Robin Phetsavong
5113014
PSTAT 174
December 10, 2013

Analysis of Annual Change in Global Temperature

Abstract

The purpose of this project was to use time series analysis techniques on a time series data set to predict future data points via forecasting. The time series in question is "Annual changes in global temperature, 1880-1985." My motivation for choosing this data set is to gain further insight on the current debate of global climate change; whether there is sufficient evidence to warrant such a debate. To analyze the time series I utilized R to plot the time series, differenced to remove trend, plot the ACF and PACF of the differenced series to narrow down potential models to AR(3), MA(2), ARMA(7,2), and ARMA(3,2). Based on the AIC criterion I ultimately decided on ARMA(3,2) as the model of best fit. Plotting the ACF and PACF of the residuals confirmed that they were within the confidence intervals. The Box-Ljung test and the McLeod-Li Q test concluded my diagnostic checking returning a p-value of .3684 and .5301 respectively. My forecasting concludes that the average annual change in global temperature from 1986 to 1995 will be minimal. Based on the forecast I cannot find enough evidence to support the claim that global warming is a real issue.

Introduction

The question I sought to address was whether or not the annual change in global temperature has been drastically increasing throughout the year to justify the controversial debate on global warming by forecasting more data points. The time series data set being analyzed is "Annual change in global temperature, 1880-1985" obtained from DataMarket.com's library of time series data. The data set records changes in global temperature in Degrees Celsius by year from 1880 to 1985 with 106 data points. The software i used to analyze the time series is R. In my analysis i used differencing to remove trend, AIC criterion, Maximum Likelihood Estimation to fit my models and the Box-Ljung test and plotting the residuals for diagnostic checking. I chose ARMA(3,2) as my model after comparing it to my other predictions of AR(3), MA(2), and ARMA(7,2). After forecasting 10 data points ahead, my data implies that from 1986 to 1995 the annual change in global temperature will be very small. This means that there isn't enough evidence to support those who argue that global warming is a current and imminent threat.

Time Series Plot

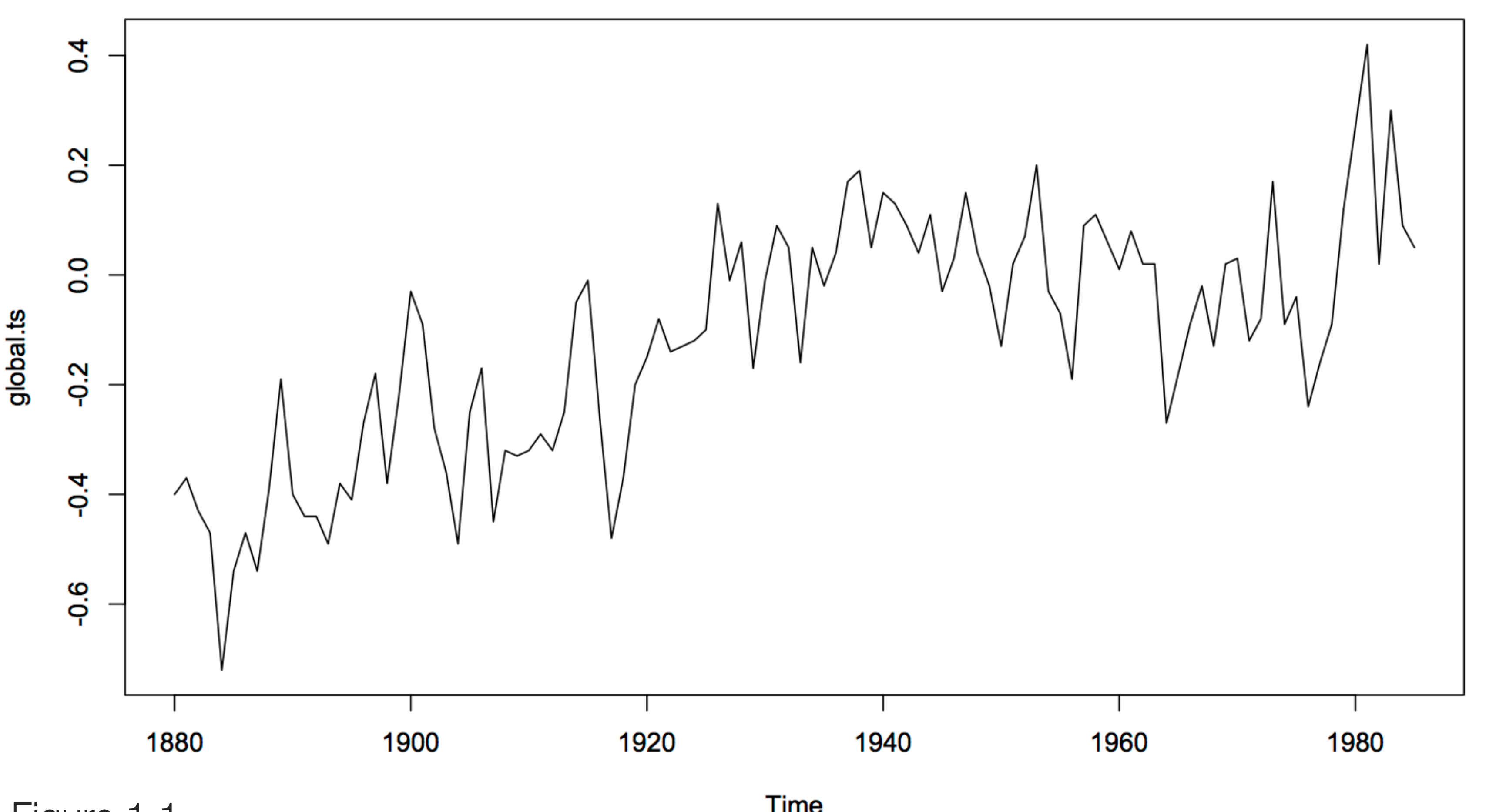

Figure 1.1

Through R i read in the data set and created a time series called “global.ts.” Figure 1.1 above is the time series plot of the data set. It is apparent that the time series is not stationary. There is an increasing trend in the mean that is particularly noticeable from circa 1885 to 1930 and then a relatively small but still discernible decreasing trend from 1940 to around 1975.

The variance which i calculated to be **0.04742253** appears to be stable throughout the entirety of the plot with the exception for a very few spikes at certain intervals, most notable are from 1915 to 1917 and 1978 to 1981.

Stationarity

The original time series data is not stationary because of the trends. In order to eliminate the trend, I differenced the data once. This resulted in the following:

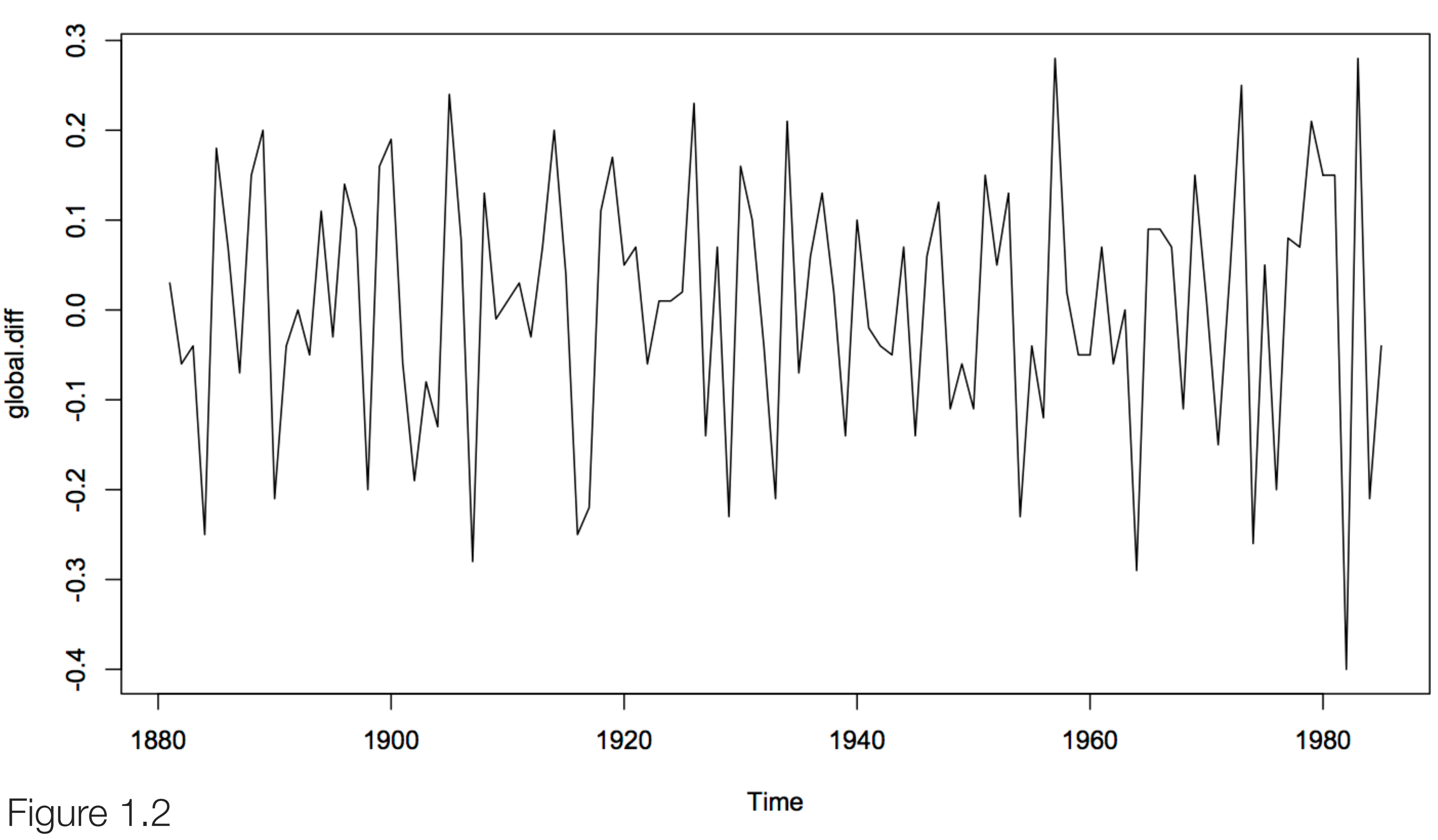

Figure 1.2

Figure 1.2 is the result of differencing the original time series once. Now any trends previously found in the original plot are gone. Since there were no apparent seasonality associated with the original plot, I decided to difference it at lag=1 to get rid of the trend.

The new time series data which i named “global.diff” now appears to be stationary. I chose not to transform the data because the variance seemed to be stable enough in the original plot. The original variance was calculated to be **0.04742253.** After differencing at lag=1, the variance **decreased** to **0.0202535**. An attempt to difference again at lag=1 resulted in over-differencing as the variance then increased to a value above that of the original variance. I found no further need to manipulate the time series since i’m content on its stationarity.

Preliminary Model Identification

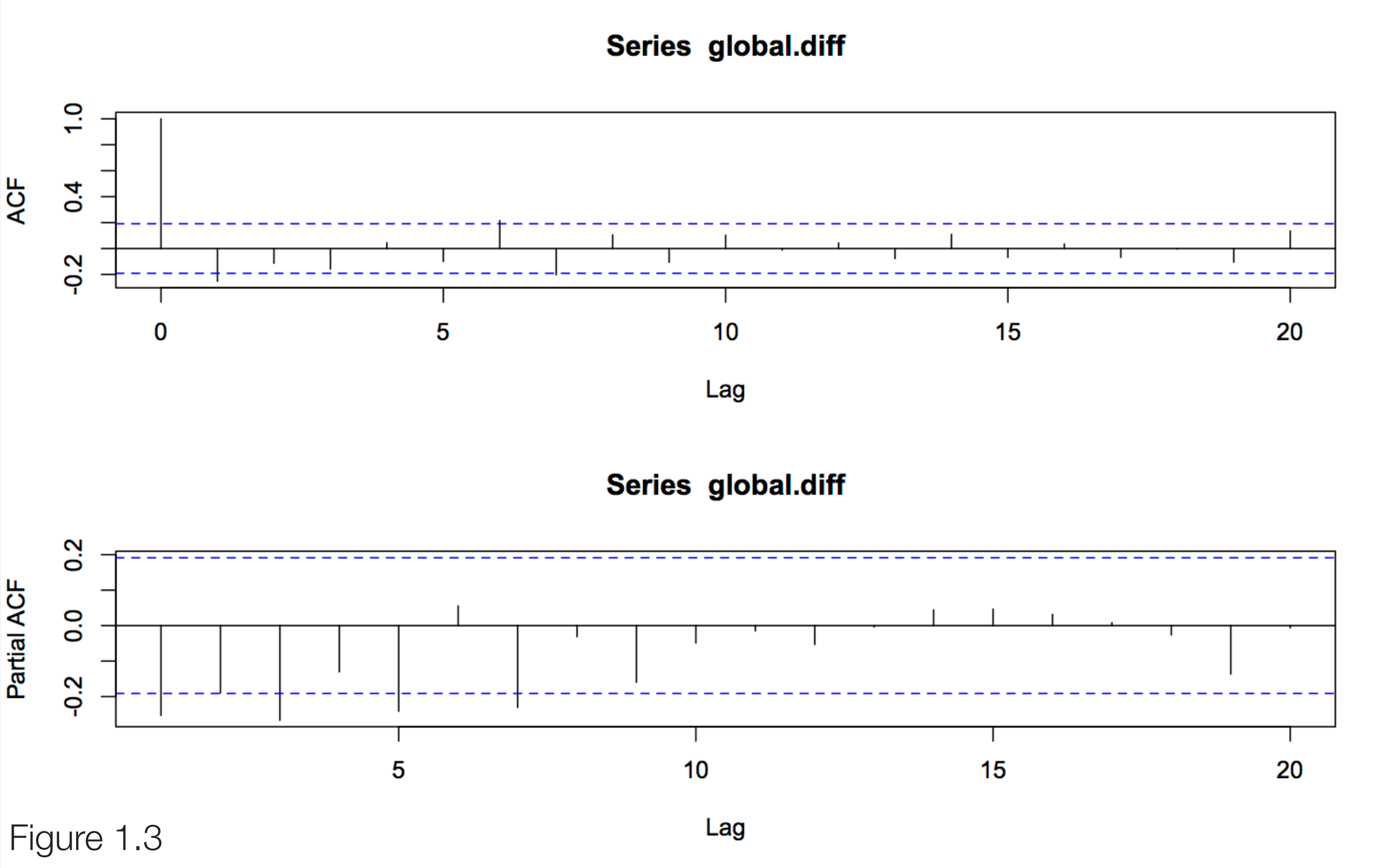

Figure 1.3

Figure 1.3 above shows the plots of the Autocorrelation function (ACF) and the Partial Autocorrelation function (PACF) of the differenced time series “global.diff”. Upon investigation i noted that the PACF cuts off at lags 3 and 7. This led me to believe AR(3) would be a possible fit, AR(7) seems unlikely since there would be too many coefficients. I then looked at the ACF and noticed that it cuts off at lag 2 indicating it could possibly be an MA(2) model. This eventually led to my presumption of ARMA(3,2) as the model of best fit with ARMA(7,2) a distant second.

In summary, based on the ACF and PACF of the differenced time series, i chose four possible models:

1. AR(3)
2. MA(2)
3. ARMA(3,2)
4. ARMA(7,2)

Fitting Models

AR(3):

I used the “Yule Walker” method to estimate the coefficients for the AR model. They were estimated to be:

AR(1): **0.6380**

AR(2): **0.0749**

AR(3): **-0.0486**

AR(4): **0.1945**,

This method gave me four coefficients but i still chose p=3 because it fits better with the PACF.

MA(2):

To estimate the coefficient for MA(2) i used the innovation algorithm. The estimates were:

MA(1): **-0.3008502**

MA(2): **-0.1135437**

ARMA(7,2)

To estimate the coefficients i used the maximum likelihood method. The results were:

AR(1): **-0.3519** AR(2): **-0.1117** AR(3): -**0.2758** AR(4): **-0.1542** AR(5): **-0.2146** AR(6): **0.0645** AR(7): **-0.1964** MA(1): **-0.003**3 MA(2): **-0.3334**

ARMA(3,2)

To estimate the coefficients i used the maximum likelihood method. The results were:

AR(1):**-0.4910**

AR(2):**0.2244**

AR(3):**-0.2323**

MA(1):**0.1032**

MA(2):**-0.6770**

AIC(C):

I used the AIC criterion to determine the model with best fit. The results were:

AR(3): -122.2246

MA(2): -128.7744

ARMA(7,2): -126.1891

ARMA(3,2): **-129.9871**

We choose the smallest AIC value when determining models. In this case, ARMA(3,2) is the best fit. This agrees with the model suggested by the ACF & PACF. The model written Algebraically is such:

$$X_t + 0.4910\, X_{t\text{-}1} - 0.2244X_{t\text{-}2} + 0.2323X_{t\text{-}3} = Z_t + 0.1032Z_{t\text{-}1} - 0.6770Z_{t\text{-}2} \qquad Zt \sim WN\left(0, \sigma_z^2\right)$$

Diagnostic Checking

I used the "*Ljung Box Test*" on the residuals of ARMA(3,2) to calculate the p-value at lag=11. The results were as follows:

Box-Ljung test

data: resfit

X-squared = 6.5119, df = 6, ***p-value = 0.3684***

I also perfumed the "*McLeod–Li Q test"* on the same residual at lag=11 and the results were as follows:

Box-Ljung test

data: resfit^2

X-squared = 10.0029, df = 11, ***p-value = 0.5301***

Since the p-value for both tests are well above .05, we do not have enough evidence to reject the null hypothesis: the data is based on an ARMA(3,2) model.

Checking Residuals

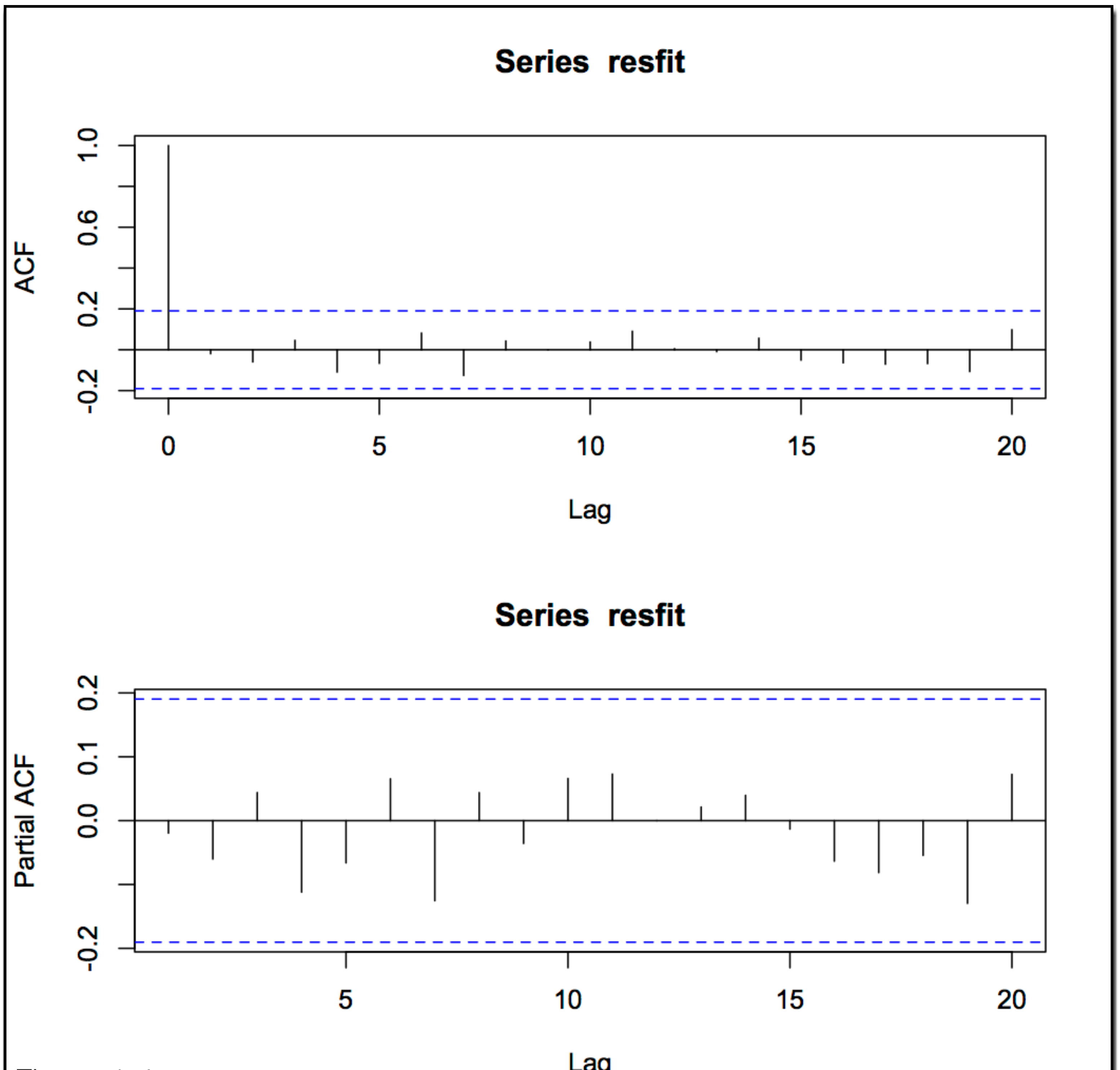

Figure 1.4

Figure 1.4 above is the ACF and PACF plot of the residuals of ARMA(3,2). All of the lags are within the confidence intervals with the exception of lag 0 of the ACF. These plots resemble the ACF and PACF of white noise.

Forecasting

Using the model ARMA(3,2) I forecasted 10 data points ahead on the original time series data “global.ts.” Since i did not use transformation, i did not need to modify the time series to revert back to the original data.

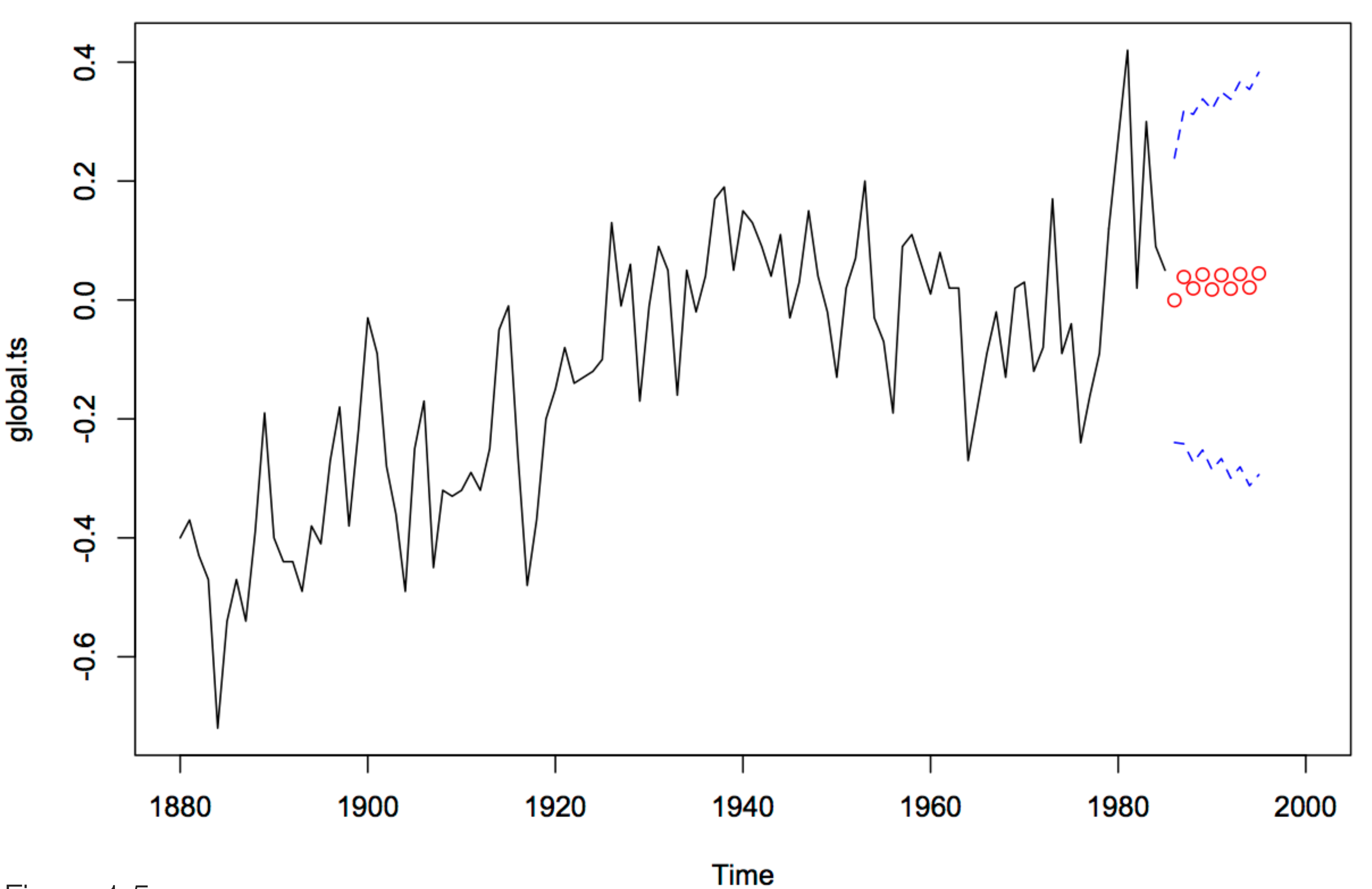

Figure 1.5

Figure 1.5 is the original time series data with 10 datapoint projections plotted (red) along with the confidence intervals (blue).

Conclusion

Based on my analysis of the time series data "Annual changes in global temperatures, 1880-1985" I was able to apply an ARMA(3,2) model. In doing so i was able to predict 10 data points ahead (i.e 10 years) from 1986 to 1995. My results show that the yearly change in global temperature will be very small ranging between .1 to minus .1 degrees celsius between my predicted interval. Based solely on this data, it is arguable that this shows global warming is not occurring as rapidly as some critics claim and that its legitimacy as a major issue is overly exaggerated.

References

1. datamarket.com "http://datamarket.com/data/set/22ku/annual-changes-in-global-temperature-1880-1985#!ds=22ku&display=line"

2. Penn State Online Course Stat 510 "https://onlinecourses.science.psu.edu/stat510/?q=node/62"

3. Professor Feldmen "Lecture notes and examples"

Appendix

```
#read in Data Set "Average Change in Global Temperature annualy"
globaltemp<-read.table("globaltemp.csv", sep=",", header=TRUE,
 nrows=107)

#Create Time Series
globalts<-ts(globaltemp[,2], start = 1880)

#Variance of Original Data:
var(global.ts)
[1] 0.04742253

#Plot Time Series
ts.plot(global.ts)

#differencing TS
global.diff=diff(global.ts, 1)
ts.plot(global.diff)
var(global.diff)
0.02025357

#Plot ACF/PACF
acf(global.diff)
pacf(global.diff)

#Fitting models
fit_ar3=arima(global.ts, order=c(3,1,0), method = "ML")
fit_arma32=arima(global.ts, order=c(3,1,2), method="ML")
fit_ma2 = arima(global.ts, order = c(0,1,2), method = "ML")
fit_arma72 = arima(global.ts, order = c(7,1,2), method = "ML")


#Calculating coefficients#
source("innovations.r")
acvf = acf(global.diff, plot=FALSE, lag.max = length(global.diff))
 $acf[,1,1] * var(global.diff)
m = length(acvf)
lh.ia = innovations.algorithm(99, acvf)

# Preliminary estimates of coefficients for MA(2)
lh.ia$thetas[2,1:2]
```

```r
#For AR(3)
ar(global.diff, method="yule-walker")

#For ARMA(3,2)
fit_arma32

#For ARMA(7,2)
fit_arma72

#AIC criterion
library(qpcR)
AICc(fit_ar3)
AICc(fit_arma32)
AICc(fit_ma2)
AICc(fit_arma72)

#fit model for arma(3,2) and confirm residuals
resfit=residuals(fit_arma32)
acf(resfit, type="correlation", plot=T)
acf(resfit, type="partial", plot=T)

#testing
Box.test(resfit, lag = 11, fitdf = 5, type = "Ljung")
Box.test(resfit^2, lag = 11, fitdf = 0, type = "Ljung")

#forecasting
pred = predict(fit_arma32, n.ahead=10)
pred.orig=pred$
pred.se=2*pred$pred*pred$se
ts.plot(global.ts,xlim=c(1880,2000))
points(1986:1995,pred.orig, col="red")
lines(1986:1995,pred.orig-1.96*pred$se,lty=2,col="blue")
lines(1986:1995,pred.orig+1.96*pred$se,lty=2,col="blue")
```