Mu Niu's Final Project Report: Peking Yearly Average Temperature

Mu Niu

2023-06-05

Abstract

In this final project, I picked the temperature data of Peking, which is the capital of China, from 1821 to 2012. Then, I eliminated the trend and checked if the data needed to be transformed. After that, I looked at the ACF and PACF to find the potential parameters of the models. Then, I constructed an AICc matrix and used for loop to fill the matrix. After I got the AICc matrix, I selected two lowest AICc and fitted the corresponding parameters to the models to estimate parameters. Following, I did diagnostic checking and used the model that passed all tests to forecast values. I also built a 95% confidence interval for the forecasted values and compared it with the actual value.

The question that I addressed was finding a model to forecast the yearly temperature of Peking based on the data from 1821 to 2012. The key result is that our data follows the final model: $\nabla_1 X_t = (1 - 0.7186_{0.0742}B)Z_t$ with $\hat{\sigma}_Z^2 = 0.2371$. This model has a very low AICc value and the fewest parameters among the models I tried. It also passed all the tests, and the test set is within the prediction interval, so it's my final model.

Introduction

(1) Problem:

- Description: I planned to use the yearly temperature data of Peking from 1821-2010 to train a model and use this model to forecast the future yearly temperature of Peking.
- Techniques: The techniques that I used to solve the problem are: data tidying; transformation and differening; parameter identification based on ACF/PACF; model selection based on AICc; Estimate model parameters; Do diagnostic checking to select the final model; Use the final model to forecast future values.

(2) Data Set:

- Source(link): Climate Change: Earth Surface Temperature Data
- Description: This data set contains many files, and the file that I used in this project is named **GlobalLandTemperaturesByMajorCity.csv**, which has monthly observations of the Global Land Temperatures By Major City. This csv file has 7 variables and 239177 observations. The variables are date, average temperature, average temperature uncertainty, city, country, latitude, and longitude. I filtered observations on Peking, average temperature, and date since they are my main focus. Then, I dealt with the missing values and transformed the monthly data to yearly data since my problem is to find a model to forecast the yearly temperature of Peking.

- Why interested: I am interested in this data set since Peking is where I was born and grew up. Moreover, the temperature has been increasingly high and the citizens are all suffered from such a high temperature. Hence, I want to take this opportunity to address a problem that I have always thought about, which is to find a model to forecast the yearly average temperature of Peking in the future based on past observations.
- Software: The software that I used is RStudio

(3) Result and Conclusion:

- Result: I successfully built a model based on the data because the model passed all diagnostic checkings and the test set is within the 95% confidence interval. Since I successfully built a model and was able to use it to forecast the future yearly temperature of Peking, I have a positive result for my final project.
- Conclusion: The final model that the original data follows is $\nabla_1 X_t = (1 - 0.7186_{0.0742}B)Z_t$ with $\hat{\sigma}_Z^2 = 0.2371$.

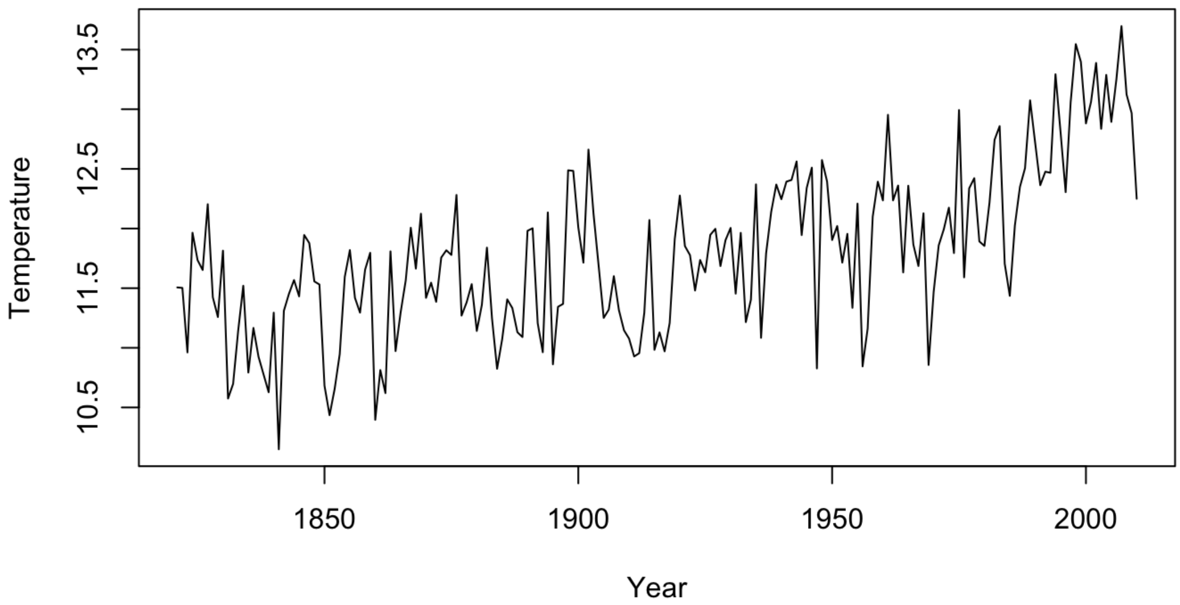
Data Preprocessing

Since I want to focus on the yearly average temperature of Peking, I first filtered the observations on Peking and removed all variables except the date and average temperature. Then, I found our that there are 14 missing values: 1832 October, 2013 September, all 12 months of 1838. For October of 1832, I filled the missing value by using the mean value of the neighbor months' average temperature. September of 2013 is the last observation in our data set, so I can't take the mean value of the neighbors' average temperature. What I did is to take the mean value of the two previous months' average temperature, since July, August, and September are in the same quarter of the year. For all 12 months of 1838, I temporarily filled them by 0 for convenience.

After solving the problems of missing values, I looked at my data and realized the data set starts in August of 1820 and ends in September of 2013, which means many consecutive months in those years are not included in our data. I only want to include years that have observations on all 12 months, so I removed observations on 1820 and 2013. Following, I convert the monthly data to yearly data by taking the average temperature for 12 months. After I got the yearly data, I filled in the yearly average temperature value of 1838 by the mean of 2 neighboring years' values.

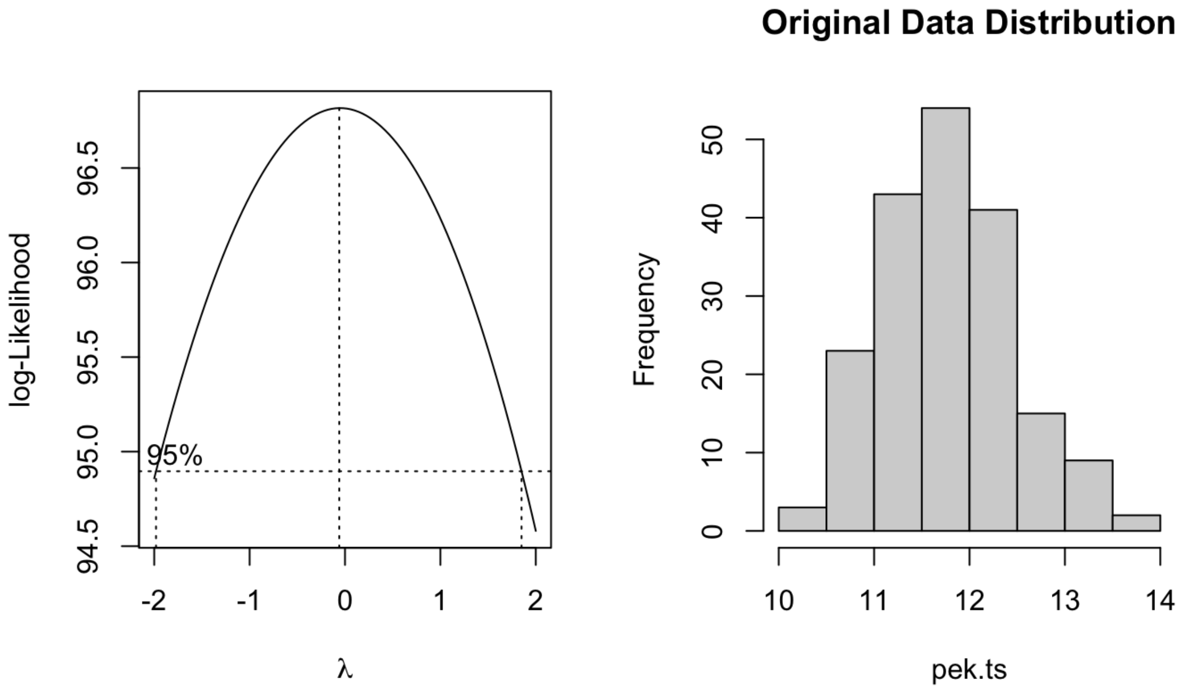
After the operations above, I have the tidied data on the yearly average temperature of Peking from 1821 to 2012. I set the data from 1821 to 2010 to be my model data, which means I will use this data to train my model. Moreover, I set the data from 2011 to 2012 to be my test set, which I will use to check my model.

Plot and Analyze



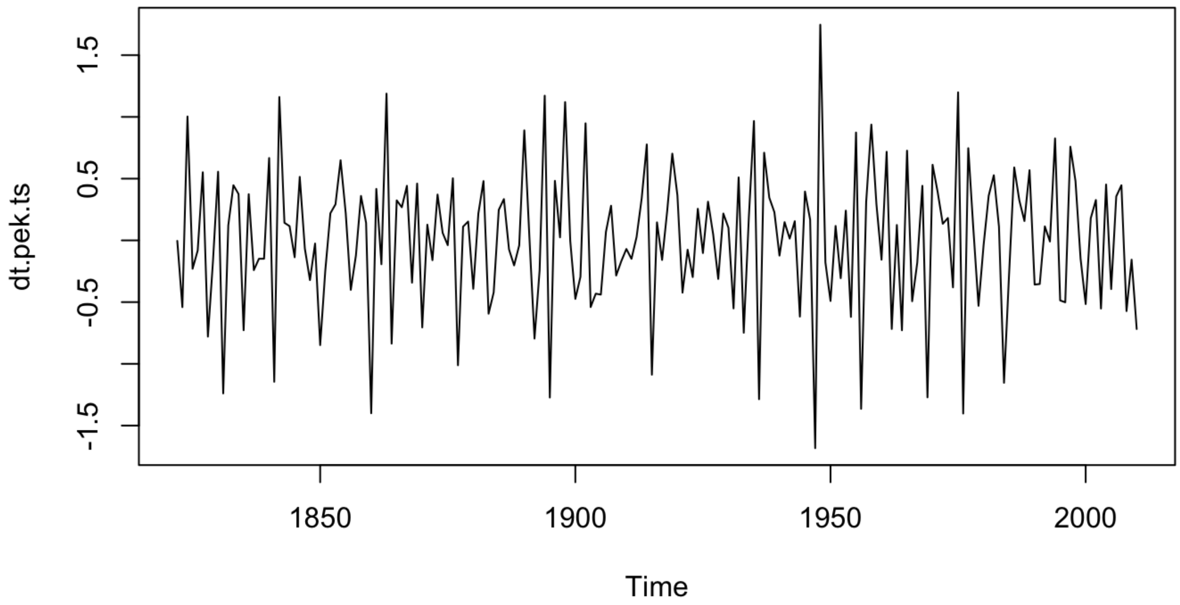
From the time series plot above, I noticed that there is an increasing trend but no seasonal component. Also, I noticed the change in trend. To be more specific, even though the overall trend of this graph is increasing, we can see that the data increased very slowly before 1950. However, the rate of increase goes up significantly after 1950. The potential reason that leads to this phenomenon is that the People's Republic of China was established in 1949, and China has started industrialization since then. Human activities, such as industrialization, released large amounts of carbon dioxide, which changed the earth's climate and increased the temperature.

Transformation and Differencing



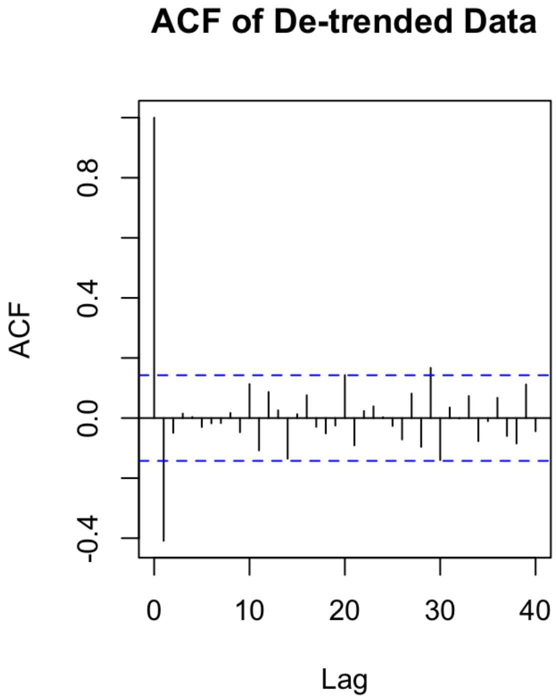
The plot on the left is the Box-Cox plot. It shows that 1 is within the 95% Confidence Interval for the value of λ. Also, the histogram on the right shows us the distribution of the original data. It appears to be normally distributed and has a bell shape. Hence, it's safe to conclude that a transformation is not necessary, and I decided not to transform the data.

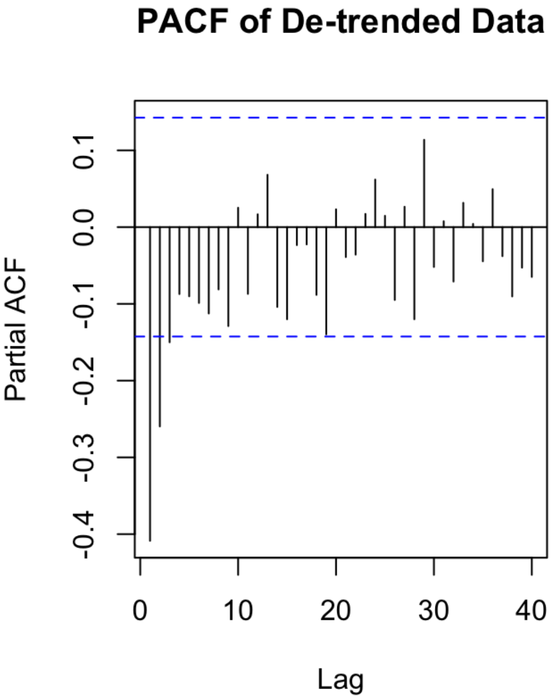
I differenced the data at lag one to eliminate the trend since I observed an increasing trend. But I don't have to difference the data at other lag since I didn't observe any seasonal component, and it's also not reasonable to have seasonality in yearly data. After I differenced the data once at lag 1, the variance decreased from 0.48 to 0.33. However, after I differenced twice at lag 1, the variance increased to 0.93, which means over differencing. Hence, I decided to only difference the data once at lag one.



The plot above is the data after differencing once at lag one, or the de-trended data. As we can see, there is no trend and the variance looks stable over time, so it's safe to conclude that my time series data is stationary now.

Preliminary Model Identification





The ACF plot shows that the ACF is statistically insignificant for lag greater than 1, and the PACF plot shows that the PACF is statistically insignificant for lag greater than 3. Also, we differenced data once at lag one. So the potential values of the parameters in my model are:

- p: 0,1,2,3
- d: 1
- q: 0,1

After knowing the potential values of the parameters, I run the for loop to try different models and compared the model fits using AICc:

	q		
p		0	1
	0	329.0586	270.5249
	1	296.4201	265.7672
	2	284.8366	265.8102
	3	281.9829	266.3851

I chose 2 models with lowest AICc from the table above: $ARIMA(1,1,1)$ with AICc = 265.7672 and $ARIMA(2,1,1)$ with AICc = 265.8102.

Coefficients Estimation and Diagnostic Checking

Coefficients Estimation

Let X_t denotes our original data,

- ARIMA(1,1,1): $(1 - 0.2373_{0.0963}B)\nabla_1 X_t = (1 - 0.8656_{0.0547}B)Z_t$ with $\hat{\sigma}_Z^2 = 0.2308$

Coefficients	ar1	ma1
value	0.2373	-0.8656
s.e.	0.0963	0.05471

- ARIMA(2,1,1): $(1 - 0.2519_{0.0890}B - 0.0870_{0.0839}B^2)\nabla_1 X_t = (1 - 0.8949_{0.0479}B)Z_t$ with $\hat{\sigma}_Z^2 = 0.2295$

Coefficients	ar1	ar2	ma1
value	0.2519	0.0870	-0.8949
s.e.	0.0890	0.0839	0.0479

However, looking at the table, I noticed that the 95% confidence interval of ϕ_2, which is (-0.0808, 0.2548), contains zero. Hence, it's better to set $\phi_2 = 0$. But doing so will give us $ARIMA(1,1,1)$, which we already have. So I decided to go back to the AICc matrix and find another model with low AICc.

	q		
p		0	1
	0	329.0586	270.5249
	1	296.4201	265.7672
	2	284.8366	265.8102
	3	281.9829	266.3851

From the table, I found out that $ARIMA(3,1,1)$ and $ARIMA(0,1,1)$ also have very low AICc in comparison with other models. So I decided to fit the data into $ARIMA(0,1,1)$ with AICc = 270.5249 and $ARIMA(3,1,1)$ with AICc = 266.3851.

- ARIMA(0,1,1): $\nabla_1 X_t = (1 - 0.7186_{0.0742}B)Z_t$ with $\hat{\sigma}_Z^2 = 0.2371$

Coefficients	ma1
value	-0.7186
s.e.	0.0742

- ARIMA(3,1,1): $(1 - 0.2641_{0.0865}B - 0.0851_{0.0818}B^2 - 0.0672_{0.0800}B^3)\nabla_1 X_t = (1 - 0.9118_{0.0443}B)Z_t$ with $\hat{\sigma}_Z^2 = 0.2286$

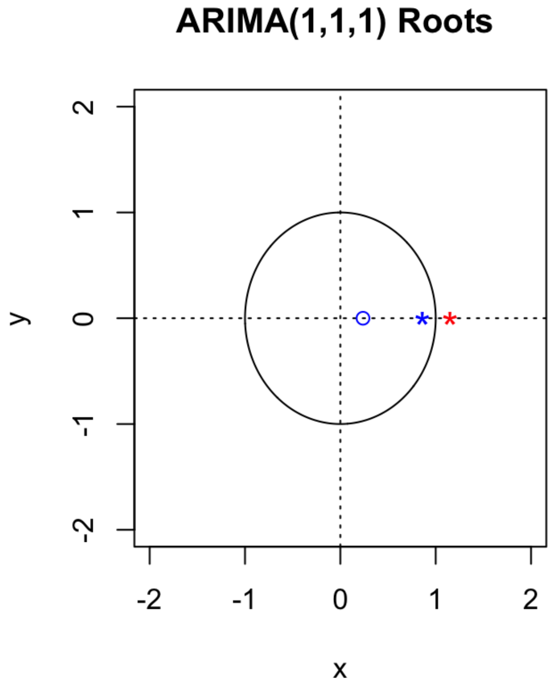
Coefficients	ar1	ar2	ar3	ma1
value	0.2641	0.0851	0.0672	-0.9118
s.e.	0.0865	0.0818	0.0800	0.0443

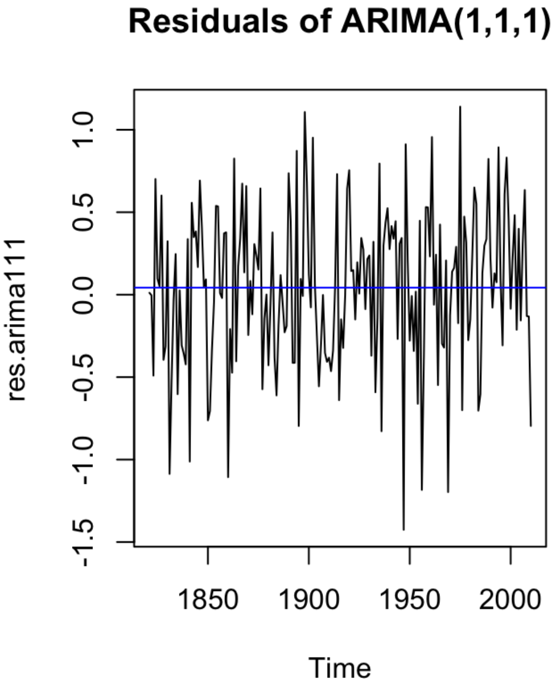
This $ARIMA(3,1,1)$ has same issue as $ARIMA(2,1,1)$. That is, the 95% confidence interval of ϕ_2, which is (-0.0785, 0.2487), contains zero and the 95% confidence interval of ϕ_3, which is (-0.0928, 0.2272), contains zero as well. Hence, it's better to set $\phi_2 = \phi_3 = 0$. But doing so will give us $ARIMA(1,1,1)$, which we already have. Thus, I decided to only take $ARIMA(1,1,1)$ and $ARIMA(0,1,1)$ to diagnostic checking.

Diagnostic Checking

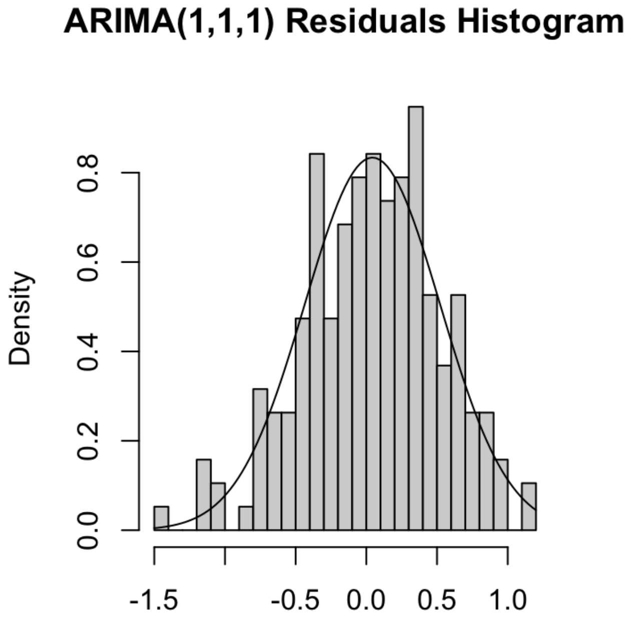
- For $ARIMA(1,1,1)$:

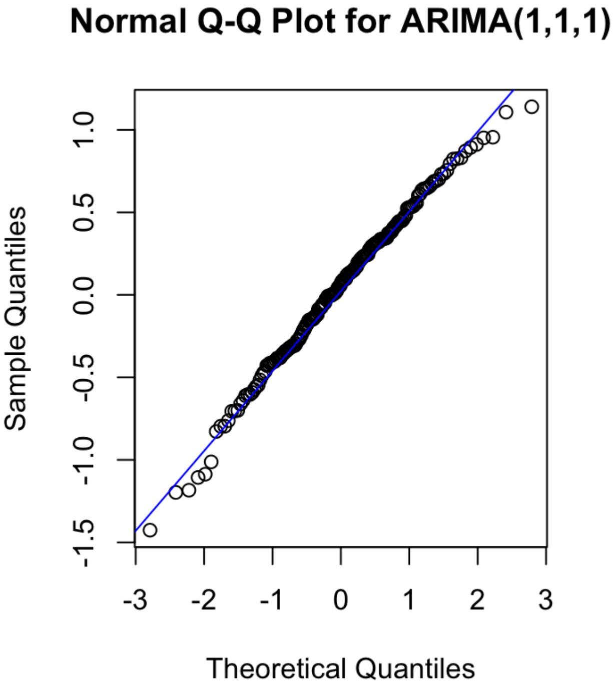
$(1 - 0.2373_{0.0963}B)\nabla_1 X_t = (1 - 0.8656_{0.0547}B)Z_t$ with $\hat{\sigma}_Z^2 = 0.2308$



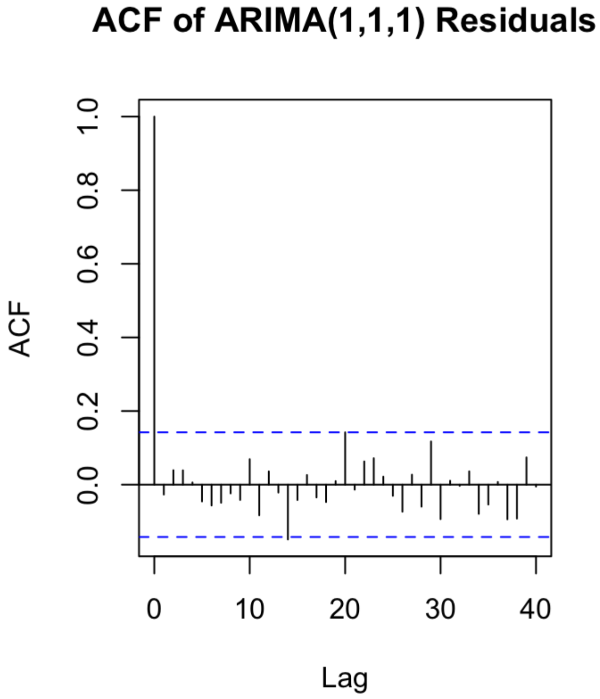


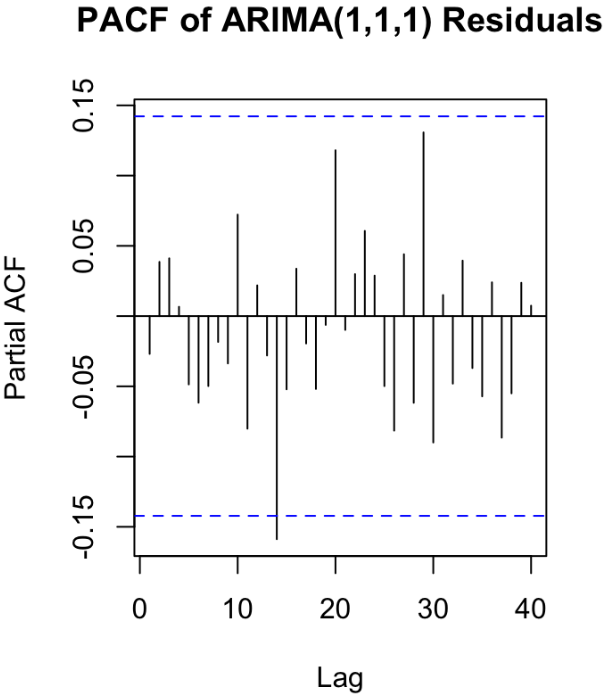
From the plots above, I found out that the roots of both Auto Regressive characteristic polynomial and Moving Average characteristic polynomial are outside of the unit circle. Also, I noticed that $|\theta_1| < 1$ and $|\phi_1| < 1$, so it's safe to conclude that this model is both stationary and invertible. The plot of the residuals shows no trend, no visible change of variance, no seasonality, and the sample mean is almost zero.


ARIMA(1,1,1) Residuals Histogram


Normal Q-Q Plot for ARIMA(1,1,1)

The histogram above indicates that the residuals appear to be normally distributed since it has a bell shape. Moreover, in the Q-Q plot, more than 95% of the points are between -2 and 2, which again indicates the residuals follow normal distribution.


ACF of ARIMA(1,1,1) Residuals


PACF of ARIMA(1,1,1) Residuals

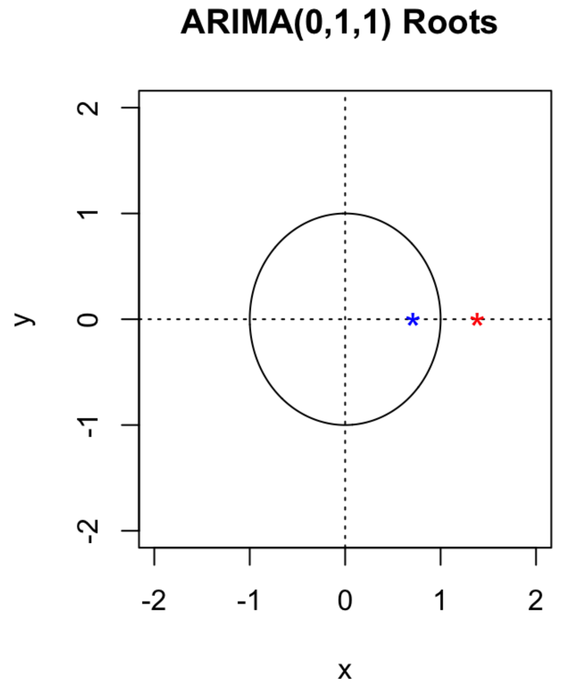
From the plots above, I found out that all the ACF of residuals are within the confidence interval, which can be counted as zeros. However, I noticed that PACF at lag 14 is slightly out of the confidence interval, and I decided to see the results of other tests. It doesn't matter if the residuals of this model pass all other tests.

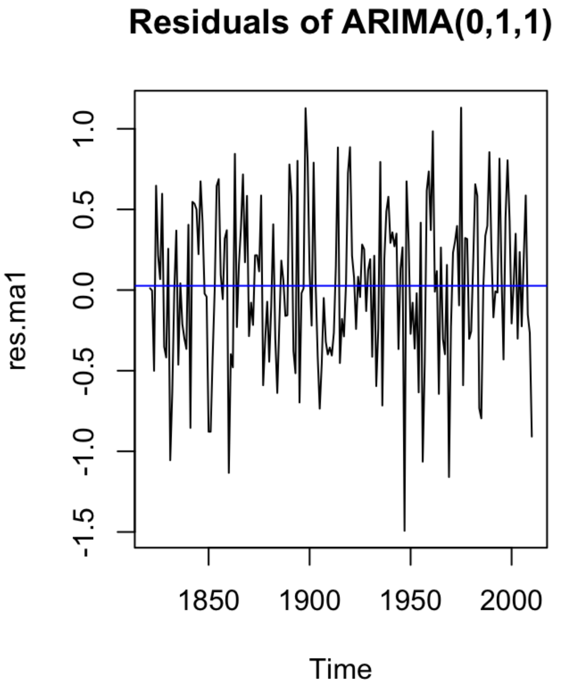
Tests	P-value/ Test Result
Shapiro-Wilk Normality Test	0.4014
Box-Pierce Test	0.6692
Ljung-Box Test	0.6104
Mcleod-Li Test	0.3185
AR Order Select	0

The table above shows the P-value of the Shapiro-Wilk normality test, Box-Pierce test, Ljung-Box test, and Mcleod-Li test. This model passed all these tests since the P-values were all larger than 0.05. Moreover, I plugged the residuals from this model into the Yule-Walker method, and it automatically selected 0, which means white noise. Since the residuals of the $ARIMA(1,1,1)$ model passed all the tests, it's safe to say that PACF at lag 14 slightly out of the confidence interval doesn't matter and we can move on to the next model.

- For $ARIMA(0,1,1)$:

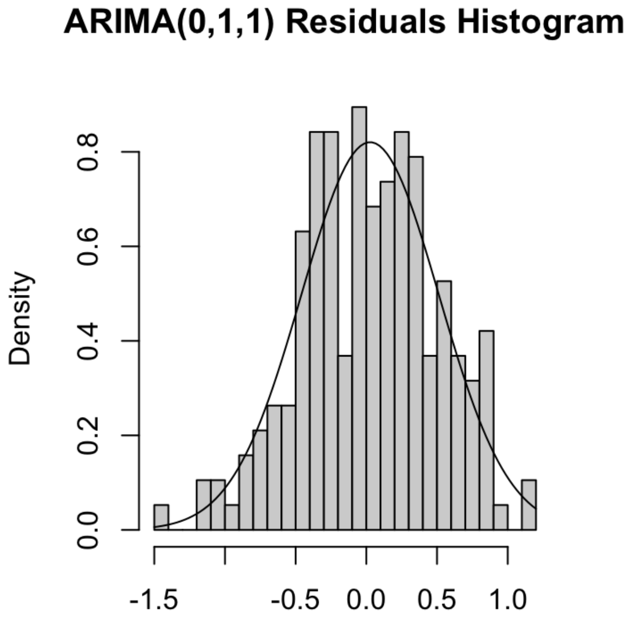
$\nabla_1 X_t = (1 - 0.7186_{0.0742}B)Z_t$ with $\hat{\sigma}_Z^2 = 0.2371$

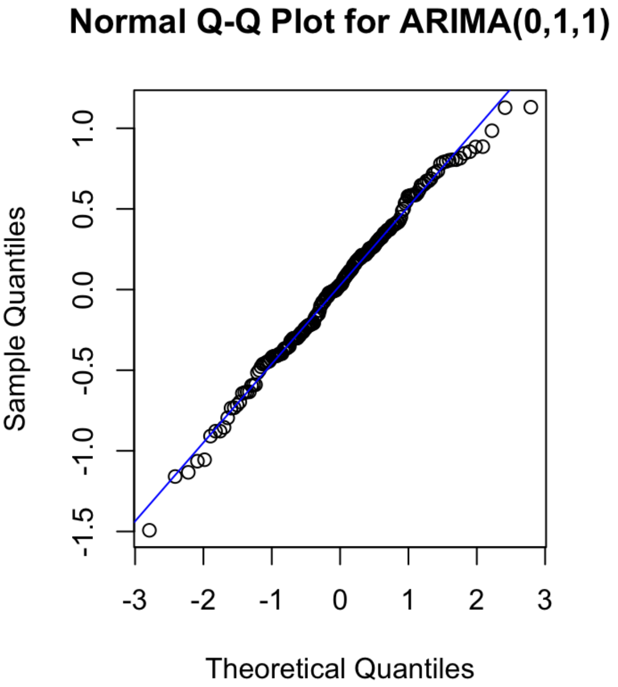




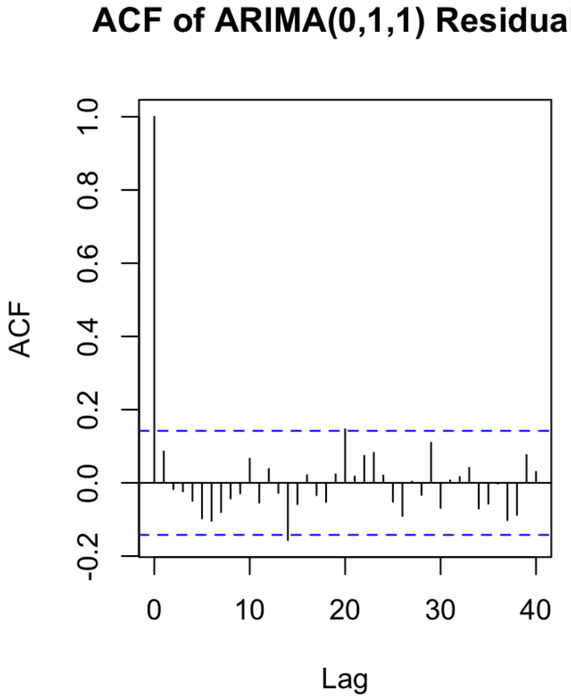
From the plots above, I found out that the root of the characteristic polynomial are outside of the unit circle. Also, I noticed that $|\theta_1| < 1$, so it's safe to conclude that this model is both stationary and invertible.

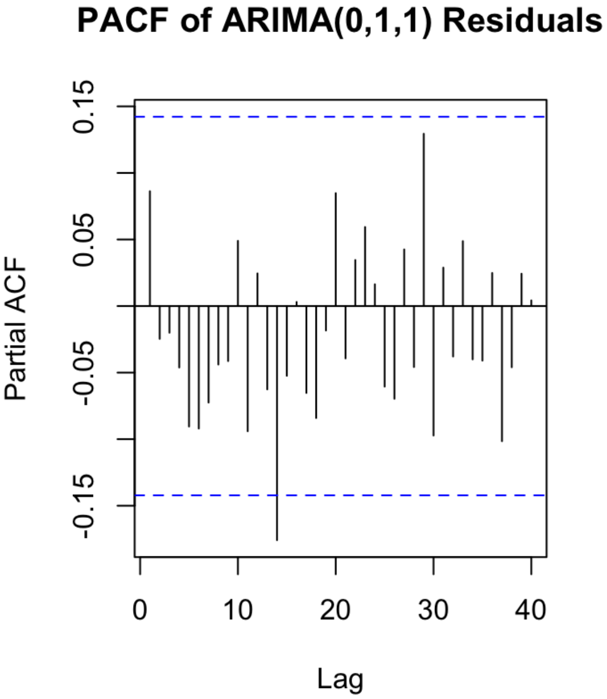
The plot of the residuals shows no trend, no visible change of variance, no seasonality, and the sample mean is almost zero.


ARIMA(0,1,1) Residuals Histogram


Normal Q-Q Plot for ARIMA(0,1,1)

The histogram above indicates that the residuals appear to be normally distributed since it has a bell shape. Moreover, in the Q-Q plot, more than 95% of the points are between -2 and 2, which again indicates the residuals follow normal distribution.





From the plots above, I found out that all the ACF of residuals are within the confidence interval, which can be counted as zeros. However, I noticed that PACF at lag 14 is slightly out of the confidence interval, and I decided to see the results of other tests. It doesn't matter if the residuals of this model pass all other tests.

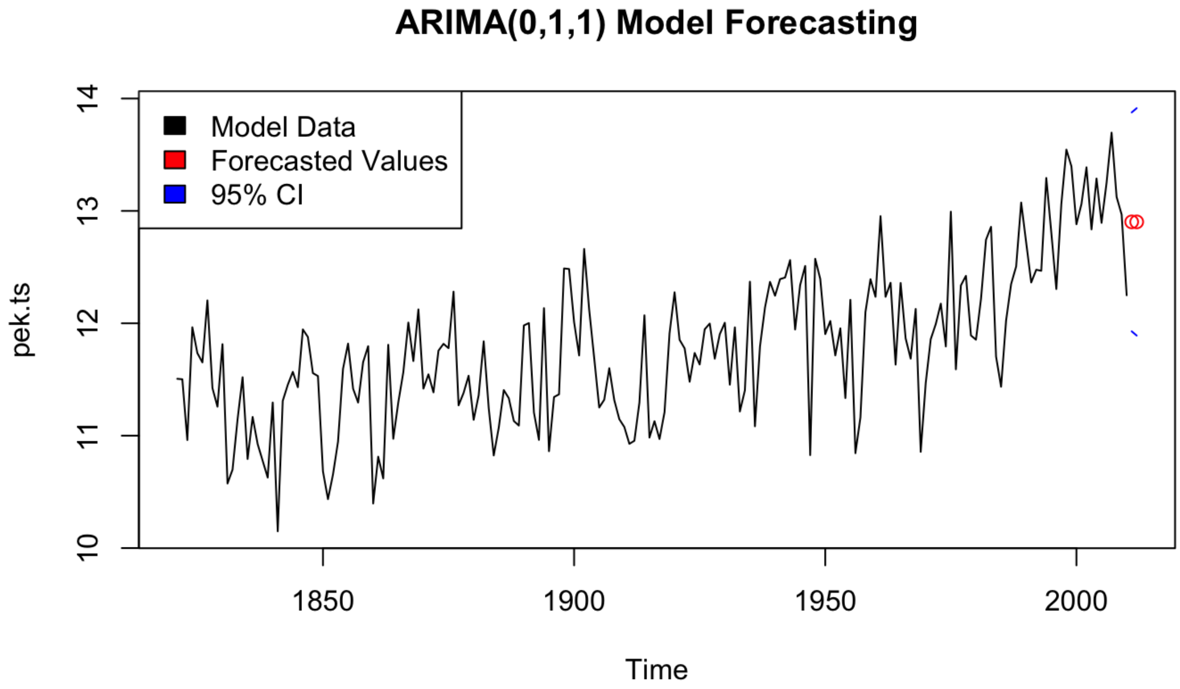
Tests	P-value/ Test Result
Shapiro-Wilk Normality Test	0.4891
Box-Pierce Test	0.3696
Ljung-Box Test	0.3136
Mcleod-Li Test	0.4206
AR Order Select	0

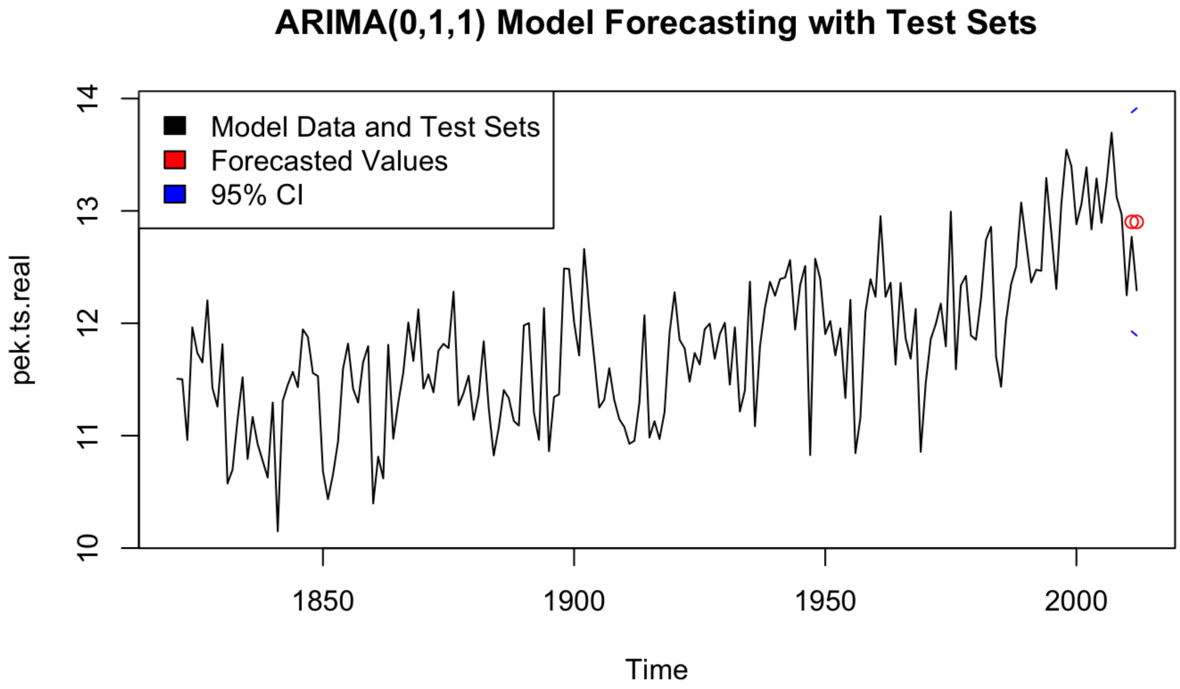
The table above shows the P-value of the Shapiro-Wilk normality test, Box-Pierce test, Ljung-Box test, and Mcleod-Li test. This model passed all these tests since the P-values were all larger than 0.05. Moreover, I plugged the residuals from this model into the Yule-Walker method, and it automatically selected 0, which means white noise. Since the residuals of the $ARIMA(0,1,1)$ model passed all the tests, it's safe to say that PACF at lag 14 slightly out of the confidence interval doesn't matter.

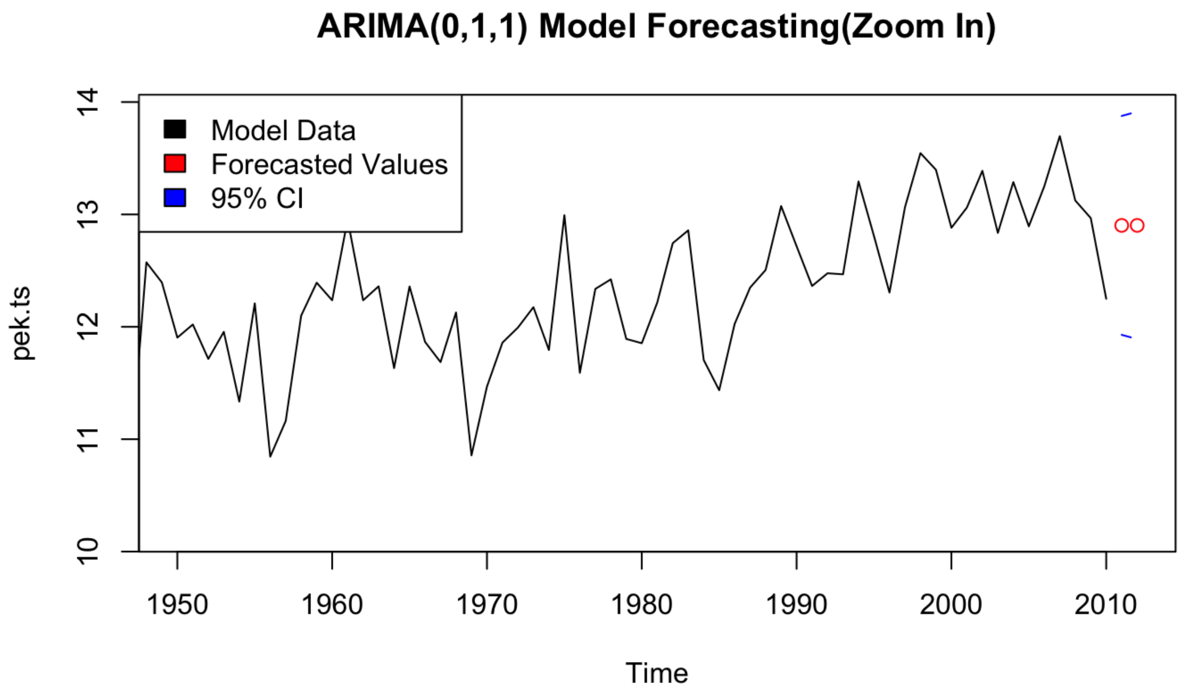
- Final Model Selection:

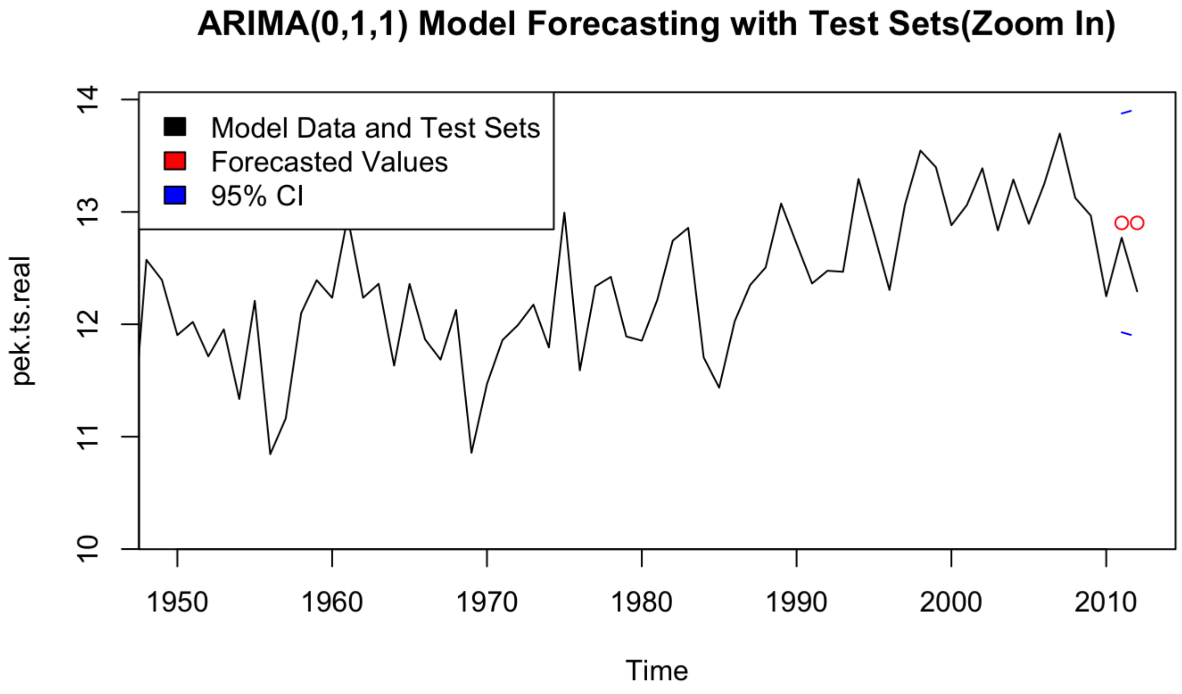
As we can see, both models passed all the tests. According to the principle of parsimony, I should choose the one with the least coefficients, which is the $ARIMA(0,1,1)$ model.This final model obtained by using AICc and the principle of parsimony is one of the models suggested by ACF and PACF. Since my final model passed all the tests on residuals of the model, I can conclude that my final model is satisfactory. Finally, my final model in algebraic form is: $\nabla_1 X_t = (1 - 0.7186_{0.0742} B) Z_t$ with $\hat{\sigma}_Z^2 = 0.2371$.

Forecasting









The first two plots include the original series, forecasted values using my final model, and the confidence interval for the forecasted values. I also included two zoom-in plots of the first two plots to be able to see the detail clearly. From the plots, I noticed that the test sets are within the confidence intervals, which means this model is satisfactory and I have successfully addressed my problem.

Conclusion

In conclusion, I successfully achieved my goal of training a model to forecast the yearly temperature of Peking. The math formula for my final model is $\nabla_1 X_t = (1 - 0.7186_{0.0742} B) Z_t$. I used the final model for forecasting, and the test sets are within the prediction intervals, which again confirmed that my model is satisfactory. Hence, I successfully achieved my goal.

Professor Feldman strongly supported me in this final project. Her lectures and slides built me a solid foundation in time series knowledge. More importantly, when I encountered detailed problems, I could always schedule office hour appointments with her and discuss the problems.

References

Berkeley, Earth. 1 May, 2017. *"Climate Change: Earth Surface Temperature Data."* Kaggle. www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data.

Feldman, Raisa. 2023. *Lecture 15 Slides-Example of a Project.*

Appendix

```r
# Data Tidying
raw <- read.csv('../data/GlobalLandTemperaturesByMajorCity.csv') # Read in the data
pek <- raw[raw$City == "Peking",] # Filter observations on Peking
pek.temp <- pek[,1:2] # Filter variables: only date and average temperature
sum(is.na(pek.temp)) # Total number of missing values
pek.temp[is.na(pek.temp$AverageTemperature) == TRUE,] # Distribution of missing values
pek.temp["176394","AverageTemperature"] <- (pek.temp["176395","AverageTemperature"]+
                                              pek.temp["176393","AverageTemperature"]) /2
# October 1832: mean of neighbor months' average temperature
pek.temp["178565","AverageTemperature"] <- (pek.temp["178563","AverageTemperature"]+
                                              pek.temp["178564","AverageTemperature"])/2
#September 2013: mean of previous two months' average temperature
for (index in as.character(seq(176457,176468))) {
  pek.temp[index, "AverageTemperature"] <- 0
} # 1838 all year fill 0 first
sum(is.na(pek.temp)) # Check missing values after the operations above
pek.temp <- pek.temp[as.character(seq(176253,178556)),]
# Remove 1820 and 2013 since missing part of the year
year.temp <- c()
for (jan in seq(1,2304,12)){
  dec <- jan+11
  year.temp <- c(year.temp,mean(pek.temp[jan:dec,2]))
} # Convert monthly data to yearly data by take the mean of 12 months
Year <- c(1821:2012) # Year list
data <- data.frame(Year = Year,
                   AverageTemperature = year.temp) # Construct a data frame of tidied yearly data
data[data$Year == 1838, "AverageTemperature"] <- (data[data$Year == 1837, "AverageTemperature"] +
                                                    data[data$Year ==1839,"AverageTemperature"])/2
# Initially, we set 1838 all year to be 0 for convenience.
# Now, after we got yearly data, we fill the value of 1838 by the mean of neighbor years
model.data <- data[Year <= 2010, ] # Model data set: 190 observations
test.data <- data[Year > 2010, ] # Test set: 2 observations
```

```r
# Plot and Analyze
pek.ts <- ts(model.data[,2], start = 1821, frequency = 1) # Transform to time series object
ts.plot(pek.ts, gpars=list(xlab="Year", ylab="Temperature")) # Plot the time series

# Transformation and Differencing
op <- par(mfrow=c(1,2)) # Box-Cox and Histogram to see if transformation is necessary
library(MASS)
t = 1:length(pek.ts)
fit = lm(pek.ts ~ t)
bcTransform = boxcox(pek.ts ~ t,plotit = T) # See what lambda value boxcox gives us
hist(pek.ts, main = 'Original Data Distribution')# Check the distribution of original data
par(op)

var(pek.ts) # Variance before Diff
dt.pek.ts <- diff(pek.ts, 1)
(var(dt.pek.ts)) # Variance after diff once at lag 1
dt2.pek.ts <- diff(dt.pek.ts, 1)
(var(dt2.pek.ts)) # Variance after diff twice at lag 1
plot(dt.pek.ts) # Plot of the de-trended data

# Preliminary Model Identification
op <- par(mfrow=c(1,2))
acf(dt.pek.ts, lag.max = 40, main = 'ACF of De-trended Data') # ACF
pacf(dt.pek.ts, lag.max = 40,  main = 'PACF of De-trended Data') # PACF
par(op)
library(qpcR)
# Construct Matrix
aiccs <- matrix(NA, nr = 4, nc = 2)
dimnames(aiccs) = list(p=0:3, q=0:1)
# Use for loop to calculate the AICc matrix
for(p in 0:3) {
  for(q in 0:1) {
    aiccs[p+1,q+1] = AICc(arima(dt.pek.ts, order = c(p,0,q), method="ML"))
  } }

# Coefficients Estimation
(arima111 <- arima(pek.ts, order=c(1,1,1), method="ML")) # Fitting ARIMA(1,1,1) model
(arima211 <- arima(pek.ts, order=c(2,1,1), method="ML")) # Fitting ARIMA(2,1,1) model
(ma1 <- arima(pek.ts, order=c(0,1,1), method="ML")) # Fitting ARIMA(0,1,1) model
(arima311 <- arima(pek.ts, order=c(3,1,1), method="ML")) # Fitting ARIMA(3,1,1) model

# Diagnostic Checking
# For ARIMA (1,1,1)
op <- par(mfrow = c(1,2))
res.arima111 <- residuals(arima111)
source('../scratch/plot.roots.R')
plot.roots(polyroot(c(1, -0.2373)),polyroot(c(1, -0.8656)), main="ARIMA(1,1,1) Roots ")
plot.ts(res.arima111, main = 'Residuals of ARIMA(1,1,1)')
abline(h=mean(res.arima111), col="blue")
hist(res.arima111, breaks = 20, xlab="", prob=TRUE,
     main = 'ARIMA(1,1,1) Residuals Histogram')
m.arima111 <- mean(res.arima111)
std.arima111 <- sqrt(var(res.arima111))
```

```
curve(dnorm(x,m.arima111,std.arima111), add=TRUE )
qqnorm(res.arima111,main= "Normal Q-Q Plot for ARIMA(1,1,1)")
qqline(res.arima111,col="blue")
acf(res.arima111, lag.max=40, main = 'ACF of ARIMA(1,1,1) Residuals')
pacf(res.arima111, lag.max=40, main = 'PACF of ARIMA(1,1,1) Residuals')
par(op)
shapiro.test(res.arima111)
Box.test(res.arima111, type=c("Box-Pierce"), lag = 14, fitdf = 2)
Box.test(res.arima111, type=c("Ljung-Box"), lag = 14, fitdf = 2)
Box.test((res.arima111)^2, type=c("Ljung-Box"), lag = 14, fitdf = 0)
ar(res.arima111, aic = TRUE, order.max = NULL, method = c("yule-walker"))
# For ARIMA(0,1,1)
op <- par(mfrow = c(1,2))
res.ma1 <- residuals(ma1)
source('plot.roots.R')
plot.roots(NULL, polyroot(c(1, -0.7186)), main="ARIMA(0,1,1) Roots ")
plot.ts(res.ma1, main='Residuals of ARIMA(0,1,1)')
abline(h=mean(res.ma1), col="blue")
hist(res.ma1, breaks = 20, xlab="", prob=TRUE,
     main = 'ARIMA(0,1,1) Residuals Histogram')
m.ma1 <- mean(res.ma1)
std.ma1 <- sqrt(var(res.ma1))
curve(dnorm(x,m.ma1,std.ma1), add=TRUE )
qqnorm(res.ma1,main= "Normal Q-Q Plot for ARIMA(0,1,1)")
qqline(res.ma1,col="blue")
acf(res.ma1, lag.max=40, main = 'ACF of ARIMA(0,1,1) Residuals')
pacf(res.ma1, lag.max=40, main = 'PACF of ARIMA(0,1,1) Residuals')
par(op)
shapiro.test(res.ma1)
Box.test(res.ma1, type=c("Box-Pierce"), lag = 14, fitdf = 1)
Box.test(res.ma1, type=c("Ljung-Box"), lag = 14, fitdf = 1)
Box.test((res.ma1)^2, type=c("Ljung-Box"), lag = 14, fitdf = 0)
ar(res.ma1, aic = TRUE, order.max = NULL, method = c("yule-walker"))

# Forecasting
library(forecast)
fit <- arima(pek.ts, order=c(0,1,1), method="ML") # Fit into model and forecast
forecast(fit)
pred<- predict(fit, n.ahead = 2) # To produce graph with 2 forecast on data:
U= pred$pred + 2*pred$se # Upper bound of prediction interval
L= pred$pred - 2*pred$se # Lower bound of prediction interval
# Plot the forecast value for 2011 and 2012 with 95% CI on original data
ts.plot(pek.ts, xlim=c(1821,2012), ylim = c(min(pek.ts),max(U)),
        main = 'ARIMA(0,1,1) Model Forecasting')
lines(2011:2012, y =  U, col="blue", lty=2)
lines(2011:2012, y =  L, col="blue", lty=2)
points(2011:2012, pred$pred, col="red")
legend("topleft",
       legend = c('Model Data', 'Forecasted Values', '95% CI'),
       fill = c('black','red','blue'),
       border = "black")
# Plot the forecast value for 2011 and 2012 with 95% CI on original data with real values
pek.ts.real <- ts(data[,2], start = 1821, frequency = 1)
```

```r
ts.plot(pek.ts.real, xlim = c(1821,2012), ylim = c(min(pek.ts),max(U)),
        main = 'ARIMA(0,1,1) Model Forecasting with Test Sets')
lines(U, col="blue", lty="dashed")
lines(L, col="blue", lty="dashed")
points(2011:2012, pred$pred, col="red")
legend("topleft",
       legend = c('Model Data and Test Sets', 'Forecasted Values', '95% CI'),
       fill = c('black','red','blue'),
       border = "black")
# Zoom in version
# Plot the forecast value for 2011 and 2012 with 95% CI on original data: Zoom In
ts.plot(pek.ts, xlim=c(1950,2012), ylim = c(min(pek.ts),max(U)),
        main = 'ARIMA(0,1,1) Model Forecasting(Zoom In)')
lines(2011:2012, y =  U, col="blue", lty=2)
lines(2011:2012, y =  L, col="blue", lty=2)
points(2011:2012, pred$pred, col="red")
legend("topleft",
       legend = c('Model Data', 'Forecasted Values', '95% CI'),
       fill = c('black','red','blue'),
       border = "black")
# Plot the forecast value for 2011 and 2012 with 95% CI on original data with real values: Zoom In
ts.plot(pek.ts.real, xlim = c(1950,2012), ylim = c(min(pek.ts),max(U)),
        main = 'ARIMA(0,1,1) Model Forecasting with Test Sets(Zoom In)')
lines(U, col="blue", lty="dashed")
lines(L, col="blue", lty="dashed")
points(2011:2012, pred$pred, col="red")
legend("topleft",
       legend = c('Model Data and Test Sets', 'Forecasted Values', '95% CI'),
       fill = c('black','red','blue'),
       border = "black")
```