**Name: Anthony Ayebiahwe Final Project work 3/27/18**

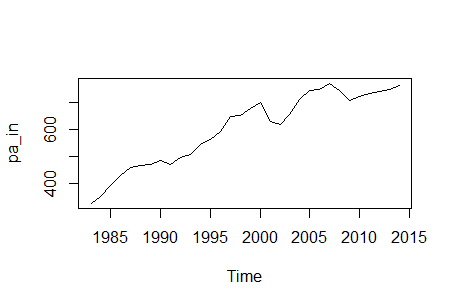
**Topic: Predicting 2015 total Airline enplanements for U.S. Air Carriers Operating Under 14 CFR 121**

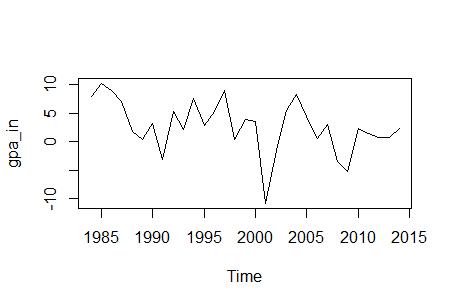
**Introduction:**

The objective of this paper is to model the time series of total passenger enplanement from 1995 through 2014, for U.S. Air Carriers Operating Under 14 CFR 121. There was a total of 32 observations from 1995 through 2014 (Yearly data).

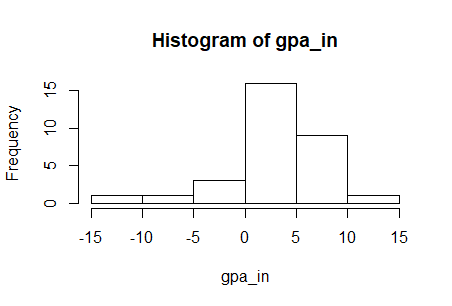
Data Source: National Transportation Safety Board

<https://catalog.data.gov/dataset/accidents-fatalities-and-rates-1995-through-2014-u-s-general-aviation>





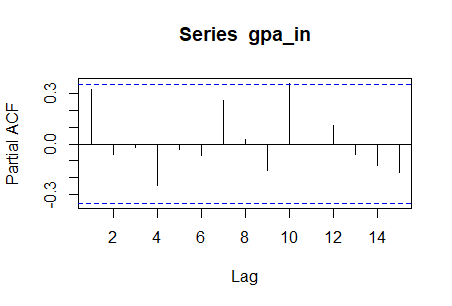
|  |
| --- |
|  |
| |  | | --- | |  | |

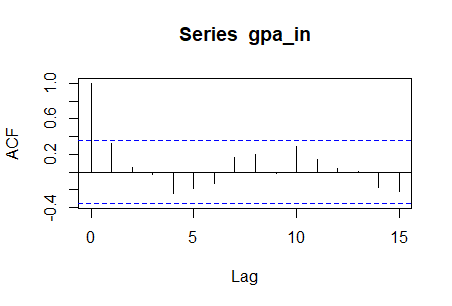


From the above graph, the lowest deployment took place in 2000 and the highest occurred somewhere during 2005. Also, the summary statistics over time are as below:

**Summary Statistics:**

|  |
| --- |
| summary  Min. 1st Qu. Median Mean 3rd Qu. Max.  -10.8377 0.6722 2.8935 2.7572 5.4535 10.2516  Standard deviation:  4.584449  kurtosis  3.937866  Skewness  -0.7434975  **Identification Stage:** |

Now, I want to see if there is time dependence in the data by calculating for the ACF and PACF with a maximum lag of 15. 



From the above ACF and partial ACF, there seems to be white noise in the data since the spikes for both the ACF and PACF are inside the blue line.

Also, I am interested in calculating the Q-stats from lag 1 through lag 15. The results are as follows. From the Q-stats below, p-values for all the lags are greater than 0.05, so we accept the null hypothesis of white noise.

Box-Ljung test

data: gpa\_in

X-squared = 3.514, df = 1, p-value = 0.06085

> Box.test(gpa\_in, lag = 2, type ="Ljung")

Box-Ljung test

data: gpa\_in

X-squared = 3.5951, df = 2, p-value = 0.1657

> Box.test(gpa\_in, lag = 3, type ="Ljung")

Box-Ljung test

data: gpa\_in

X-squared = 3.6163, df = 3, p-value = 0.306

> Box.test(gpa\_in, lag = 4, type ="Ljung")

Box-Ljung test

data: gpa\_in

X-squared = 5.7567, df = 4, p-value = 0.2181

> Box.test(gpa\_in, lag = 5, type ="Ljung")

Box-Ljung test

data: gpa\_in

X-squared = 7.0776, df = 5, p-value = 0.2149

> Box.test(gpa\_in, lag = 6, type ="Ljung")

Box-Ljung test

data: gpa\_in

X-squared = 7.7155, df = 6, p-value = 0.2597

> Box.test(gpa\_in, lag = 7, type ="Ljung")

Box-Ljung test

data: gpa\_in

X-squared = 8.8051, df = 7, p-value = 0.267

> Box.test(gpa\_in, lag = 8, type ="Ljung")

Box-Ljung test

data: gpa\_in

X-squared = 10.522, df = 8, p-value = 0.2303

> Box.test(gpa\_in, lag = 9, type ="Ljung")

Box-Ljung test

data: gpa\_in

X-squared = 10.534, df = 9, p-value = 0.309

> Box.test(gpa\_in, lag = 10, type ="Ljung")

Box-Ljung test

data: gpa\_in

X-squared = 14.682, df = 10, p-value = 0.1441

> Box.test(gpa\_in, lag = 11, type ="Ljung")

Box-Ljung test

data: gpa\_in

X-squared = 15.78, df = 11, p-value = 0.1495

> Box.test(gpa\_in, lag = 11, type ="Ljung")

Box-Ljung test

data: gpa\_in

X-squared = 15.78, df = 11, p-value = 0.1495

> Box.test(gpa\_in, lag = 12, type ="Ljung")

Box-Ljung test

data: gpa\_in

X-squared = 15.869, df = 12, p-value = 0.1973

> Box.test(gpa\_in, lag = 13, type ="Ljung")

Box-Ljung test

data: gpa\_in

X-squared = 15.872, df = 13, p-value = 0.2561

> Box.test(gpa\_in, lag = 14, type ="Ljung")

Box-Ljung test

data: gpa\_in

X-squared = 17.693, df = 14, p-value = 0.2211

> Box.test(gpa\_in, lag = 15, type ="Ljung")

Box-Ljung test

data: gpa\_in

X-squared = 20.829, df = 15, p-value = 0.1424

**Model Selection:**

From the PACF, there are 2 positive spikes so I could entertain AR(2). Also, from the ACF, there are 3 positive prominent spikes, so I can entertain MA(3). Also, from ACF, I can entertain as far as MA(5).Even though, there is no smooth decay towards zero in both ACF and PACF, I can entertain ARMA(2,3) and ARMA(2,5)

AR(2)

z test of coefficients:

Estimate Std. Error z value Pr(>|z|)

ar1 0.337765 0.177749 1.9002 0.057403 .

ar2 -0.045374 0.184180 -0.2464 0.805404

intercept 2.809246 1.072462 2.6194 0.008808 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

MA(3)

z test of coefficients:

Estimate Std. Error z value Pr(>|z|)

ma1 0.38160 0.16703 2.2847 0.02233 \*

ma2 0.15009 0.22958 0.6537 0.51327

ma3 0.21698 0.23918 0.9072 0.36431

intercept 2.87268 1.28567 2.2344 0.02546 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

ARMA(2,3)

z test of coefficients:

Estimate Std. Error z value Pr(>|z|)

ar1 0.3665928 0.7957907 0.4607 0.64504

ar2 -0.5968401 0.4672988 -1.2772 0.20153

ma1 -0.0064784 0.8205370 -0.0079 0.99370

ma2 0.6445594 0.3098092 2.0805 0.03748 \*

ma3 0.3702523 0.2682717 1.3801 0.16754

intercept 2.8649327 1.1835541 2.4206 0.01549 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

ARMA(2,5)

z test of coefficients:

Estimate Std. Error z value Pr(>|z|)

ar1 0.812518 0.705364 1.1519 0.24936

ar2 -0.516754 0.506412 -1.0204 0.30753

ma1 -0.505177 0.742804 -0.6801 0.49644

ma2 0.338147 0.390362 0.8662 0.38636

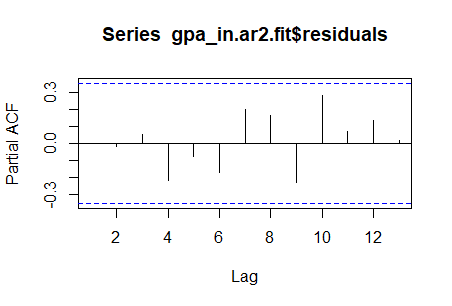
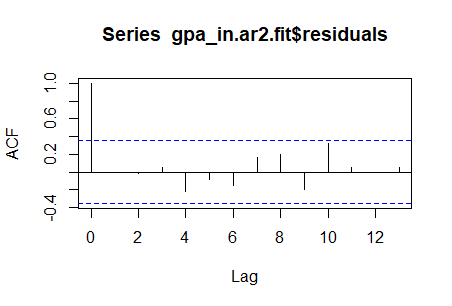
ma3 0.221372 0.215278 1.0283 0.30380

ma4 -0.272365 0.221956 -1.2271 0.21978

ma5 -0.010259 0.283594 -0.0362 0.97114

intercept 2.683955 0.824314 3.2560 0.00113 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ # Getting the ACF and PACF up to 13 lags of residuals of AR(2) and getting the Q-stats from 3 through 13. 

Box-Ljung test

data: gpa\_in.ar2.fit$residuals

X-squared = 0.10773, df = 1, p-value = 0.7427

> Box.test(gpa\_in.ar2.fit$residuals, lag = 4, type = c("Ljung"), fitdf = 2)

Box-Ljung test

data: gpa\_in.ar2.fit$residuals

X-squared = 1.9229, df = 2, p-value = 0.3823

> Box.test(gpa\_in.ar2.fit$residuals, lag = 5, type = c("Ljung"), fitdf = 2)

Box-Ljung test

data: gpa\_in.ar2.fit$residuals

X-squared = 2.164, df = 3, p-value = 0.5391

> Box.test(gpa\_in.ar2.fit$residuals, lag = 6, type = c("Ljung"), fitdf = 2)

Box-Ljung test

data: gpa\_in.ar2.fit$residuals

X-squared = 3.0893, df = 4, p-value = 0.543

> Box.test(gpa\_in.ar2.fit$residuals, lag = 7, type = c("Ljung"), fitdf = 2)

Box-Ljung test

data: gpa\_in.ar2.fit$residuals

X-squared = 4.2637, df = 5, p-value = 0.5121

> Box.test(gpa\_in.ar2.fit$residuals, lag = 8, type = c("Ljung"), fitdf = 2)

Box-Ljung test

data: gpa\_in.ar2.fit$residuals

X-squared = 5.953, df = 6, p-value = 0.4285

> Box.test(gpa\_in.ar2.fit$residuals, lag = 9, type =c("Ljung"), fitdf = 2)

Box-Ljung test

data: gpa\_in.ar2.fit$residuals

X-squared = 7.7472, df = 7, p-value = 0.3554

> Box.test(gpa\_in.ar2.fit$residuals, lag = 10, type = c("Ljung"), fitdf = 2)

Box-Ljung test

data: gpa\_in.ar2.fit$residuals

X-squared = 12.839, df = 8, p-value = 0.1175

> Box.test(gpa\_in.ar2.fit$residuals, lag = 11, type = c("Ljung"), fitdf = 2)

Box-Ljung test

data: gpa\_in.ar2.fit$residuals

X-squared = 12.978, df = 9, p-value = 0.1636

> Box.test(gpa\_in.ar2.fit$residuals, lag = 12, type = c("Ljung"), fitdf = 2)

Box-Ljung test

data: gpa\_in.ar2.fit$residuals

X-squared = 12.979, df = 10, p-value = 0.2249

> Box.test(gpa\_in.ar2.fit$residuals, lag = 13, type = c("Ljung"), fitdf = 2)

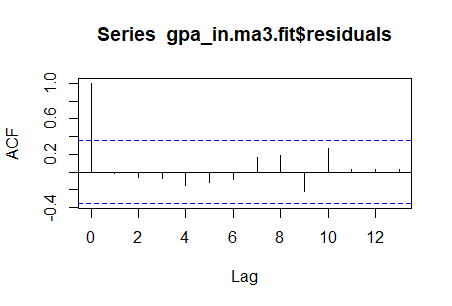
Box-Ljung test

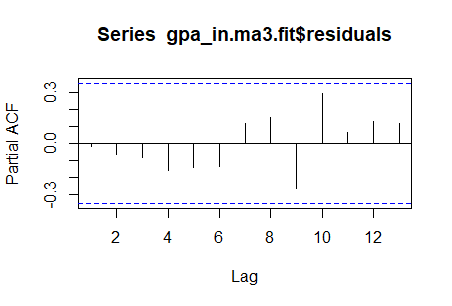
data: gpa\_in.ar2.fit$residuals

X-squared = 13.147, df = 11, p-value = 0.2838

#From the Q stats, all the p-values are greater than 0.05, so none of the lags are significant at the 5% level.

# Getting the ACF and PACF up to 13 lags of residuals of MA (3) and getting the Q-stats from 3 through 13.





Box-Ljung test

data: gpa\_in.ma3.fit$residuals

X-squared = 1.207, df = 1, p-value = 0.2719

> Box.test(gpa\_in.ma3.fit$residuals, lag = 5, type = c("Ljung"), fitdf = 3)

Box-Ljung test

data: gpa\_in.ma3.fit$residuals

X-squared = 1.7561, df = 2, p-value = 0.4156

> Box.test(gpa\_in.ma3.fit$residuals, lag = 6, type = c("Ljung"), fitdf = 3)

Box-Ljung test

data: gpa\_in.ma3.fit$residuals

X-squared = 2.0738, df = 3, p-value = 0.5572

> Box.test(gpa\_in.ma3.fit$residuals, lag = 7, type = c("Ljung"), fitdf = 3)

Box-Ljung test

data: gpa\_in.ma3.fit$residuals

X-squared = 3.297, df = 4, p-value = 0.5094

> Box.test(gpa\_in.ma3.fit$residuals, lag = 8, type = c("Ljung"), fitdf = 3)

Box-Ljung test

data: gpa\_in.ma3.fit$residuals

X-squared = 4.8401, df = 5, p-value = 0.4357

> Box.test(gpa\_in.ma3.fit$residuals, lag = 9, type = c("Ljung"), fitdf = 3)

Box-Ljung test

data: gpa\_in.ma3.fit$residuals

X-squared = 7.1628, df = 6, p-value = 0.3061

> Box.test(gpa\_in.ma3.fit$residuals, lag = 10, type =c("Ljung"), fitdf = 3)

Box-Ljung test

data: gpa\_in.ma3.fit$residuals

X-squared = 10.765, df = 7, p-value = 0.1492

> Box.test(gpa\_in.ma3.fit$residuals, lag = 11, type = c("Ljung"), fitdf = 3)

Box-Ljung test

data: gpa\_in.ma3.fit$residuals

X-squared = 10.817, df = 8, p-value = 0.2123

> Box.test(gpa\_in.ma3.fit$residuals, lag = 12, type = c("Ljung"), fitdf = 3)

Box-Ljung test

data: gpa\_in.ma3.fit$residuals

X-squared = 10.88, df = 9, p-value = 0.284

> Box.test(gpa\_in.ma3.fit$residuals, lag = 13, type = c("Ljung"), fitdf = 3)

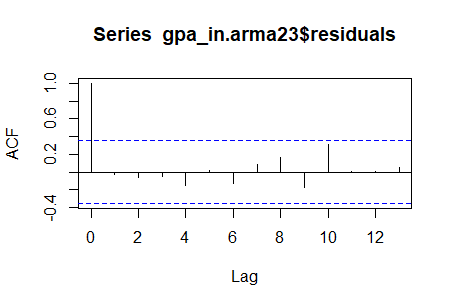
Box-Ljung test

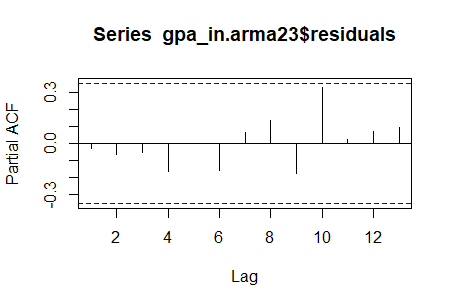
data: gpa\_in.ma3.fit$residuals

X-squared = 10.924, df = 10, p-value = 0.3635

#From the Q stats, all the p-values are greater than 0.05, so none of the lags are significant at the 5% level.

#Getting the ACF and PACF up to 13 lags of residuals of ARMA (2,3) and getting the Q-stats from 5 through 13.





Box-Ljung test

data: gpa\_in.arma23$residuals

X-squared = 1.8544, df = 1, p-value = 0.1733

> Box.test(gpa\_in.arma23$residuals, lag = 7, type = c("Ljung"), fitdf = 5)

Box-Ljung test

data: gpa\_in.arma23$residuals

X-squared = 2.1864, df = 2, p-value = 0.3351

> Box.test(gpa\_in.arma23$residuals, lag = 8, type = c("Ljung"), fitdf = 5)

Box-Ljung test

data: gpa\_in.arma23$residuals

X-squared = 3.4514, df = 3, p-value = 0.3271

> Box.test(gpa\_in.arma23$residuals, lag = 9, type = c("Ljung"), fitdf = 5)

Box-Ljung test

data: gpa\_in.arma23$residuals

X-squared = 4.9088, df = 4, p-value = 0.2968

> Box.test(gpa\_in.arma23$residuals, lag = 10, type = c("Ljung"), fitdf =5)

Box-Ljung test

data: gpa\_in.arma23$residuals

X-squared = 9.6448, df = 5, p-value = 0.08595

> Box.test(gpa\_in.arma23$residuals, lag = 11, type = c("Ljung"), fitdf =5)

Box-Ljung test

data: gpa\_in.arma23$residuals

X-squared = 9.6465, df = 6, p-value = 0.1403

> Box.test(gpa\_in.arma23$residuals, lag = 12, type =c("Ljung"), fitdf = 5)

Box-Ljung test

data: gpa\_in.arma23$residuals

X-squared = 9.6469, df = 7, p-value = 0.2095

> Box.test(gpa\_in.arma23$residuals, lag = 13, type =c("Ljung"), fitdf = 5)

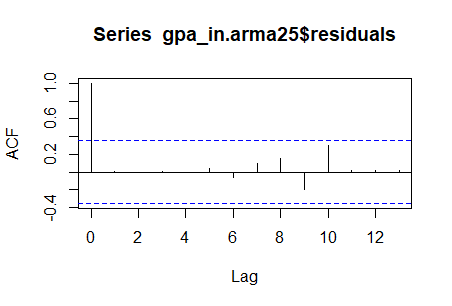
Box-Ljung test

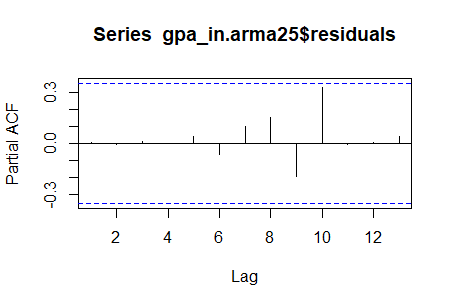
data: gpa\_in.arma23$residuals

X-squared = 9.7847, df = 8, p-value = 0.2805

#From the Q stats, all the p-values are greater than 0.05, so none of the lags are significant at the 5% level.

## Getting the ACF and PACF up to 13 lags of residuals of ARMA (2,5) and getting the Q-stats from 7 through 13.





Box-Ljung test

data: gpa\_in.arma25$residuals

X-squared = 1.7337, df = 1, p-value = 0.1879

> Box.test(gpa\_in.arma25$residuals, lag = 9, type = c("Ljung"), fitdf = 7)

Box-Ljung test

data: gpa\_in.arma25$residuals

X-squared = 3.4293, df = 2, p-value = 0.18

> Box.test(gpa\_in.arma25$residuals, lag = 10, type = c("Ljung"), fitdf = 7)

Box-Ljung test

data: gpa\_in.arma25$residuals

X-squared = 7.851, df = 3, p-value = 0.04919

> Box.test(gpa\_in.arma25$residuals, lag = 11, type = c("Ljung"), fitdf = 7)

Box-Ljung test

data: gpa\_in.arma25$residuals

X-squared = 7.8773, df = 4, p-value = 0.09618

> Box.test(gpa\_in.arma25$residuals, lag = 12, type = c("Ljung"), fitdf =7)

Box-Ljung test

data: gpa\_in.arma25$residuals

X-squared = 7.8912, df = 5, p-value = 0.1623

> Box.test(gpa\_in.arma25$residuals, lag = 13, type = c("Ljung"), fitdf =7)

Box-Ljung test

data: gpa\_in.arma25$residuals

X-squared = 7.9071, df = 6, p-value = 0.245

# From the Q stats, all the p-values are greater than 0.05, so none of the lags are significant at the 5% level.

# Finding the residual variance for each model

# gaa.rv.ar2

[1] 18.11059

> gaa.rv.ma3 = gpa\_in.ma3.fit$sigma2

> gaa.rv.ma3

[1] 17.55273

> gaa.rv.arma23 = gpa\_in.arma23$sigma2

> gaa.rv.arma23

[1] 16.58427

> gaa.rv.arma25=gpa\_in.arma25$sigma2

> gaa.rv.arma25

[1] 16.05368

# Finding the AIC and BIC of the models

> AIC(gpa\_in.ar2.fit)

[1] 185.88

> BIC(gpa\_in.ar2.fit)

[1] 191.6159

> AIC(gpa\_in.ma3.fit)

[1] 187.0748

> BIC(gpa\_in.ma3.fit)

[1] 194.2448

> AIC(gpa\_in.arma23)

[1] 189.675

> BIC(gpa\_in.arma23)

[1] 199.7129

> AIC(gpa\_in.arma25)

[1] 192.7126

> BIC(gpa\_in.arma25)

[1] 205.6185

# From the above, the competing models are the AR(2) and ARMA(2,5). AR(2) has the largest variance residuals but the lowest AIC and BIC; while the ARMA(2,5) has a lower AIC but bigger than that of AR(2). Moreover, ARMA (2,5) has the lowest residual variance. Hence, I will choose ARMA(2,5) for the forecast.

**One Step Ahead forecast:**

|  |
| --- |
| Forecast method: ARIMA(2,0,0) with non-zero mean  Model Information:  Call:  arima(x = gpa\_in, order = c(2, 0, 0))  Coefficients:  ar1 ar2 intercept  0.3378 -0.0454 2.8092  s.e. 0.1777 0.1842 1.0725  sigma^2 estimated as 18.11: log likelihood = -88.94, aic = 185.88  Error measures:  ME RMSE MAE MPE MAPE MASE ACF1  Training set -0.04473901 4.255654 3.275818 -54.07427 146.2733 0.7871413 -2.242259e-05  Forecasts:  Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  2015 2.756513 -2.697327 8.210352 -5.584416 11.09744  In 2015, I expect the total enplanement on major airlines to increase by approximately 2.756513%. |
|  |
| |  | | --- | |  | |

with about 17.85 uncertainty around this forecast.

