Economics 144 - Project 1

Ian Chisholm & Niyant Gurwara

4/26/2017

I: Introduction

Living in California, our group decided that we wanted to forecast a metric that was relevant to ourselves as well as the region. Rainfall is never far from mind when living in arid, drought-stricken Southern California, so we decided to forecast precipitation in this region. We acquired our data from the National Oceanic and Atmospheric Administration (see references). It consists of monthly observations ranging from 1895 to 2017.

Upon inspection, we found that the data for precipitation did not contain of any discernible trend. As a result, we turned our attention to temperature, particularly average monthly temperatures. This data set showed more promise, as we could easily discern an upwards trend, as well as the obvious seasonality of climate data.

Average temperatures around the world have been rising as a result of human activities. Climate change is set to disrupt the global economy in a variety of ways, from disrupting food production to massive property damage and refugee crises resulting from rising sea levels. Being able to accurately predict these climatic changes will allow governments to better prepare for these events.

II: Results
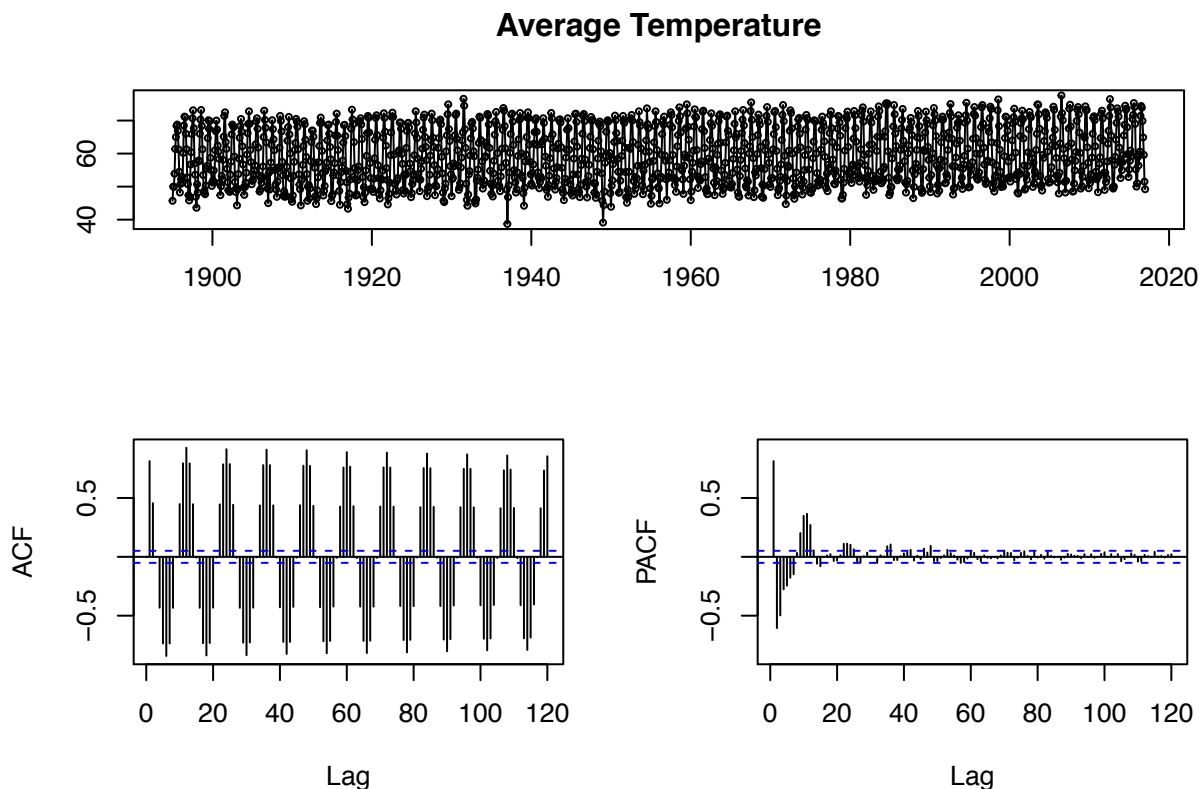
(1) Modeling and Forecasting Trend
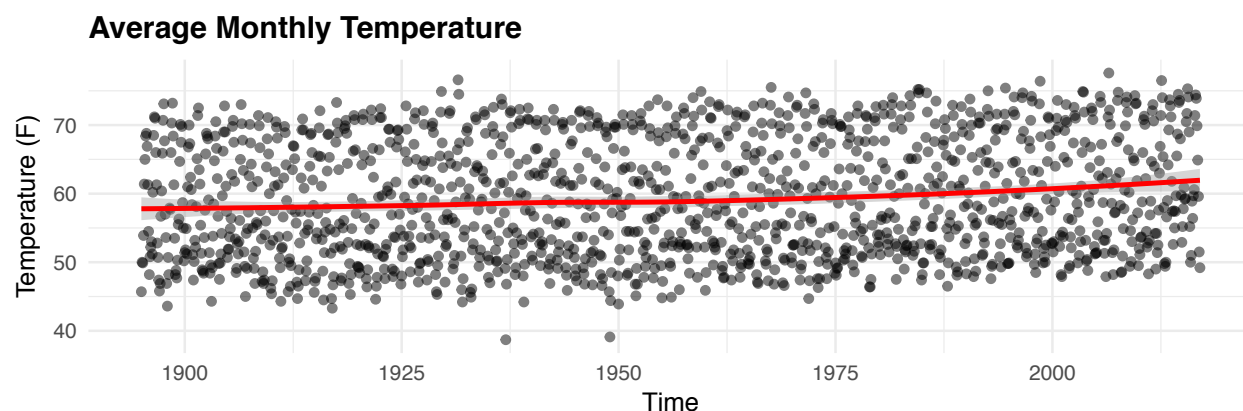
(a) Time Series Plot

(b) Covariance Stationarity

The plot in (a) suggests that the data is not covariance stationary. There is a discernible upwards trend in the data. It can be clearly seen by overlaying a Loess smoother (red line) on a time series plot of the data.

(c) Autocorrelation and Partial Autocorrelation

Average Temperature

ACF: Looking at samples of the data going back 1, 5 and 10 years, the ACF plot suggests that there is a pattern within the data. The acf values rise and fall after a regular number of bands. This would make sense considering that the data represents average temperature at a monthly frequency. We would expect the the direction of correlation to change when moving from summer to winter months and vice versa. For example, the acf plot shows that average temperature in January is most closely correlated to average temperatures in December and February. It is also perfectly correlated to average temperature in January of other years in the data. The plot also shows that the average temperature June is the most negatively correlated to that in January.

PACF: The Partial Autocorrelation function controls for the value of all lags prior to a specific one. In our data, the pacf shows the autocorrelation between the the zero lag month (January) with any of the later months in the year while controlling the values for the lags in between. As expected, the correlations are less pronounced that those in the acf plot. If we look at June for example, the correlation is less than before since it does not account for the correlation with all the months between January and June as well as those after June. The correlation also ultimately reduced to 0 when go back beyond a certain number of lags.

(d) Model Fitting

```
##
## Call:
## lm(formula = tavg.ts ~ yearmonth, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.925  -7.482  -1.170   7.956  18.156
##
```
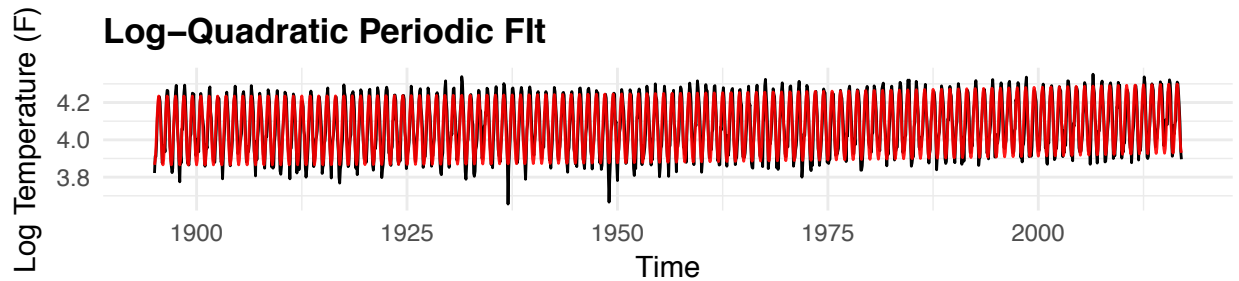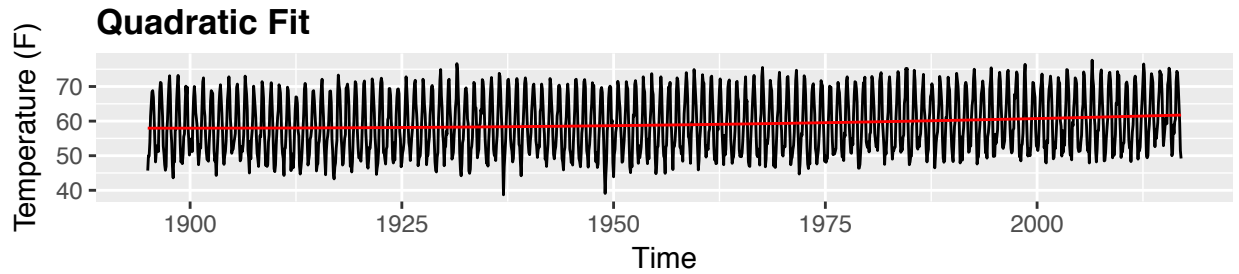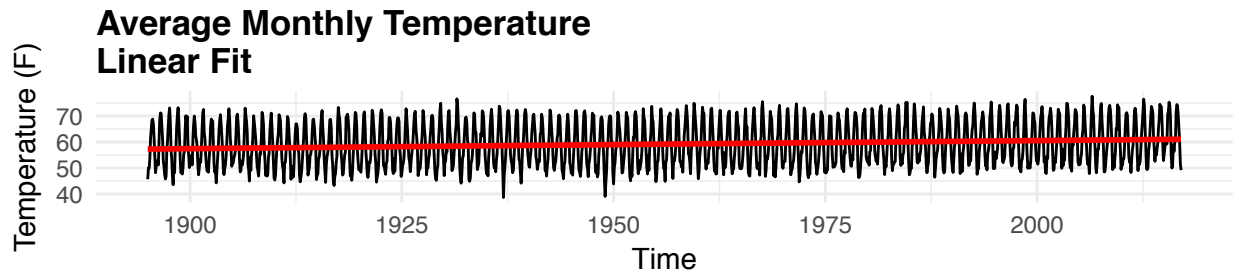
```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.2850403 12.2840179  -0.023    0.981
## yearmonth    0.0003041  0.0000628   4.842 1.42e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.472 on 1463 degrees of freedom
## Multiple R-squared:  0.01578,    Adjusted R-squared:  0.0151
## F-statistic: 23.45 on 1 and 1463 DF,  p-value: 1.418e-06

##
## Call:
## lm(formula = tavg ~ t + I(t^2), data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.665  -7.405  -1.103   7.939  18.341
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.024e+03  7.617e+02   1.345    0.179
## t           -1.018e+00  7.790e-01  -1.307    0.191
## I(t^2)       2.682e-04  1.991e-04   1.347    0.178
##
## Residual standard error: 8.467 on 1462 degrees of freedom
## Multiple R-squared:  0.01759,    Adjusted R-squared:  0.01625
## F-statistic: 13.09 on 2 and 1462 DF,  p-value: 2.318e-06

##
## Call:
## lm(formula = ltemp ~ t + t2 + cos.t)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.215802 -0.042620  0.004386  0.044147  0.144337
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.186e+01  5.271e+00   4.147 3.56e-05 ***
## t           -1.872e-02  5.391e-03  -3.473 0.000529 ***
## t2           4.922e-06  1.378e-06   3.571 0.000367 ***
## cos.t       -1.845e-01  2.164e-03 -85.227  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0586 on 1461 degrees of freedom
## Multiple R-squared:  0.8356, Adjusted R-squared:  0.8352
## F-statistic:  2474 on 3 and 1461 DF,  p-value: < 2.2e-16
```

Average Monthly Temperature
Linear Fit

Quadratic Fit

Log–Quadratic Periodic FIt

The periodic element relating to the changing seasons, and hence average temperatures during the year, is best emulated by a periodic model. Particularly, a Log-Quadratic Periodic Trend. The Linear and Quadratic Trend Models do not capture this element, and hence cannot accurately emulate the process that drives these values. These plots suggest that the third variation of trend, the log-quadratic periodic one, is the most appropriate model for our data. We will confirm this in the steps ahead.

4

(e) Residuals vs. Fitted Values

Linear Model


Polynomial Model

Quadratic Periodic Model
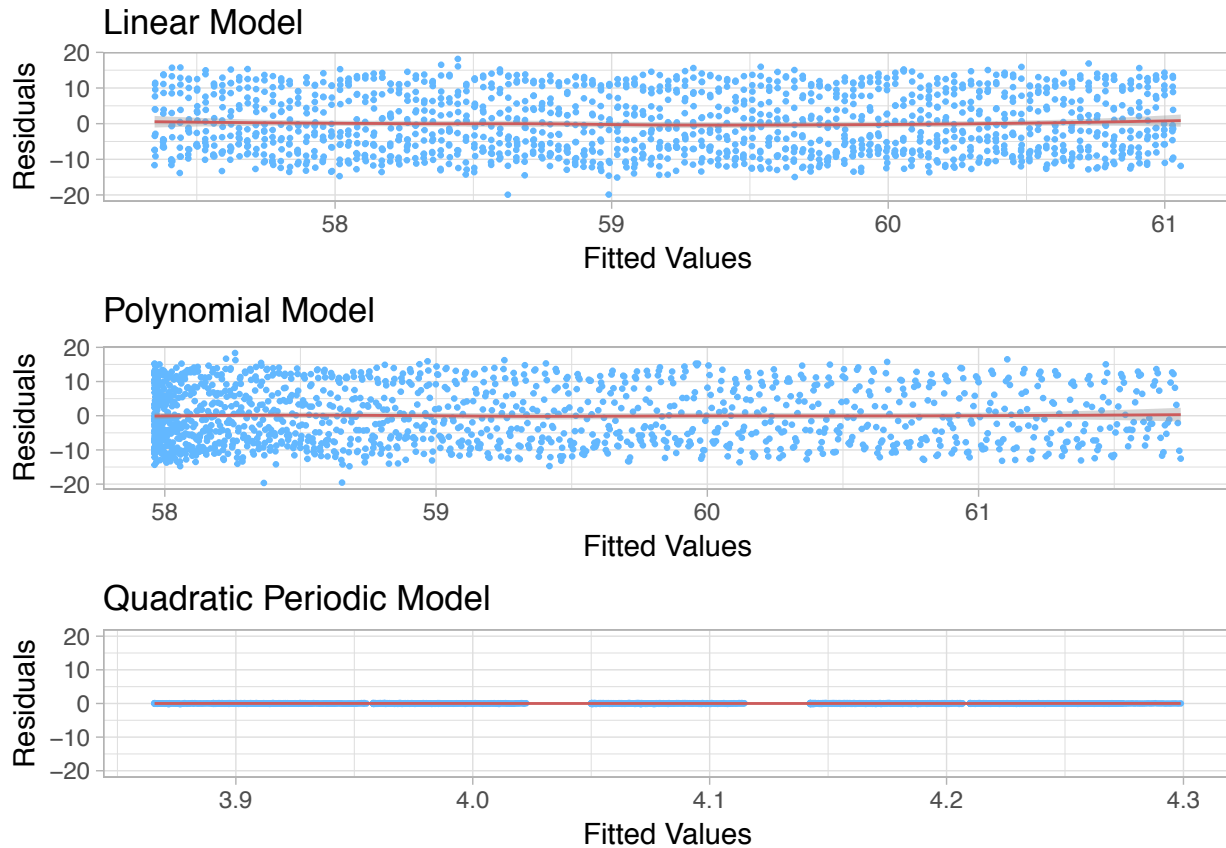
Linear Model: The plot suggests that the linear model is not a good fit for the data. The residual values are spread over a fairly large range (-10 to +15).

Quadratic Model: The plot is quite similar to that of the linear model, indicating that is not a good fit for the data. The residuals are spread over an almost identical range. The only difference is that the fitted values are more concentrated at the lower levels. This is because the shape of the quadratic model accounts for a larger number of lower fitted values as it gradually curves upwards.

Log Quadratic Periodic Model: The plot suggests that this model is the best fit for the data. The residuals are spread over a significantly smaller range (0.15 to +0.15).

Observing the the Residuals vs. Fitted Values plots of the three fits, it is clear tht the Loq Quadratic Periodic fit most closely models the actual values of the Average Temperature time series.

(f) Histogram of Residuals

Linear Model

Quadratic Model

Quadratic Periodic Model
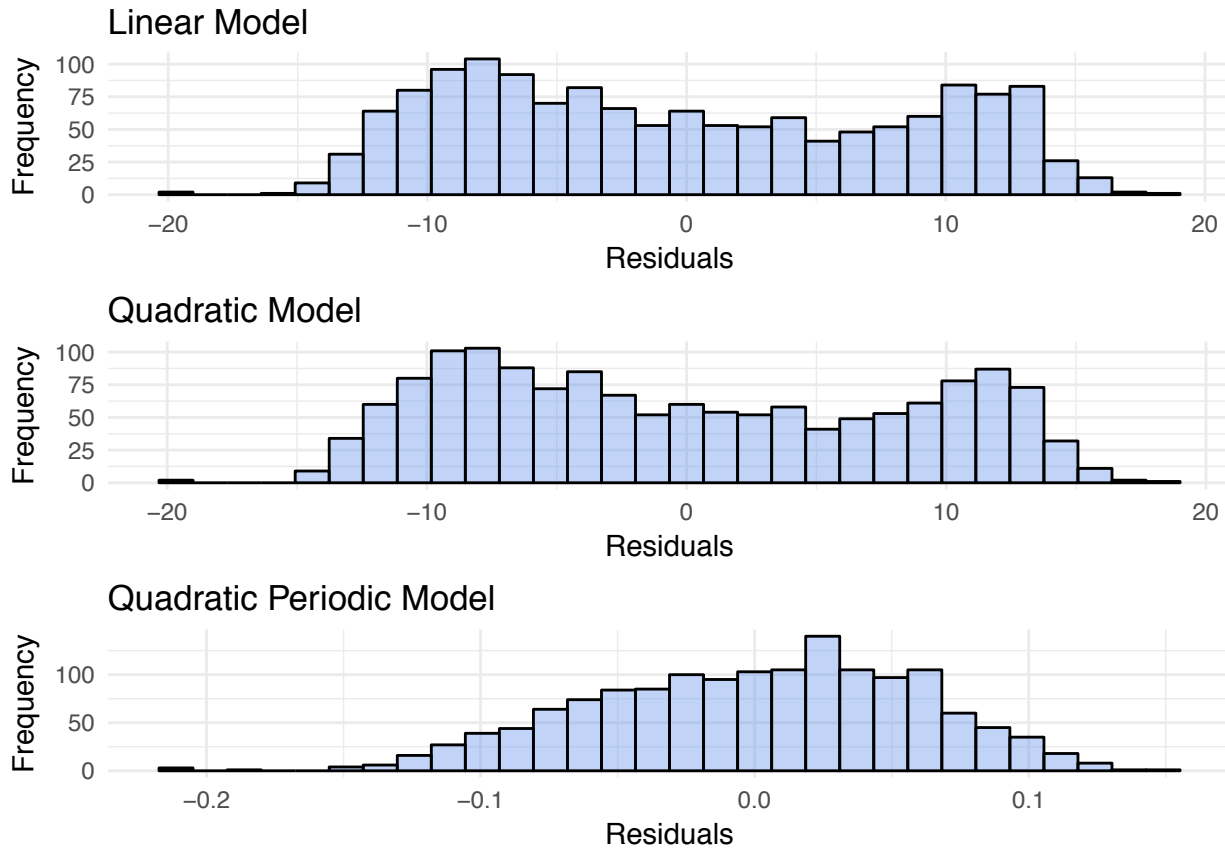
Linear Model: The frequency of the residuals have observable peaks at -8 and +10-12. This means that the model is over and underestimating the data at different points in the time series. However, it is clearly undershooting more often.

Quadratic Model: The histogram of residuals is very similar to that of the linear model. This once again indicates that the linear and quadratic models fit this data almost the same.

Log Quadratic Periodic Model: Frequency of residuals peaks very close to 0 on the positive side. This means that the model, while fitting the data extremely well, marginally overestimates the values. Seeing that the residuals are concentrated so close to 0, we can assume that this model is likely to be the best fit.

(g) Diagnostic Statistics

Linear Model: $R2 = 0.0151$. This means that only 1.5% of the change in y is explained by the regression. That is a very low value. F-statistic $= 23.45$. This value exceed the F critical value, which means that we can reject the null hypothesis that both, the intercept and the coefficient for time, are 0. T-statistic $= 4.842$. Mean Annual Percentage Error $= 12.77331$

Quadratic Model: $R2 = 0.01701$. This means that only 1.7% of the change in y is explained by the regression. That is a very low value. F-statistic $= 12.65$. This value exceed the F critical value, which means that we can reject the null hypothesis that both, the intercept and the coefficient for time, are 0. T-values $= -1.317, 1.356$. Mean Annual Percentage Error $= 12.7787$

Log Quadratic Periodic Model: $R2 = 0.8352$. This means that 83.52% of the change in y is explained by the regression. That value is significantly high. F-statistic $= 2474$. This value exceed the F critical value,

which means that we can reject the null hypothesis that both, the intercept and the coefficient for time, are 0. T-values = -3.473, 3.571, -85.227 Mean Annual Percentage Error = 1.1881

Judging by the values of R2 and Mean Annual Percentage Error, one would conclude that the Log Quadratic Periodic Model is the best fit.

(h) Model Selection

```
##            df        AIC
## lin.model   3 10422.070
## quad.model  4 10421.362
## qp.model    5 -4149.226
```

```
##            df        BIC
## lin.model   3 10437.939
## quad.model  4 10442.520
## qp.model    5 -4122.778
```

```
##                        ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 4.253305e-16 8.465796 7.431419 -2.065473 12.77331 0.9948527
```

```
##                        ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 2.421887e-16 8.457976 7.432949 -2.061427 12.77418 0.9950576
```

```
##                       ME       RMSE        MAE         MPE     MAPE
## Training set -1.33396e-17 0.05851516 0.04827229 -0.02074074 1.188109
##                  MASE
## Training set 0.3825223
```

The AIC and BIC statistics both suggest that the Log-Quadratic-Periodic Model is the best fit for the chosen data. Additionally, the LQP model outperformed the other models in every available mean error statistic. This is what we had expected based off of all the plots we examined prior to these statistics.

(i) Forecast
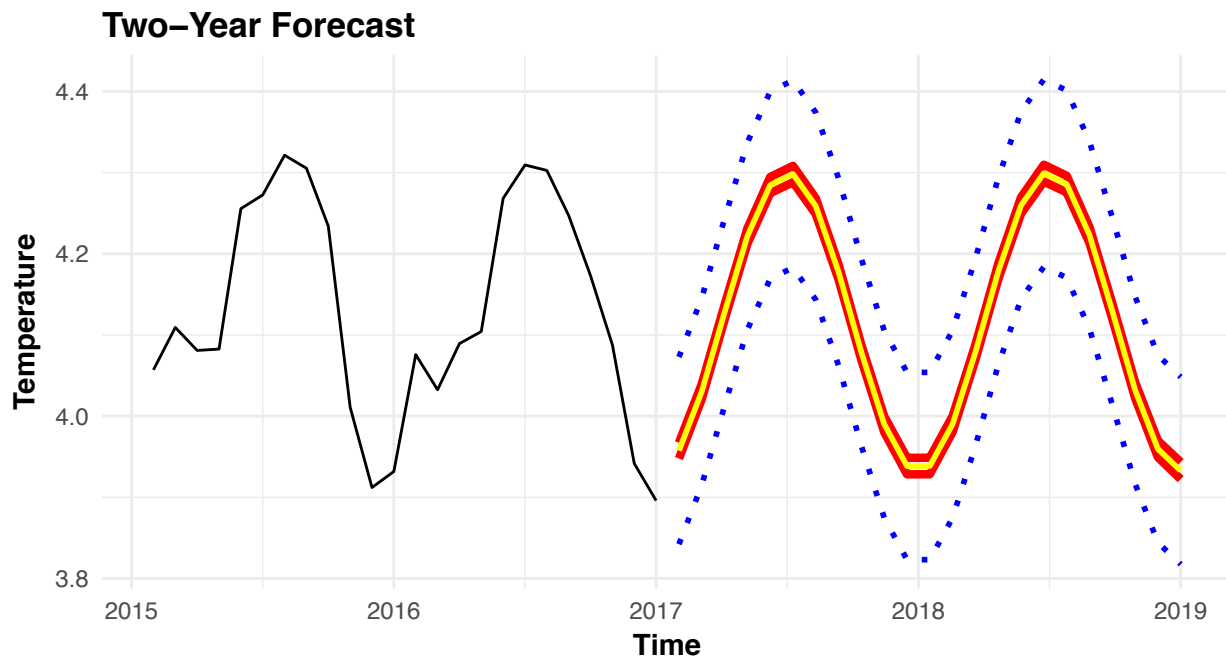
Our 2-year forecast strongly resembles the values in the estimation sample. This would imply that, since this series has a regularly repeating values, we have managed to captre the underlying dynamics of the series. It also means that our forecast values are reasonably accurate condiering our current knowledge base of forecasting theory and application.

(2) Modelling and Forecasting Seasonality

(a) Model with Seasonal Dummies

```
##
## Call:
## tslm(formula = ltemp.ts ~ season)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.235449 -0.028125  0.000405  0.027889  0.164092
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.891289   0.004028 966.152  < 2e-16 ***
## season2     0.020460   0.005708   3.585 0.000349 ***
## season3     0.058409   0.005708  10.234  < 2e-16 ***
## season4     0.118921   0.005708  20.836  < 2e-16 ***
## season5     0.199816   0.005708  35.009  < 2e-16 ***
## season6     0.291882   0.005708  51.139  < 2e-16 ***
## season7     0.375936   0.005708  65.866  < 2e-16 ***
## season8     0.380818   0.005708  66.721  < 2e-16 ***
## season9     0.336163   0.005708  58.898  < 2e-16 ***
## season10    0.233635   0.005708  40.934  < 2e-16 ***
## season11    0.118524   0.005708  20.766  < 2e-16 ***
## season12    0.015762   0.005708   2.762 0.005825 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04467 on 1453 degrees of freedom
## Multiple R-squared:  0.905,  Adjusted R-squared:  0.9042
## F-statistic:  1258 on 11 and 1453 DF,  p-value: < 2.2e-16
```

Each seasonal dummy has a statisitcially significant coefficient. This implies that each season of the year (in this case 12, 1 for each month) has an effect on the average temperature in that month.

(b) Seasonal Factors

Seasonal Model

Seasonal Factors
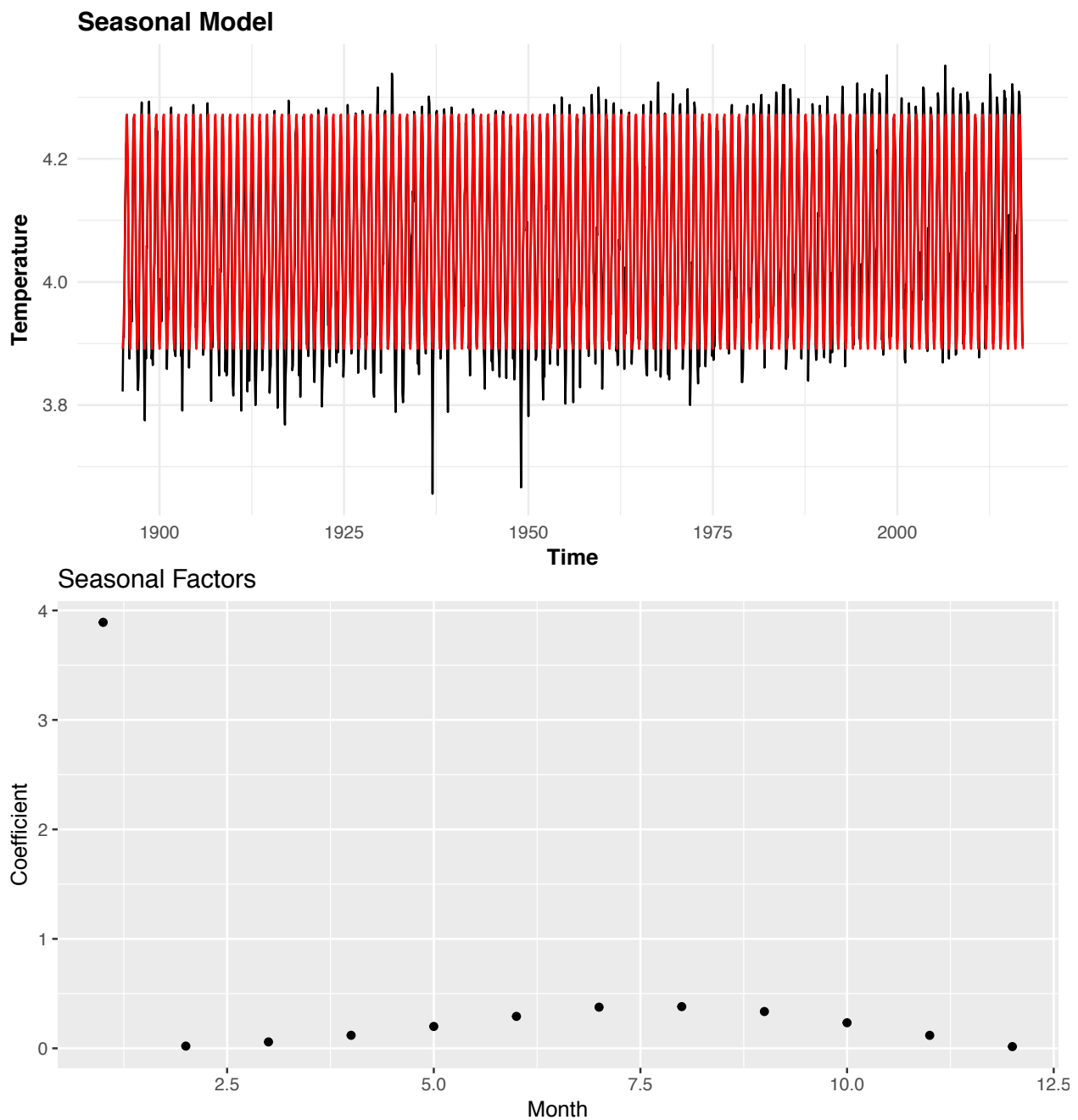
As to be expected, although the seasonal model predicts the timing of changes in temperatures well, the lack of any underlying trend causes it to be covariance stationary, and thus a poor fit for the data.

The seasonal coefficients come as no surprise; we would expect the largest coeffficients during the middle months of the year as that is when temperatures are highest in this (western) hemisphere. The outlying first seasonal coefficient does not completely isolate the effect of season1, which is actually much closer to zero.

—

(c) Full Model (Trend + Seasonality)

```
##
## Call:
## tslm(formula = ltemp.ts ~ season + t + I(t^2) + cos(2 * pi *
##     t))
##
## Residuals:
##       Min       1Q    Median       3Q       Max
## -0.221071 -0.025610  0.000721  0.025772  0.139464
##
## Coefficients: (1 not defined because of singularities)
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2.166e+01  3.625e+00   5.976 2.87e-09 ***
## season2         2.078e-02  5.149e-03   4.036 5.73e-05 ***
## season3         5.868e-02  5.149e-03  11.398  < 2e-16 ***
## season4         1.192e-01  5.149e-03  23.143  < 2e-16 ***
## season5         2.000e-01  5.149e-03  38.846  < 2e-16 ***
## season6         2.920e-01  5.149e-03  56.720  < 2e-16 ***
## season7         3.760e-01  5.149e-03  73.037  < 2e-16 ***
## season8         3.809e-01  5.149e-03  73.977  < 2e-16 ***
## season9         3.362e-01  5.149e-03  65.295  < 2e-16 ***
## season10        2.336e-01  5.149e-03  45.373  < 2e-16 ***
## season11        1.184e-01  5.149e-03  23.006  < 2e-16 ***
## season12        1.564e-02  5.149e-03   3.038  0.00242 **
## t              -1.870e-02  3.707e-03  -5.045 5.12e-07 ***
## I(t^2)          4.915e-06  9.476e-07   5.186 2.45e-07 ***
## cos(2 * pi * t)        NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04029 on 1451 degrees of freedom
## Multiple R-squared:  0.9228, Adjusted R-squared:  0.9221
## F-statistic:  1334 on 13 and 1451 DF,  p-value: < 2.2e-16

## Don't know how to automatically pick scale for object of type ts. Defaulting to continuous.
## Don't know how to automatically pick scale for object of type ts. Defaulting to continuous.
```

Full Model

Residuals vs. Fitted Values – Full Model
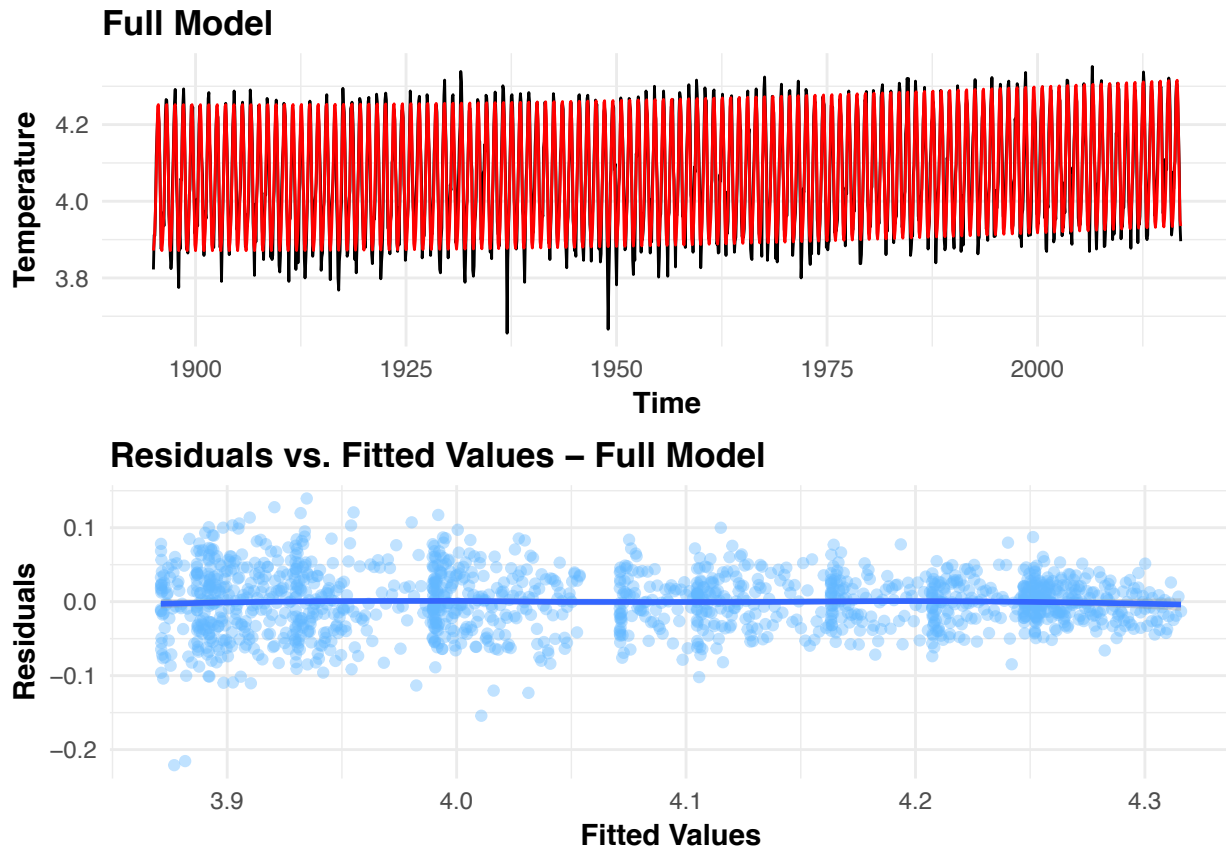
The residuals appear to be roughly randomly distributed around zero, and are within a fairly small range which is encouraging. However, the range of residuals decreases steadily as the fitted values increase, meaning that our model is increasingly accurate as temperatures rise. This makes sense, as we observed that in our previous best model (log-quadratic periodic, see problem 1). This tendency to overestimate temperature is due to the positive trend in our model.

(d) Summary Statistics

```
##
## Call:
## tslm(formula = ltemp.ts ~ season + t + I(t^2) + cos(2 * pi *
##     t))
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.221071 -0.025610  0.000721  0.025772  0.139464
##
## Coefficients: (1 not defined because of singularities)
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.166e+01  3.625e+00   5.976 2.87e-09 ***
## season2        2.078e-02  5.149e-03   4.036 5.73e-05 ***
## season3        5.868e-02  5.149e-03  11.398  < 2e-16 ***
## season4        1.192e-01  5.149e-03  23.143  < 2e-16 ***
## season5        2.000e-01  5.149e-03  38.846  < 2e-16 ***
## season6        2.920e-01  5.149e-03  56.720  < 2e-16 ***
## season7        3.760e-01  5.149e-03  73.037  < 2e-16 ***
```

11

```
## season8           3.809e-01  5.149e-03  73.977  < 2e-16 ***
## season9           3.362e-01  5.149e-03  65.295  < 2e-16 ***
## season10          2.336e-01  5.149e-03  45.373  < 2e-16 ***
## season11          1.184e-01  5.149e-03  23.006  < 2e-16 ***
## season12          1.564e-02  5.149e-03   3.038  0.00242 **
## t                -1.870e-02  3.707e-03  -5.045 5.12e-07 ***
## I(t^2)            4.915e-06  9.476e-07   5.186 2.45e-07 ***
## cos(2 * pi * t)         NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04029 on 1451 degrees of freedom
## Multiple R-squared:  0.9228, Adjusted R-squared:  0.9221
## F-statistic:  1334 on 13 and 1451 DF,  p-value: < 2.2e-16

##                 df       AIC
## full.model      15 -5236.478
## qp.model         5 -4149.226
## seasonal.model  13 -4936.430

##                 df       BIC
## full.model      15 -5157.133
## qp.model         5 -4122.778
## seasonal.model  13 -4867.665

##                        ME       RMSE        MAE         MPE      MAPE
## Training set -1.33396e-17 0.05851516 0.04827229 -0.02074074 1.188109
##                   MASE
## Training set 0.3825223

##                         ME       RMSE        MAE     MPE      MAPE
## Training set -7.884129e-18 0.04010031 0.03121494 -0.0101 0.7748892
##                   MASE
## Training set 0.2473554
```

The fact that all of the coefficients (minus the periodic element) of the model are statistically significant is comforting, along with the very high adjusted R-squared of 0.92. The AIC and BIC both confirm that the addition of a seasonal element improved the original LQP model. Additionally, the full model outperformed the original LQP model according to every error metric with the exception of mean error. Overall, we believe that the full model is the best fit, and likely a very strong predictor of future temperature levels.

(e) Full Model Forecast

Two–Year Forecast
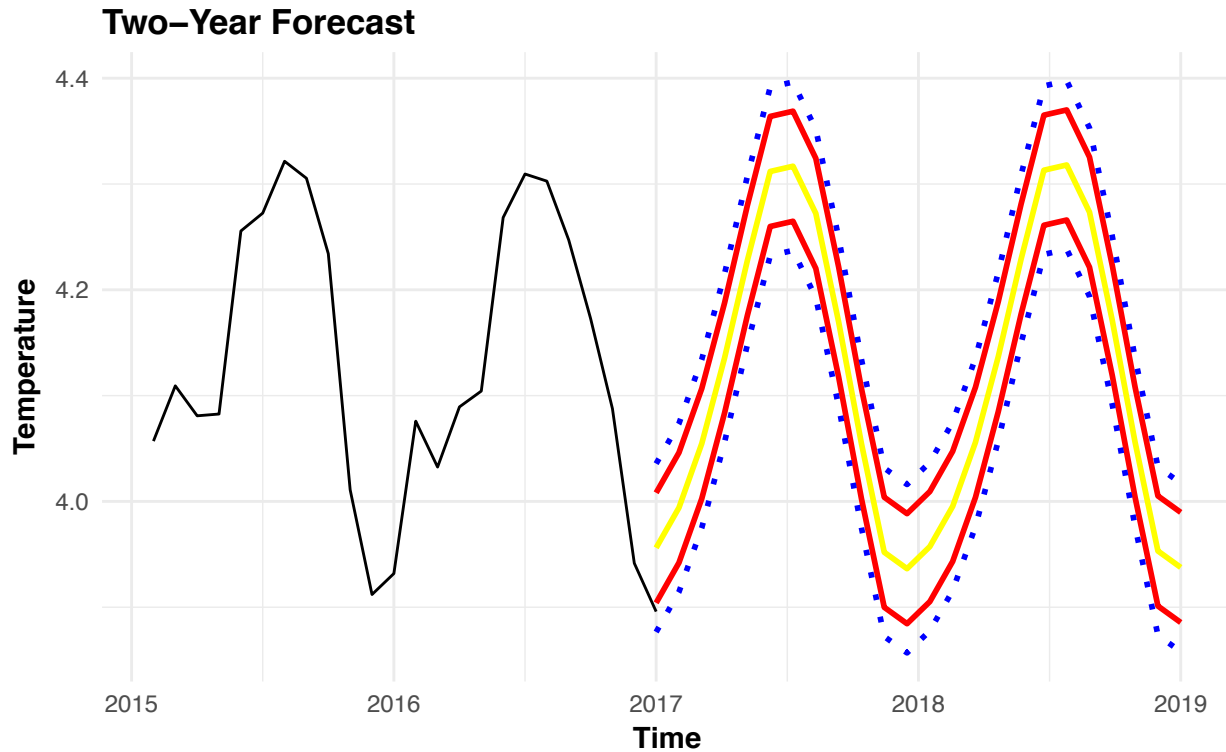
III: Conclusions and Future Work

Given our current level of understanding, this seems to be the best model we can create. Using climatic, and thus truly seasonal data no doubt made it easier to model and thus improved the quality of our forecast. Going forward, the two issues with our model are its tendency to slightly overestimate temperature (as seen in the residuals vs. fitted values and histogram of residuals) and its limited amplitude relative to the data. In the future we would like to find some way to solve these issues.

IV: References

1. "Divisional Data Select". National Oceanic and Atmospheric Administration. Www7.ncdc.noaa.gov. Web.

V: Source Code

```
library(timeSeries)
library(forecast)
library(dplyr)
library(ggplot2)
library(ggthemes)
library(car)
library(gridExtra)
```

```
rm(list=ls())

setwd('/Users/NiyantG/Desktop/College/Senior/Spring/Econ 144/Project 1/')

# -----------Setup----------

d <- read.csv('/Users/NiyantG/Desktop/College/Senior/Spring/Econ 144/Project 1/south_coast_climate.csv')
names(d) <- tolower(names(d))
d$yearmonth <- ts(d$yearmonth)

# II
# (1) ----- Modelling and Forecasting Trend
# (a) Time Series Plot

t<-seq(1895,2017,length=length(d$tavg))

ts.plot <- ggplot(d, aes(x=t))
ts.plot + geom_point(aes(y=tavg),alpha=0.5)  +
  geom_smooth(aes(y=tavg),method='loess',color='red') + theme_minimal() +
  labs(x='Time',y='Temperature (F)',title='Average Monthly Temperature') +
  theme(title=element_text(face='bold'),axis.title = element_text(face='plain'))

# (c) Autocorrelation and Partial Autocorrelation

d <- mutate(d, tavg.ts=ts(tavg,start=1895,frequency=12))
tsdisplay(d$tavg.ts,lag.max=120, main='Average Temperature')
t<-seq(1895,2017,length=length(d$tavg))
t2<-t^2
d <- mutate(d,tavg.ts=ts(d$tavg))

# (d) Model Fitting

#--------Linear Fit ---------
t<-seq(1895,2017,length=length(d$tavg))
lin.model <- lm(tavg.ts ~ yearmonth,d)
summary(lin.model)
lin.plot <- ggplot(d, aes(x=t))
lin.plot <- lin.plot + geom_line(aes(y=tavg)) + geom_smooth(aes(y=tavg),method='lm',se=F,color='red') +
  labs(x='Time',y='Temperature (F)',title='Average Monthly Temperature\nLinear Fit') +
  theme(title=element_text(face='bold'),axis.title = element_text(face='plain'))

#--------Quadratic Fit--------
quad.model <- lm(tavg~ t + I(t^2),d)
summary(quad.model)

quad.plot <- ggplot(d,aes(x=t))
quad.plot <- quad.plot + geom_line(aes(y=tavg)) +
  geom_line(aes(y=quad.model$fitted.values),color='red') + labs(x='Time',y='Temperature (F)',title='Quad
  theme(title=element_text(face='bold'),axis.title = element_text(face='plain'))

#--------Log-Quadratic-Periodic Fit --------
t<-seq(1895,2017,length=length(d$tavg))
t2<-t^2
d <- mutate(d,tavg.ts=ts(d$tavg))
```

```
ltemp<-log(d$tavg)
cos.t<-cos(2*pi*t)

log.df <- data.frame(t=t,t2=t2,cos.t=cos.t,ltemp=ltemp,yearmonth=d$yearmonth)

qp.model=lm(ltemp~t+t2+cos.t)
summary(qp.model)

qp.plot <- ggplot(log.df, aes(x=t))
qp.plot <- qp.plot + geom_line(aes(y=ltemp)) + geom_line(aes(y=qp.model$fitted.values),color='red',alpha
  theme_minimal() + labs(x='Time',y='Log Temperature (F)',title='Log-Quadratic Periodic FIt') +
  theme(title=element_text(face='bold'),axis.title = element_text(face='plain'))

grid.arrange(lin.plot,quad.plot,qp.plot)

# (e) Residuals vs. Fitted Values

# Linear Model
res.vs.fit1<- ggplot(d, aes(x=lin.model$fitted.values,y=lin.model$residuals)) + geom_point(color='steel
  labs(x='Fitted Values',y='Residuals') +
  ggtitle('Linear Model') +
  theme_light()

# Quadratic Model
res.vs.fit2<- ggplot(d, aes(x=quad.model$fitted.values,y=quad.model$residuals)) +  geom_point(color='st
  labs(x='Fitted Values',y='Residuals') +
  ggtitle('Polynomial Model') +
  theme_light()

# Quadratic - Preiodic Model
res.vs.fit3<- ggplot(log.df, aes(x=qp.model$fitted.values,y=qp.model$residuals)) + geom_point(color='st
  labs(x='Fitted Values',y='Residuals') + ylim(-20,20) +
  ggtitle('Quadratic Periodic Model') +
  theme_light()

grid.arrange(res.vs.fit1,res.vs.fit2,res.vs.fit3)

# (f) Histogram of Residuals

# Linear Model
lin.hist<- ggplot(d, aes(x=lin.model$residuals)) + geom_histogram(fill='cornflowerblue',col='black',bins
  labs(title='Linear Model',x='Residuals', y='Frequency') +
  theme_minimal()

# Quadratic Model
quad.hist<- ggplot(d, aes(x=quad.model$residuals)) + geom_histogram(fill='cornflowerblue',col='black',b
  labs(title='Quadratic Model',x='Residuals', y='Frequency') +
  theme_minimal()

# Quadratic Periodic Model
qp.hist<- ggplot(log.df, aes(x=qp.model$residuals)) + geom_histogram(fill='cornflowerblue',col='black',b
  labs(title='Quadratic Periodic Model',x='Residuals', y='Frequency') +
  theme_minimal()
```

```
grid.arrange(lin.hist,quad.hist,qp.hist)

# (h) Model Selection

AIC(lin.model,quad.model,qp.model)
BIC(lin.model,quad.model,qp.model)

accuracy(lin.model)
accuracy(quad.model)
accuracy(qp.model)

# (i) Forecast

ltemp.ts <- ts(log.df$ltemp)

h <- 24 # forecast h-steps ahead

tn=data.frame(t=seq(2017,2019,length.out=h)) # new data

pred <- predict(lm(ltemp.ts~t+I(t^2)+cos(2*pi*t)),newdata=tn,se.fit=T)
pred.plim <- predict(lm(ltemp.ts~t+I(t^2)+cos(2*pi*t)),tn,level=0.95,interval='prediction')
pred.clim <-predict(lm(ltemp.ts~t+I(t^2)+cos(2*pi*t)),tn,level=0.95,interval='confidence')

# the next several lines create a new data frame, combining the historic data with the forecasts

fc <- data.frame(t=tn,hist=NA,p.lwr=pred.plim[,2],p.upr=pred.plim[,3],c.lwr=pred.clim[,2],c.upr=pred.cl
s=dim(fc)[1]
fc <- fc[2:s,]

hist.fc <- data.frame(t=log.df$t,hist=log.df$ltemp,p.lwr=NA,p.upr=NA,c.lwr=NA,c.upr=NA,point=NA)
fc <- rbind(hist.fc,fc)

# fore plot

p16 <- subset(fc, t > 2015)
fc.plot <- ggplot(p16,aes(x=t,y=hist) )
fc.plot + geom_line()  + geom_line(aes(y=c.lwr),color='red',size=1.5) +
    geom_line(aes(y=c.upr),color='red',size=1.5) + geom_line(aes(y=point),color='yellow',size=1) + geom
  geom_line( aes(y=p.upr),color='blue',size=1,linetype='dotted') + theme_minimal() + labs(x='Time',y='T
  theme(title = element_text(face='bold'),axis.text = element_text(face='plain'))

# (2)--------Modelling and Forecasting Seasonality---------
# (a) Model with Seasonal Dummies

ltemp.ts = ts(log.df$ltemp, frequency=12) # create time series object

seasonal.model=tslm(ltemp.ts ~ season) # fit a seasonal model
summary(seasonal.model)

# (b) Seasonal Factors

seasonal.plot <- ggplot(log.df, aes(x=t,y=ltemp))
seasonal.plot + geom_line() + geom_line(aes(y=seasonal.model$fitted.values),color='red') +
```

```r
  theme_minimal() + labs(x='Time',y='Temperature',title='Seasonal Model') +
  theme(title = element_text(face='bold'), axis.text = element_text(face='plain'))

seasonal.factors <- qplot(seq(1,12),seasonal.model$coefficients,ylab = 'Coefficient',xlab = 'Month',mai
seasonal.factors

# (c) Full Model (Trend + Seasonality)

# add log-quadratic-period model from 1. to the seasonal model

full.model=tslm(ltemp.ts ~ season + t+I(t^2)+cos(2*pi*t))
summary(full.model)

seasonal.plot <- ggplot(log.df, aes(x=t,y=ltemp)) +
geom_line() + geom_line(aes(y=full.model$fitted.values),color='red') +
  theme_minimal() + labs(x='Time',y='Temperature',title='Full Model') +
  theme(title = element_text(face='bold'), axis.text = element_text(face='plain'))

# plot residuals vs. fitted values of new model

full.resvfit <- ggplot(log.df,aes(x=full.model$fitted.values,y=full.model$residuals)) +
geom_point(color='steelblue1',alpha=0.4) +
theme_minimal() + labs(x='Fitted Values',y='Residuals',title='Residuals vs. Fitted Values - Full Model'
  theme(title = element_text(face='bold'), axis.text = element_text(face='plain'))

grid.arrange(seasonal.plot,full.resvfit)

# (d) Summary Statistics

summary(full.model)
AIC(full.model,qp.model,seasonal.model)
BIC(full.model,qp.model,seasonal.model)
accuracy(qp.model)
accuracy(full.model)

# (e) Full Model Forecast

full.forecast <- forecast(tslm(ltemp.ts~t+I(t^2)+season),newdata=tn,level=c(80,95),h=12)
full.forecast <- data.frame(t=tn$t,full.forecast)

# combine historic data with forecast

hist.fc <- data.frame(t=log.df$t,point=NA,p.lwr=NA,p.upr=NA,c.lwr=NA,c.upr=NA,historic=log.df$ltemp)
full.forecast <- mutate(full.forecast,historic=NA)
names(full.forecast) = names(hist.fc)

full.forecast<- rbind(hist.fc,full.forecast)

# plot two-year forecast

p16 <- subset(full.forecast, t > 2015)
fc.plot <- ggplot(p16,aes(x=t,y=historic) )
fc.plot + geom_line()  + geom_line(aes(y=p.lwr),color='red',size=1) +
  geom_line(aes(y=p.upr),color='red',size=1) + geom_line(aes(y=point),color='yellow',size=1) + geom_lin
```

```
geom_line( aes(y=c.upr),color='blue',size=1,linetype='dotted') + theme_minimal() + labs(x='Time',y='T
theme(title = element_text(face='bold'),axis.text = element_text(face='plain'))
```