````
---
title: "05_Diagnostic_Tests"
author: "Callum Weinberg"
date: "December 1, 2021"
output: pdf_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

## Libraries

```{r}
library(dplyr, warn.conflicts = FALSE) #Using
library(tidyr) #Using
library(knitr) #Using
library(lubridate, warn.conflicts = FALSE) #Using
library(ggplot2) #Using
library(MASS) #Uncertain
library(qpcR) #Using
library(forecast) #Using
library(cowplot) #Using
library(TSA) #Using
```

## QQplot Function
```{r}
#Function for QQ Plot in GGPLOT
# Source: https://stackoverflow.com/questions/4357031/qqnorm-and-qqline-in-ggplot2
qqplot_residuals <- function (vec) # argument: vector of numbers
{
  # following four lines from base R's qqline()
  y <- quantile(vec[!is.na(vec)], c(0.25, 0.75))
  x <- qnorm(c(0.25, 0.75))
  slope <- diff(y)/diff(x)
  int <- y[1L] - slope * x[1L]

  d <- data.frame(resids = vec)

  ggplot(d, aes(sample = resids)) +
    stat_qq(color = "blue") +
    geom_abline(slope = slope, intercept = int) +
  theme(plot.title = element_text(hjust = 0.5))
}
```

##### SEASONAL ONLY MODELS ######

```{r}
load(file="Data/landings_transformed_season_only.Rdata")
landing_ts_so = landings_transformed_season_only$pounds_transformed
```

```{r}
# Model 40 AKA Model 1 in the Report
model40 = arima(landing_ts_so, order=c(2,0,2), seasonal = list(order = c(1,1,1), period = 12),
                method = "ML",fixed = c(NA,NA,0,NA,NA,NA))
model40
AICc(model40)


# Diagnostics
Box.test(residuals(model40),lag = 11, type = ("Box-Pierce"), fitdf = 5) # Only 5 since one term fixed at 0
Box.test(residuals(model40),lag = 11, type = ("Ljung-Box"), fitdf = 5)
Box.test(residuals(model40)^2,lag = 11, type = ("Ljung-Box"), fitdf = 0)
shapiro.test(residuals(model40))


## acf
acf_resid = acf(residuals(model40),main = "Autocorrelation", lag.max = 60)
# Put into Dataframe
sample_acf = as.data.frame(do.call(cbind, acf_resid))
# Confidence Interval Line
conf.level = 0.95
ciline_resid = qnorm((1 - conf.level)/2)/sqrt(132)

# Plot
ACF_residual_mod1= ggplot(data = sample_acf, mapping = aes(x = as.numeric(lag), y = as.numeric(acf))) +
  geom_hline(aes(yintercept = 0)) +
````
````

```
  geom_segment(mapping = aes(xend = as.numeric(lag), yend = 0)) +
  geom_hline(aes(yintercept = ciline_resid), linetype = 2, color = 'darkblue') +
  geom_hline(aes(yintercept = -ciline_resid), linetype = 2, color = 'darkblue') +
  labs(x = "lag", y = "ACF") +
  #labs(title = "") +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 10),
    legend.text = element_text(size = 10),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 12),
    axis.text.x = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))

## pacf
pacf_resid = pacf(residuals(model40),main = "Autocorrelation", lag.max = 60)
# Put into Dataframe
sample_pacf = as.data.frame(do.call(cbind, pacf_resid))
# Plot
PACF_residual_mod1= ggplot(data = sample_pacf, mapping = aes(x = as.numeric(lag), y = as.numeric(acf))) +
  geom_hline(aes(yintercept = 0)) +
  geom_segment(mapping = aes(xend = as.numeric(lag), yend = 0)) +
  geom_hline(aes(yintercept = ciline_resid), linetype = 2, color = 'darkblue') +
  geom_hline(aes(yintercept = -ciline_resid), linetype = 2, color = 'darkblue') +
  labs(x = "lag", y = "PACF") +
  #labs(title = "") +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 10),
    legend.text = element_text(size = 10),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 12),
    axis.text.x = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))

# Histogram
hist_df_mod = data.frame(x = residuals(model40))
histogram_resid1 = ggplot(data = hist_df, aes(x = x/1000)) +
  geom_histogram(aes(y = ..density..)) +
  geom_density(alpha = 0.1, fill = "red") +
  labs(x = "Residuals (Divided by 1000)", y = "Frequency") +
  #scale_x_continuous(label = comma) +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 15),
    legend.text = element_text(size = 15),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 12),
    axis.text.x = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))
histogram_resid

# q-q plot
qq_resid_mod1 = qqplot_residuals(residuals(model40))

# Plot All Together
#png(filename = "Images/diagnostics_mod1.png", width = 960, height = 960)
model1_grid = plot_grid(qq_resid_mod1,histogram_resid1,ACF_residual_mod1,PACF_residual_mod1, labels = NULL, label_size =
12, ncol = 2, nrow = 2)
#dev.off()

```

```{r}
# Model 43 AKA Model 3 in the Report
model43 = arima(landing_ts_so, order=c(1,0,1), seasonal = list(order = c(1,1,1), period = 12),
                method = "ML",fixed = c(NA,NA,NA,NA))
model43
AICc(model43)



# Diagnostics
Box.test(residuals(model43),lag = 11, type = ("Box-Pierce"), fitdf = 4)
Box.test(residuals(model43),lag = 11, type = ("Ljung-Box"), fitdf = 4)
Box.test(residuals(model43)^2,lag = 11, type = ("Ljung-Box"), fitdf = 0)
shapiro.test(residuals(model43))

## acf
acf_resid = acf(residuals(model43),main = "Autocorrelation", lag.max = 60)
# Put into Dataframe
```

```
sample_acf = as.data.frame(do.call(cbind, acf_resid))
# Confidence Interval Line
conf.level = 0.95
ciline_resid = qnorm((1 - conf.level)/2)/sqrt(132)

# Plot
ACF_residual_mod3= ggplot(data = sample_acf, mapping = aes(x = as.numeric(lag), y = as.numeric(acf))) +
  geom_hline(aes(yintercept = 0)) +
  geom_segment(mapping = aes(xend = as.numeric(lag), yend = 0)) +
  geom_hline(aes(yintercept = ciline_resid), linetype = 2, color = 'darkblue') +
  geom_hline(aes(yintercept = -ciline_resid), linetype = 2, color = 'darkblue') +
  labs(x = "lag", y = "ACF") +
  #labs(title = "") +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 10),
    legend.text = element_text(size = 10),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 12),
    axis.text.x = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))

## pacf
pacf_resid = pacf(residuals(model43),main = "Autocorrelation", lag.max = 60)
# Put into Dataframe
sample_pacf = as.data.frame(do.call(cbind, pacf_resid))
# Plot
PACF_residual_mod3= ggplot(data = sample_pacf, mapping = aes(x = as.numeric(lag), y = as.numeric(acf))) +
  geom_hline(aes(yintercept = 0)) +
  geom_segment(mapping = aes(xend = as.numeric(lag), yend = 0)) +
  geom_hline(aes(yintercept = ciline_resid), linetype = 2, color = 'darkblue') +
  geom_hline(aes(yintercept = -ciline_resid), linetype = 2, color = 'darkblue') +
  labs(x = "lag", y = "PACF") +
  #labs(title = "") +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 10),
    legend.text = element_text(size = 10),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 12),
    axis.text.x = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))

# Histogram
hist_df_mod = data.frame(x = residuals(model43))
histogram_resid3 = ggplot(data = hist_df, aes(x = x/1000)) +
  geom_histogram(aes(y = ..density..)) +
  geom_density(alpha = 0.1, fill = "red") +
  labs(x = "Residuals (Divided by 1000)", y = "Frequency") +
  #scale_x_continuous(label = comma) +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 15),
    legend.text = element_text(size = 15),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 12),
    axis.text.x = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))

# q-q plot
qq_resid_mod3 = qqplot_residuals(residuals(model43))

# Plot All Together
png(filename = "Images/diagnostics_mod3.png", width = 960, height = 960)
plot_grid(qq_resid_mod3,histogram_resid3,ACF_residual_mod3,PACF_residual_mod3, label_size = 12, ncol = 2, nrow = 2)
dev.off()

```



```{r}
# Model 44 AKA Model 2 in Report
model44 = arima(landing_ts_so, order=c(2,0,3), seasonal = list(order = c(1,1,1), period = 12),
                method = "ML",fixed = c(0,NA,0,0,NA,NA,NA))
model44
AICc(model44)

# Diagnostics
Box.test(residuals(model44),lag = 11, type = ("Box-Pierce"), fitdf = 4) # Only 4 coefficients, since 3 are set to 0
Box.test(residuals(model44),lag = 11, type = ("Ljung-Box"), fitdf = 4)
```

```
Box.test(residuals(model44)^2,lag = 11, type = ("Ljung-Box"), fitdf = 4)
shapiro.test(residuals(model44))


## acf
acf_resid = acf(residuals(model44),main = "Autocorrelation", lag.max = 60)
# Put into Dataframe
sample_acf = as.data.frame(do.call(cbind, acf_resid))
# Confidence Interval Line
conf.level = 0.95
ciline_resid = qnorm((1 - conf.level)/2)/sqrt(132)

# Plot
ACF_residual_mod4= ggplot(data = sample_acf, mapping = aes(x = as.numeric(lag), y = as.numeric(acf))) +
  geom_hline(aes(yintercept = 0)) +
  geom_segment(mapping = aes(xend = as.numeric(lag), yend = 0)) +
  geom_hline(aes(yintercept = ciline_resid), linetype = 2, color = 'darkblue') +
  geom_hline(aes(yintercept = -ciline_resid), linetype = 2, color = 'darkblue') +
  labs(x = "lag", y = "ACF") +
  #labs(title = "") +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 10),
    legend.text = element_text(size = 10),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 12),
    axis.text.x = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))

## pacf
pacf_resid = pacf(residuals(model44),main = "Autocorrelation", lag.max = 60)
# Put into Dataframe
sample_pacf = as.data.frame(do.call(cbind, pacf_resid))
# Plot
PACF_residual_mod4= ggplot(data = sample_pacf, mapping = aes(x = as.numeric(lag), y = as.numeric(acf))) +
  geom_hline(aes(yintercept = 0)) +
  geom_segment(mapping = aes(xend = as.numeric(lag), yend = 0)) +
  geom_hline(aes(yintercept = ciline_resid), linetype = 2, color = 'darkblue') +
  geom_hline(aes(yintercept = -ciline_resid), linetype = 2, color = 'darkblue') +
  labs(x = "lag", y = "PACF") +
  #labs(title = "") +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 10),
    legend.text = element_text(size = 10),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 12),
    axis.text.x = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))

# Histogram
hist_df_mod = data.frame(x = residuals(model44))
histogram_resid4 = ggplot(data = hist_df, aes(x = x/1000)) +
  geom_histogram(aes(y = ..density..)) +
  geom_density(alpha = 0.1, fill = "red") +
  labs(x = "Residuals (Divided by 1000)", y = "Frequency") +
  #scale_x_continuous(label = comma) +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 15),
    legend.text = element_text(size = 15),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 12),
    axis.text.x = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))

# q-q plot
qq_resid_mod4 = qqplot_residuals(residuals(model44))

# Plot All Together
model2_grid = plot_grid(qq_resid_mod4,histogram_resid4,ACF_residual_mod4,PACF_residual_mod4, labels = NULL, label_size =
12, ncol = 2, nrow = 2)

png(filename = "Images/diagnostics_mod1_2.png", width = 960, height = 720)
plot_grid(model1_grid,model2_grid, label_size = 12, ncol = 2, nrow = 1, labels = c("Model1","Model2"), label_x = .2)
dev.off()
```

````
##### SEASONAL and TREND MODELS - Not used in report #####

## Model 0

```{r}
model0 = arima(landing_ts, order=c(1,1,0), seasonal = list(order = c(1,1,0), period = 12), method = "ML")
model0
AICc(model0)



# Diagnostics
Box.test(residuals(model0),lag = 11, type = ("Box-Pierce"), fitdf = 1)

plot(residuals(model0))

acf(residuals(model0))
pacf(residuals(model0))

```


## Model 1

```{r}
model1 = arima(landing_ts, order=c(2,1,0), seasonal = list(order = c(1,1,0), period = 12), method = "ML")
model1
AICc(model1)



# Diagnostics
Box.test(residuals(model1),lag = 11, type = ("Box-Pierce"), fitdf = 3)

plot(residuals(model1))

acf(residuals(model1))
pacf(residuals(model1))

```




###### TREND DIFFERENCE ONLY MODELS - Not used in report #####

```{r}
# THIS WORKS, BUT PROBABLY NOT A GREAT MODEL
model21 = arima(landing_ts_ns, order=c(1,1,0), method = "ML")
model21
AICc(model21)

# Phi (corresponsing to AR)
AR = polyroot(c(1,-0.6605))
roots_AR = c("Root 1")
root_model5 = data.frame(Root = roots_AR, Value = AR)
kable(root_model5, caption = "Phi(B) Roots")


# Diagnostics
Box.test(residuals(model21),lag = 11, type = ("Box-Pierce"), fitdf = 1)

plot(residuals(model21))

acf(residuals(model21), lag.max = 50)
pacf(residuals(model21), lag.max = 50)
```
````

```{r}
# Rerun the Model (defined in 04)
model22 = arima(landing_ts_ns, order=c(15,1,0), method = "ML")
model22
AICc(model22)


# Diagnostics
Box.test(residuals(model22),lag = 11, type = ("Box-Pierce"), fitdf = 15) # Fails, not enough df
Box.test(residuals(model22),lag = 11, type = ("Ljung-Box"), fitdf = 15) # Fails, not enough df
Box.test(residuals(model22)^2,lag = 11, type = ("Ljung-Box"), fitdf = 0) # Fine, no evidence of non-linear dependence

# Residuals
plot(residuals(model22))

## QQPlot
qqnorm(residuals(model22))
qqline(residuals(model22),col ="blue")

# Plot diagnostics of residuals
par(mfrow=c(1,2),oma=c(0,0,2,0))
op <- par(mfrow=c(2,2))
# acf
acf(residuals(model22),main = "Autocorrelation")
# pacf
pacf(residuals(model22),main = "Partial Autocorrelation")
# Histogram
hist(residuals(model22),main = "Histogram")
# q-q plot
qqnorm(residuals(model22))
qqline(residuals(model22),col ="blue")
```

```{r}
# Try a different version
model5c = arima(landing_ts_ns, order=c(7,1,0), method = "ML")
model5c
AICc(model5c)


# Diagnostics
Box.test(residuals(model5c),lag = 11, type = ("Box-Pierce"), fitdf = 7)


Box.test(residuals(model5c)^2,lag = 11, type = ("Ljung-Box"), fitdf = 0)


plot(residuals(model5c))

# Plot diagnostics of residuals
par(mfrow=c(1,2),oma=c(0,0,2,0))
op <- par(mfrow=c(2,2))
# acf
acf(residuals(model5c),main = "Autocorrelation")
# pacf
pacf(residuals(model5c),main = "Partial Autocorrelation")
# Histogram
hist(residuals(model5c),main = "Histogram")
# q-q plot
qqnorm(residuals(model5c))
qqline(residuals(model5c),col ="blue")
```