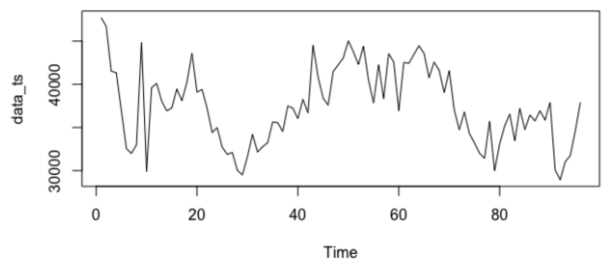
BANA7050 Forecasting and Time Series Methods (Case Study #1)

SAGAR SAHOO
MS BUSINESS ANALYTICS
M13433382

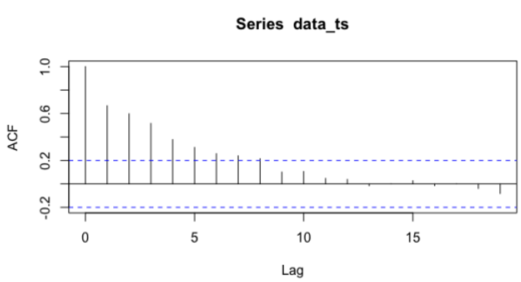
The data is monthly bituminous coal production in the US from January 1952 through December 1959, a total of 96 observations. The data is seasonally adjusted (i.e., no need to consider seasonality).

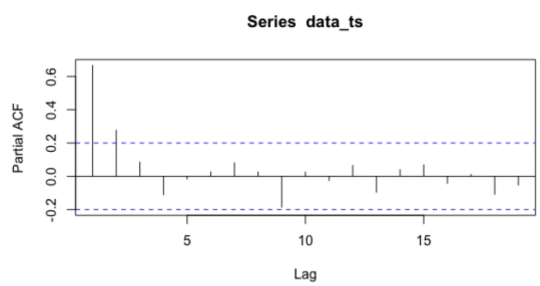
The Data was imported into R workspace and using plot method the time series was plotted.



We would now need to find some models based on the above realization. We can use ACF and PACF of the series to estimate some models.

Please note, we are assuming that the data has been seasonally adjusted. As well there seems no increasing or decreasing trend of the mean, hence the mean can be assumed to be stationary.





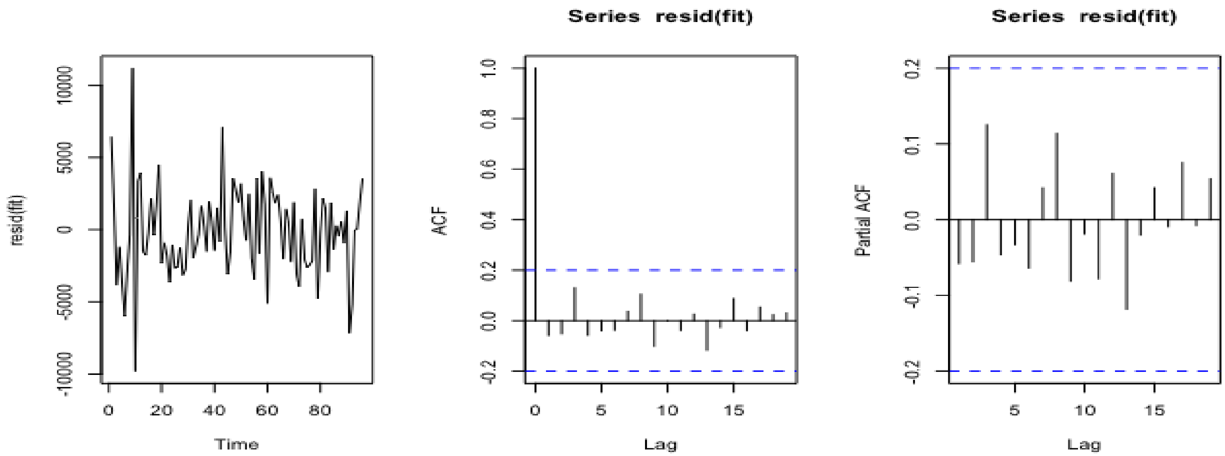
The decaying pattern of the ACF suggests that we should start with an AR Model (we have no evidence to start with MA Model as the ACFs do not cut off sharply at any lag). The PACF pattern is closely consistent with an AR(2) model.

So, we apply AR(2) to our time series:

```
> fit <- arima(data_ts,order=c(2,0,0))
> AIC(fit)
[1] 1822.801
> BIC(fit)
[1] 1833.059
> fit

Call:
arima(x = data_ts, order = c(2, 0, 0))

Coefficients:
         ar1     ar2  intercept
      0.4840  0.3224  37981.397
s.e.  0.0965  0.0988   1543.223

sigma^2 estimated as 9403405:  log likelihood = -907.4,  aic = 1822.8
```

```
> coeftest(fit)

z test of coefficients:

           Estimate Std. Error z value  Pr(>|z|)
ar1       4.8399e-01 9.6492e-02  5.0158  5.28e-07 ***
ar2       3.2236e-01 9.8805e-02  3.2626  0.001104 **
intercept 3.7981e+04 1.5432e+03 24.6117 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residuals' ACF reveals much about the lack of fit. If the model fits the data well, then residual will behave like white noises.

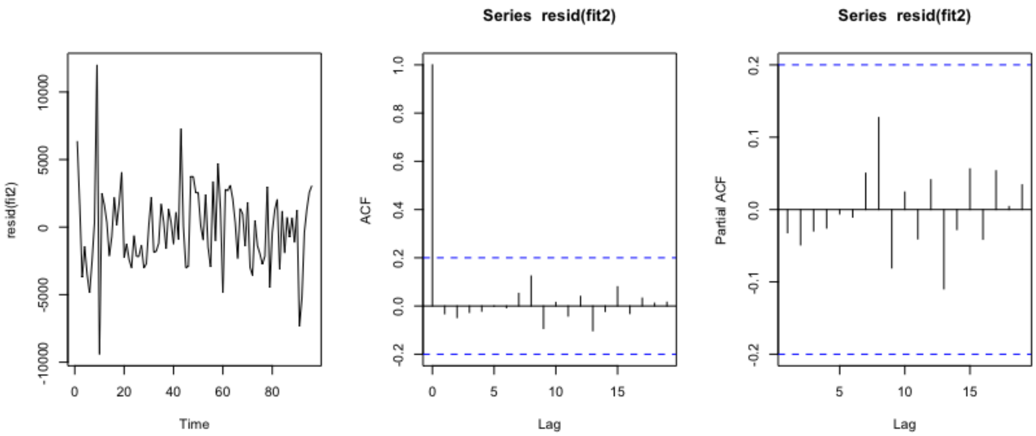


From the ACF plot we could see that, there is no spike at lags k > 0. It means that the residuals are approximately white noise.

However, we can still check for other models and see if they give us a better model than AR(2). From the residual plot, although there is no spike, still there is an exponential decay pattern of ACF. An MA model has such ACF trend. The lag looks from k = 3, so we try fitting MA(3) to the existing fitted AR(2) model i.e we can now apply ARMA(2,3).

```
> fit2 <- arima(data_ts,order=c(2,0,3))
> AIC(fit2)
[1] 1825.522
> BIC(fit2)
[1] 1843.473
> fit2

Call:
arima(x = data_ts, order = c(2, 0, 3))

Coefficients:
         ar1     ar2      ma1     ma2     ma3  intercept
      0.5574  0.1759  -0.0966  0.1225  0.2006  37963.241
s.e.  0.6997  0.6139   0.6875  0.3219  0.1682   1377.236

sigma^2 estimated as 9073092:  log likelihood = -905.76,  aic = 1825.52
```

We could see that the AIC and BIC of the above fitted model increased. AR(2) fits better as compared to ARMA(2,3). However the plot of the residuals of ARMA(2,3) looks a white noise:



The ARMA(2,3) is a promising model but since the AIC/BIC is more as compared to AR(2), we should ideally go with **AR(2) Model**.

The coefficients of the **AR(2)** Model are **0.4840** and **0.3224**. The coefficients satisfies the below stationarity conditions as well:

$\phi_1+\phi_2<1$

$\phi_2-\phi_1<1$

$-1<\phi_2<1$

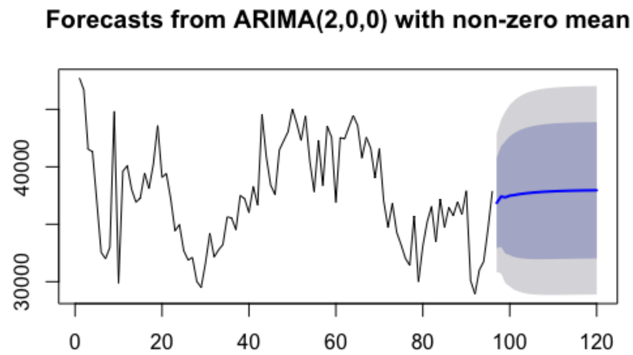
Portmanteau Lack of Fit Test: This test helps us detect lack of fit by analysing the residuals. Here, the Box-Ljung test gives a significant p-value of 0.9439, hence we cannot reject the Null Hypothesis that the model is fine.

```
> Box.test(resid(fit), lag = 10, type = "Ljung-Box", fitdf = 2)

        Box-Ljung test

data:  resid(fit)
X-squared = 2.8408, df = 8, p-value = 0.9439
```

The Final AR(2) Model is :

$$Z_t = 37981.397 + 0.4840\,\tilde{Z}_{t-1} + 0.3224\,\tilde{Z}_{t-2} + a_t$$

Since AR(2) is a good fit for the model, we will now forecast using AR(2) for the next 24 Months. We can achieve the same using forecast function available in R *"forecast(fit,h=24)"*



And below are the forecasted values:

```
> forecast(fit,h=24)
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
97	36856.64	32926.77	40786.52	30846.42	42866.87
98	37411.11	33045.16	41777.07	30733.97	44088.26
99	37342.81	32459.55	42226.07	29874.51	44811.11
100	37488.49	32326.99	42650.00	29594.66	45382.33
101	37536.98	32157.94	42916.02	29310.45	45763.52
102	37607.41	32080.07	43134.76	29154.06	46060.76
103	37657.13	32020.67	43293.59	29036.91	46277.35
104	37703.90	31988.91	43418.89	28963.58	46444.22
105	37742.56	31970.07	43515.05	28914.30	46570.82
106	37776.35	31961.85	43590.85	28883.84	46668.86
107	37805.16	31959.82	43650.51	28865.48	46744.84
108	37830.00	31962.00	43698.00	28855.67	46804.33
109	37851.31	31966.64	43735.98	28851.49	46851.14
110	37869.64	31972.69	43766.58	28851.04	46888.23
111	37885.37	31979.38	43791.36	28852.94	46917.81
112	37898.89	31986.24	43811.55	28856.26	46941.53
113	37910.51	31992.94	43828.09	28860.36	46960.67
114	37920.49	31999.29	43841.70	28864.79	46976.20
115	37929.07	32005.19	43852.95	28869.28	46988.86
116	37936.44	32010.58	43862.29	28873.62	46999.25
117	37942.77	32015.46	43870.08	28877.73	47007.81
118	37948.21	32019.82	43876.60	28881.52	47014.90
119	37952.88	32023.70	43882.07	28884.98	47020.79
120	37956.90	32027.13	43886.67	28888.10	47025.70