7. ARIMA & SARIMA

author: ECONOMIC FORECASTING date: SUMMER 2017 autosize: true

Overview

ARIMA and the **Box-Jenkins Methodology** - Technique for univariate data - can be extended by ARIMAX (X is for exogenous variables) - Does not impose any structure on the data generating process - less restrictive than many other models - Model is "done" when the error term contains no more useful information

== So far, we have assumed that the data has been stationary, so that we can use an ARMA methodology

. . . but what if the data is NOT stationary?

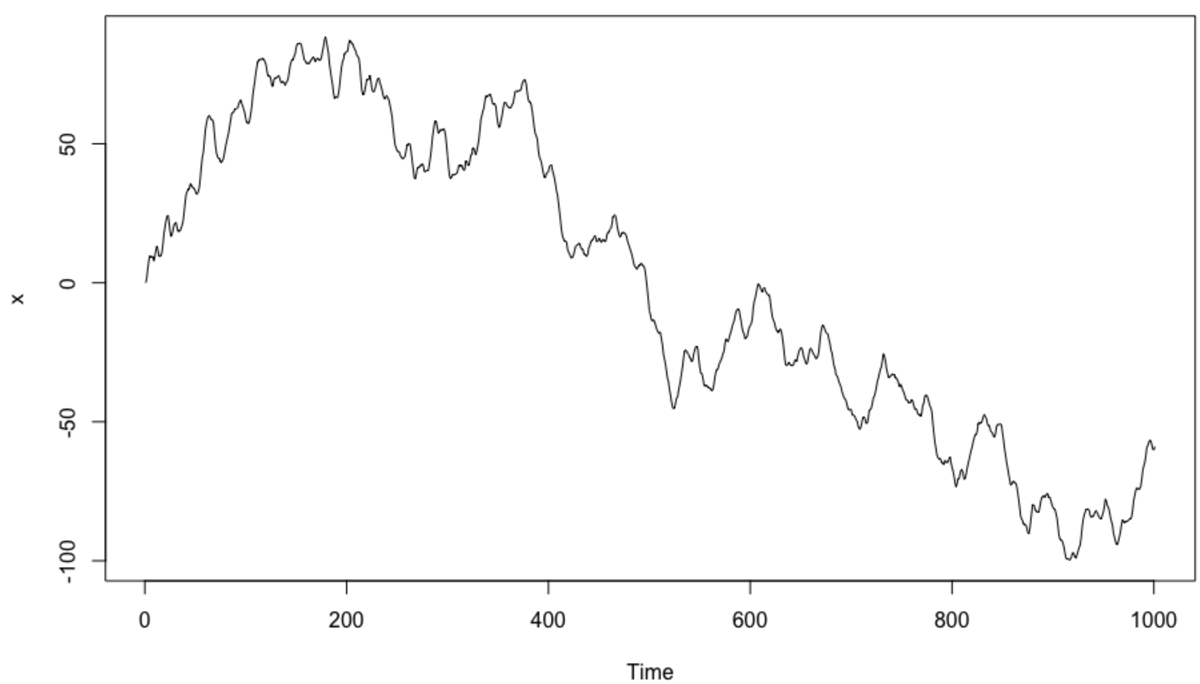


Figure 1:

==

==

== If the plot of the data wasn't indicative enough that the data were not stationary, the ACF clearly shows that it is not stationary

. . . but what about the difference of the data?

==

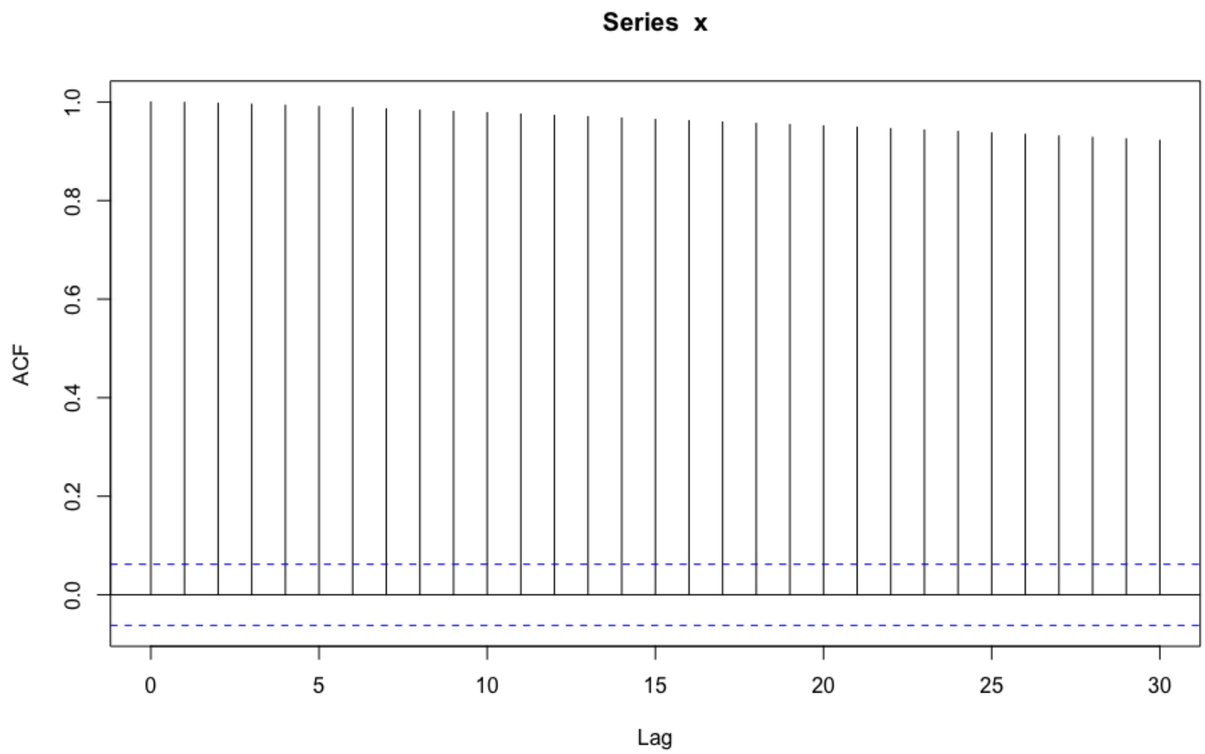


Figure 2:

==

== Actual Simulation: `arima.sim(list(order=c(1,1,1), ar=0.7, ma=0.2), n=1000)`

ARIMA(2,1,0) which was suggested by R - $AIC = 2790.36$ - $\hat{\alpha}_1 = 0.8819$ (.0314) and $\hat{\beta}$ was not estimated

ARIMA(4,1,0) which is what it looks like… - $AIC = 2794.31$ - $\hat{\alpha}_1 = 0.8822$ (.0317) and $\hat{\beta}$ not estimated

ARIMA(1,1,1) which is what it is… - $AIC = 2790.67$ - $\hat{\alpha}_1 = 0.7192$ (.0276) and $\hat{\beta} = 0.1613$ (0.039)

Box-Jenkins Methodology

== **Step 1.** Postulate a general class of model - Choose between univariate and multivariate method - Currently, we are looking at a univariate method

Step 2a. Identify the Model - Determine whether or not the series is stationary - Visual inspections - If stationary, move to **Step 2b: Choosing Order of Model** - If nonstationary, data must be made stationary via differencing/detrending

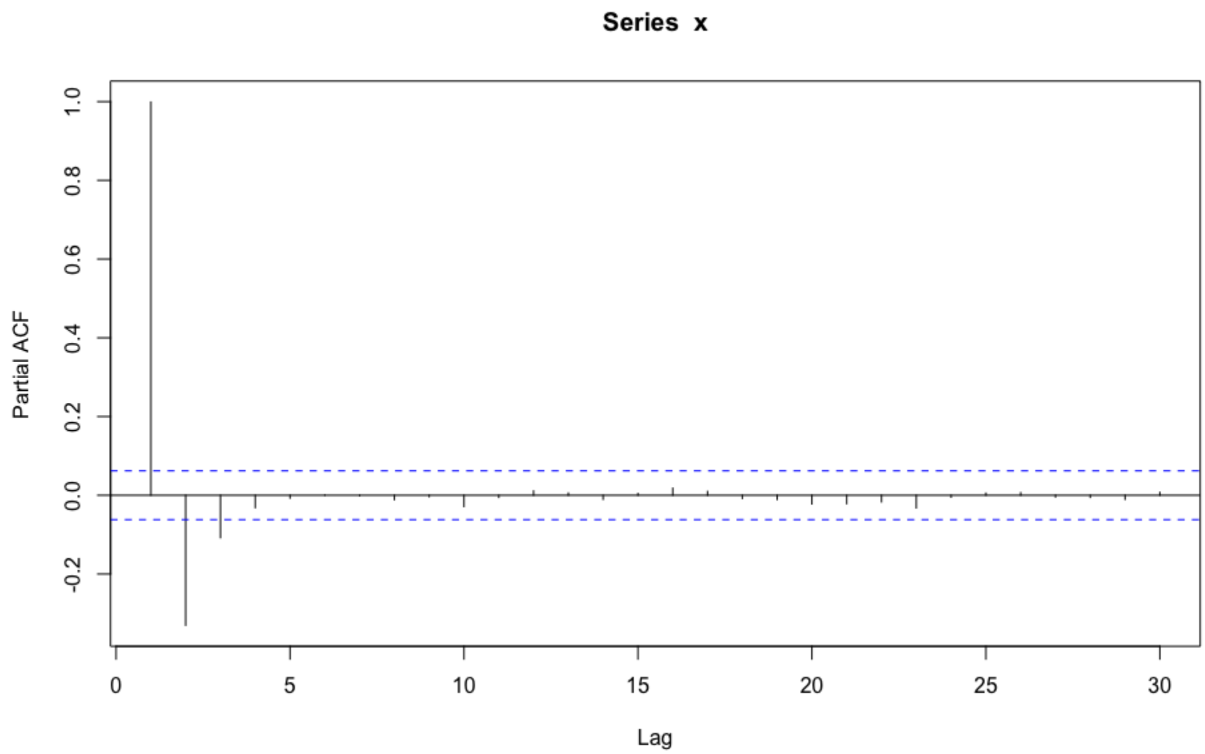


Figure 3:

Definition

Integrated Data > A time series $\{x_t\}$ is **integrated of order** d, denoted $I(d)$ if the dth difference of $\{x_t\}$ is white noise $\{w_t\}$. That is $\nabla^d x_t = w_t$

In backshift notation...

$$(1 - \mathbf{B})^d x_t = w_t$$

Integrated data (i.e. contains at least one unit root) must be differenced to be made stationary. - Detrending a unit root, by adding time measure variables, will not make the series stationary. - must be differenced. - However, differencing a series that needs to be detrended will introduce autocorrelation of a high order

Continued...

- To make decisions regarding trend stationary vs difference stationary, inspection of the ACFs of the residuals of a time series polynomial is a good place to start.
  - Slow decay of the ACF $\longrightarrow$ Requires differencing
  - Quick decay of the ACF $\longrightarrow$ Requires detrending

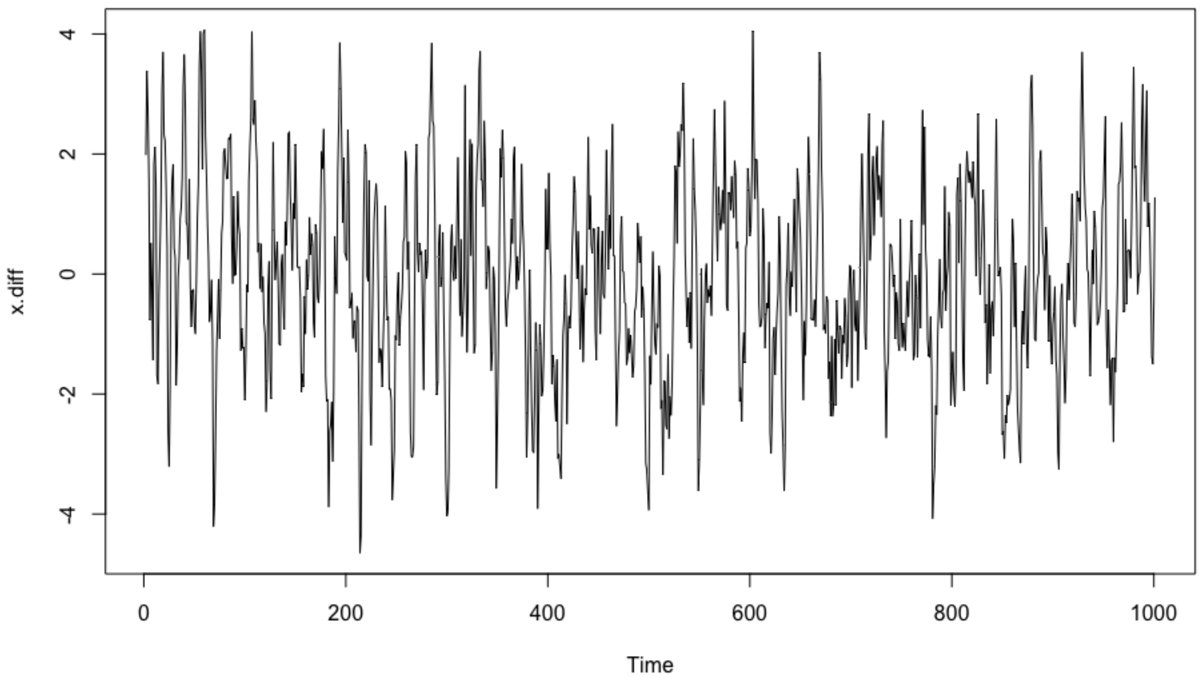


Figure 4:

Definition

ARIMA > A time series $\{x_t\}$ follows an ARIMA(p, d, q) process if the dth difference of the $\{x_t\}$ series follows an ARMA(p, q) process.

In backshift notation…

$$\theta_p(\mathbf{B})(1-\mathbf{B})^d x_t = \theta_q(\mathbf{B}) w_t$$

== **Step 2b.** Choose Order of Model - Inspect ACF and PACF of the stationary series. - Estimate Model - Data driven approach - try multiple specifications - Keep terms that are statistically significant - Compare AIC and AIC_c - Remember to consider parsimony

== **Step 3.** Determine Adequacy - Visually inspect normality of errors - `plot(density(<residual object>))` - `qqnorm(<residual object>)` then `qqline(<residual object>)` - majority of points should fall on the line - Individual residual autocorrelation coefficients should be small - Residual autocorrelation, as a group, should be white noise. - LM test

If model is indadequate, resturn to **Step 2a** and re-identify; otherwise move on to **Step 4. Forecasting**

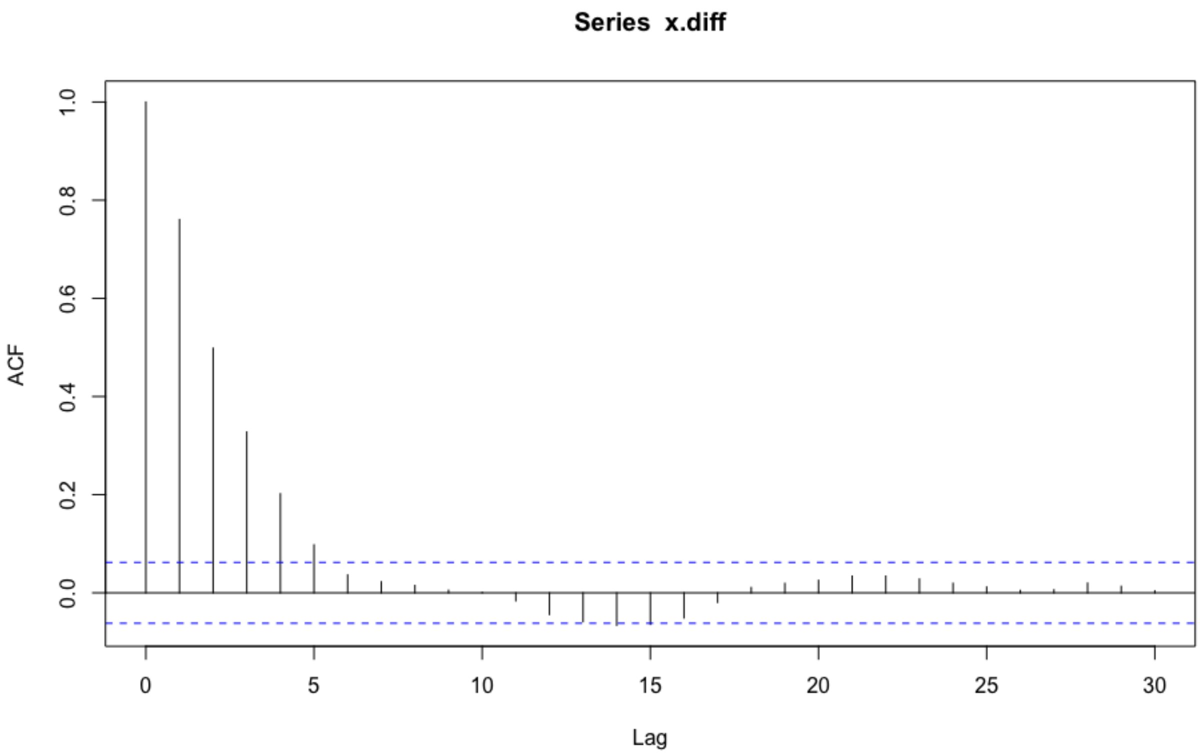


Figure 5:

Density Plot of Residuals

Density Plot of Residuals

== **Step 4.** Forecasting

Monitoring the forecast - ARIMA models are appropriate for short- and medium-run forecasts - As time passes, the analyst should assess the model and forecast to make sure it is still appropriate - Parameters may need to be re-estimated - If a pattern emerges in the residuals, the models will need to be redone

Seasonal ARIMA

SARIMA$(p, d, q)(P, D, Q)_s$ accounts for data with seasonality.

In backshit notation…

$$\Theta_P(\mathbf{B}^s)\theta_p(\mathbf{B})(1-\mathbf{B}^s)^D(1-\mathbf{B})^d x_t = \Phi_Q(\mathbf{B}^s)\phi_q(\mathbf{B})w_t$$

Again, to assume stationarity, the roots of the characteristic equation must all be greater than 1 in absolute value

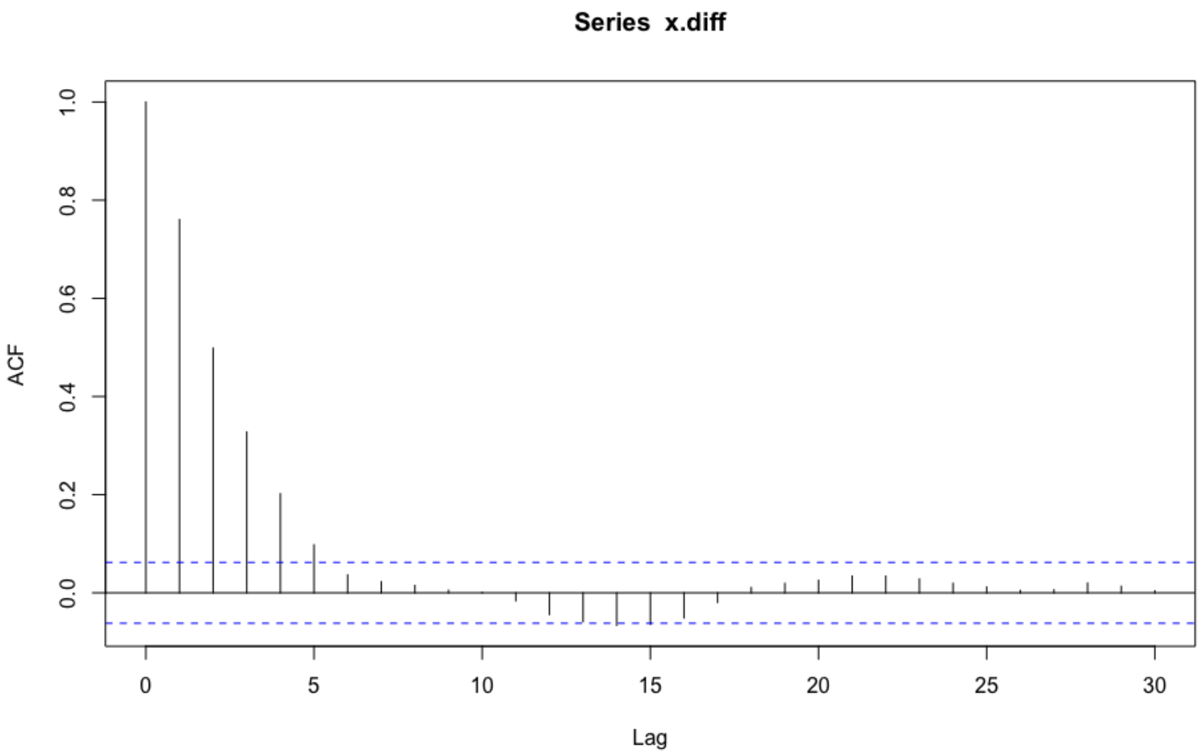


Figure 6:

Arrivals to Hawaii

Log Arrivals to Hawaii

==

==

Difference Arrivals to Hawaii

==

==

In R

Call

```
arima(x=arrivals, order=c(1,1,1), seasonal=list(order=c(1,1,1), frequency=12))
```

	ar1	ma1	sar1	sma1
Coefficients	0.1109	-0.9402	0.0192	-0.7039

	ar1	ma1	sar1	sma1
s.e.	0.0735	0.0239	0.1017	0.0829

sigmaˆ2 estimated as 224247640: log likelihood = -2377.43, aic = 4764.85

Density of Residuals

QQ Plot of Residuals

SARIMA Stationarity with B Operator

$$\begin{aligned}
z_t = \frac{1}{2}x_{t-1} + \frac{1}{4}x_{t-4} - \frac{1}{8}z_{t-5} + \\
w_t - \frac{1}{2}w_{t-1} - \frac{1}{2}w_{t-4} + \frac{1}{4}w_{t-5} \\
z_t - \frac{1}{2}x_{t-1} - \frac{1}{4}x_{t-4} + \frac{1}{8}z_{t-5} &= w_t - \frac{1}{2}w_{t-1} - \frac{1}{2}w_{t-4} + \frac{1}{4}w_{t-5} \\
z_t\left(1 - \frac{1}{2}\mathbf{B} - \frac{1}{4}\mathbf{B}^4 + \frac{1}{8}\mathbf{B}^5\right) &= w_t\left(1 - \frac{1}{2}\mathbf{B} - \frac{1}{2}\mathbf{B}^4 + \frac{1}{4}\mathbf{B}^5\right) \\
z_t\left(1 - \frac{1}{2}\mathbf{B}\right)\left(1 - \frac{1}{4}\mathbf{B}^4\right) &= w_t\left(1 - \frac{1}{2}\mathbf{B}\right)\left(1 - \frac{1}{2}\mathbf{B}^4\right)
\end{aligned}$$

This is an ARIMA(1,0,1)(1,0,1)$_4$ … or ARIMA(0,0,0)(1,0,1)$_4$?

(S)ARIMA Wrap-Up

PROS - Easily forecast nonstationary data in terms of original series - Data driven approach that is flexible - Typically outperforms econometric models inthe shortrun

CONS - (Possibly) Time Costly - Large amount of data is required - Especially true if seasonal - "Black Box" - No Economic interpretation - However, - system may be too complex to understand - Objective of user may be to do a forecast, and NOT determine causality.

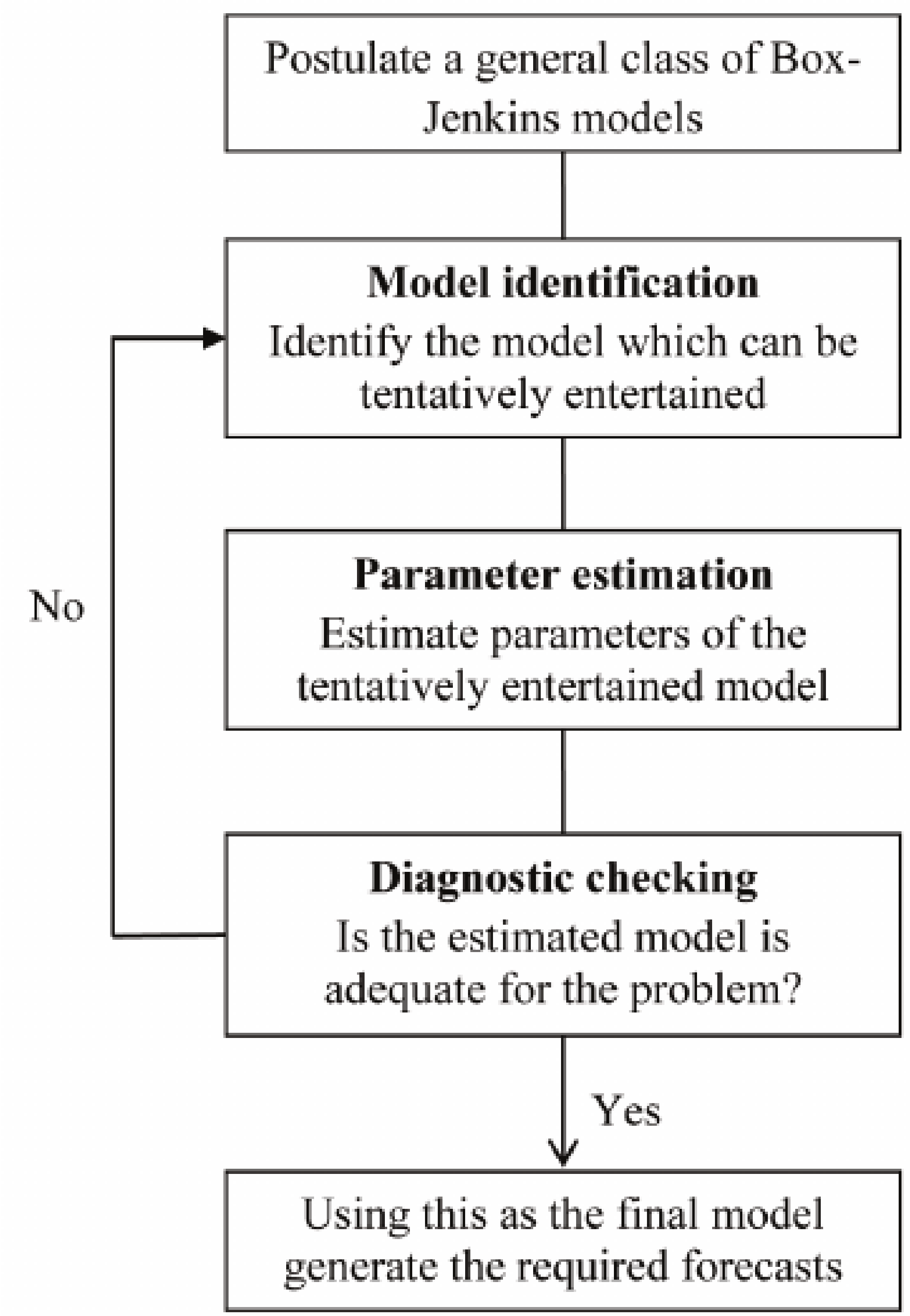

Postulate a general class of Box-Jenkins models
Model identification
Identify the model which can be tentatively entertained
Parameter estimation
Estimate parameters of the tentatively entertained model
Diagnostic checking
Is the estimated model is adequate for the problem?
No
Yes
Using this as the final model generate the required forecasts

Figure 7:

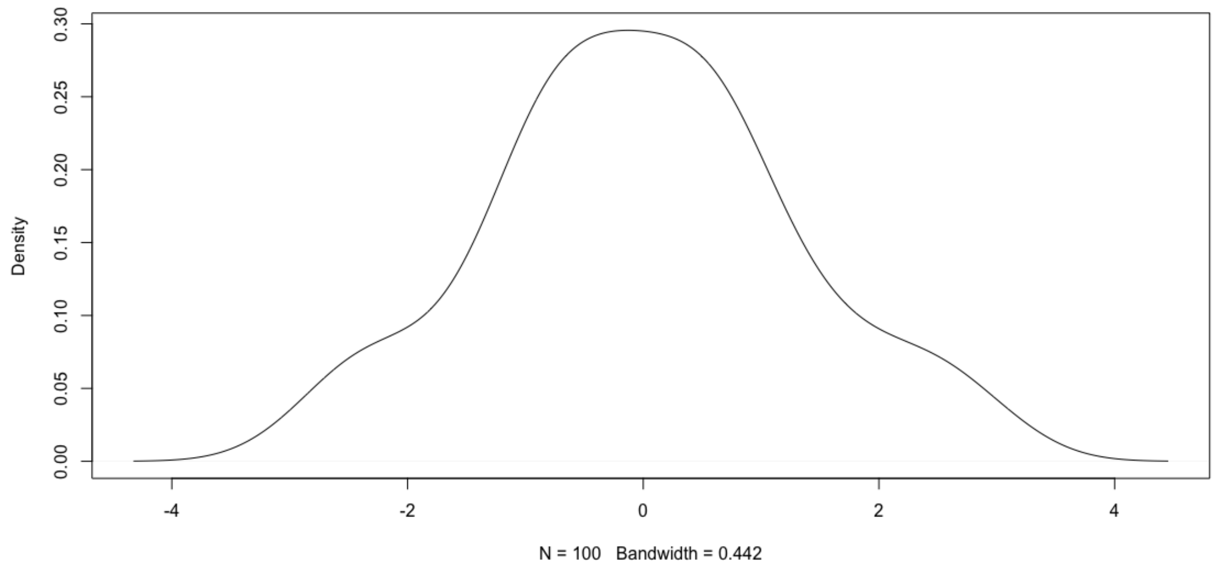


Figure 8:

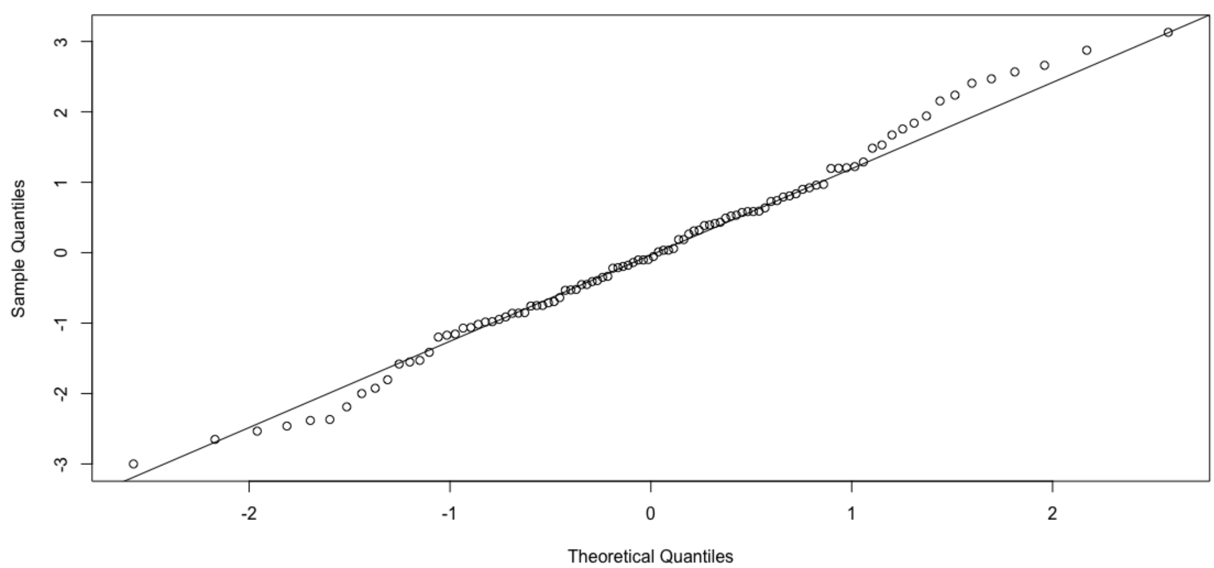


Figure 9:

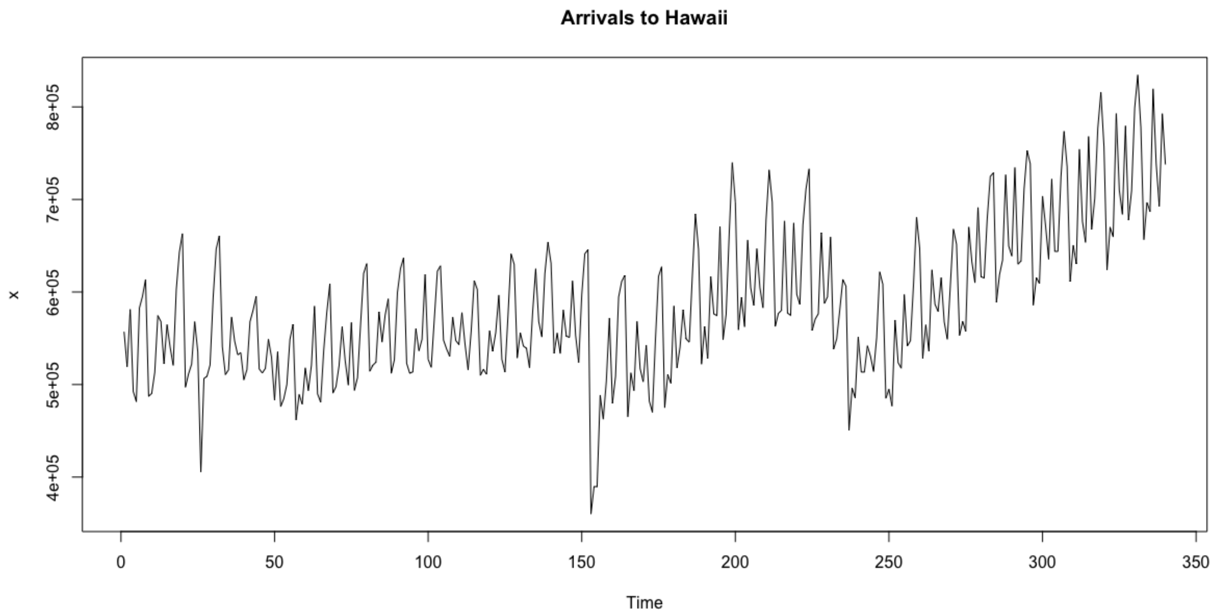


Figure 10:

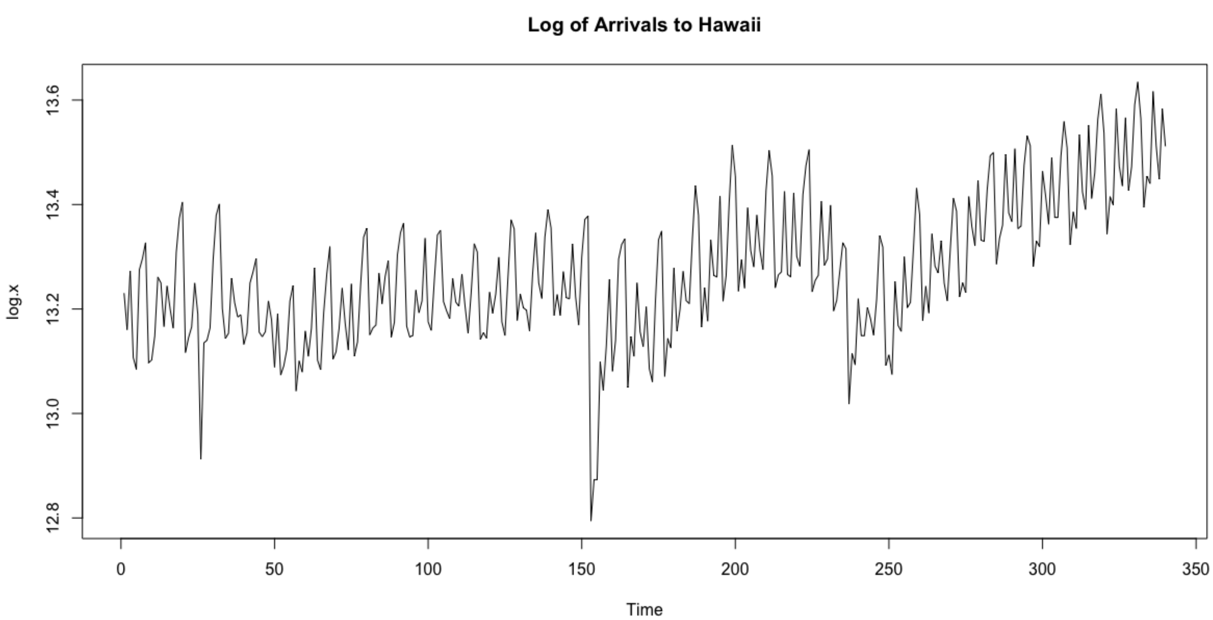


Figure 11:

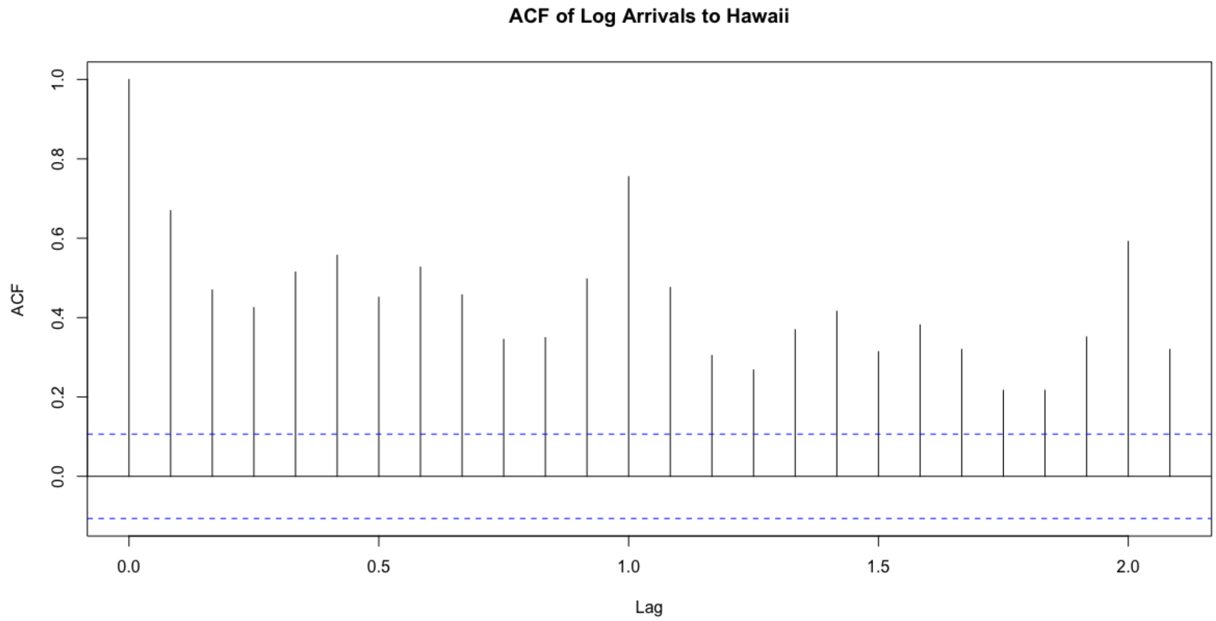


Figure 12:

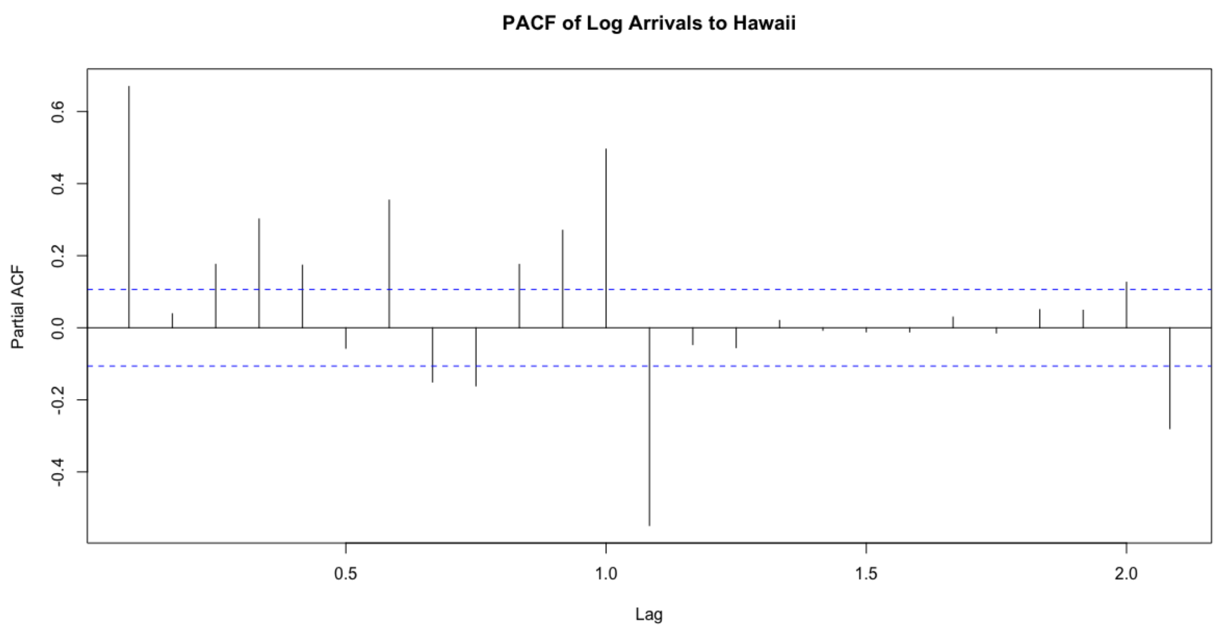


Figure 13:

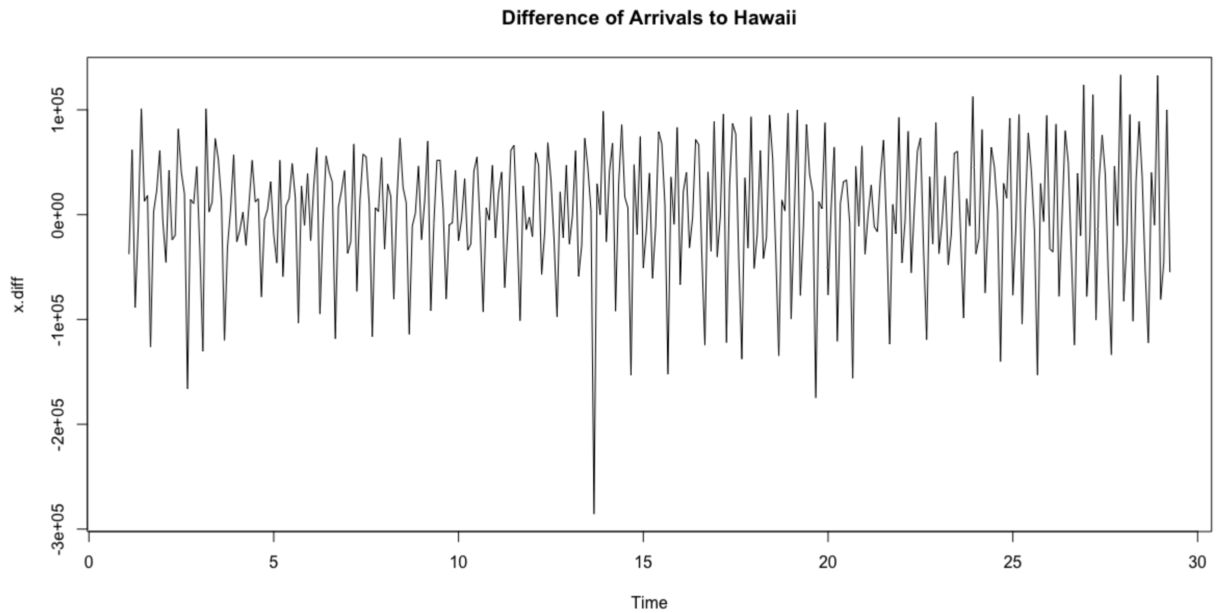


Figure 14:

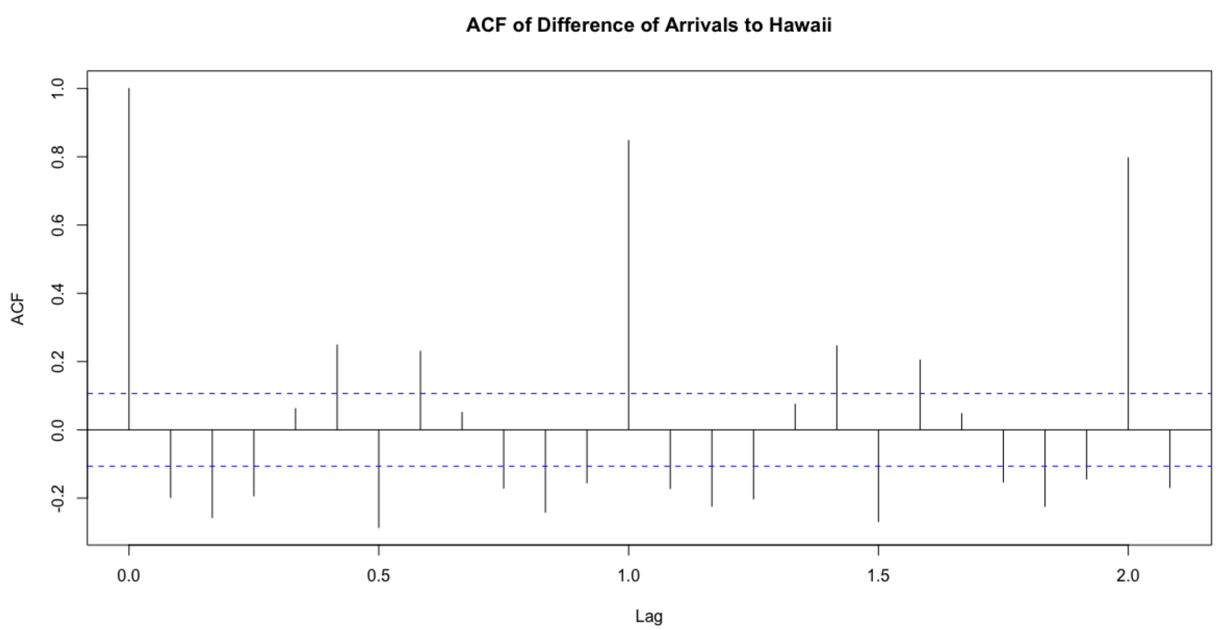


Figure 15:

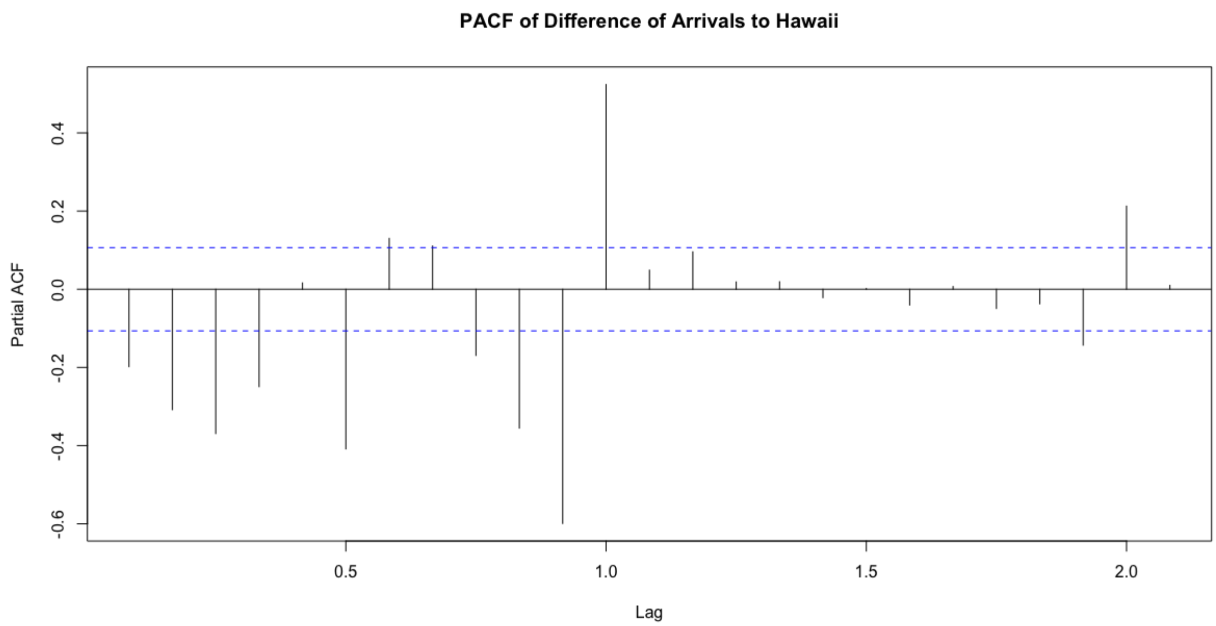


Figure 16:

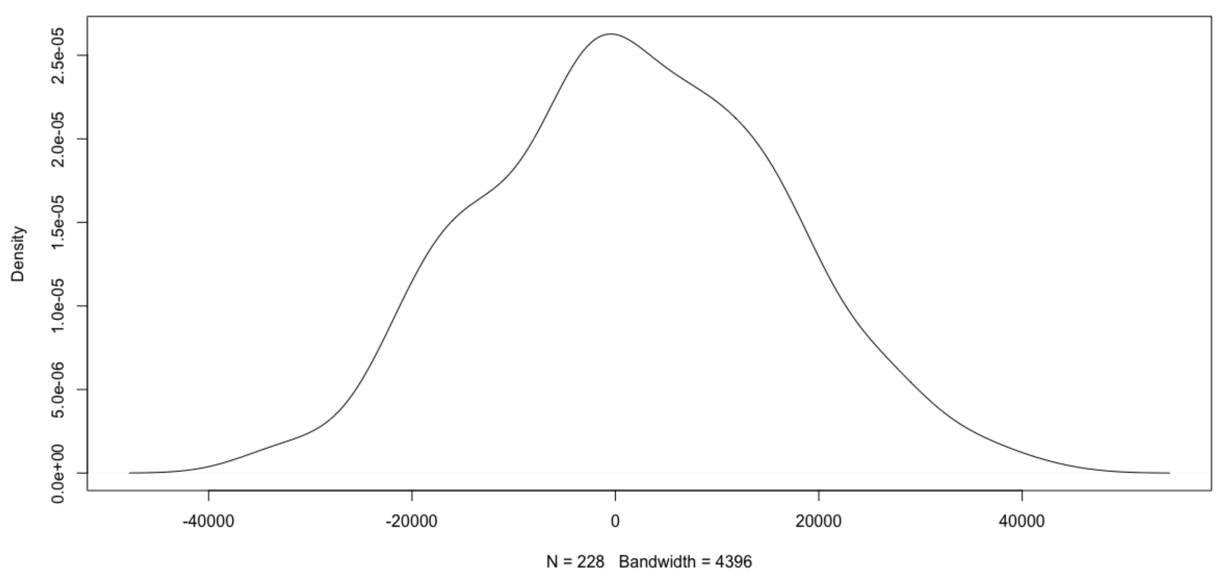
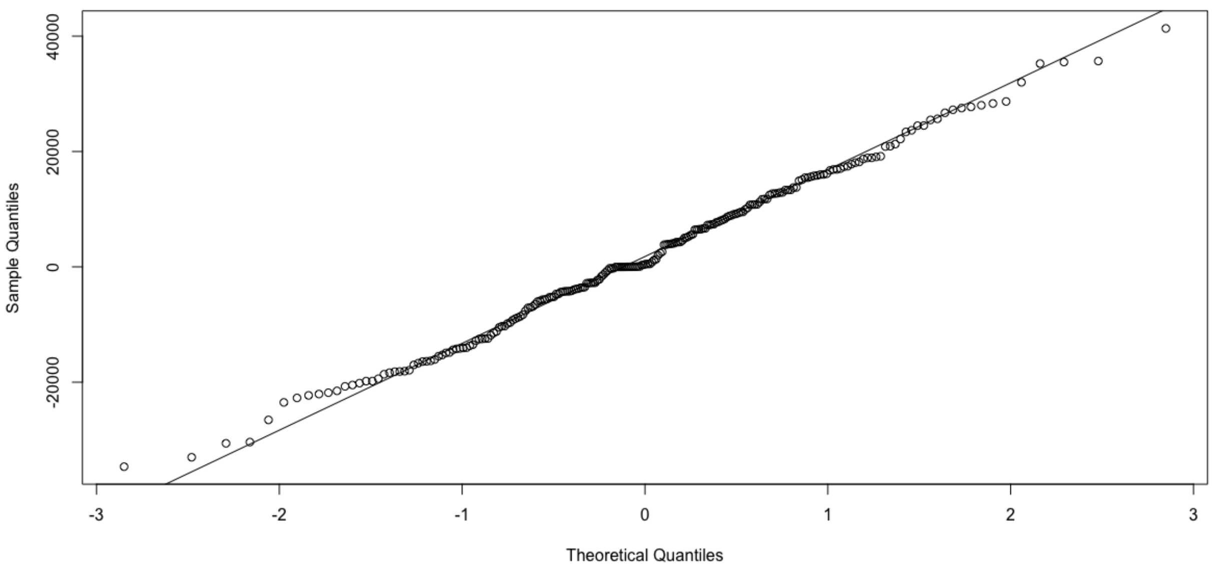


Figure 17:



40000
20000
0
-20000
Sample Quantiles
-3
-2
-1
0
1
2
3
Theoretical Quantiles

Figure 18: