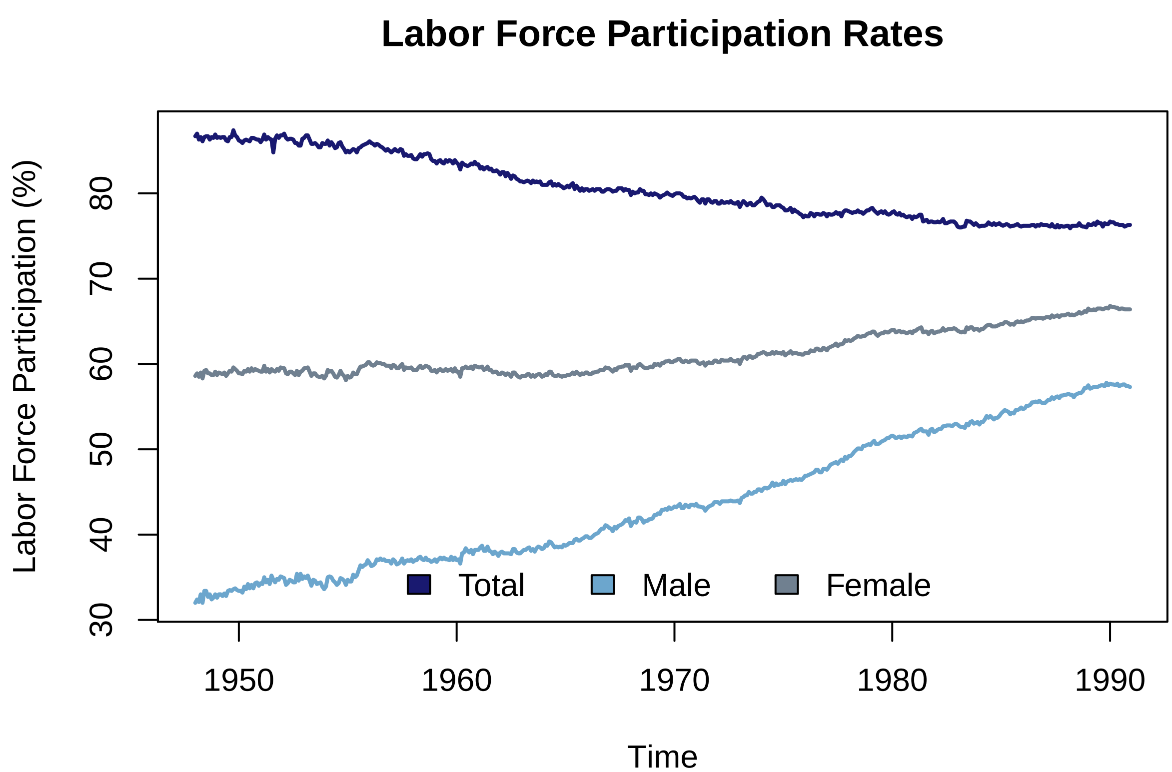
Econ 144: Homework #2

Problem 1.

The data file `labor.dat` consists of the labor force participation rates by gender (including the total, i.e. sum of male and female) for the years 1948 to 1991. Each observation is a monthly data point. The objective is to fit a trend to the time series data, and based on the best fit model, make a 10 year forecast.

1a. Show a time series plot of your data.



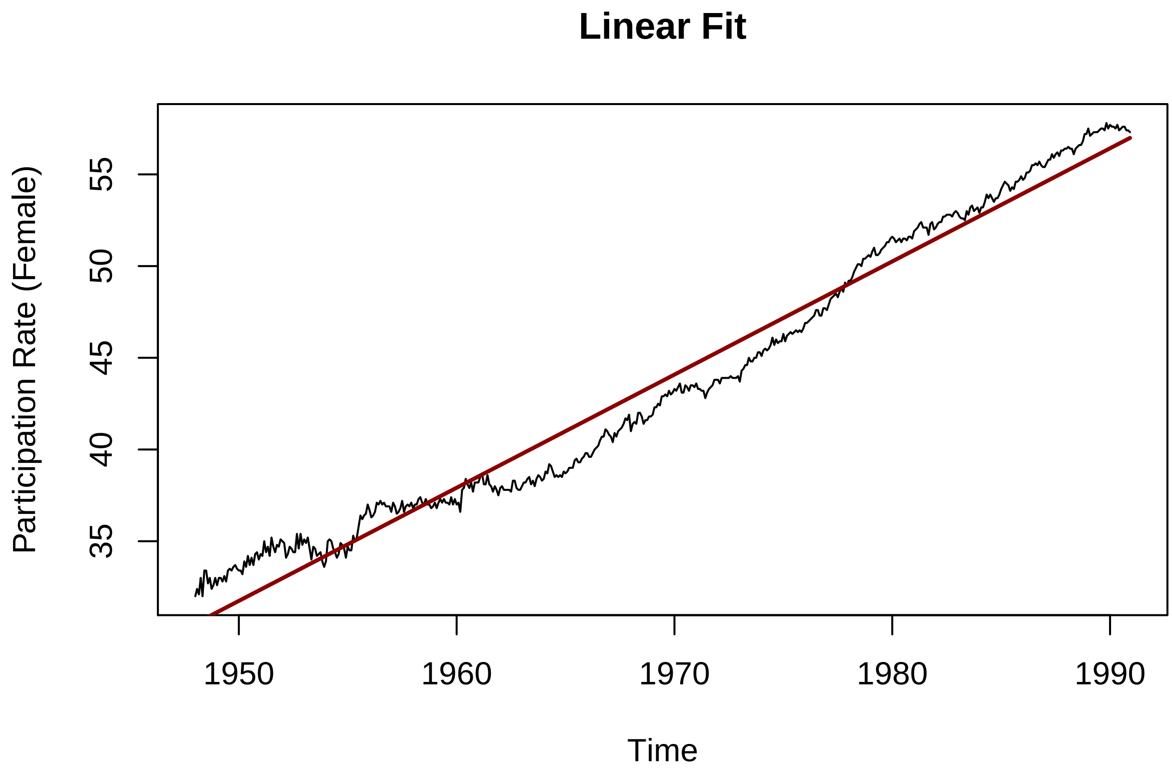
1b. Fit a linear, polynomial, and exponential model to the female labor force participation rate.

We begin by fitting a linear model of the following form:

$$female_t = \alpha_t + \beta_0 time_t + \varepsilon_t$$

From just the adjusted R-squared value, the linear model appears to fit the female labor force participation rate data pretty well. However, from the graph, we can see that there is a slight curvature to the data overall that the linear model does not accomodate.

```r
m1 <- lm(female_ts ~ t)
plot(female_ts, main = "Linear Fit", ylab = "Participation Rate (Female)",
    xlab = "Time")
lines(t, m1$fit, col = "red4", lwd = 2)
```




```r
summary(m1)
```

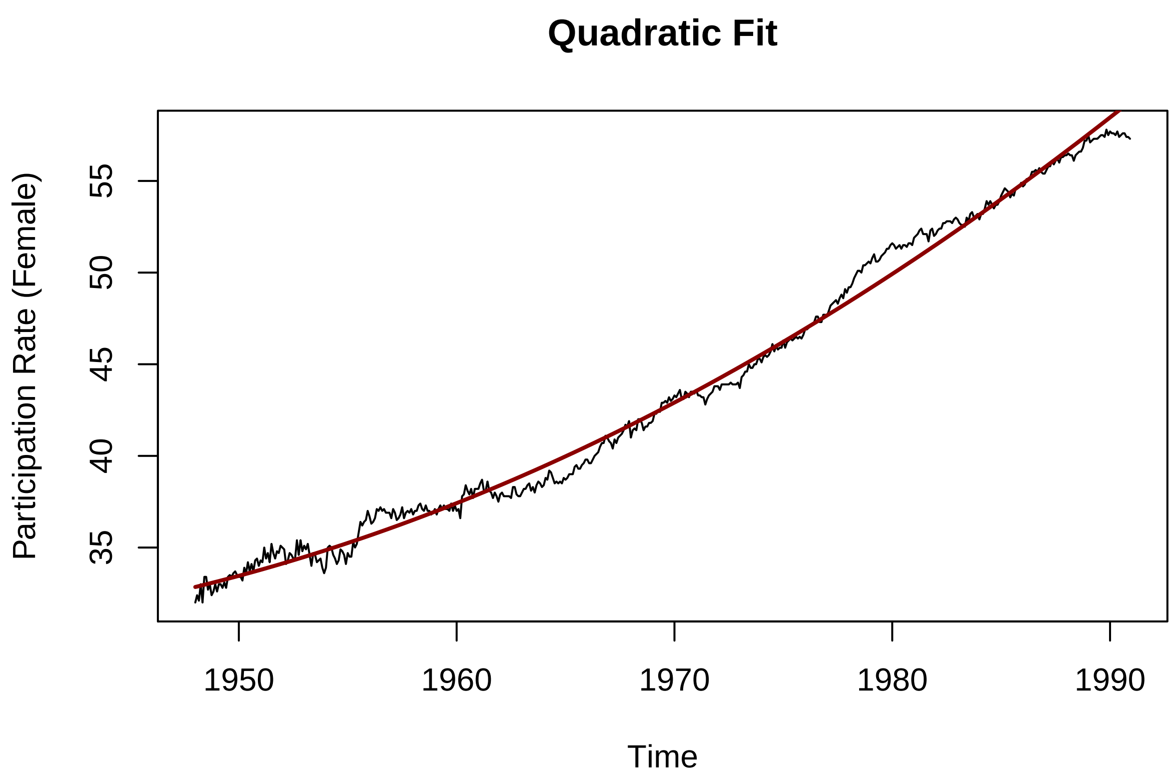
```
##
## Call:
## lm(formula = female_ts ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3959 -1.1994  0.2523  1.0177  2.6321
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.171e+03  8.772e+00  -133.5   <2e-16 ***
## t            6.169e-01  4.454e-03   138.5   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.256 on 514 degrees of freedom
## Multiple R-squared:  0.9739, Adjusted R-squared:  0.9739
## F-statistic: 1.919e+04 on 1 and 514 DF,  p-value: < 2.2e-16
```

For the polynomial model, we try both a quadratic and a cubic model. Our quadratic model takes the following form:

$$female_t = \alpha_t + \beta_0 time_t + \beta_1 time_t^2 + \varepsilon_t$$

The quadratic model has an R-squared value of 0.9922, which indicates that it also accomodates for the data pretty well. The quadratic model does accomodate for the curvature in the data; however, the degree of curvature/concavity that is present in the data is not perfectly estimated.

```
m2 <- lm(female_ts ~ t + I(t^2))
plot(female_ts, main = "Quadratic Fit", ylab = "Participation Rate (Female)",
    xlab = "Time")
lines(t, m2$fit, col = "red4", lwd = 2)
```




```
summary(m2)
```

```
##
## Call:
## lm(formula = female_ts ~ t + I(t^2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.02400 -0.47184 -0.05572  0.42321  1.71762
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.841e+04  8.488e+02   33.47   <2e-16 ***
## t           -2.942e+01  8.620e-01  -34.13   <2e-16 ***
```

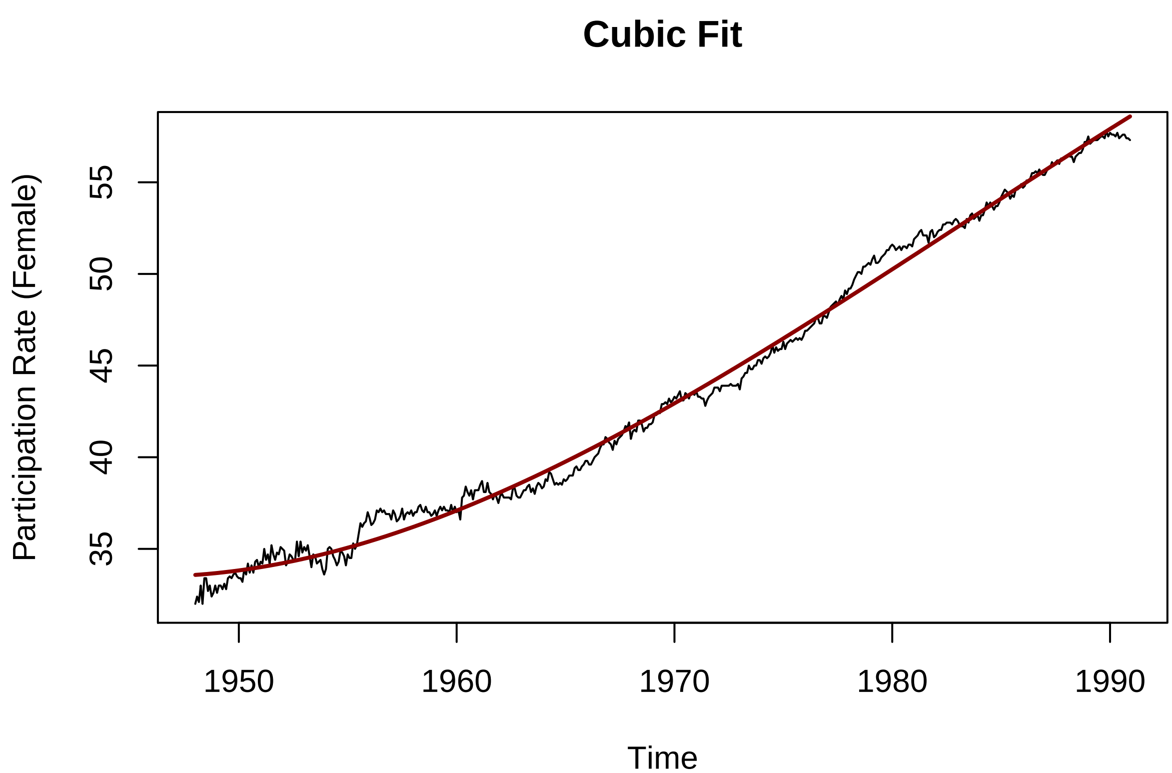
```
## I(t^2)       7.626e-03  2.188e-04   34.85   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6851 on 513 degrees of freedom
## Multiple R-squared:  0.9923, Adjusted R-squared:  0.9922
## F-statistic: 3.285e+04 on 2 and 513 DF,  p-value: < 2.2e-16
```

The cubic model is as follows:

$$female_t = \alpha_t + \beta_0 time_t + \beta_1 time_t^2 + \beta_3 time_t^3 + \varepsilon_t$$

The cubic model (which also includes a quadratic term) does accomodate for the curvature, and does so in a closer manner than the quadratic model does.

```
m3 <- lm(female_ts ~ t + I(t^2) + I(t^3))
plot(female_ts, main = "Cubic Fit", ylab = "Participation Rate (Female)",
    xlab = "Time")
lines(t, m3$fit, col = "red4", lwd = 2)
```




```
summary(m3)
```

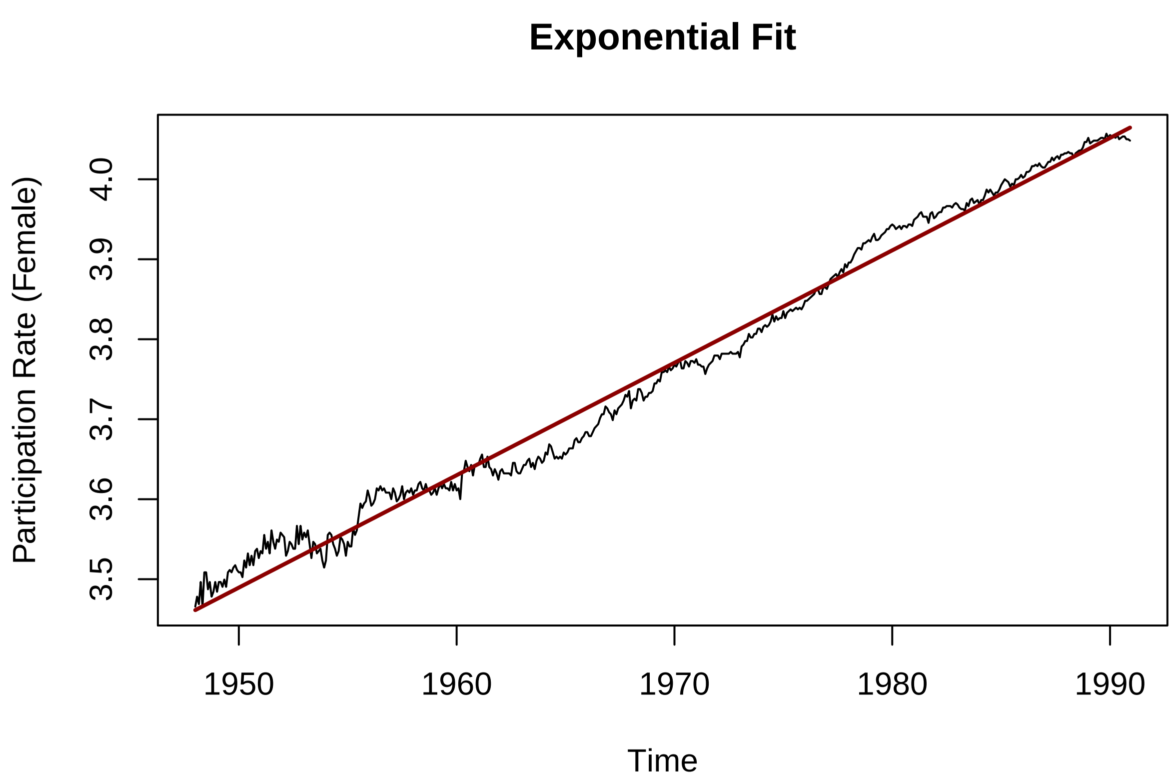
```
##
## Call:
## lm(formula = female_ts ~ t + I(t^2) + I(t^3))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -1.6100 -0.4593 -0.0761  0.4111  1.6190
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.453e+06  1.400e+05   10.38   <2e-16 ***
## t           -2.199e+03  2.132e+02  -10.31   <2e-16 ***
## I(t^2)       1.109e+00  1.083e-01   10.25   <2e-16 ***
## I(t^3)      -1.865e-04  1.832e-05  -10.18   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6254 on 512 degrees of freedom
## Multiple R-squared:  0.9936, Adjusted R-squared:  0.9935
## F-statistic: 2.631e+04 on 3 and 512 DF,  p-value: < 2.2e-16
```

Hypothetically, we could use higher orders, but this would run the risk of overfitting to the data. Additionally, in testing for a quartic model, there was a large degree of collinearity within the quadratic term and the quartic term. Therefore, no further polynomial models were tested.

For the exponential model, there are several ways to go about fitting such a model. To bypass dealing with potential singularities, we take the natural log of our dependent variable ($female_t$) first, and then fit a linear model to $log(female_t)$.

```
m4 <- lm(log(female_ts) ~ t)
plot(log(female_ts), main = "Exponential Fit", ylab = "Participation Rate (Female)",
    xlab = "Time")
lines(t, m4$fit, col = "red4", lwd = 2)
```

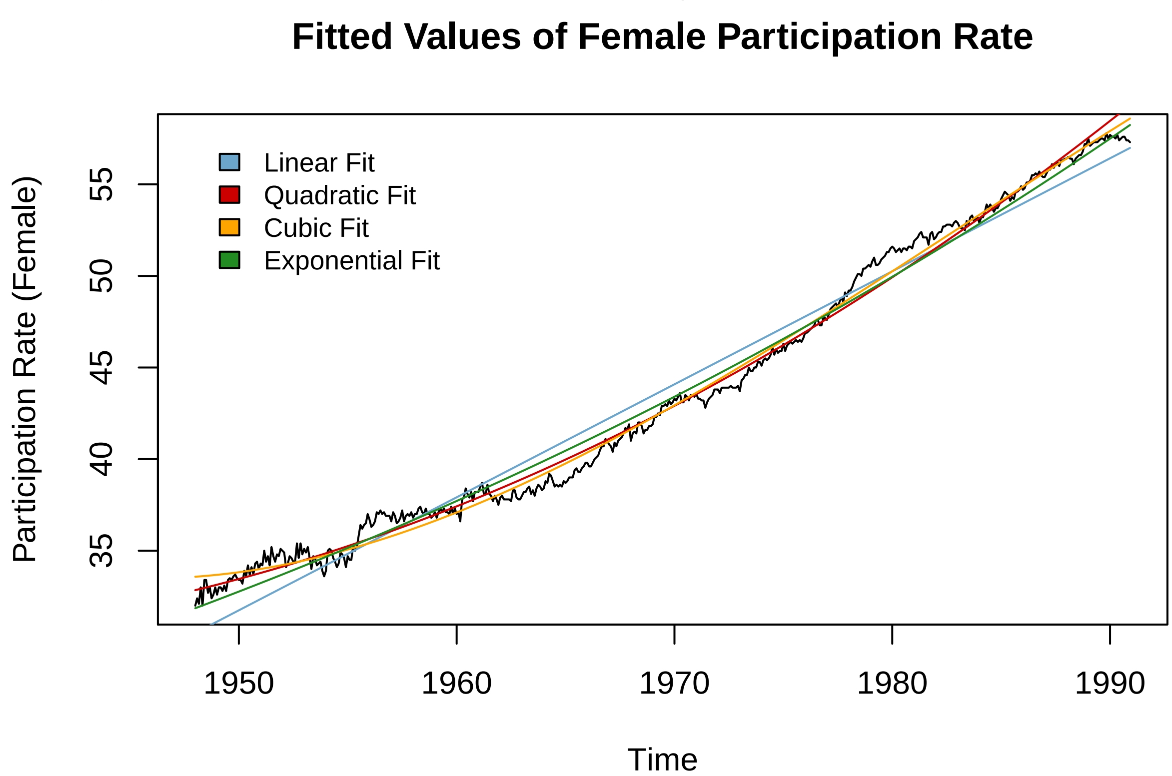


```
summary(m4)

##
## Call:
## lm(formula = log(female_ts) ~ t)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.047288 -0.014909  0.003055  0.014656  0.050510
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.392e+01  1.424e-01  -168.0   <2e-16 ***
## t            1.406e-02  7.231e-05   194.4   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02039 on 514 degrees of freedom
## Multiple R-squared:  0.9866, Adjusted R-squared:  0.9866
## F-statistic: 3.778e+04 on 1 and 514 DF,  p-value: < 2.2e-16
```

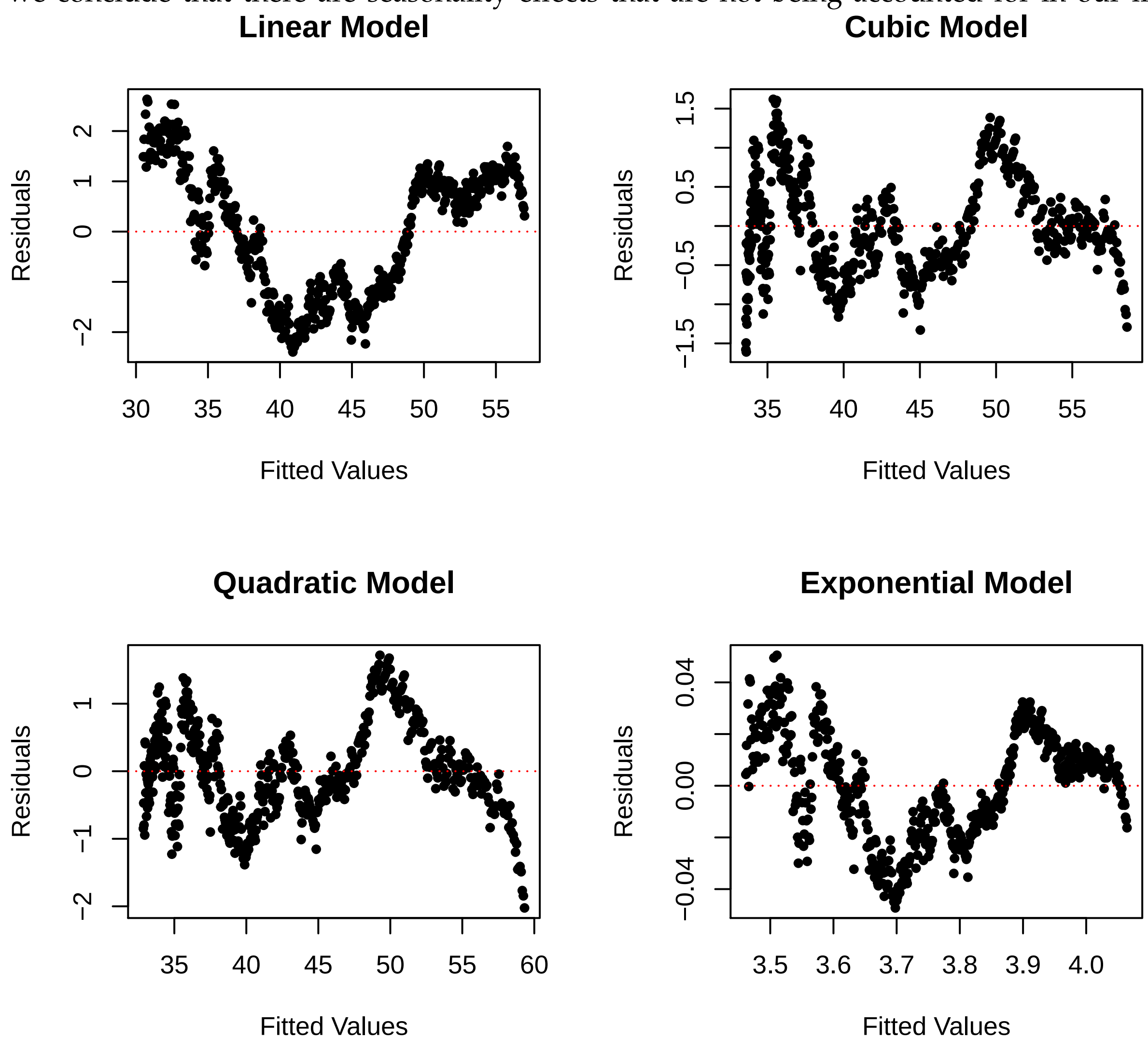
Note: if we wanted to compare the fitted values from the exponential model to that of the fitted values from our previous polynomial models, we can simply take the exponential of the fitted values (i.e., $exp(log(\widehat{female_ts})))$.

We compare the fitted values from all four models together:



1c. Plot the residuals v. fitted values for each of the models in the previous section.

For the linear model, we see somewhat of a parabolic pattern in the residuals, indicating that perhaps a polynomial fit might be better for the model. Likewise, the exponential model's plot of residuals also appears to have a parabolic pattern in the residuals. However, looking at the residuals of the polynomial models, we see that while the patterns are not as apparent as that of the residuals from the linear and exponential models, there is still a large degree of structure present; therefore, we conclude that there are seasonality effects that are not being accounted for in our model.



1d. Based on AIC/BIC values, choose the best fit model.

Based on the AIC/BIC values, the cubic model appears to be the best fit; the two values converge to the same result.

```
AIC(m1, m2, m3, m4)
```

```
##    df       AIC
## m1  3  1703.4624
```

```
## m2  4  1079.0142
## m3  5   985.9713
## m4  3 -2548.9657
```

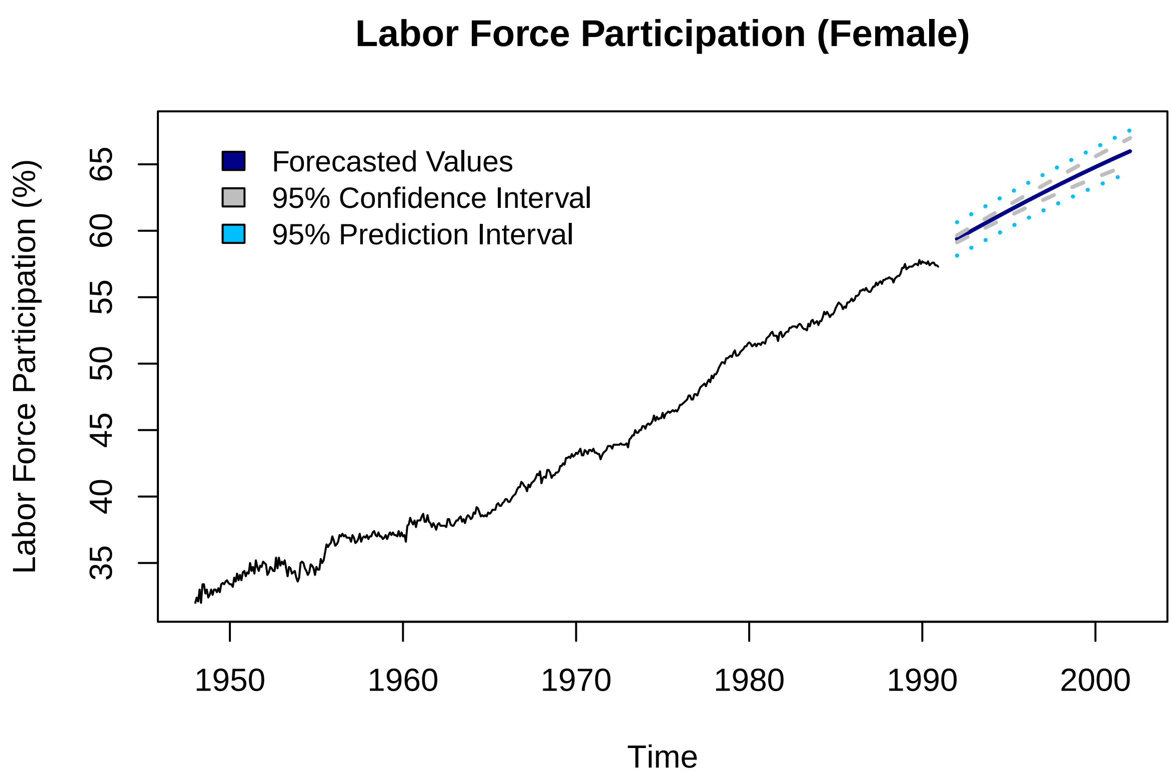
```r
BIC(m1, m2, m3, m4)
```

```
##     df       BIC
## m1   3  1716.201
## m2   4  1095.999
## m3   5  1007.202
## m4   3 -2536.227
```

1e. Using your selected best fit model, forecast and plot your estimated female labor force participation rate for the years 1992 to 2002.

Because the cubic model appears to be the best fit (based on our AIC/BIC values), we use this to forecast the female labor force participation rate. The dotted lines represent the prediction intervals, while the solid gray lines represent the 95% confidence interval. From the graph, we can observe the 95% confidence interval becoming wider as we attempt to forecast further into the future. Using this trend, we can observe that there our forecast predicts that the female labor force participation to increase over time. However, we can see that most of the observed fluctuations present in the series are not being accounted for with a simple trend model.

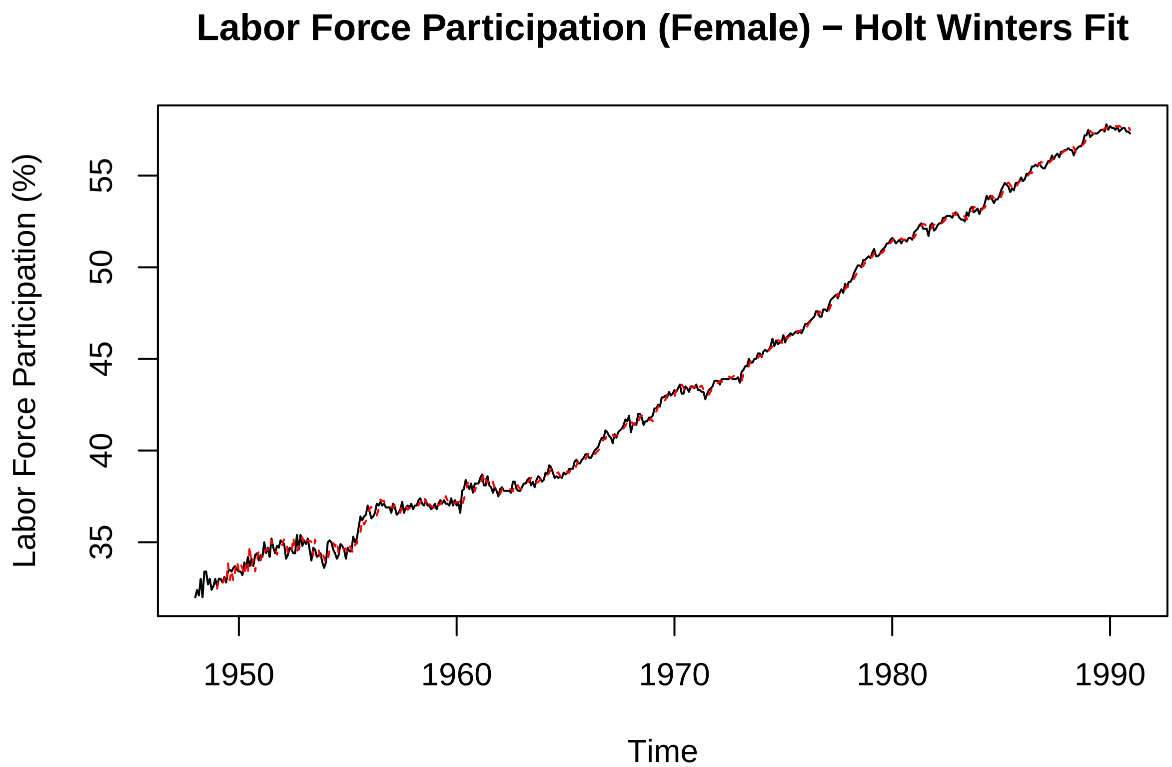
```r
tn = data.frame(t = seq(1992, 2002))
pred.plim = predict(m3, tn, level = 0.95, interval = "prediction")
pred.clim = predict(m3, tn, level = 0.95, interval = "confidence")
```



1f. Fit a Holt-Winters filter to oyur data, and show the fit. How does this model compare to your best fit model?

Looking at the Holt-Winters fit compared to the historical values, we see that the Holt-Winters fit is a lot closer to the actual values than the polynomial best fit model.

```
hwfit<-HoltWinters(female_ts)
```

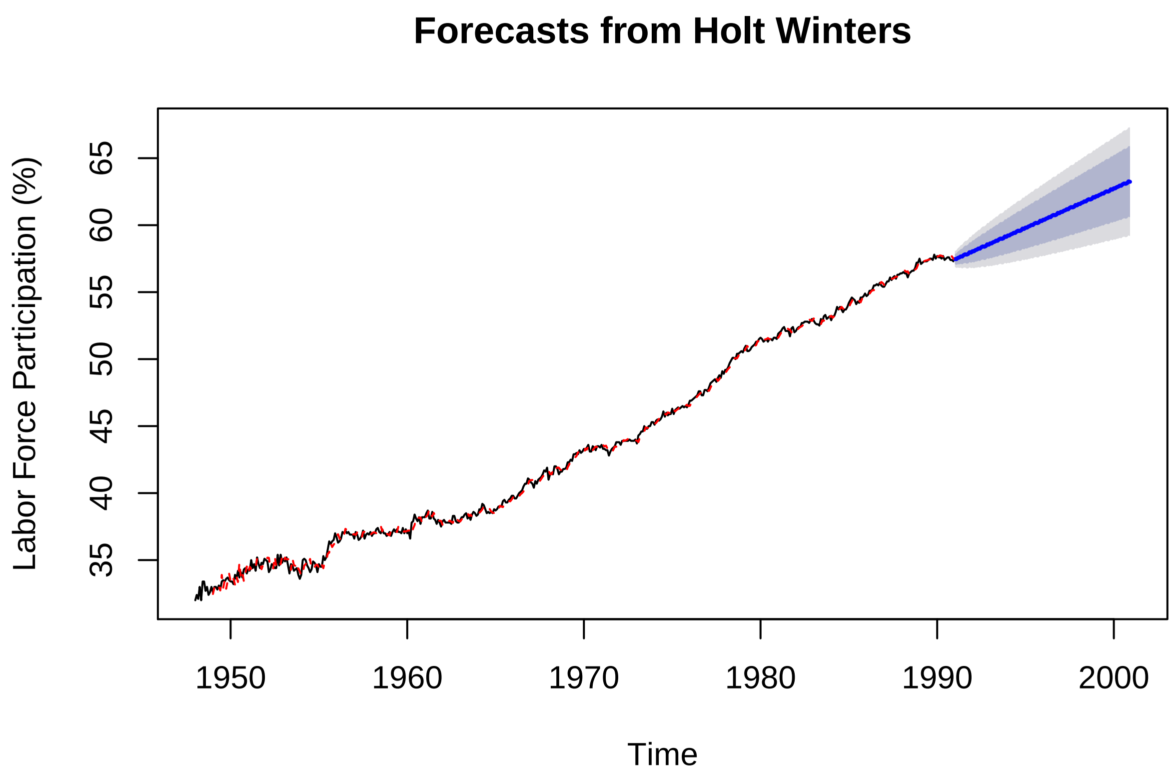


1g. Based on the Holt-Winters fit, forecast and plot your estimated female labor force participation. How does this model compare to your best fit model?

While the fitted values correctly account for seasonal/cyclical behavior in the historical data, the forecast appears to have a rather deterministic nature about the fluctuations that are being modeled. However, it does a better job modeling some of the behaviors present in the series than the polynomial models from before.

Note: There are several ways to forecast models; in this case, we use the `forecast` function. (`predict` would also work. As an added note, `predict` has more robustness when it comes to adding in new data into your predictions, which is necessary for when you are forecasting models that are not entirely univariate, i.e., traditional linear regression models.)

```
plot(forecast(hwfit, h = 120), ylab = "Labor Force Participation (%)",
    xlab = "Time", main = "Forecasts from Holt Winters")
lines(t[-(1:12)], hwfit$fitted[, 1], col = "red", lty = 2)
```



Problem 2.

We begin by downloading our data, using the `getSymbols` function from the library `quantmod` to access FRED data:

```
library(dplyr)

# CPI
getSymbols("CPALTT01USM661S", src = "FRED")
cpi <- as.data.frame(CPALTT01USM661S)
data_all <- data.frame(dates = as.Date(row.names(cpi)), cpi = cpi[,
    1])

#

# 3 Month Treasury Bill
getSymbols("TB3MS", src = "FRED")
tb <- as.data.frame(TB3MS)
data_all <- left_join(data_all, data.frame(dates = as.Date(row.names(tb)),
    tb = tb[, 1]), by = "dates")
data_all <- data_all[60:(length(data_all[, 1]) - 1), ]

# Starts in 1960
cpi_ts <- ts(data_all$cpi, start = 1960, freq = 12)  #starts in 1960
```

```
tb_ts <- ts(data_all$tb, start = 1960, freq = 12)
```

Something to note is that if we are using built in functions to grab data, the starting time of the data may not all necesssarily line up. For example, above, we see that the three month treasury bill contains data prior to 1960, while the CPI data is only provided, starting from 1960. We can do several things to account for this. One would be to interpolate missing data points. Another would be to simply truncate our series so that they are all starting from the same point. Because there are many years included in the interval, and losing a few years is not going to alter the number of observations we have (and for the sake of simplicity), we simply truncate the 3 month Treasury Bill series so that it is of the same length.

We calculate the monthly inflation (which equates to the growth rate of monthly CPI):

```
infl <- diff(log(cpi_ts))
real_int <- tb_ts[-1] - infl
```

We now add in the real interest rate into our model estimated from last homework:

```
# From last time:
getSymbols("PCE", src = "FRED")
pce <- as.data.frame(PCE)
pce_ts <- ts(pce$PCE[13:(length(pce$PCE))], start = 1960, freq = 12)
# Growth rates:
log.pce <- log(pce_ts)
r_pce <- diff(log.pce)

# Disposable personal income:
getSymbols("DSPIC96", src = "FRED")
dpi <- as.data.frame(DSPIC96)
dpi_ts <- ts(dpi$DSPIC96[13:length(dpi$DSPIC96)], start = 1960,
    freq = 12)
# Growth rates:
log.dpi <- log(dpi_ts)
r_dpi <- diff(log.dpi)

model<-lm(r_pce~r_dpi+real_int)
summary(model)
```

```
##
## Call:
## lm(formula = r_pce ~ r_dpi + real_int)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0261648 -0.0028247 -0.0000727  0.0030049  0.0247749
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.277e-03  3.586e-04   9.140  < 2e-16 ***
## r_dpi       1.015e-01  2.756e-02   3.681  0.00025 ***
```

```
## real_int     3.992e-04  6.288e-05   6.348 3.94e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.005264 on 694 degrees of freedom
## Multiple R-squared:  0.07245,    Adjusted R-squared:  0.06977
## F-statistic:  27.1 on 2 and 694 DF,  p-value: 4.641e-12
```

From the regression results, we see that there is a statistically significant negative relationship between real interest rates and personal consumption expenditure. This is consistent with economic literature, which has argued that as interest rates increase, consumption decreases, as the potential opportunity cost of not saving money goes up.

Problem 3.

We use the same data from last homework on quarterly and annual housing prices and interest rates, as well as their respective growth rates.

We set up a model for quarterly price growth, using lags of one, two, three, and four:

$$price_t = \alpha_t + \beta_1 price_{t-1} + \beta_2 price_{t-2} + \beta_3 price_{t-3} + \beta_4 price_{t-4} + \varepsilon_t$$

Note: this is essentially an autoregressive model of lag 4 (i.e., AR(4)).

```
#Quarterly:
m_q<-read.csv('Mortgage_quarterly.csv')
names(m_q)<-c("time", "price", "rates")
r_price_q<-diff(log(m_q$price))
r_rates_q<-diff(log(m_q$rates))

#Run seveal regression models w/ one, two, three, and four lags of price growth
#Lag 1
lag.price1<-r_price_q[-length(r_price_q)]
lag.price2<-lag.price1[-length(lag.price1)]
lag.price3<-lag.price2[-length(lag.price2)]
lag.price4<-lag.price3[-length(lag.price3)]
lag.price1<-append(NA, lag.price1)
lag.price2<-append(c(NA, NA), lag.price2)
lag.price3<-append(c(NA, NA, NA), lag.price3)
lag.price4<-append(c(NA, NA, NA, NA), lag.price4)
model<-lm(r_price_q~lag.price1+lag.price2+lag.price3+lag.price4)
summary(model)

##
## Call:
## lm(formula = r_price_q ~ lag.price1 + lag.price2 + lag.price3 +
##     lag.price4)
##
```

```
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.034072 -0.003491  0.000696  0.003942  0.021627
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.883e-05  9.383e-04   0.105   0.9163
## lag.price1   7.815e-01  9.217e-02   8.479 8.39e-14 ***
## lag.price2  -6.370e-01  1.061e-01  -6.001 2.29e-08 ***
## lag.price3   6.087e-01  1.081e-01   5.630 1.28e-07 ***
## lag.price4   1.745e-01  9.618e-02   1.814   0.0723 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.008413 on 116 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.7316, Adjusted R-squared:  0.7223
## F-statistic: 79.04 on 4 and 116 DF,  p-value: < 2.2e-16
```

Based on the regression output, we see that the the first three lags are statistically significant at the $\alpha = 5\%$ level. Both the first lag and the third lag are positively related with the price growth. However, the second lag is negatively related. The fourth lag is not statistically significant, with a p-value of 0.0723. Because these are quarterly observations, this could mean that there exists not as much statistical significance in the relationship between observations 4 quarters apart, i.e., one year.

We compare this to the **annual** models implemented before. We should expect there to be some differences in the results, as these are now based on annual lags. The first model is as follows:

$$price_t = \alpha_t + \beta_1 price_{t-1} + \beta_2 price_{t-2} + \varepsilon_t$$

```
# Load Annual Data
m <- read.csv("Mortgage.csv")
head(m)
```

```
##   Year        P  R..in...
## 1 1975 25.54491  9.038333
## 2 1976 27.40669  8.863333
## 3 1977 30.48179  8.841667
## 4 1978 34.70104  9.631667
## 5 1979 39.14988 11.192500
## 6 1980 42.76916 13.769167
```

```
names(m) <- c("year", "price", "rates")
# Growth:
r_price <- diff(log(m$price))
r_rates <- diff(log(m$rates))

lag.price_annual1 <- r_price[-length(r_price)]
lag.price_annual2 <- lag.price_annual1[-length(lag.price_annual1)]
```

```r
lag.price_annual1 <- append(NA, lag.price_annual1)
lag.price_annual2 <- append(c(NA, NA), lag.price_annual2)

model1 <- lm(r_price ~ lag.price_annual1 + lag.price_annual2)
summary(model1)

##
## Call:
## lm(formula = r_price ~ lag.price_annual1 + lag.price_annual2)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.075902 -0.008761  0.006968  0.015660  0.031019
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.006284   0.006204   1.013  0.31894
## lag.price_annual1  1.269847   0.157396   8.068 4.13e-09 ***
## lag.price_annual2 -0.485851   0.161879  -3.001  0.00527 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02556 on 31 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.7791, Adjusted R-squared:  0.7649
## F-statistic: 54.68 on 2 and 31 DF,  p-value: 6.823e-11
```

The second model includes lagged values from interest rate growth as well:

$$price_t = \alpha_t + \beta_1 price_{t-1} + \beta_2 price_{t-2} + \beta_3 int_{t-1} + \beta_4 int_{t-2} + \varepsilon_t$$

```r
lag.rates_annual1 <- r_rates[-length(r_rates)]
lag.rates_annual2 <- lag.rates_annual1[-length(lag.rates_annual1)]
lag.rates_annual1 <- append(NA, lag.rates_annual1)
lag.rates_annual2 <- append(c(NA, NA), lag.rates_annual2)

model2 <- lm(r_price ~ lag.price_annual1 + lag.price_annual2 +
    lag.rates_annual1 + lag.rates_annual2)
summary(model2)

##
## Call:
## lm(formula = r_price ~ lag.price_annual1 + lag.price_annual2 +
##     lag.rates_annual1 + lag.rates_annual2)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.064875 -0.004436  0.005939  0.014799  0.030749
```

```
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        9.328e-05  7.031e-03   0.013   0.9895
## lag.price_annual1  1.324e+00  1.707e-01   7.755  1.5e-08 ***
## lag.price_annual2 -4.387e-01  1.962e-01  -2.236   0.0332 *
## lag.rates_annual1 -1.185e-01  4.743e-02  -2.497   0.0184 *
## lag.rates_annual2  4.381e-02  5.134e-02   0.853   0.4005
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02386 on 29 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.8199, Adjusted R-squared:  0.795
## F-statistic:    33 on 4 and 29 DF,  p-value: 2.07e-10
```

We see that the there is statistical significance in including lagged values of interest rate growth. The adjusted R^2 of the second model (with the interest rate growth lagged values) shows an increase from 0.7649 to 0.795, with respect to the first model, with only lagged values of price growth. However, we would need to see whether the inclusion of the interest rate growth lagged values is statistically significant in its increase in model performance.

As a note, it is incorrect to use AIC and BIC to compare the quarterly model with the annual models, because our dependent variable is different (one is on a quarterly frequency, and one is on an annual frequency).

We pick the quarterly model to test for parameter stability. We begin by using the recursive scheme. We partition the data into 2/3rds training and 1/3rd testing; at each iteration, we add in data from the testing set into the training set, and re-estimate the coefficients.

```r
# Recursive scheme for the model proposed with quarterly
# data: 2/3rds:
n <- round(length(r_price_q) * 2/3)
df1 <- data.frame(r_price_q, lag.price1, lag.price2, lag.price3,
    lag.price4)
coefficients <- c()

for (i in n:length(r_price_q)) {
    df2 <- df1[1:i, ]
    model <- lm(r_price_q ~ ., data = df2)
    coefficients <- rbind(coefficients, model$coefficients)
}
```

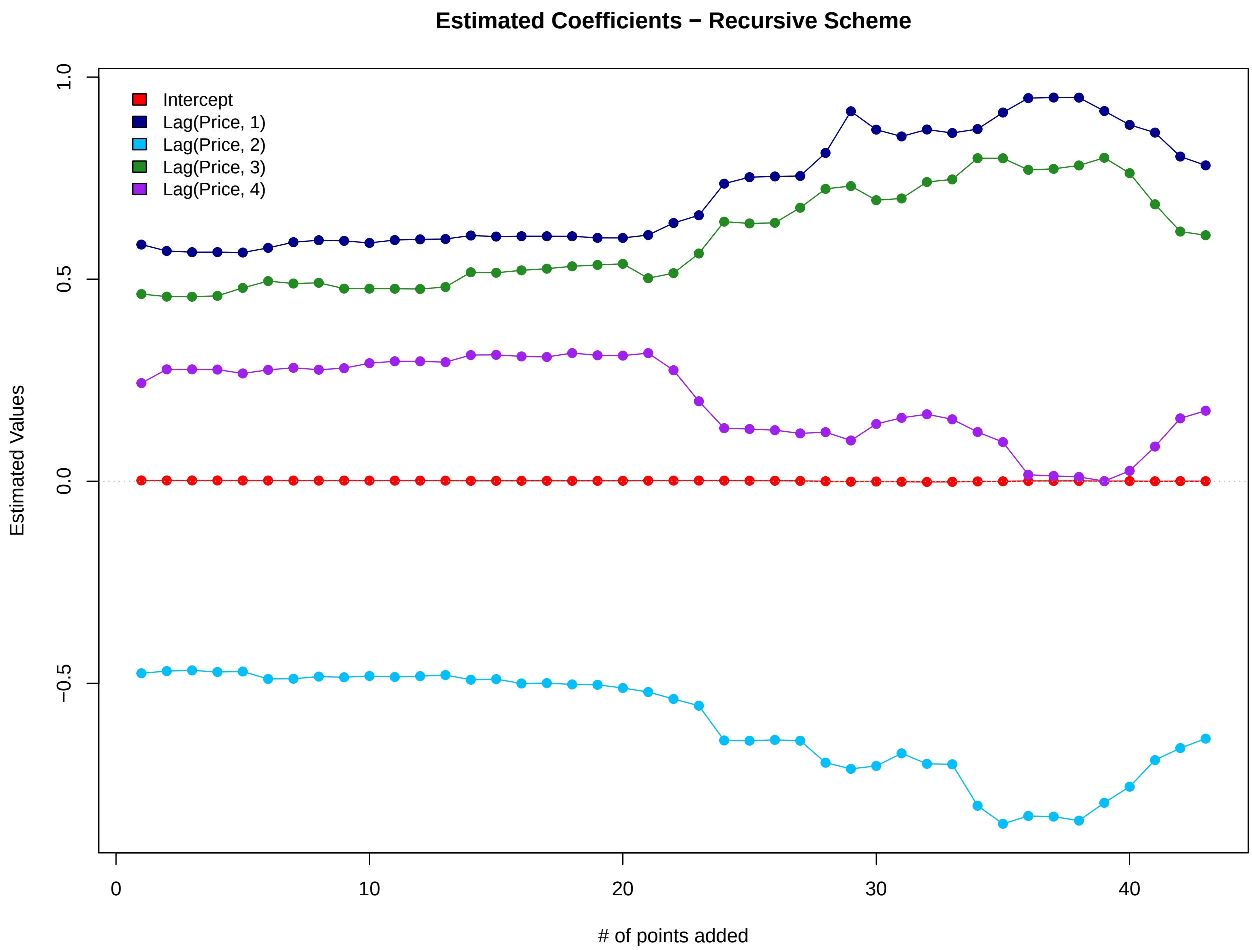
We also use a moving window.

```r
# Moving window:
coefficients2 <- c()
for (i in 0:(length(r_price_q) - n)) {
    df2 <- df1[1 + i:n + i, ]
    model <- lm(r_price_q ~ ., data = df2)
```

```
    coefficients2 <- rbind(coefficients2, model$coefficients)
}
```

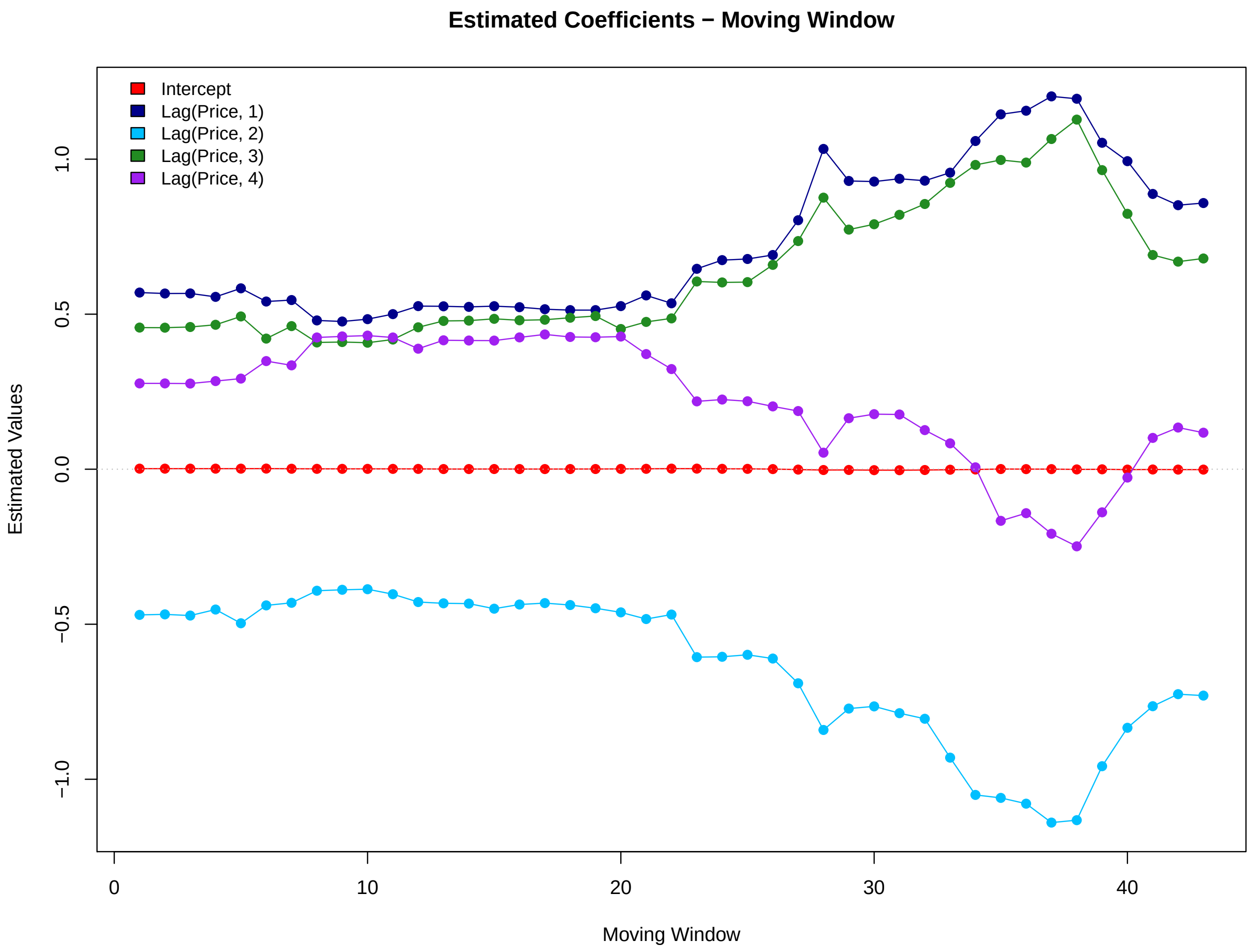
In both cases, we want to check whether the inclusion of new data points at a period of time will lead to large changes in our estimated parameters. If this is the case, that would lead to some problems, because the estimated effects of the independent variables on the dependent variable would be dependent upon the time frame we are looking at. Therefore, if we find that the parameters are not robust across time, then that would lead to concerns as to whether or not the forecasts we generate from them are robust and accurate as well.

Plotting out the estimated coefficients from our recursive scheme over time, we see the following:



In all four parameters, we find that for the first 20 points added into our model, the parameters remain largely constant. However, as we start adding in more data points, the estimated effect from price growth observations at a lag of four diminishes, while the other estimated effects appear to be exacerbated.

We look at the estimated coefficients from the moving window:

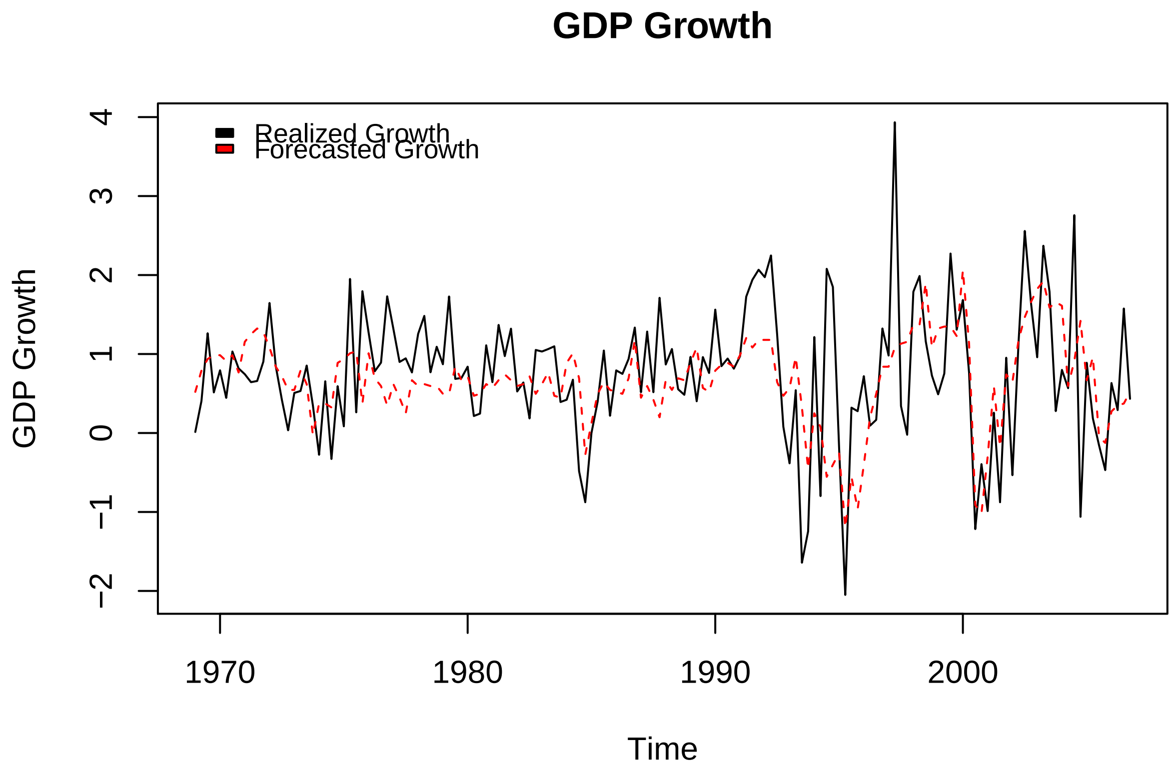


The results seem to be similar in behavior with the recursive testing scheme, but with slightly more fluctuation.

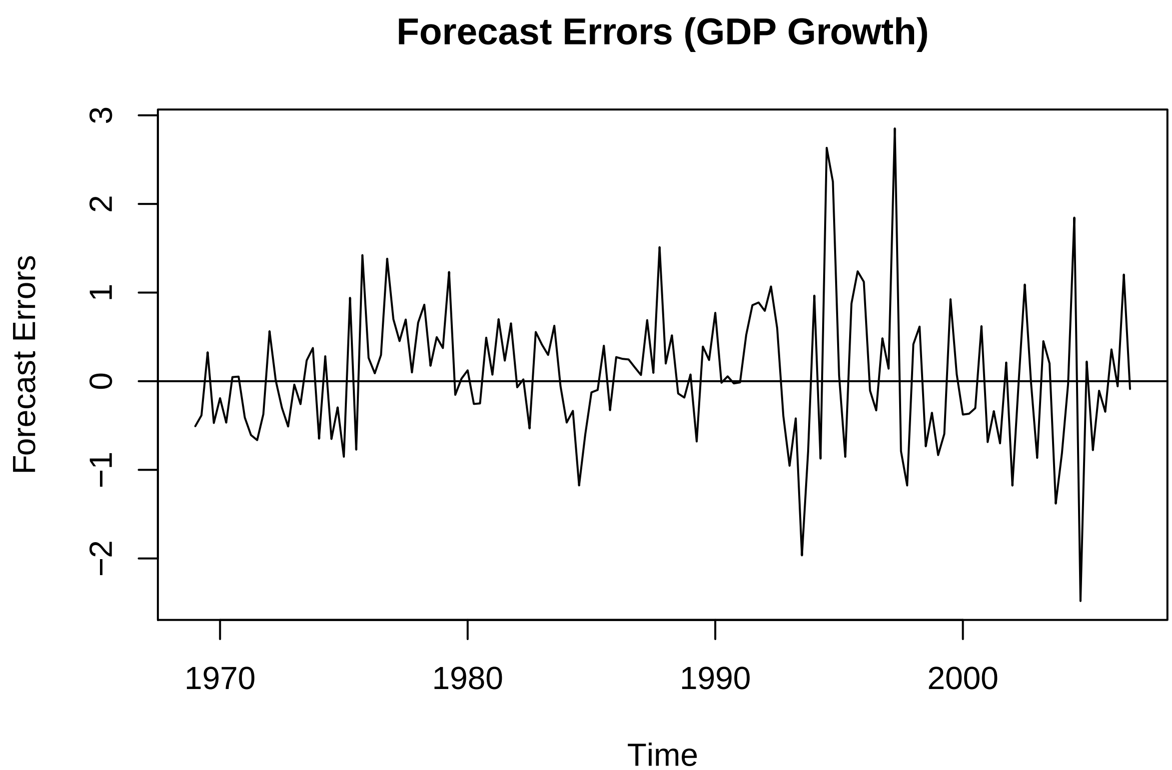
For completeness, it would be best to also check the statistical significance of the estimated coefficients through time (It is pretty clear that $price_{t-4}$ appears to fluctuate between being statistically significant, as the estimated coefficients fluctuate between being positive and negative.) We could do that pretty easily by also plotting the confidence intervals associated with the coefficient estimates, but... we'll save that for next time.

Problem 4.

Compare real GDP growth with that of the forecasted GDP growth values. We begin by looking at a plot of the two series:



Plotting the forecast errors, we see the following:

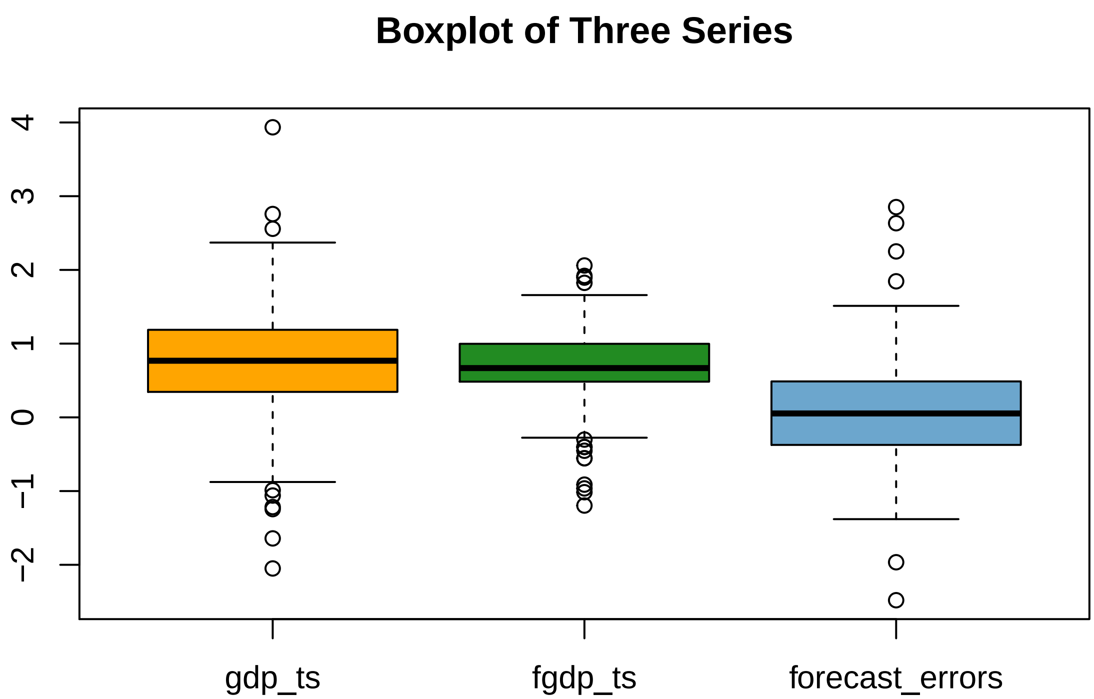


When comparing the plot of the realized values versus the forecasts, the forecasted values appear to relatively close to actual values of GDP growth; however, looking at the plot of forecast errors, it appears that the forecasts underestimates the real values. This is particularly apparent during the period of time between 1995-2000. There also appears to be much more error in the forecasts after 1990, as seen by the increased fluctuations in forecast errors after 1990.

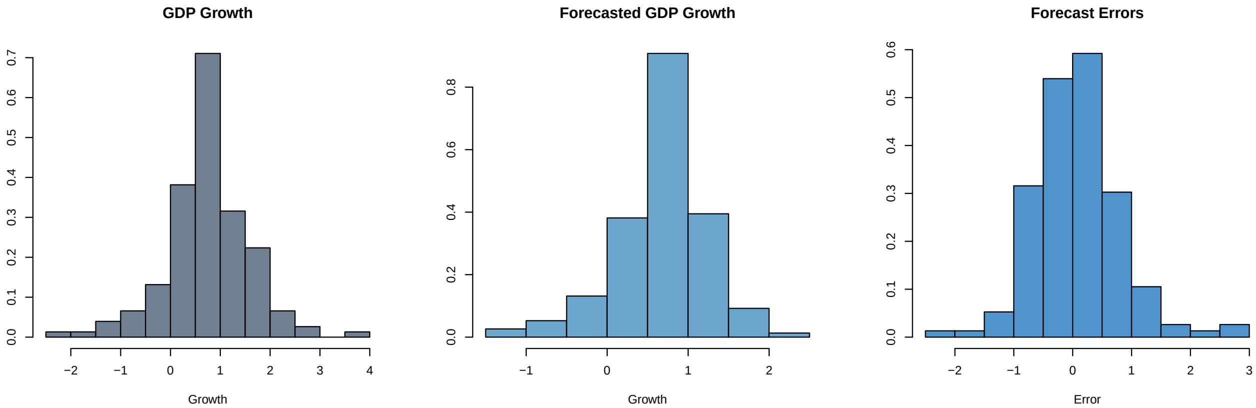
Summary statistics are as follows:

```
##           Series          Min.        1st Qu.       Median
## 1            GDP -2.049666294   0.347508868 0.7678891755
## 2 Forecasted GDP -1.196296175 0.49013303075  0.668271308
## 3         Errors -2.480299237 -0.3723629135 0.0532554015
##                 Mean       3rd Qu.         Max.
## 1  0.753937694848684 1.17407228825 3.934342233
## 2  0.676474517815789  0.9914072985 2.060439584
## 3 0.0774631770328947  0.4858156035 2.852040026
```

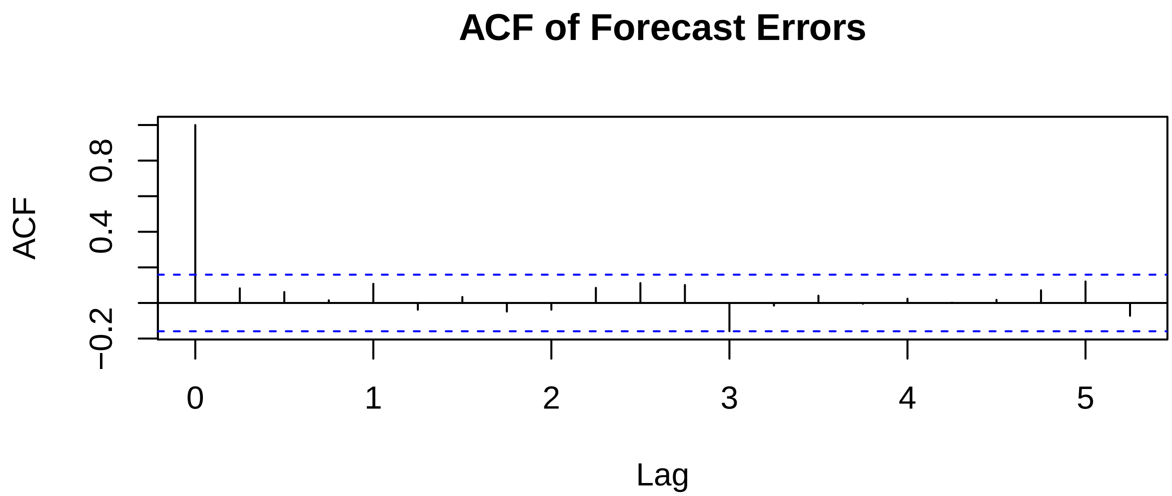
Looking at the summary statistics, the forecasted GDP growth values are centered at 0.67, while the actual GDP growth values are centered ot 0.75; this is consistent with our earlier observations that the forecasted values slightly underestimate the actual values. Also, looking at the combined histograms of the actual GDP growth values and the forecasted GDP growth values, there is a larger spread in the actual values, whereas the forecasted values have less of a spread. This can be visualized by looking at a boxplot:

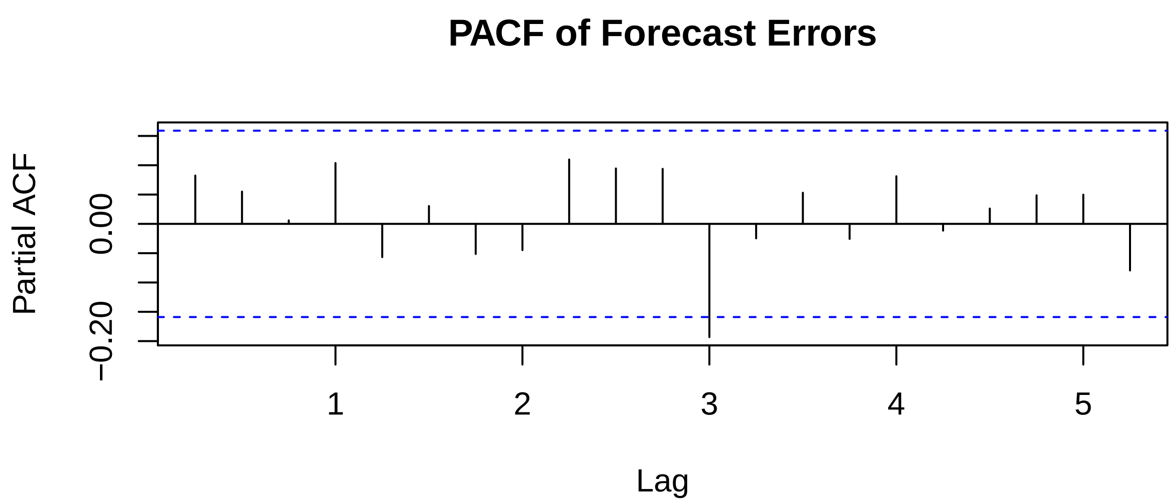


From the histograms, we see that the distributions may be slightly skewed.



We look at the autocorrelation functions for the forecast errors. We see that it has mostly been reduced to white noise, which means that there is very little structure left in what are essentially the residuals of the model:

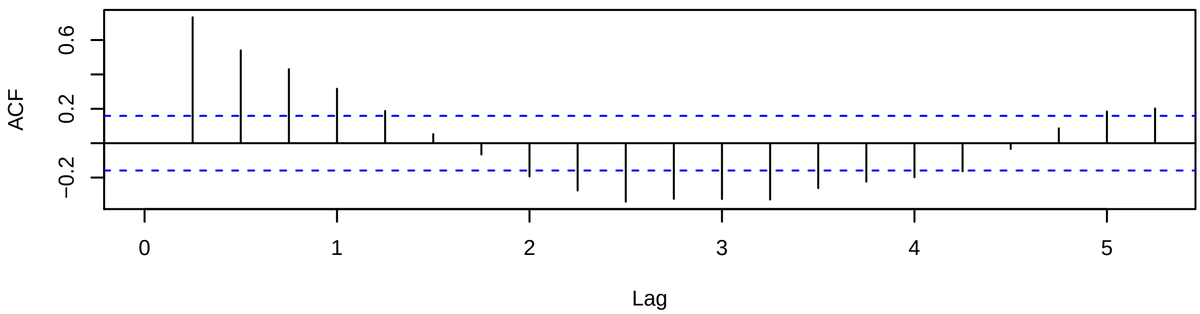




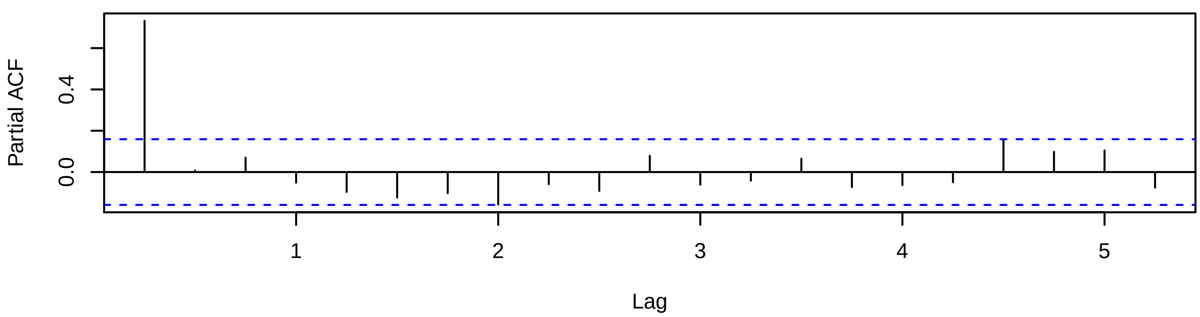
Based on the plots of the autocorrelation functions, realized values of GDP growth are statistically significantly correlated with lagged values of 1, 2, 8, and 12; the PACF of realized values of GDP growth indicates that there is some relation with lagged values of 1, 8, and 12. The autocorrelation functions for forecasted values of GDP growth show a sinusoidal pattern. The autocorrelation functions for the forecast errors show little statistically significant values; however, the PACF has a statistically significant value at a lag of 12. There are signs of time dependence.

Note: We refer to lags in terms of period lags. However, on the graph, the axis is in terms of years (i.e., a lag of 1 on the graph is actually a lag of 1 year, not 1 quarter).

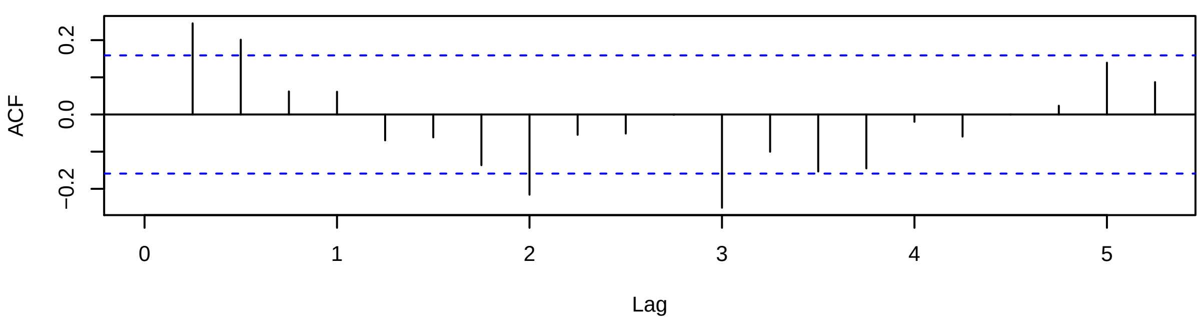
ACF of Forecasted GDP (Lag 0 removed)



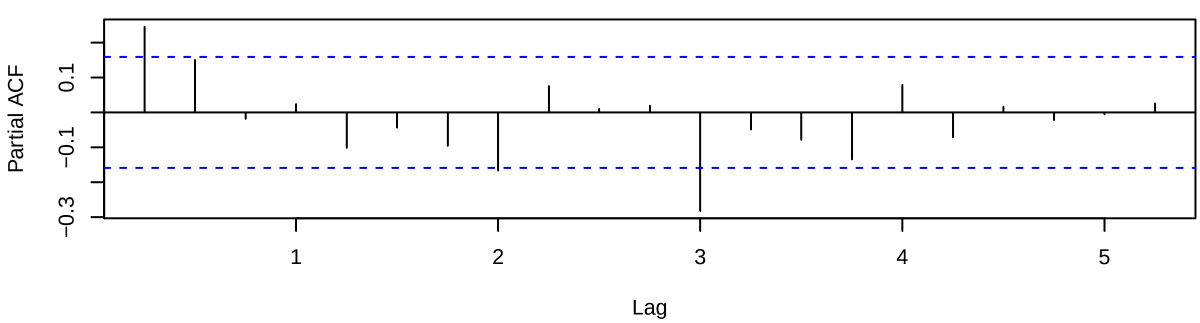
PACF of Forecasted GDP



ACF of Realized Values of GDP (Lag 0 removed)



PACF of Realized Values of GDP

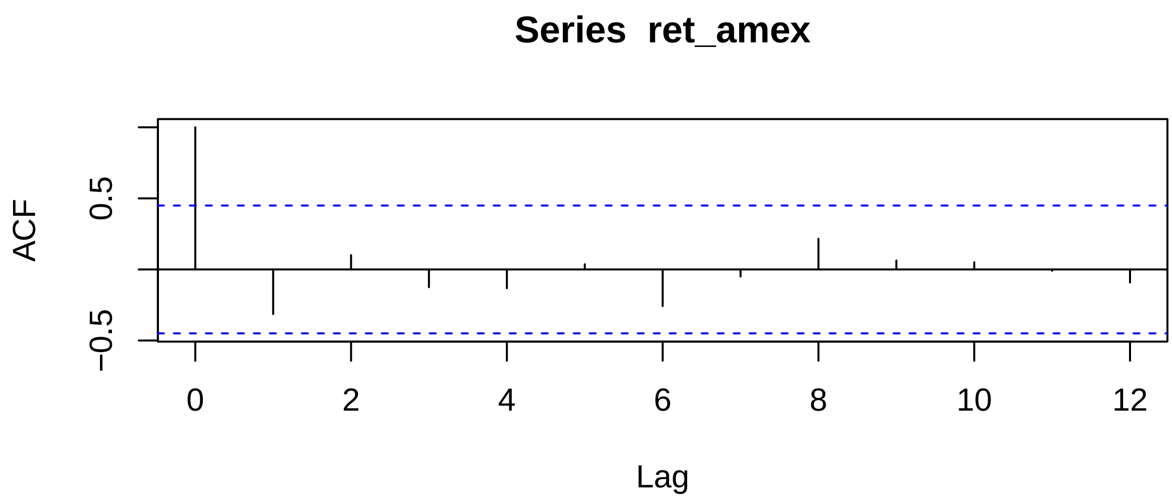


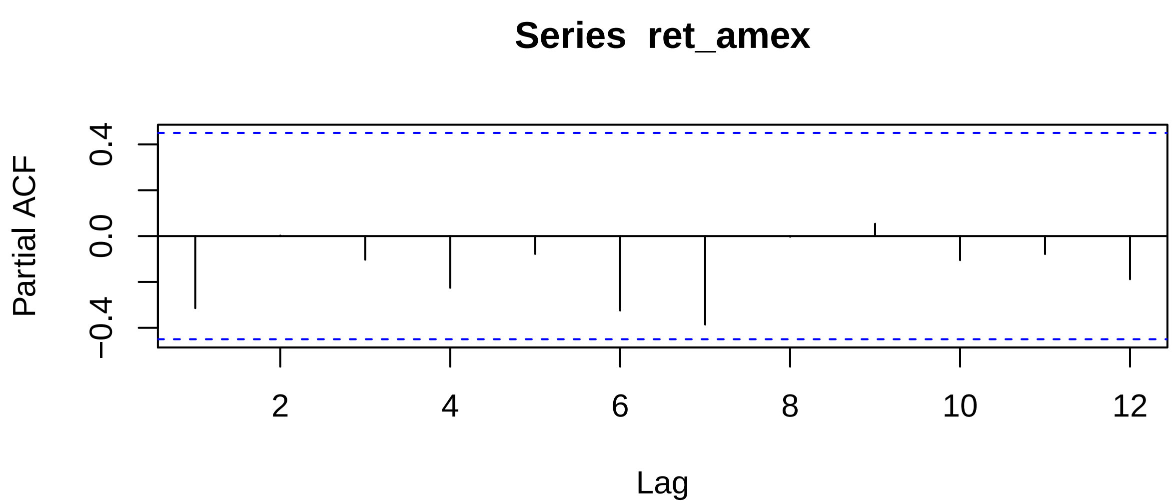
Problem 5.

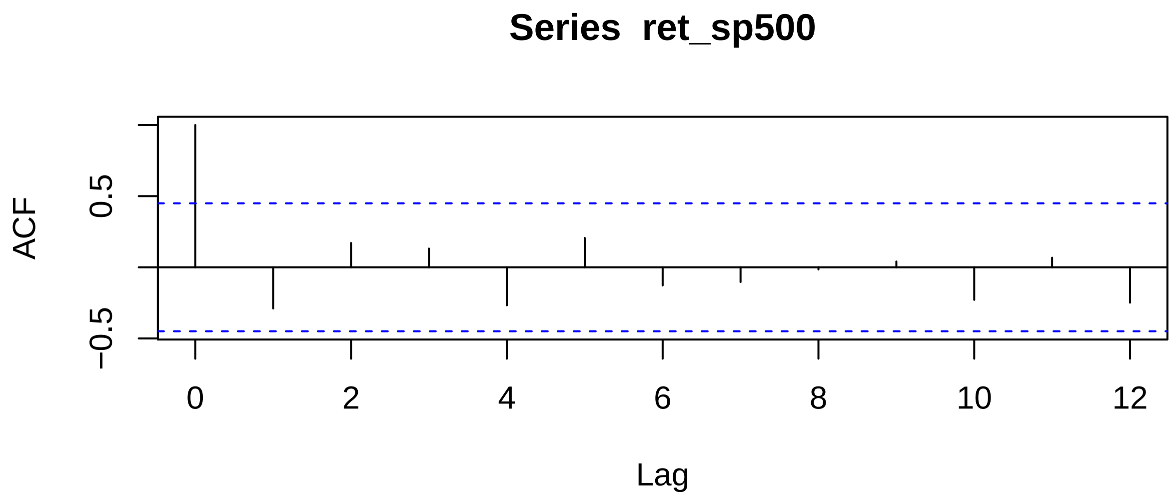
In the original figure provided of the autocorrelation functions, we find that a spike in the ACF at lag 1, and a spike in the PACF at lag 1 for the AMEX portfolio. However, upon updating the observations, we find that there exists no spikes the ACF/PACF. This is probably due to the fact that as time has gone on, markets have grown increasingly efficient, which means that hypothetically speaking, stock returns should not be displaying dependencies upon its past values, and exhibit behavior much closer to that of a random walk.

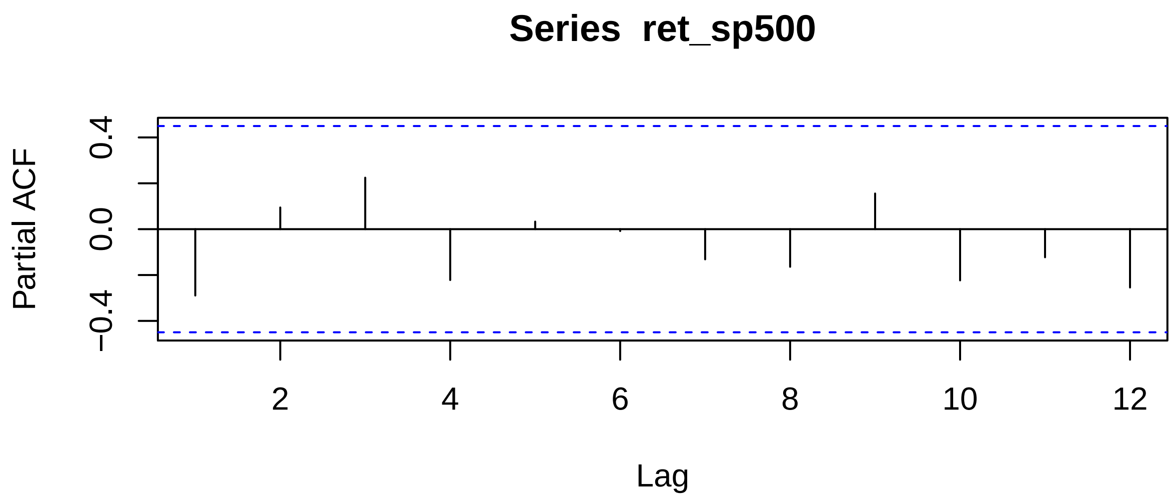
The same is seen in the S&P 500, though the original figure does not appear to show statistically significant spikes of large magnitudes either.

```
ret_sp500<-getReturns("^GSPC", start="1995-01-03")
ret_amex<-getReturns("^XAX", start="1995-01-03")
```






Series ret_sp500
ACF
0.5
−0.5
0
2
4
6
8
10
12
Lag


Series ret_sp500
Partial ACF
0.4
0.0
−0.4
2
4
6
8
10
12
Lag

R-Code:

```
rm(list = ls(all = TRUE))

# Load Libraries
library(lattice)
library(foreign)
library(MASS)
library(car)
require(stats)
require(stats4)
library(KernSmooth)
library(fastICA)
library(cluster)
library(leaps)
library(mgcv)
library(rpart)
library(pan)
library(mgcv)
library(DAAG)
library("TTR")
library(tis)
require("datasets")
require(graphics)
library("forecast")
require(astsa)
library(xtable)
library(stats)
library(quantmod)
library(stockPortfolio)

# Problem 1.

# Loading the data:

data <- read.table("~/Documents/Econ 144/Data Sets/HW 2/labordata.dat",
    quote = "\"", comment.char = "")
colnames(data) <- c("total", "female", "male")
male_ts <- ts(data$male, start = 1948, freq = 12)
female_ts <- ts(data$female, start = 1948, freq = 12)
total_ts <- ts(data$total, start = 1948, freq = 12)
all_ts <- ts(data, start = 1948, freq = 12)
t <- seq(1948, length = length(male_ts), by = 1/12)

# a.  Plot
quartz()
plot(all_ts, plot.type = "s", col = c("midnightblue", "skyblue3",
```

```r
    "slategray"), lwd = 2, ylab = "Labor Force Participation (%)",
    xlab = "Time", main = "Labor Force Participation Rates")
legend("bottom", c("Total", "Male", "Female"), fill = c("midnightblue",
    "skyblue3", "slategray"), cex = 1, bty = "n", horiz = T)

# b.
m1 <- lm(female_ts ~ t)
quartz()
plot(female_ts, main = "Linear Fit", ylab = "Participation Rate (Female)",
    xlab = "Time")
lines(t, m1$fit, col = "red4", lwd = 2)
summary(m1)

# Polynomial Quadratic
m2 <- lm(female_ts ~ t + I(t^2))
quartz()
plot(female_ts, main = "Quadratic Fit", ylab = "Participation Rate (Female)",
    xlab = "Time")
lines(t, m2$fit, col = "red4", lwd = 2)
summary(m2)

# Cubic
m3 <- lm(female_ts ~ t + I(t^2) + I(t^3))
quartz()
plot(female_ts, main = "Cubic Fit", ylab = "Participation Rate (Female)",
    xlab = "Time")
lines(t, m3$fit, col = "red4", lwd = 2)
summary(m3)

# Exponential
m4 <- lm(log(female_ts) ~ t)
quartz()
plot(log(female_ts), main = "Exponential Fit", ylab = "Participation Rate (Female)",
    xlab = "Time")
lines(t, m4$fit, col = "red4", lwd = 2)
summary(m4)

plot(female_ts, main = "Fitted Values of Female Participation Rate",
    ylab = "Participation Rate (Female)", xlab = "Time")
lines(t, m1$fit, col = "skyblue3")
lines(t, m2$fit, col = "red3")
lines(t, m3$fit, col = "orange")
lines(t, exp(m4$fit), col = "forestgreen")
legend(1948, 58, c("Linear Fit", "Quadratic Fit", "Cubic Fit",
    "Exponential Fit"), fill = c("skyblue3", "red3", "orange",
    "forestgreen"), bty = "n")
```

```r
# c.
quartz()
par(mfcol = c(2, 2))
plot(m1$fit, m1$res, ylab = "Residuals", pch = 20, xlab = "Fitted Values",
    main = "Linear Model")
plot(m2$fit, m2$res, ylab = "Residuals", pch = 20, xlab = "Fitted Values",
    main = "Quadratic Model")
plot(m3$fit, m3$res, ylab = "Residuals", pch = 20, xlab = "Fitted Values",
    main = "Cubic Model")
plot(m4$fit, m4$res, ylab = "Residuals", pch = 20, xlab = "Fitted Values",
    main = "Exponential Model")

# d. AIC/BIC
AIC(m1, m2, m3, m4)
BIC(m1, m2, m3, m4)
# Best fit here appears to be the cubic.

# e.
tn = data.frame(t = seq(1992, 2002))
pred.plim = predict(m3, tn, level = 0.95, interval = "prediction")
pred.clim = predict(m3, tn, level = 0.95, interval = "confidence")

plot(t, female_ts, xlim = range(t, tn), type = "l", ylim = range(female_ts,
    pred.plim), ylab = "Labor Force Participation (%)", xlab = "Time",
    main = "Labor Force Participation (Female)")
lines(tn[, 1], pred.plim[, 1], lwd = 2, col = "darkblue")
lines(tn[, 1], pred.plim[, 2], lty = 3, col = "deepskyblue",
    lwd = 2)
lines(tn[, 1], pred.plim[, 3], lty = 3, col = "deepskyblue",
    lwd = 2)
lines(tn[, 1], pred.clim[, 2], lty = 2, col = "gray", lwd = 2)
lines(tn[, 1], pred.clim[, 3], lty = 2, col = "gray", lwd = 2)
legend(1948, 68, c("Forecasted Values", "95% Confidence Interval",
    "95% Prediction Interval"), fill = c("darkblue", "gray",
    "deepskyblue"), bty = "n")

# f.
hwfit <- HoltWinters(female_ts)
quartz()
plot(t, female_ts, type = "l", ylab = "Labor Force Participation (%)",
    xlab = "Time", main = "Labor Force Participation (Female) - Holt Winters Fit")
lines(t[-(1:12)], hwfit$fitted[, 1], col = "red", lty = 2)

# g.
quartz(width = 11, height = 8.5)
plot(forecast(hwfit, h = 120), ylab = "Labor Force Participation (%)",
    xlab = "Time", main = "Forecasts from Holt Winters")
```

```r
lines(t[-(1:12)], hwfit$fitted[, 1], col = "red", lty = 2)


# Problem 2 (3.2 from the Textbook)

# CPI
getSymbols("CPALTT01USM661S", src = "FRED")
cpi <- as.data.frame(CPALTT01USM661S)
cpi_ts <- ts(cpi$CPALTT01USM661S, start = 1960, freq = 12)  #starts in 1960

getSymbols("TB3MS", src = "FRED")
tb <- as.data.frame(TB3MS)
which(row.names(tb) == "1960-01-01")
tb_ts <- ts(tb$TB3MS[31:length(tb$TB3MS)], start = 1960, freq = 12)  #starts in 1960

# Calculate monthly inflation (growth rate of monthly CPI):
infl <- diff(log(cpi_ts))
real_int <- tb_ts - infl

# Add the real interest rate to the equation from 1b. From
# last time:
getSymbols("PCE", src = "FRED")
pce <- as.data.frame(PCE)
pce_ts <- ts(pce$PCE[13:length(pce$PCE)], start = 1960, freq = 12)
# Growth rates:
log.pce <- log(pce_ts)
r_pce <- diff(log.pce)

# Disposable personal income:
getSymbols("DSPIC96", src = "FRED")
dpi <- as.data.frame(DSPIC96)
dpi_ts <- ts(dpi$DSPIC96[13:length(dpi$DSPIC96)], start = 1960,
    freq = 12)
# Growth rates:
log.dpi <- log(dpi_ts)
r_dpi <- diff(log.dpi)

model <- lm(r_pce ~ r_dpi + real_int)
summary(model)

# Problem 3

# Load all the data again... Quarterly:
m_q <- read.csv("~/Documents/Econ 144 (S17)/Homework 1 Solutions/Mortgage_quarterly.csv")
names(m_q) <- c("time", "price", "rates")
r_price_q <- diff(log(m_q$price))
r_rates_q <- diff(log(m_q$rates))
```

```r
# Setting up lags:
lag.price1 <- r_price_q[-length(r_price_q)]
lag.price2 <- lag.price1[-length(lag.price1)]
lag.price3 <- lag.price2[-length(lag.price2)]
lag.price4 <- lag.price3[-length(lag.price3)]
lag.price1 <- append(NA, lag.price1)
lag.price2 <- append(c(NA, NA), lag.price2)
lag.price3 <- append(c(NA, NA, NA), lag.price3)
lag.price4 <- append(c(NA, NA, NA, NA), lag.price4)
model <- lm(r_price_q ~ lag.price1 + lag.price2 + lag.price3 +
    lag.price4)
summary(model)

# 4.1.1 Equations Annual
m <- read.csv("~/Documents/Econ 144/Data Sets/HW 1/Mortgage.csv")
head(m)
names(m) <- c("year", "price", "rates")
# Growth:
r_price <- diff(log(m$price))
r_rates <- diff(log(m$rates))

lag.price_annual1 <- r_price[-length(r_price)]
lag.price_annual2 <- lag.price_annual1[-length(lag.price_annual1)]
lag.price_annual1 <- append(NA, lag.price_annual1)
lag.price_annual2 <- append(c(NA, NA), lag.price_annual2)

model1 <- lm(r_price ~ lag.price_annual1 + lag.price_annual2)
summary(model1)


lag.rates_annual1 <- r_rates[-length(r_rates)]
lag.rates_annual2 <- lag.rates_annual1[-length(lag.rates_annual1)]
lag.rates_annual1 <- append(NA, lag.rates_annual1)
lag.rates_annual2 <- append(c(NA, NA), lag.rates_annual2)

model2 <- lm(r_price ~ lag.price_annual1 + lag.price_annual2 +
    lag.rates_annual1 + lag.rates_annual2)
summary(model2)

# Recursive scheme for the model proposed with quarterly
# data:

# 2/3rds:
n <- round(length(r_price_q) * 2/3)
df1 <- data.frame(r_price_q, lag.price1, lag.price2, lag.price3,
    lag.price4)
```

```r
head(df1)
coefficients <- c()
for (i in n:length(r_price_q)) {
    df2 <- df1[1:i, ]
    model <- lm(r_price_q ~ ., data = df2)
    coefficients <- rbind(coefficients, model$coefficients)
}

plot(coefficients[, 1], ylim = range(coefficients), type = "l",
    col = "red", main = "Estimated Coefficients - Recursive Scheme",
    ylab = "Estimated Values", xlab = "# of points added")
lines(coefficients[, 2], col = "darkblue")
lines(coefficients[, 3], col = "deepskyblue")
lines(coefficients[, 4], col = "forestgreen")
lines(coefficients[, 5], col = "purple")
legend(0, 1, c("Intercept", "Lag(Price, 1)", "Lag(Price, 2)",
    "Lag(Price, 3)", "Lag(Price, 4)"), fill = c("red", "darkblue",
    "deepskyblue", "forestgreen", "purple"), bty = "n", cex = 0.95)
abline(h = 0, col = "gray", lty = 3)

coefficients2 <- c()
for (i in 0:(length(r_price_q) - n)) {
    df2 <- df1[1 + i:n + i, ]
    model <- lm(r_price_q ~ ., data = df2)
    coefficients2 <- rbind(coefficients2, model$coefficients)
}

plot(coefficients2[, 1], ylim = range(coefficients2), type = "l",
    col = "red", main = "Estimated Coefficients - Moving Window",
    ylab = "Estimated Values", xlab = "Moving Window")
lines(coefficients2[, 2], col = "darkblue")
lines(coefficients2[, 3], col = "deepskyblue")
lines(coefficients2[, 4], col = "forestgreen")
lines(coefficients2[, 5], col = "purple")
legend(0, 1.3, c("Intercept", "Lag(Price, 1)", "Lag(Price, 2)",
    "Lag(Price, 3)", "Lag(Price, 4)"), fill = c("red", "darkblue",
    "deepskyblue", "forestgreen", "purple"), bty = "n", cex = 0.95)
abline(h = 0, col = "gray", lty = 3)

# Problem 4 (4.6 from textbook)
g <- read.csv("~/Documents/Econ 144/Data Sets/HW 2/gdp_fed.csv",
    quote = "\"", comment.char = "")
g <- g[nrow(g):1, ]
colnames(g) <- c("Quarter", "gdp", "forecasted_gdp")

gdp_ts <- ts(g$gdp, start = 1969, frequency = 4)
fgdp_ts <- ts(g$forecasted_gdp, start = 1969, frequency = 4)
```

```r
# g_ts <- ts(g[c(-1)], start = 1969, frequency = 4)
t <- seq(1969, 2006, length = length(gdp_ts))

# Compute the 1-quarter ahead forecast errors
forecast_errors <- ts(gdp_ts - fgdp_ts, start = 1969, freq = 4)

# Plot the time series of realized values, forecasts, and
# forecast errors. What do you observe?
quartz()
plot(gdp_ts, type = "l", main = "GDP Growth", ylab = "GDP Growth")
lines(fgdp_ts, lty = 2, col = "red")
legend(1969, 4, c("Realized Growth", "Forecasted Growth"), fill = c("black",
    "red"), bty = "n")

# Forecast errors:
plot(forecast_errors, type = "l", main = "Forecast Errors (GDP Growth)",
    xlab = "Time", ylab = "Forecast Errors")
abline(h = 0)

# Compute the descriptive statistics of the three series.
# Real GDP
summary(gdp_ts)

# Forecasted GDP
summary(fgdp_ts)

# Forecast errors
summary(forecast_errors)

# Forecast Errors
summary(forecast_errors)

as.data.frame(cbind(Series = c("GDP", "Forecasted GDP", "Errors"),
    rbind(summary(gdp_ts), summary(fgdp_ts), summary(forecast_errors))))

boxplot(data.frame(gdp_ts, fgdp_ts, forecast_errors), main = "Boxplot of Three Series")

quartz()
par(mfcol = c(1, 3))
truehist(gdp_ts, col = "slategray", main = "GDP Growth", xlab = "Growth")
truehist(fgdp_ts, col = "skyblue3", main = "Forecasted GDP Growth",
    xlab = "Growth")
truehist(forecast_errors, col = "steelblue3", main = "Forecast Errors",
    xlab = "Error")


# Compute the ACF/PACF Forecast errors:
```

```
acf(forecast_errors, main = "ACF of Forecast Errors")
pacf(forecast_errors, main = "PACF of Forecast Errors")

# Forecasted Values
fgdp_acf <- acf(fgdp_ts, main = "ACF of Forecasted GDP", plot = FALSE)
fgdp_acf$acf[1] <- NA
par(mfrow = c(4, 1))
plot(fgdp_acf, xaxt = "n", main = "ACF of Forecasted GDP (Lag 0 removed)")
axis(1:12)
fgdp_pacf <- pacf(fgdp_ts, main = "PACF of Forecasted GDP")

# Realized Values
gdp_acf <- acf(gdp_ts, main = "ACF of Realized Values of GDP",
    plot = FALSE)
# Removing lag 0:
gdp_acf$acf[1] <- NA
plot(gdp_acf, xaxt = "n", main = "ACF of Realized Values of GDP (Lag 0 removed)")
axis(1:12)
gdp_pacf <- pacf(gdp_ts, main = "PACF of Realized Values of GDP")

# Forecast Errors
quartz()
par(mfcol = c(2, 1))
gdp_acf <- acf(gdp_ts, main = "ACF of Realized Values of GDP")
gdp_pacf <- pacf(gdp_ts, main = "PACF of Realized Values of GDP")

# Problem 5
sp500 <- getReturns("^GSPC", start = "1995-01-03")
amex <- getReturns("^XAX", start = "1995-01-03")

par(mfrow = c(2, 1))
acf(ret_sp500)
pacf(ret_sp500)

par(mfrow = c(2, 1))
acf(ret_amex)
pacf(ret_amex)
```