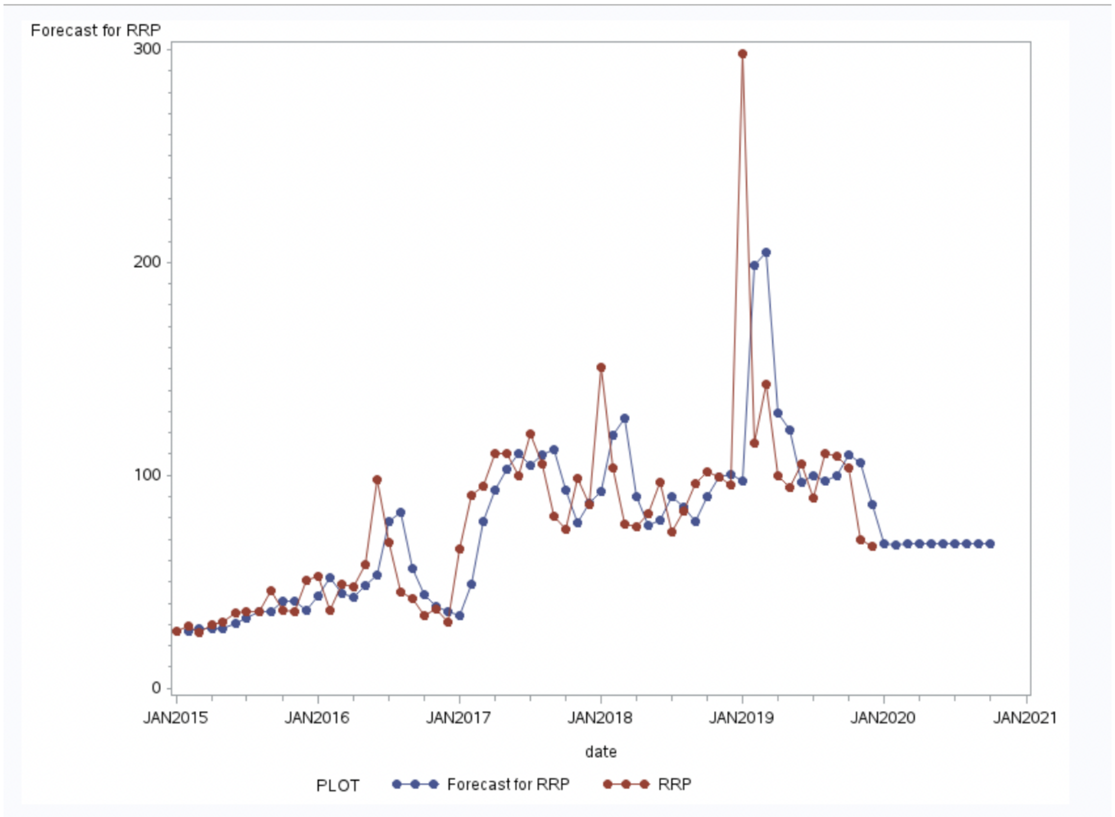
In the forecasting output below, the first 13 observations are not present because we are calculating the moving average of the first few months to forecast next month.

| Obs | date | demand | FORECAST |
|---|---|---|---|
| 1 | JAN2015 | 120576.86032 | . |
| 2 | FEB2015 | 128754.10768 | . |
| 3 | MAR2015 | 117730.70484 | . |
| 4 | APR2015 | 119606.6035 | . |
| 5 | MAY2015 | 125072.10194 | . |
| 6 | JUN2015 | 134137.91783 | . |
| 7 | JUL2015 | 137856.05839 | . |
| 8 | AUG2015 | 132113.75339 | . |
| 9 | SEP2015 | 123427.208 | . |
| 10 | OCT2015 | 117135.67516 | . |
| 11 | NOV2015 | 115927.91133 | . |
| 12 | DEC2015 | 123824.12887 | . |
| 13 | JAN2016 | 120292.74081 | . |
| 14 | FEB2016 | 124767.84 | 128469.99 |
| 15 | MAR2016 | 124549.92 | 116730.55 |
| 16 | APR2016 | 116731.0185 | 120118.80 |
| 17 | MAY2016 | 119950.54968 | 124929.06 |
| 18 | JUN2016 | 133713.50667 | 133031.98 |
| 19 | JUL2016 | 131618.32839 | 136881.94 |
| 20 | AUG2016 | 131178.28258 | 130121.59 |
| 21 | SEP2016 | 122656.68317 | 121639.42 |
| 22 | OCT2016 | 115029.60839 | 115544.64 |
| 23 | NOV2016 | 115241.02417 | 114237.26 |
| 24 | DEC2016 | 108336.59758 | 122327.62 |
| 25 | JAN2017 | 112818.7271 | 116090.22 |
| 26 | FEB2017 | 114614.89036 | 122111.82 |
| 27 | MAR2017 | 117311.88613 | 114083.35 |
| 28 | APR2017 | 107725.24933 | 112595.65 |
| 29 | MAY2017 | 120448.6479 | 116195.26 |
| 30 | JUN2017 | 131431.60467 | 128015.94 |
| 31 | JUL2017 | 129783.94484 | 130003.35 |
| 32 | AUG2017 | 129554.92629 | 126399.73 |
| 33 | SEP2017 | 119511.055 | 118391.29 |
| 34 | OCT2017 | 111114.91984 | 111766.94 |
| 35 | NOV2017 | 119519.94383 | 111016.86 |
| 36 | DEC2017 | 113257.60129 | 114469.31 |

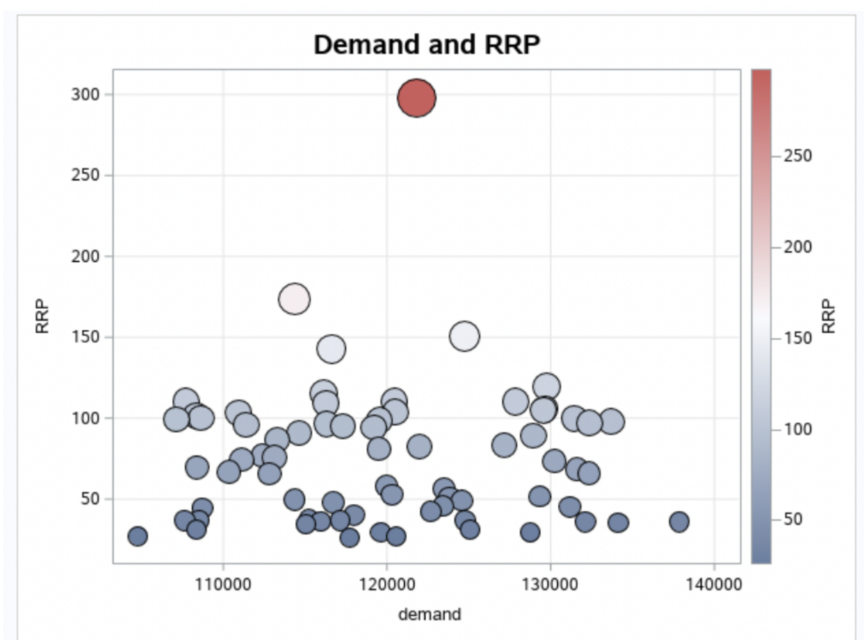



We have used similar methodology and found out the forecasting for RRP variables too.

```
*checking which model is better for rrp;
proc arima data = work.filter1 plots (only)=(series(corr crosscorr) residual(corr normal)
        forecast(forecastonly)) out=work.rrp ;
identify var = RRP(1) nlag=25 ;
estimate P= (1) ma=(1.0) method=cls NOINT  printall plot;
forecast lead=10 back=0 alpha=0.05 id=date interval=month;
run;
```




As the demand for electricity increases, the price of the electricity is also decreasing. Over years, the price of electricity has increased with decreasing demand for electricity.



## Conclusion:

- The forecasted values and the original values almost seem to overlap each other.

- There is been few differences in the forecasting values and the original values because Victoria witnessed slashing of electricity demand.

- The model used for estimating and forecasting time series is adequate.

- The p values of white noise autocorrelation is checked, in our model the values are significant which means we will reject the null hypothesis stating that our model doesn't fit appropriately and follow the alternative hypothesis which states that the model is adequate and fits appropriately.

## References:


1. https://www.sciencedirect.com/science/article/abs/pii/S0301421512004387
2. https://www.sciencedirect.com/science/article/abs/pii/S1474034617301477
3. https://www.sciencedirect.com/science/article/abs/pii/S0360544217320820
4. https://www.sciencedirect.com/science/article/pii/S1364032121001829 - :~:text=The analysis results indicate the full 2020 lockdown period.

5. https://www.sciencedirect.com/science/article/abs/pii/S0301421508007143
6. https://towardsdatascience.com/electricity-demand-in-victoria-during-coronavirus-pandemic-30c23ff4e0d.
7. https://support.sas.com/documentation/onlinedoc/ets/132/arima.pdf