# STAT 560 Final

*Yuchi Hu*

*November 21, 2018*

## 1.

Analyze the attached final_data1 by three methods: ARIMA, exponential smoothing, and linear regression. The first column of final_data1 is the year; the second column is the coal production.

### 1.1 ARIMA model

#### 1.1.1

Fit an ARIMA model to this time series, excluding the last 10 observations.

### Answer

First, we examine the time series plot, ACF, and PACF of final_data1 (excluding the last 10 observations).

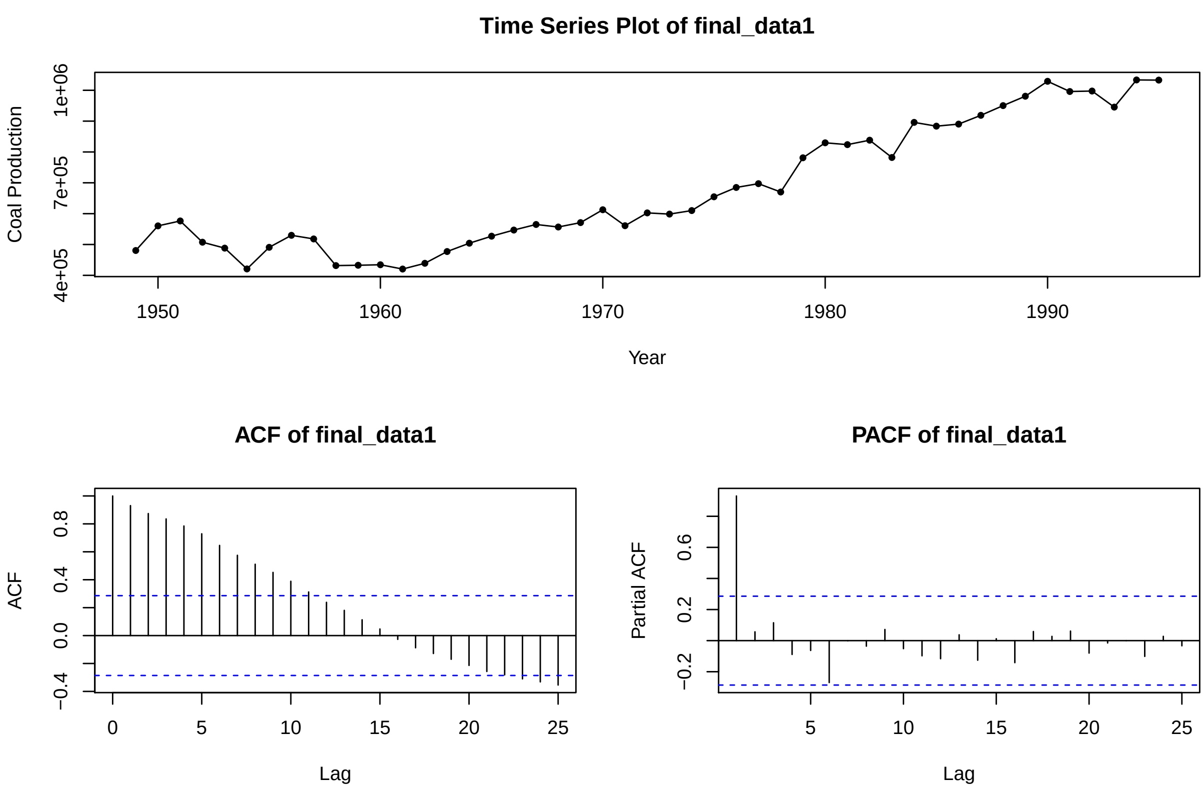


Figure 1: Time series plot, ACF, and PACF of final_data1.

**Figure 1** shows the time series plot, ACF, and PACF of final_data1. The data exhibits an upward trend (nonstationary). The ACF decays slowly, and the PACF cuts off after lag 1. Thus, we could fit an **AR(1)** model. Since the data appears nonstationary, we proceed with taking the first difference of the data.

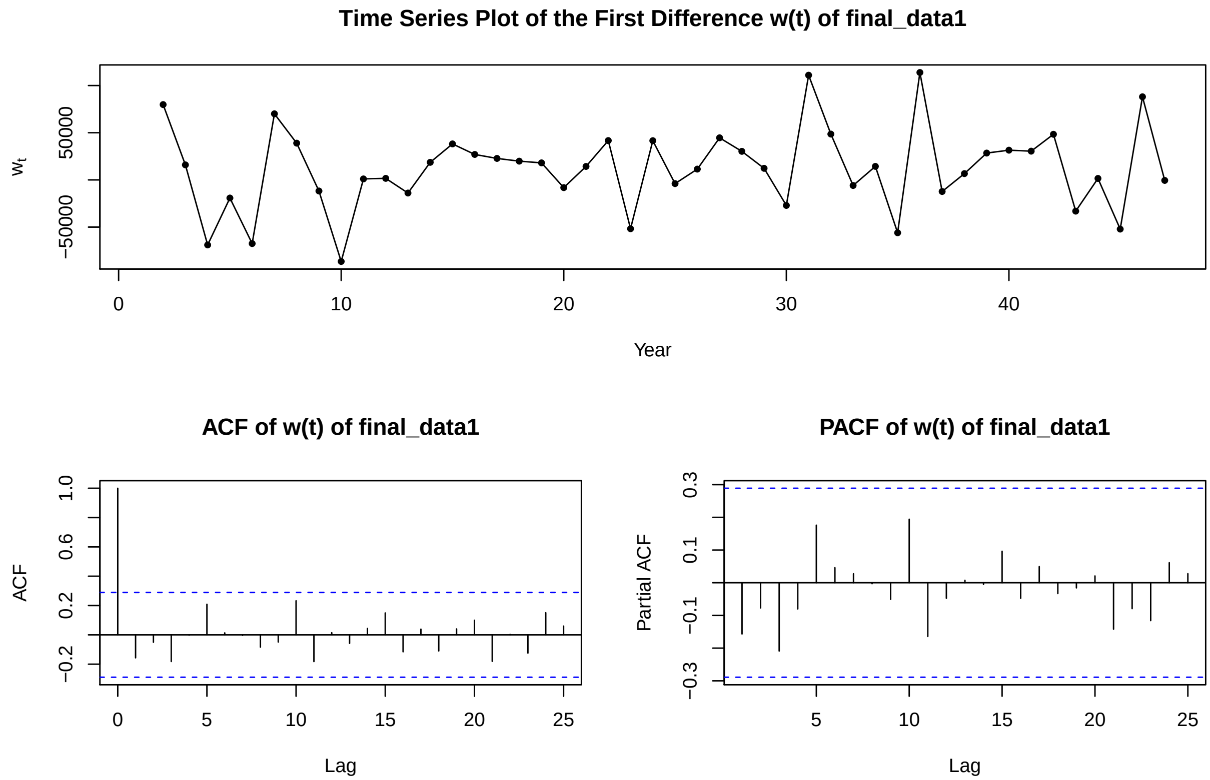


Figure 2: Time series plot, ACF, and PACF of the first difference $w_t$ of final_data1.

**Figure 2** shows the time series plot, ACF, and PACF of the first difference $w_t$ of final_data1. The first difference of the data appears stationary, and the ACF and PACF do not indicate any significant autocorrelation. Thus, we could fit an **ARIMA(0,1,0)** model as well.

Putting it all together, we fit an AR(1) and an ARIMA(0,1,0) model to the data (excluding the last 10 observations) then compare their AIC's.

Table 1: AIC Comparison

| | AIC |
|---|---|
| AR(1) | 1149.371 |
| ARIMA(0,1,0) | 1116.514 |

**Table 1** shows the AIC's of the fitted models. We can see that the ARIMA(0,1,0) model has the lowest AIC. The output for the ARIMA(0,1,0) model is below:

```
## Series: final_data1$coal_production[1:47]
## ARIMA(0,1,0) with drift
##
## Coefficients:
##          drift
##      12008.778
## s.e.   6369.539
##
## sigma^2 estimated as 1.908e+09:  log likelihood=-556.26
## AIC=1116.51   AICc=1116.79   BIC=1120.17
```

The parameter estimate is $\hat{\delta} = 12008.778$, so the fitted ARIMA(0,1,0) model is

$$y_t = 12008.778 + y_{t-1} + \epsilon_t$$

```r
# Fitted values
fit.arima010 <- as.vector(fitted(arima010))

# ARIMA fitted values plot
plot(final_data1$coal_production ~ year, data=final_data1, type='p', pch=16,
    main='ARIMA Fitted Values Plot for final_data1', ylab='Coal Production', xlab='Year')
lines(1949:1995, fit.arima010)
legend('topleft', c('Actual', 'Fitted'), pch=c(16, NA), lty=c(NA, 1))
```

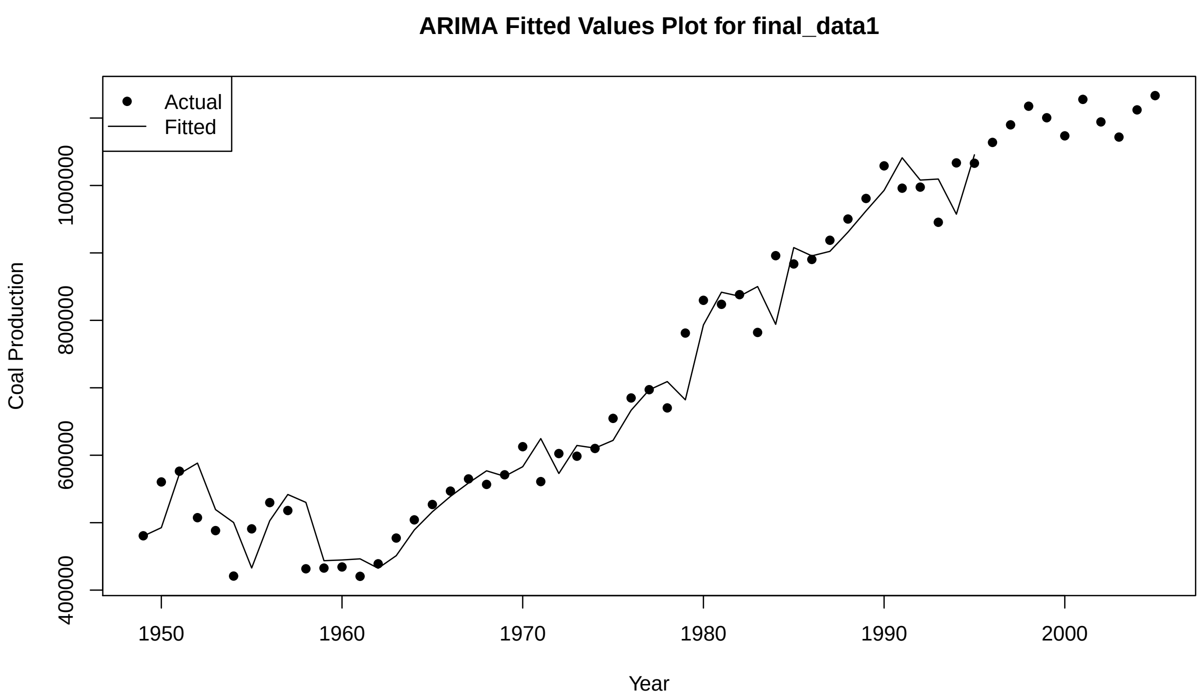


Figure 3: Fitted values for the ARIMA(0,1,0) model of final_data1.

**Figure 3** shows the fitted values for the ARIMA(0,1,0) model of final_data1.

### 1.1.2

Forecast the last 10 observations. Calculate the mean squared error (MSE) and the mean absolute percent forecast error (MAPE).

### Answer

At year 1995, we make the one-step-ahead forecast for 1996. Then when data for 1996 is available, we make the one-step-ahead forecast for 1997, and so on. At each year from 1996-2004, we also update the ARIMA model as we update the actual values.

```
arima010 <- Arima(final_data1$coal_production[1:47], order=c(0, 1, 0), include.drift=T)

# One-step ahead ARIMA forecast and prediction limits for year 1996
forecast.arima010 <- forecast(arima010, h=1)$mean
lpl <- forecast(arima010, h=1)$lower[,2]
upl <- forecast(arima010, h=1)$upper[,2]

a <- final_data1$coal_production[1:47]
# One-step ahead ARIMA forecasts and prediction limits for years 1997-2005
for (i in 1:9) {
  # Updated actual value
  a[47+i] <- final_data1$coal_production[47+i]
  # Updated ARIMA model
  arima010 <- Arima(a, order=c(0, 1, 0), include.drift=T)
  # Vector of forecasted values
  forecast.arima010 <- c(forecast.arima010, forecast(arima010, h=1)$mean)
  # Vectors of prediction limits
  lpl <- c(lpl, forecast(arima010, h=1)$lower[,2])
  upl <- c(upl, forecast(arima010, h=1)$upper[,2])
}

# ARIMA forecast plot
plot(final_data1$coal_production[1:47] ~ year[1:47], data=final_data1, type='p', pch=16,
     xlim=c(1949, 2005), ylim=c(420000, 1223000),
     main='ARIMA Forecasts for final_data1\n(1996-2005)', ylab='Coal Production', xlab='Year')
points(1996:2005, final_data1$coal_production[48:57])
lines(1996:2005, forecast.arima010)
lines(1996:2005, lpl, lty=2)
lines(1996:2005, upl, lty=2)
legend('topleft', pch=c(16,1,NA,NA), lty=c(NA,NA,1,2), bty='n',
       legend=c('Actual (1949-1995)','Actual (1996-2005)',
                'Forecast', '95% Prediction Intervals'))
```

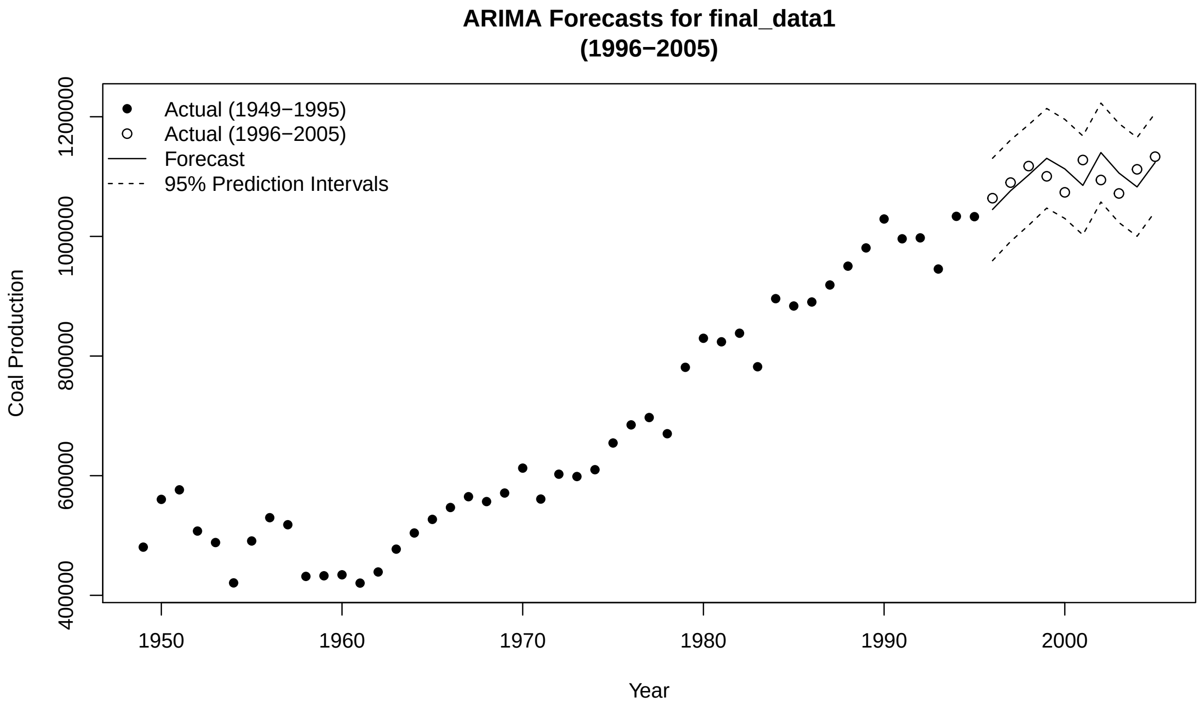


Figure 4: One-step ahead ARIMA forecasts for final_data1 from 1996-2005.

**Figure 4** shows the one-step ahead ARIMA forecasts for final_data1 from 1996-2005.

Next, we calculate the mean squared error (MSE) and mean absolute percentage error (MAPE) of the ARIMA forecasts.

```
# ARIMA forecast errors
error.arima010 <- final_data1$coal_production[48:57] - forecast.arima010

# MSE and MAPE of ARIMA forecasts
MSE.arima011 <- sum(error.arima010^2) / 10
MAPE.arima011 <- 100*sum(abs(error.arima010/final_data1$coal_production[48:57])) / 10

dt <- rbind(MSE=MSE.arima011, MAPE=MAPE.arima011)
kable(dt, caption='MSE and MAPE of ARIMA Forecasts')
```

Table 2: MSE and MAPE of ARIMA Forecasts

| | |
|---|---|
| MSE | 9.234861e+08 |
| MAPE | 2.538415e+00 |

**Table 2** shows the MSE and MAPE of the ARIMA forecasts.

### 1.1.3

Show how to obtain the 95% prediction intervals (PIs) for the forecasts in part 1.1.2). Your answer should include how to obtain the linear filter $\psi_i$s for $y_{t+\tau}$, the formula of calculating PIs and the calculated intervals.

## Answer

The product of the required polynomials for the ARIMA(0,1,0) model is

$$(\psi_0 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + ...)(1 - B) = 1$$

Equating like power of B, we find that

$B^0 : \psi_0 = 1$

$B^1 : \psi_1 - \psi_0 = 0 \rightarrow \psi_1 = 1$

$B^2 : \psi_2 - \psi_1 = 0 \rightarrow \psi_2 = 1$

$B^3 : \psi_3 - \psi_2 = 0 \rightarrow \psi_3 = 1$

In general, we can show for the ARIMA(0,1,0) model that $\psi_j = \psi_{j-1} = 1$.

The variance of the forecast error is

$Var[e_T(\tau)] = \sigma^2 \sum_{i=0}^{\tau-1} \psi_i^2 = \sigma^2 \tau$

Thus, the $100(1-\alpha)$ percent prediction interval for $y_{T+\tau}$ is

$\hat{y}_{T+\tau}(T) \pm Z_{\alpha/2} * \sqrt{Var[e_T(\tau)]}$ or $\hat{y}_{T+\tau}(T) \pm Z_{\alpha/2} * \sigma\sqrt{\tau}$

For $\tau = 1$ the prediction interval becomes

$\hat{y}_{T+1}(T) \pm Z_{\alpha/2} * \sigma$

The calculated intervals are displayed in the table below:

```r
# Table of ARIMA forecasts and prediction intervals from 1996-2005
dt <- cbind(Year=1996:2005, Forecast=forecast.arima010, '95% LPL'=lpl, '95% UPL'=upl)
kable(dt, row.names=F, caption='ARIMA Forecasts and 95% Prediction Intervals (1996-2005)')
```

Table 3: ARIMA Forecasts and 95% Prediction Intervals (1996-2005)

| Year | Forecast | 95% LPL | 95% UPL |
|---|---|---|---|
| 1996 | 1044983 | 959375.5 | 1130590 |
| 1997 | 1076266 | 991422.7 | 1161109 |
| 1998 | 1102627 | 1018602.1 | 1186652 |
| 1999 | 1130534 | 1047284.9 | 1213784 |
| 2000 | 1112829 | 1030011.6 | 1195646 |
| 2001 | 1085240 | 1002551.6 | 1167928 |
| 2002 | 1140133 | 1057450.9 | 1222816 |
| 2003 | 1105863 | 1023053.8 | 1188671 |
| 2004 | 1082700 | 1000173.6 | 1165227 |
| 2005 | 1123581 | 1041453.8 | 1205709 |

**Table 3** shows the ARIMA forecasts and 95% prediction intervals for final_data1 from 1996-2005.

### 1.2 Exponential Smoothing method

#### 1.2.1

Use an exponential smoothing with the optimum value of $\lambda$ to smooth the data, excluding the last 10 observations. Let the range of $\lambda$ be [0.3, 1].

#### Answer

First, we need to find the optimal $\lambda$. The optimal $\lambda$ for a linear trend process is the $\lambda$ value that minimizes the sum of squared prediction errors (SSE) of the second smoothing.

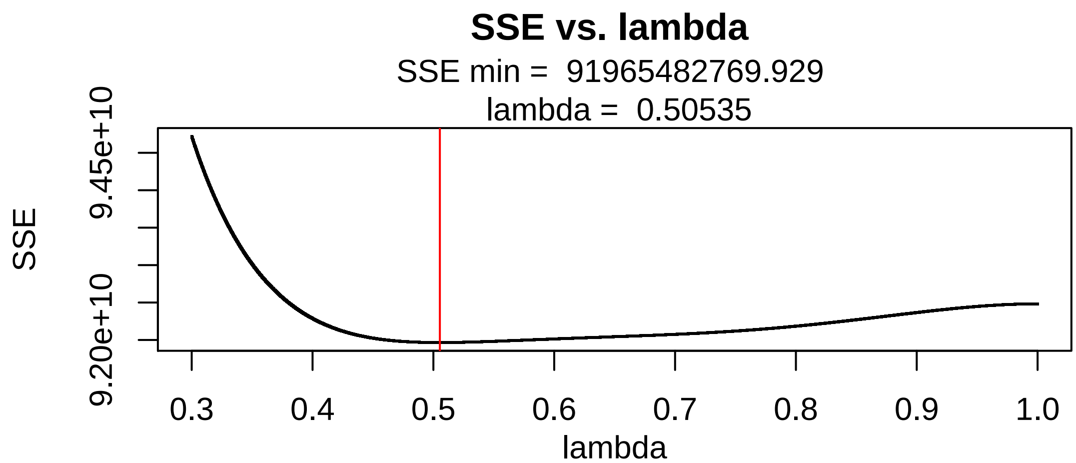


Figure 5: SSE vs. $\lambda$.

**Figure 5** shows the SSE vs. $\lambda$. We can see that the optimal $\lambda$ is 0.50535.

We use second-order exponential smoothing on the data (excluding the last 10 observations) with $\lambda$ = 0.50535.

```r
y <- final_data1$coal_production[1:47]
smooth1 <- firstsmooth(y=y, lambda=0.50535)
smooth2 <- firstsmooth(y=smooth1, lambda=0.50535)
# Fitted values
fit.smooth <- 2*smooth1-smooth2

# Exponential smoothing fitted values plot
plot(final_data1$coal_production ~ year, data=final_data1, type='p', pch=16,
    main='Exponential Smoothing Fitted Values Plot for final_data1 \n(lambda = 0.50535)',
    ylab='Coal Production', xlab='Year')
lines(1949:1995, fit.smooth)
legend('topleft', c('Actual', 'Fitted'), pch=c(16, NA), lty=c(NA, 1))
```

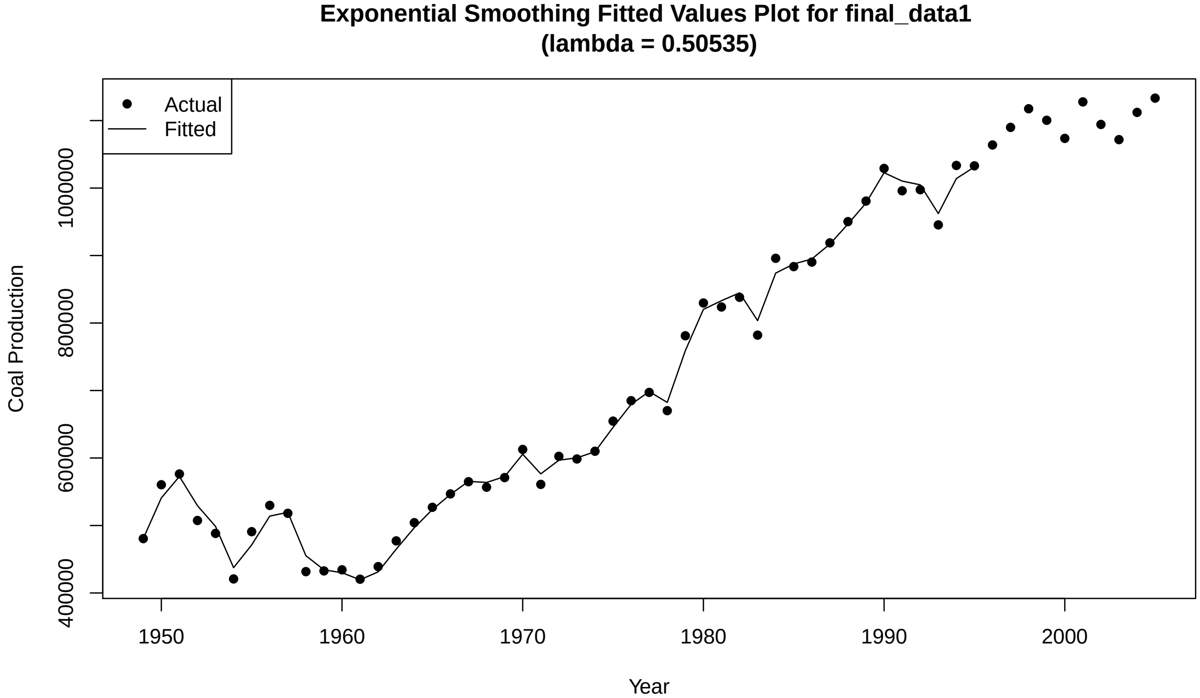


Figure 6: Fitted values for the second-order exponential smoothing of final_data1.

**Figure 6** shows the fitted values for the second-order exponential smoothing of final_data1.

### 1.2.2

Forecast the last 10 observations. Calculate the mean squared error (MSE) and the mean absolute percent forecast error (MAPE).

## Answer

At year 1995, we make the one-step-ahead forecast for 1996. Then when data for 1996 is available, we make the one-step-ahead forecast for 1997, and so on.

```
# One-step ahead exponential smoothing forecasts for years 1996-2005
lambda <- 0.50535
T <- 47
tau <- 10
alpha.lev <- .05
forecast.smooth <- rep(0, tau)
cl <- rep(0, tau)
smooth1 <- rep(0, T+tau)
smooth2 <- rep(0, T+tau)
for (i in 1:tau) {
  smooth1[1:(T+i-1)] <- firstsmooth(y=final_data1$coal_production[1:(T+i-1)], lambda=lambda)
  smooth2[1:(T+i-1)] <- firstsmooth(y=smooth1[1:(T+i-1)], lambda=lambda)
  forecast.smooth[i] <- (2+(lambda/(1-lambda)))*smooth1[T+i-1]-(1+(lambda/(1-lambda)))*smooth2[T+i-1]
```

```
  y.hat <- 2*smooth1[1:(T+i-1)]-smooth2[1:(T+i-1)]
  sig.est <- sqrt(var(final_data1$coal_production[2:(T+i-1)]- y.hat[1:(T+i-2)]))
  cl[i] <- qnorm(1-alpha.lev/2)*sig.est
}

# Exponential smoothing forecast plot
plot(final_data1$coal_production[1:T] ~ year[1:T], data=final_data1, type='p', pch=16,
     xlim=c(1949,2005), ylim=c(420000, 1223000),
     main='Exponential Smoothing Forecasts for final_data1\n(1996-2005, lambda = 0.50535)',
     ylab='Coal Production', xlab='Year')
points(1996:2005, final_data1$coal_production[(T+1):(T+tau)])
lines(1996:2005, forecast.smooth)
lines(1996:2005, forecast.smooth+cl, lty=2)
lines(1996:2005, forecast.smooth-cl, lty=2)
legend('topleft', pch=c(16,1,NA,NA), lty=c(NA,NA,1,2), bty='n',
       legend=c('Actual (1949-1995)','Actual (1996-2005)',
                'Forecast', '95% Prediction Intervals'))
```

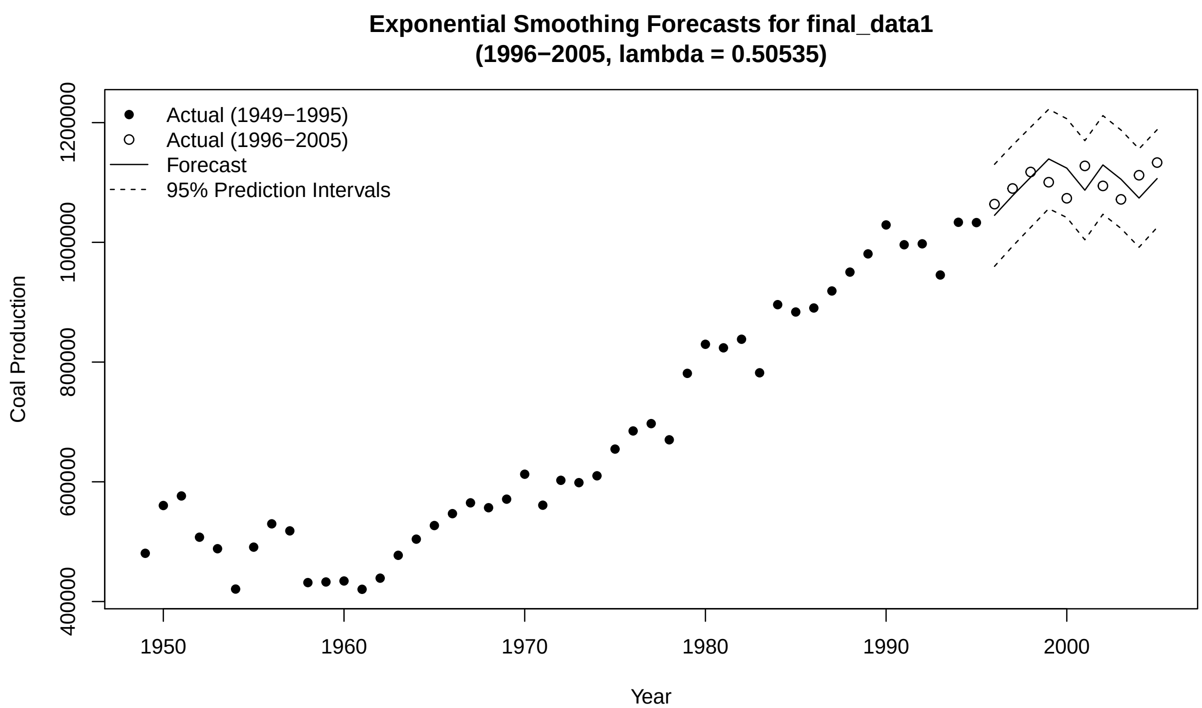


Figure 7: One-step ahead second-order exponential smoothing forecasts for final_data1 from 1996-2005.

**Figure 7** shows the one-step ahead second-order exponential smoothing forecasts for final_data1 from 1996-2005.

Next, we calculate the mean squared error (MSE) and mean absolute percentage error (MAPE) of the exponential smoothing forecasts.

```
# Exponential smoothing forecast errors
error.smooth <- final_data1$coal_production[48:57] - forecast.smooth
```

```r
# MSE and MAPE of exponential smoothing forecasts
MSE.smooth <- sum(error.smooth^2) / 10
MAPE.smooth <- 100*sum(abs(error.smooth/final_data1$coal_production[48:57])) / 10

dt <- rbind(MSE=MSE.smooth, MAPE=MAPE.smooth)
kable(dt, caption='MSE and MAPE of Exponential Smoothing Forecasts')
```

Table 4: MSE and MAPE of Exponential Smoothing Forecasts

| | |
|---|---|
| MSE | 1.07986e+09 |
| MAPE | 2.76123e+00 |

**Table 4** shows the MSE and MAPE of the exponential smoothing forecasts.

### 1.2.3

Show how to obtain the 95% prediction intervals for the forecasts in part 1.2.2). Your answer should include the formula of calculating PIs and the calculated intervals.

## Answer

The $100(1-\alpha)$ percent prediction interval for any lead time $\tau$ is

$(2+\frac{\lambda}{1-\lambda}\tau)\hat{y}_T^{(1)}-(1+\frac{\lambda}{1-\lambda}\tau)\hat{y}_T^{(2)} \pm Z_{\alpha/2}\frac{c_\tau}{c_1}\hat{\sigma}_e$ or $\hat{y}_{T+\tau}(T) \pm Z_{\alpha/2}\frac{c_\tau}{c_1}\hat{\sigma}_e$

where $c_i^2 = 1+\frac{\lambda}{(2-\lambda)^3}[(10-14\lambda+5\lambda^2)+2i\lambda(4-3\lambda)+2i^2\lambda^2]$

For $\tau = 1$ the prediction interval becomes

$(2+\frac{\lambda}{1-\lambda})\hat{y}_T^{(1)}-(1+\frac{\lambda}{1-\lambda})\hat{y}_T^{(2)} \pm Z_{\alpha/2} * \hat{\sigma}_e$ or $\hat{y}_{T+1}(T) \pm Z_{\alpha/2} * \hat{\sigma}_e$

The calculated intervals are displayed in the table below:

```r
# Table of exponential smoothing forecasts and prediction intervals from 1996-2005
dt <- cbind(Year=1996:2005, Forecast=forecast.smooth,
            '95% LPL'=forecast.smooth-cl, '95% UPL'=forecast.smooth+cl)
kable(dt, row.names=F,
      caption='Exponential Smoothing Forecasts and 95% Prediction Intervals (1996-2005)')
```

Table 5: Exponential Smoothing Forecasts and 95% Prediction Intervals (1996-2005)

| Year | Forecast | 95% LPL | 95% UPL |
|---|---|---|---|
| 1996 | 1045133 | 959924.6 | 1130341 |
| 1997 | 1077885 | 993409.1 | 1162362 |
| 1998 | 1108672 | 1024944.5 | 1192399 |
| 1999 | 1139317 | 1056322.4 | 1222312 |
| 2000 | 1123966 | 1041447.4 | 1206485 |
| 2001 | 1087093 | 1004295.5 | 1169891 |
| 2002 | 1129284 | 1046901.2 | 1211667 |
| 2003 | 1105437 | 1023278.2 | 1187595 |

| Year | Forecast | 95% LPL | 95% UPL |
|---|---|---|---|
| 2004 | 1073982 | 991802.3 | 1156161 |
| 2005 | 1106494 | 1024888.2 | 1188100 |

**Table 5** shows the exponential smoothing forecasts and 95% prediction intervals for final_data1 from 1996-2005.

## 1.3 Linear regression model

### 1.3.1

Use a linear regression model to fit the data, excluding the last 10 observations.

### Answer

First, we convert the data (excluding the last 10 observations) to a time-series object using the **ts** function. Then, we fit a linear model with trend component to the time-series object using the **tslm** function from the **forecast** package. The summary of the model is below:

```r
# Convert data to ts object
y <- ts(final_data1$coal_production[1:47], start=1949, end=1995)
# Fitted linear model with trend component
lm <- tslm(y ~ trend)
# Model summary
summary(lm)
```

```
##
## Call:
## tslm(formula = y ~ trend)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -105853  -63630  -10117   52524  186302
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 349159.7    22567.4   15.47   <2e-16 ***
## trend        13624.3      818.6   16.64   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 76130 on 45 degrees of freedom
## Multiple R-squared:  0.8602, Adjusted R-squared:  0.8571
## F-statistic:   277 on 1 and 45 DF,  p-value: < 2.2e-16
```

Thus, the fitted linear regression model is

$$\hat{y} = 349159.7 + 13624.3x$$

where $\hat{y}$ = predicted value of coal production and $x$ = time ($x = 1$ for 1949, $x = 2$ for 1950, etc.).

```r
# Fitted values
fit.lm <- lm$fitted.values

# Linear regression fitted values plot
plot(final_data1$coal_production ~ year, data=final_data1, type='p', pch=16,
    main='Linear Regression Fitted Values Plot for final_data1',
    ylab='Coal Production', xlab='Year')
lines(1949:1995, fit.lm)
legend('topleft', c('Actual', 'Fitted'), pch=c(16, NA), lty=c(NA, 1))
```

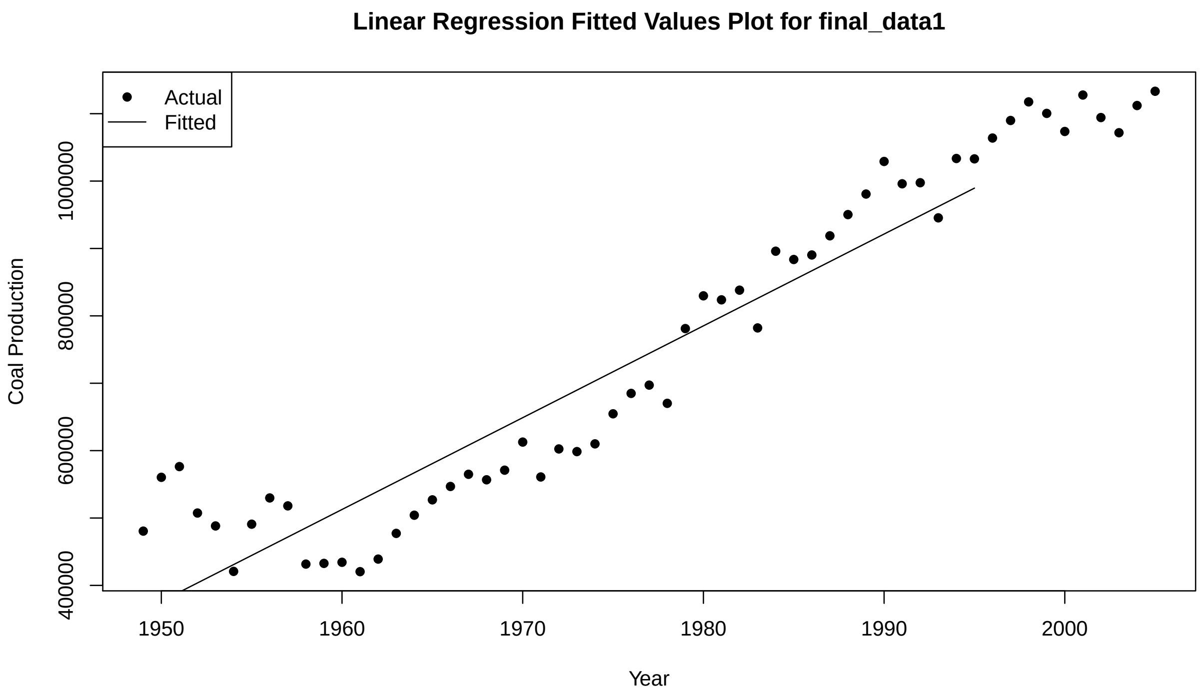


Figure 8: Fitted values for the linear regression model of final_data1.

**Figure 8** shows the fitted values for the linear regression model of final_data1.

### 1.3.2

Forecast the last 10 observations. Calculate the mean squared error (MSE) and the mean absolute percent forecast error (MAPE).

## Answer

At year 1995, we make the one-step-ahead forecast for 1996. Then when data for 1996 is available, we make the one-step-ahead forecast for 1997, and so on. At each year from 1996-2004, we also update the linear regression model as we update the actual values.

```r
# One-step ahead linear regression forecast and prediction limits for year 1996
forecast.lm <- forecast(lm, h=1)$mean
lpl <- forecast(lm, h=1)$lower[,2] ; upl <- forecast(lm, h=1)$upper[,2]

# One-step ahead linear regression forecasts and prediction limits for years 1997-2005
for (i in 1:9) {
  # Updated actual value
  y <- ts(final_data1$coal_production[1:(47+i)], start=1949, end=(1995+i))
  # Updated linear model
  lm <- tslm(y ~ trend)
  # Vector of forecasted values
  forecast.lm <- c(forecast.lm, forecast(lm, h=1)$mean)
  # Vectors of prediction limits
  lpl <- c(lpl, forecast(lm, h=1)$lower[,2])
  upl <- c(upl, forecast(lm, h=1)$upper[,2])
}

# Linear regression forecast plot
plot(final_data1$coal_production[1:47] ~ year[1:47], data=final_data1, type='p', pch=16,
xlim=c(1949, 2005), ylim=c(420000, 1295000),
main='Linear Regression Forecasts for final_data1\n(1996-2005)', ylab='Coal Production', xlab='Year')
points(1996:2005, final_data1$coal_production[48:57])
lines(1996:2005, forecast.lm) ; lines(1996:2005, lpl, lty=2) ; lines(1996:2005, upl, lty=2)
legend('topleft', pch=c(16,1,NA,NA), lty=c(NA,NA,1,2), bty='n',
       legend=c('Actual (1949-1995)','Actual (1996-2005)',
                'Forecast', '95% Prediction Intervals'))
```

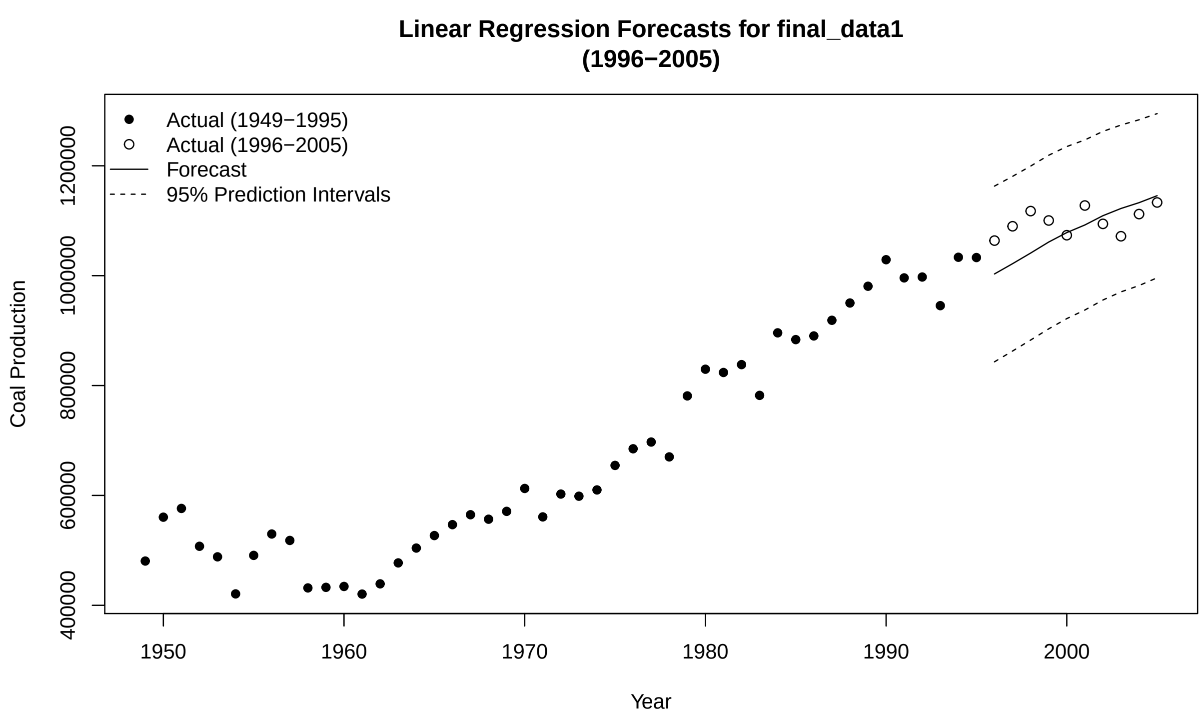


Figure 9: One-step ahead linear regression forecasts for final_data1 from 1996-2005.

**Figure 9** shows the one-step ahead linear regression forecasts for final_data1 from 1996-2005.

Next, we calculate the mean squared error (MSE) and mean absolute percentage error (MAPE) of the linear regression forecasts.

```
# Linear regression forecast errors
error.lm <- final_data1$coal_production[48:57] - forecast.lm
# MSE and MAPE of linear regression forecasts
MSE.lm <- sum(error.lm^2) / 10
MAPE.lm <- 100*sum(abs(error.lm/final_data1$coal_production[48:57])) / 10

dt <- rbind(MSE=MSE.lm, MAPE=MAPE.lm)
kable(dt, caption='MSE and MAPE of Linear Regression Forecasts')
```

Table 6: MSE and MAPE of Linear Regression Forecasts

| | |
|---|---|
| MSE | 2.033958e+09 |
| MAPE | 3.497526e+00 |

**Table 6** shows the MSE and MAPE of the linear regression forecasts.

We can now compare the MSE's and MAPE's of all three methods.

```
dt1 <- cbind('MSE', signif(MSE.arima011, 3), signif(MSE.smooth, 3), signif(MSE.lm, 3))
dt2 <- cbind('MAPE', round(MAPE.arima011, 3), round(MAPE.smooth, 3), round(MAPE.lm, 3))
dt <- rbind(dt1, dt2)

# Table of MSE and MAPE comparison
kable(dt, col.names=c('', 'ARIMA', 'Exponential Smoothing', 'Linear Regression'),
      caption='MSE and MAPE Comparison')
```

Table 7: MSE and MAPE Comparison

| | ARIMA | Exponential Smoothing | Linear Regression |
|---|---|---|---|
| MSE | 9.23e+08 | 1.08e+09 | 2.03e+09 |
| MAPE | 2.538 | 2.761 | 3.498 |

**Table 7** shows the MSE's and MAPE's of the three methods: ARIMA, exponential smoothing, and linear regression. The ARIMA forecasts are the most accurate (smallest errors), while the linear regression forecasts are the least accurate (largest errors).

### 1.3.3

Show how to obtain the 95% prediction intervals for the forecasts in part 1.3.2). Your answer should include the formula of calculating PIs and the calculated intervals.

## Answer

The $100(1-\alpha)$ percent prediction interval for $y_{n+1}$ when the predictor $x = x_{n+1}$ is

$$\hat{y}_{n+1} \pm t_{(\alpha/2,n-2)} * \sqrt{MSE}\sqrt{1 + \frac{1}{n} + \frac{(x_{n+1}-\bar{x})^2}{\sum(x_i-\bar{x})^2}}$$

The calculated intervals are displayed in the table below:

```
# Table of linear regression forecasts and prediction intervals from 1996-2005
dt <- cbind(Year=1996:2005, Forecast=forecast.lm, '95% LPL'=lpl, '95% UPL'=upl)
kable(dt, row.names=F,
      caption='Linear Regression Forecasts and 95% Prediction Intervals (1996-2005)')
```

Table 8: Linear Regression Forecasts and 95% Prediction Intervals (1996-2005)

| Year | Forecast | 95% LPL | 95% UPL |
|---|---|---|---|
| 1996 | 1003127 | 843206.0 | 1163048 |
| 1997 | 1021812 | 862847.7 | 1180777 |
| 1998 | 1041152 | 882844.2 | 1199460 |
| 1999 | 1061209 | 903209.8 | 1219208 |
| 2000 | 1078411 | 921830.3 | 1234992 |
| 2001 | 1092256 | 937436.4 | 1247076 |
| 2002 | 1109135 | 955700.3 | 1262569 |
| 2003 | 1122313 | 970480.0 | 1274146 |
| 2004 | 1132884 | 982013.8 | 1283754 |
| 2005 | 1145549 | 996134.9 | 1294964 |

**Table 8** shows the linear regression forecasts and 95% prediction intervals for final_data1 from 1996-2005.

## 2.

The data final_data_560 contains 7 years of monthly data on the number of airline miles flown in the United Kingdom. This is a seasonal data. Develop an appropriate ARIMA model and a procedure for these data. And calculate the prediction intervals for the eighth year (next 12 months).

## Answer

First, we examine the time series plot, ACF, and PACF of the data.

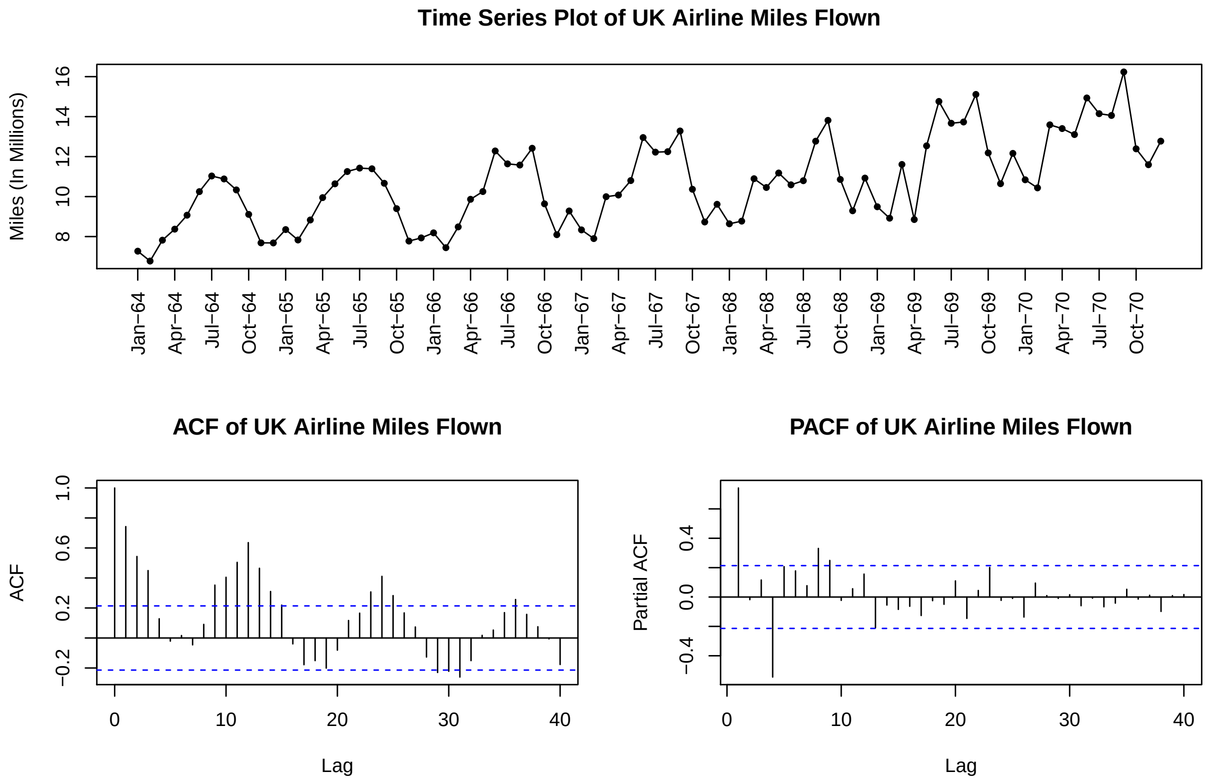


Figure 10: Time series plot, ACF, and PACF of UK airline miles flown.

**Figure 10** shows the time series plot, ACF, and PACF of UK airline miles flown. The data exhibits an upward trend as well as a seasonal pattern. ACF values at lags 12, 24, 36 are significant and slowly decreasing, which indicate a monthly seasonality (s=12). We proceed with taking the seasonal difference and first difference of the data.

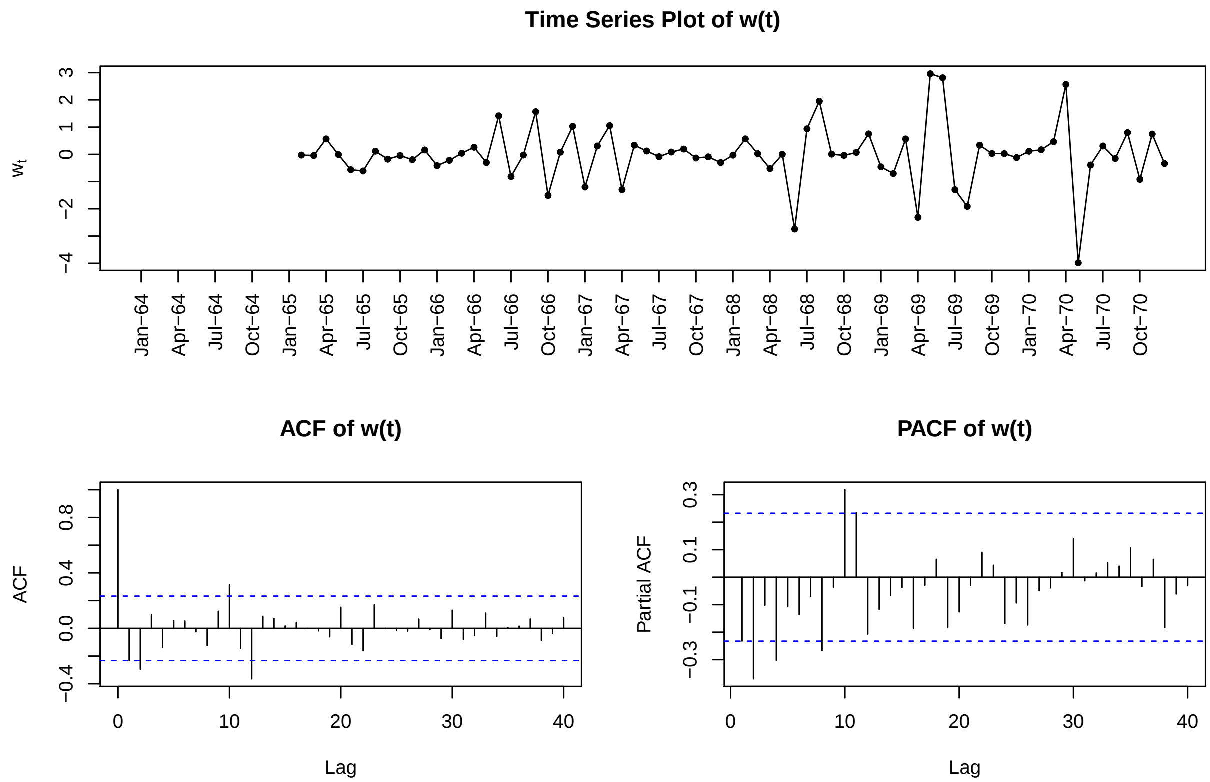


Figure 11: Time series plot, ACF, and PACF of $w_t = (1 - B)(1 - B^{12})y_t$.

**Figure 11** shows the time series plot, ACF, and PACF of $w_t = (1 - B)(1 - B^{12})y_t$. Applying seasonal differencing $D = 12$ and first differencing appears to have removed the trend. The ACF seems to be significant at the first two lags, and the PACF decays at early lags. This suggests a nonseasonal MA(2) component in the model. The ACF having a significant value at lag 12 and the PACF decaying at lags 12, 24, and 36 also suggest a seasonal MA(1) component. Thus, we fit an ARIMA(0,1,2) x $(0, 1, 1)_{12}$ model to the data.

The parameter estimates are:

```
## Series: final_data_560$Miles
## ARIMA(0,1,2)(0,1,1)[12]
##
## Coefficients:
##           ma1      ma2     sma1
##       -0.5845  -0.2429  -0.5425
## s.e.   0.1528   0.1664   0.1204
##
## sigma^2 estimated as 0.6389:  log likelihood=-86.01
## AIC=180.02   AICc=180.63   BIC=189.08
```

Thus, the fitted model is

$$(1 - B)(1 - B)^{12}y_t = (1 - 0.5845B - 0.2429B^2)(1 - 0.5425B^{12})\epsilon_t$$

Next, we plot the ACF and PACF of the residuals.

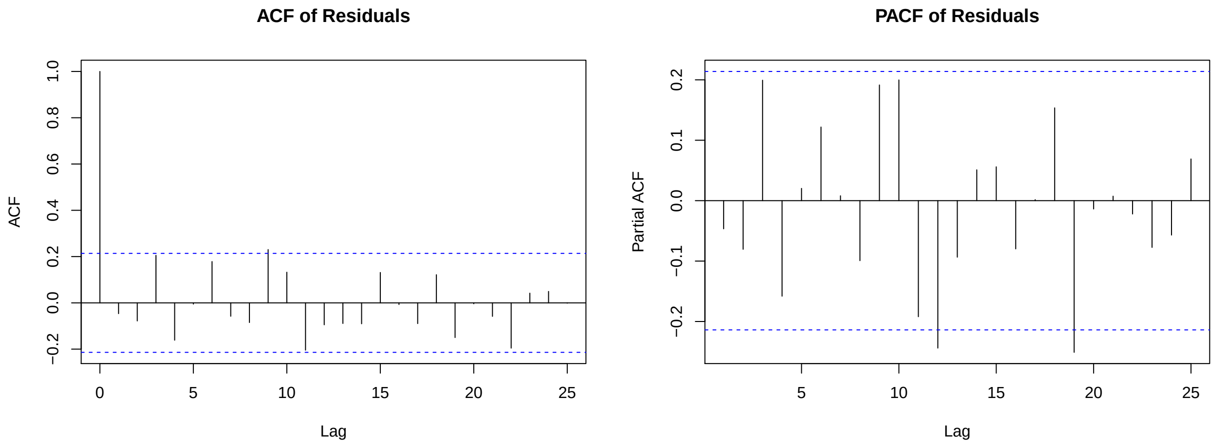


Figure 12: ACF and PACF of residuals from the ARIMA(0,1,2) x $(0,1,1)_{12}$ model.

**Figure 12** shows the ACF and PACF of residuals from the ARIMA(0,1,2) x $(0,1,1)_{12}$ model. There are still some small significant values, but most of the autocorrelation has been modeled out.

Next, we look at the the 4-in-1 residual plots (normal Q-Q, residual vs. fitted value, histogram, and residual vs. observation order).

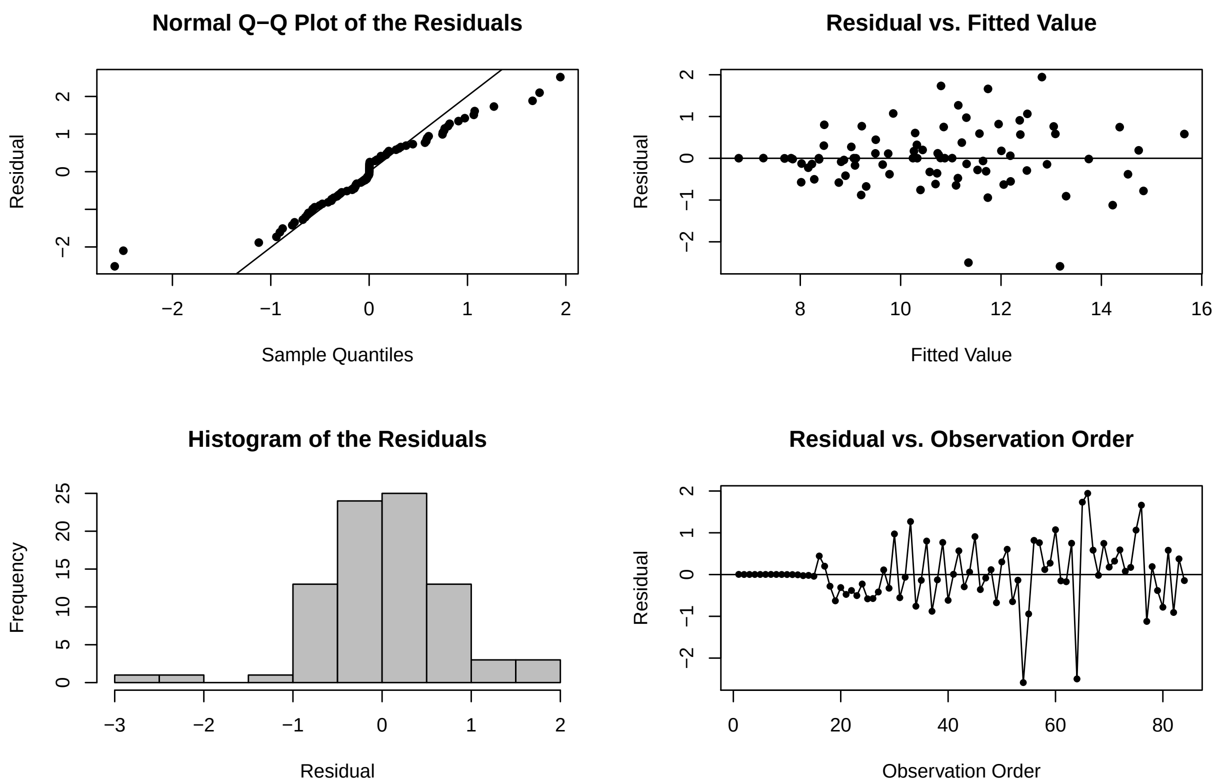


Figure 13: Residual plots from the ARIMA(0,1,2) x $(0,1,1)_{12}$ model.

**Figure 13** shows the residual plots from the ARIMA(0,1,2) x $(0,1,1)_{12}$ model. The normal q-q plot of the residuals shows that most of the residuals follow a normal distribution except at the tails. The histogram of the residuals looks mostly normal. The residual vs. fitted value plot does not show any obvious patterns in the residuals, so the equal variance assumption does not appear to be violated. The residual vs. observation order plot does not indicate any autocorrelation.

Finally, we plot the fitted values and forecast the eighth year (next 12 months).

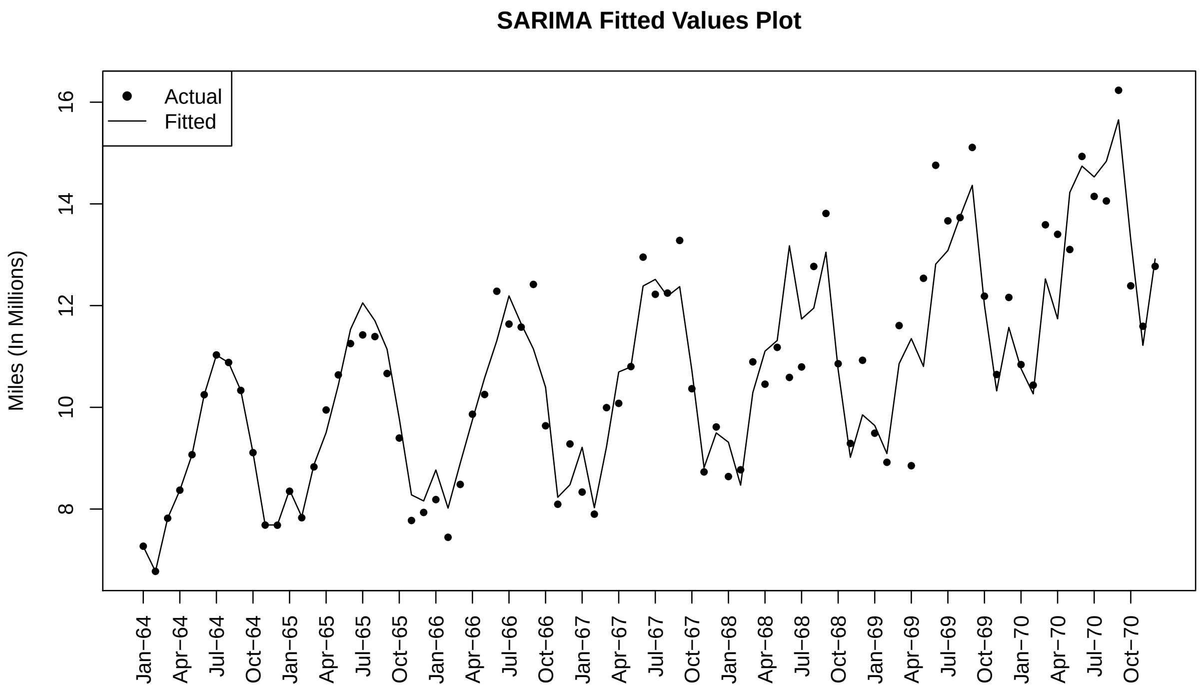


Figure 14: Fitted values for the ARIMA(0,1,2) x $(0,1,1)_{12}$ model.

**Figure 14** shows the fitted values for the ARIMA(0,1,2) x $(0,1,1)_{12}$ model. The model provides a reasonable fit to the data.

```
# One- to 12-step ahead SARIMA forecasts and prediction limits for the next 12 months
forecast.sarima <- forecast(sarima, h=12)$mean
lpl <- forecast(sarima, h=12)$lower[,2]
upl <- forecast(sarima, h=12)$upper[,2]

# SARIMA forecast plot
plot(final_data_560$Miles, type='p', pch=16, cex=0.8, xlim=c(1, 96), ylim=c(6, 19),
     main='SARIMA Forecasts for the Eighth Year (Next 12 Months)',
     ylab='Miles (In Millions)', xlab='', xaxt='n')
axis(1, seq(1, 96, 3), labels=final_data_560$Month[seq(1, 96, 3)], las=2)
lines(85:96, forecast.sarima)
lines(85:96, lpl, lty=2)
lines(85:96, upl, lty=2)
legend('topleft', pch=c(16,NA,NA), lty=c(NA,1,2), bty='n',
       legend=c('Actual', 'Forecast', '95% Prediction Intervals'))
```

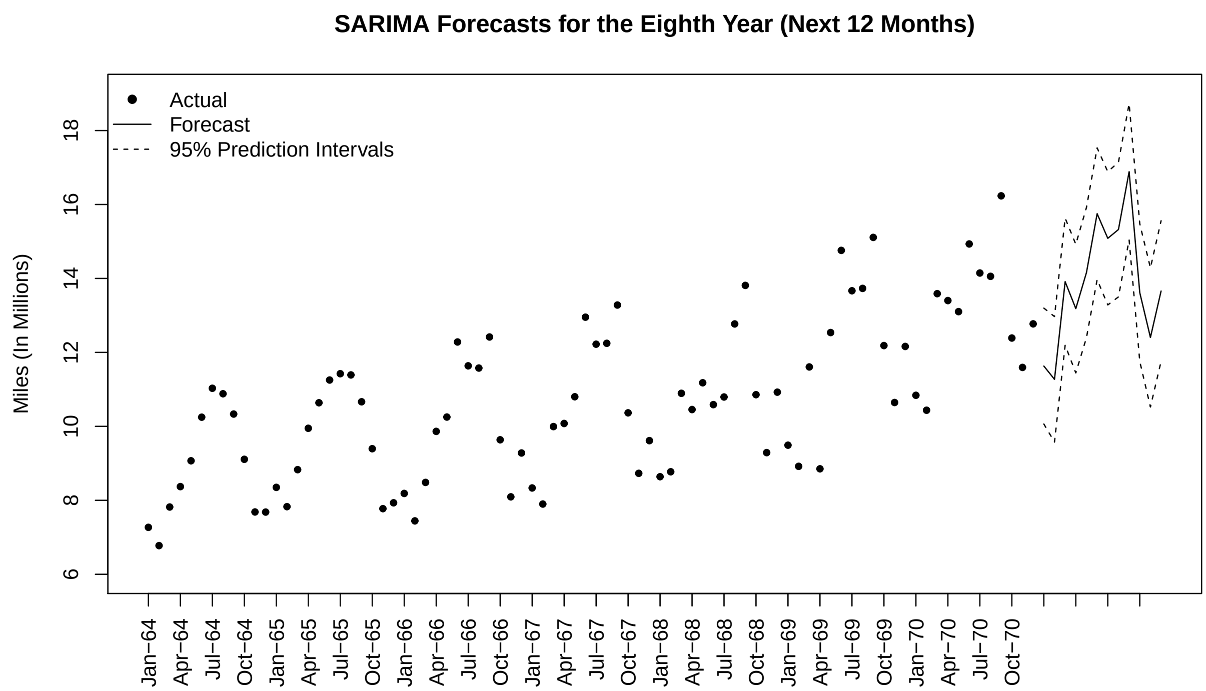


Figure 15: One- to 12-step ahead SARIMA forecasts for the next 12 months.

**Figure 15** shows the one- to 12-step ahead SARIMA forecasts for the next 12 months.

```
dt <- cbind(Forecast=forecast.sarima, '95% LPL'=lpl, '95% UPL'=upl)
kable(dt, caption='SARIMA Forecasts and 95% Prediction Intervals for the Eighth Year (Next 12 Months)')
```

Table 9: SARIMA Forecasts and 95% Prediction Intervals for the Eighth Year (Next 12 Months)

| Forecast | 95% LPL | 95% UPL |
|---|---|---|
| 11.63127 | 10.064325 | 13.19821 |
| 11.27238 | 9.575575 | 12.96919 |
| 13.91280 | 12.194590 | 15.63101 |
| 13.18440 | 11.445043 | 14.92375 |
| 14.16068 | 12.400439 | 15.92092 |
| 15.75091 | 13.970035 | 17.53179 |
| 15.08497 | 13.283681 | 16.88625 |
| 15.32265 | 13.501188 | 17.14411 |
| 16.88642 | 15.045006 | 18.72784 |
| 13.61482 | 11.753664 | 15.47598 |
| 12.40506 | 10.524369 | 14.28575 |
| 13.66080 | 11.760756 | 15.56084 |

**Table 9** shows the SARIMA forecasts and 95% prediction intervals for the next 12 months.