Econ 493 B1 - Winter 2023

Homework 4 - Solution

Exercise 1

Let y_t follow an ARIMA(1,1,0) process

$$y'_t - \mu = \phi(y'_{t-1} - \mu) + \varepsilon_t, \ \varepsilon_t \sim WN(0, \sigma^2)$$

where $|\phi| < 1$. Recall the h-step ahead optimal forecast of y_{T+h} is

$$y_{T+h|T} = y_T + h\mu + (y'_T - \mu) \sum_{s=1}^{h} \phi^h.$$

a. Find the forecast errors, $\varepsilon_{T+h|T}$, for $h = 1, 2, 3$.

$$\begin{aligned} y_{T+1} &= y_T + \mu + \phi(y'_T - \mu) + \varepsilon_{T+1} \\ y_{T+1|T} &= y_T + \mu + \phi(y'_T - \mu) \\ \varepsilon_{T+1|T} &= \varepsilon_{T+1} \end{aligned}$$

$$\begin{aligned} y_{T+2} &= y_{T+1} + \mu + \phi(y'_{T+1} - \mu) + \varepsilon_{T+2} \\ y_{T+2|T} &= y_{T+1|T} + \mu + \phi(y'_{T+1|T} - \mu) \\ \varepsilon_{T+2|T} &= \varepsilon_{T+2} + (1 + \phi)\varepsilon_{T+1} \end{aligned}$$

$$\begin{aligned} y_{T+3} &= y_{T+2} + \mu + \phi(y'_{T+2} - \mu) + \varepsilon_{T+3} \\ y_{T+3|T} &= y_{T+2|T} + \mu + \phi(y'_{T+2|T} - \mu) \\ \varepsilon_{T+3|T} &= \varepsilon_{T+3} + (1 + \phi)\varepsilon_{T+2} + (1 + \phi + \phi^2)\varepsilon_{T+1} \end{aligned}$$

b. Find the forecast error variances for $h = 1, 2, 3$.

$$\begin{aligned} Var(\varepsilon_{T+1|T}) &= \sigma^2 \\ Var(\varepsilon_{T+2|T}) &= \sigma^2[1 + (1 + \phi)^2] \\ Var(\varepsilon_{T+3|T}) &= \sigma^2[1 + (1 + \phi)^2 + (1 + \phi + \phi^2)^2] \end{aligned}$$

Exercise 2

Let y_t follow an ARIMA(0,1,1) process

$$y_t' - \mu = \varepsilon_t + \theta\varepsilon_{t-1},\ \ \varepsilon_t \sim WN(0, \sigma^2)$$

where $|\theta| < 1$.

a. Find the h-step ahead optimal forecast of y_{T+h} for $h = 1, 2, 3$.
b. Find the forecast errors, $\varepsilon_{T+h|T}$, for $h = 1, 2, 3$.

$$\begin{aligned}
y_{T+1} &= y_T + \mu + \varepsilon_{T+1} + \theta\varepsilon_T \\
y_{T+1|T} &= y_T + \mu + \theta\varepsilon_T \\
\varepsilon_{T+1|T} &= \varepsilon_{T+1}
\end{aligned}$$

$$\begin{aligned}
y_{T+2} &= y_{T+1} + \mu + \varepsilon_{T+2} + \theta\varepsilon_{T+1} \\
y_{T+2|T} &= y_{T+1|T} + \mu \\
\varepsilon_{T+2|T} &= \varepsilon_{T+2} + (1+\theta)\varepsilon_{T+1}
\end{aligned}$$

$$\begin{aligned}
y_{T+3} &= y_{T+2} + \mu + \varepsilon_{T+3} + \theta\varepsilon_{T+2} \\
y_{T+3|T} &= y_{T+2|T} + \mu \\
\varepsilon_{T+3|T} &= \varepsilon_{T+3} + (1+\theta)\varepsilon_{T+2} + (1+\theta)\varepsilon_{T+1}
\end{aligned}$$

c. Find the forecast error variances for $h = 1, 2, 3$.

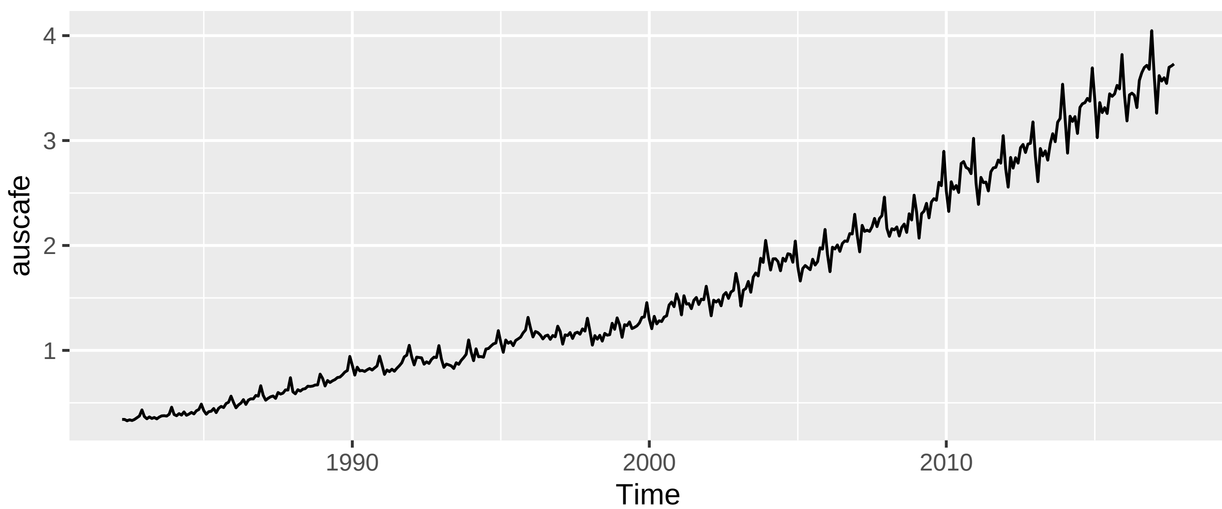
$$\begin{aligned}
Var(\varepsilon_{T+1|T}) &= \sigma^2 \\
Var(\varepsilon_{T+2|T}) &= \sigma^2[1 + (1+\theta)^2] \\
Var(\varepsilon_{T+3|T}) &= \sigma^2[1 + (1+\theta)^2 + (1+\theta)^2]
\end{aligned}$$

Exercise 3 (R)

Consider the total monthly expenditure on cafes, restaurants, and takeaway food services in Australia ($billion) for the sample April 1982 to September 2017 (data set `auscafe`).

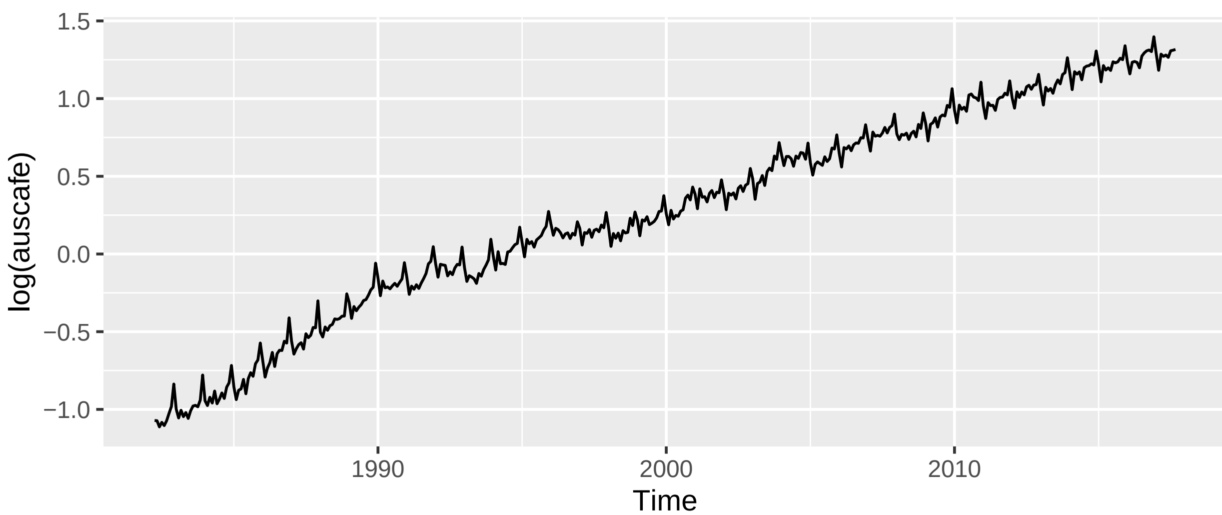
a. Do the data need transforming? If so, find a suitable transformation.

```
autoplot(auscafe)
```



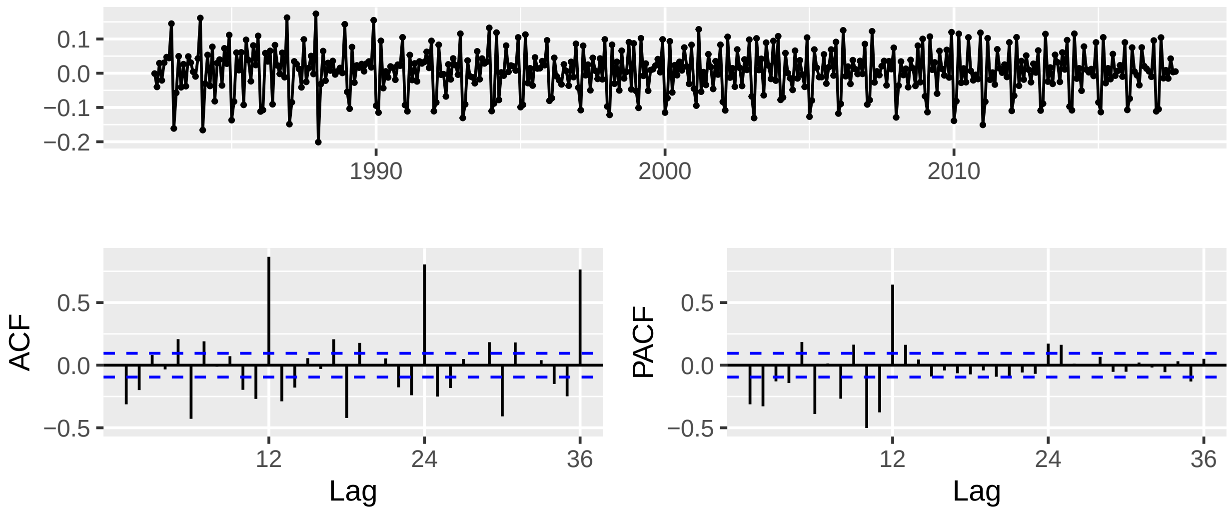
A logarithmic transformation seems appropriate.

```
autoplot(log(auscafe))
```

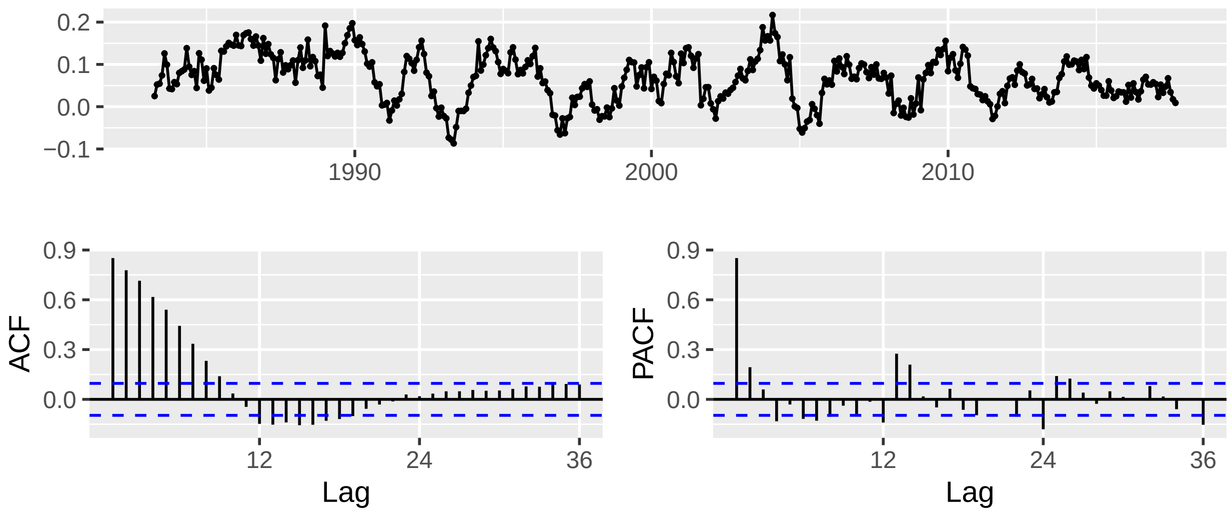


b. Are the data stationary? If not, find an appropriate differencing which yields stationary data.

```
ggtsdisplay(diff(log(auscafe),lag=1))
```

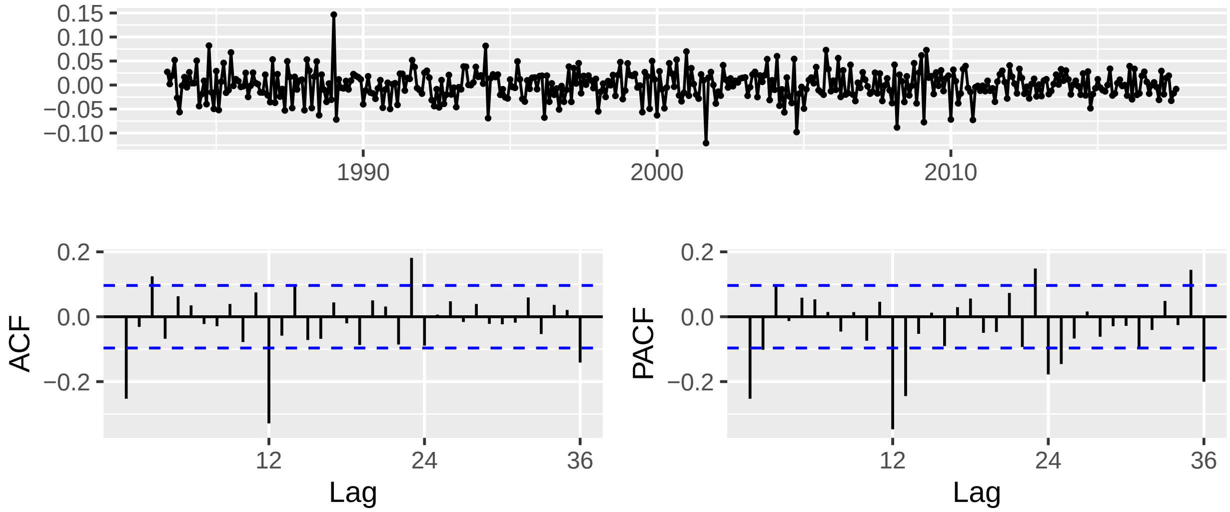



```
ggtsdisplay(diff(log(auscafe),lag=12))
```



After a first difference or a seasonal difference substantial autocorrelation remains in the data. As a result, both differences may be necessary.

```
ggtsdisplay(diff(diff(log(auscafe),lag=12),lag=1))
```



c. Identify a couple of ARIMA models that might be useful in describing the time series. Which of your models is the best according to their AIC values?

Since we don't observe any cut-offs, an ARIMA(1,1,1)(1,1,1)[12] could be a good starting point. Other models similar to this one could also be accurate.

```
AIC(Arima(log(auscafe), order=c(1,1,1), seasonal=c(1,1,1)))
```

```
## [1] -1895.3
```

```
AIC(Arima(log(auscafe), order=c(2,1,1), seasonal=c(2,1,1)))
```

```
## [1] -1901
```

```
AIC(Arima(log(auscafe), order=c(2,1,2), seasonal=c(2,1,2)))
```

```
## [1] -1901.1
```

Of the three models I tried, an ARIMA(2,1,2)(2,1,2)[12] is preferred. Did you find a better model?

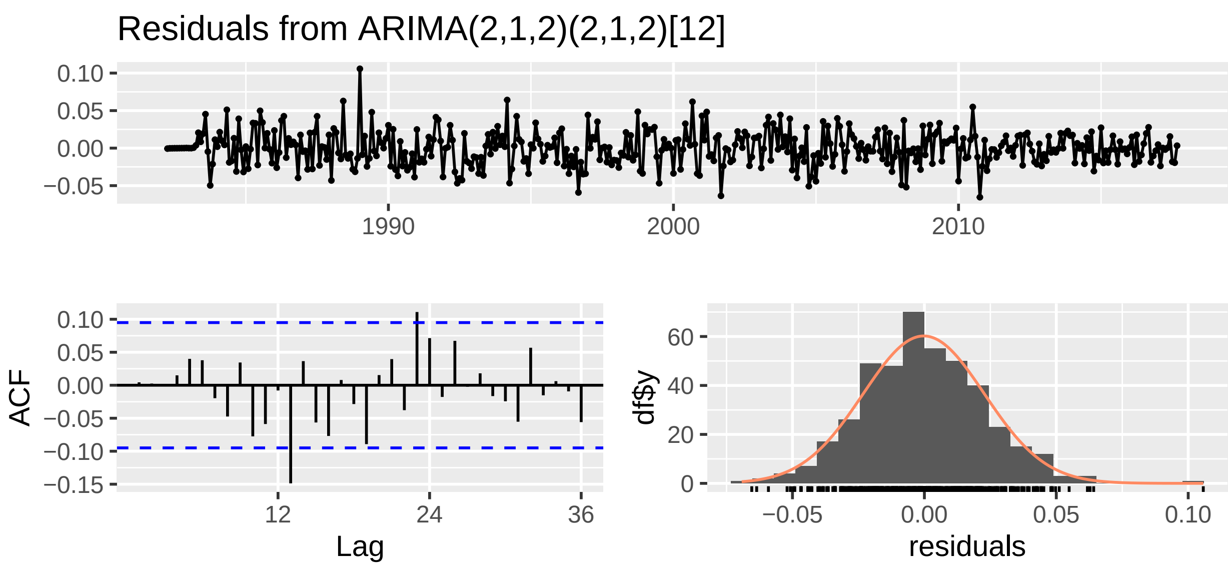
d. Estimate the parameters of your best model and do diagnostic testing on the residuals. Do the residuals resemble white noise? If not, try to find another ARIMA model which fits better.

```
(fit1 <- Arima(log(auscafe), order=c(2,1,2), seasonal=c(2,1,2)))
```

```
## Series: log(auscafe)
## ARIMA(2,1,2)(2,1,2)[12]
##
## Coefficients:
##          ar1     ar2    ma1    ma2   sar1    sar2    sma1   sma2
##       -0.912  -0.348  0.576  0.044  0.722  -0.215  -1.434  0.555
## s.e.   0.232   0.152  0.237  0.211  0.175   0.075   0.171  0.150
##
```

```
## sigma^2 = 0.000555:  log likelihood = 959.53
## AIC=-1901   AICc=-1900.6   BIC=-1864.8
```

```
checkresiduals(fit1)
```

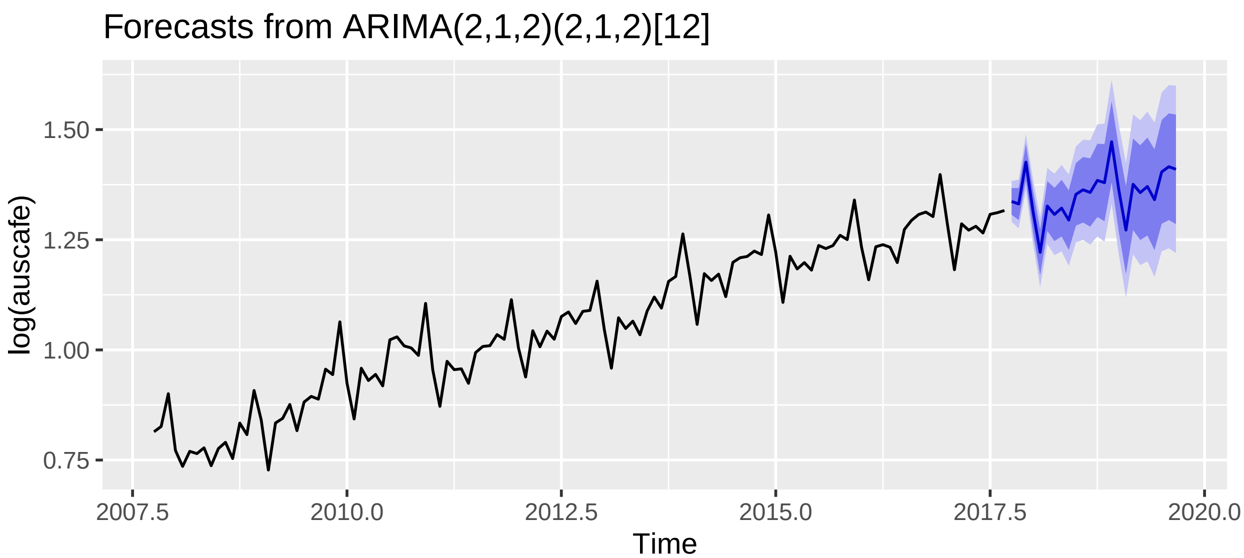



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,1,2)(2,1,2)[12]
## Q* = 35, df = 16, p-value = 0.004
##
## Model df: 8.   Total lags used: 24
```

The residuals are almost WN. However, some seasonal autocorrelations remain.

e. Forecast the next 24 months of data using your preferred model.

```
fore1 <- forecast(fit1)
autoplot(fore1, include=120)
```



Forecasts look quite good!

f. Refit the model using `auto.arima()`. How different are the two models?

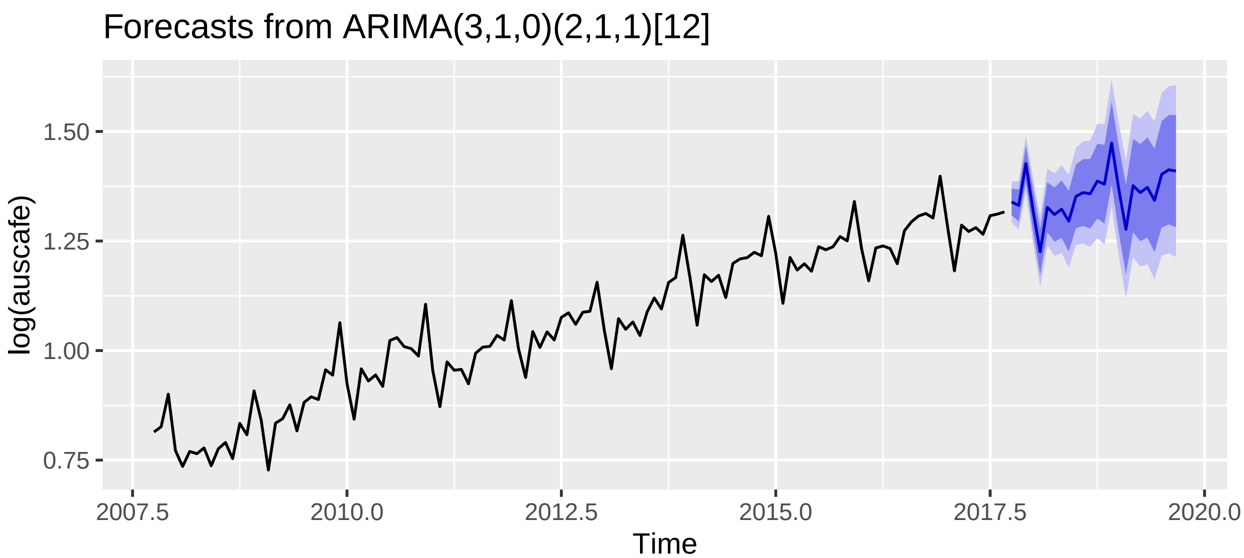
```
(fit2 <- auto.arima(log(auscafe)))
```

```
## Series: log(auscafe)
## ARIMA(3,1,0)(2,1,1)[12]
##
## Coefficients:
##          ar1     ar2    ar3   sar1    sar2    sma1
##       -0.341  -0.109  0.096  0.122  -0.053  -0.830
## s.e.   0.051   0.052  0.049  0.065   0.059   0.043
##
## sigma^2 = 0.000561:  log likelihood = 956.56
## AIC=-1899.1   AICc=-1898.8   BIC=-1871
```

The preferred model is similar to the one we selected: ARIMA(2,1,1)(2,1,2)[12].

g. Compare the forecasts obtained using `auto.arima`.

```
fore2 <- forecast(fit2)
autoplot(fore2, include=120)
```




```
# compute the difference in point forecasts
fore1$mean-fore2$mean
```

```
##                Jan         Feb         Mar         Apr         May         Jun
## 2017
## 2018 -5.5885e-03 -4.0340e-03 -6.0324e-04 -2.9672e-03 -9.4327e-04 -8.6389e-04
## 2019 -4.7512e-03 -4.6841e-03 -3.7674e-04 -3.5740e-03 -1.5777e-03 -1.7128e-03
##                Jul         Aug         Sep         Oct         Nov         Dec
## 2017                                     -2.3137e-03 -3.8048e-05 -5.1250e-04
## 2018  1.2063e-03  2.8887e-03 -6.6406e-04 -2.2369e-03 -1.2609e-04 -1.2042e-03
## 2019  1.5410e-03  3.1917e-03  3.3892e-04
```

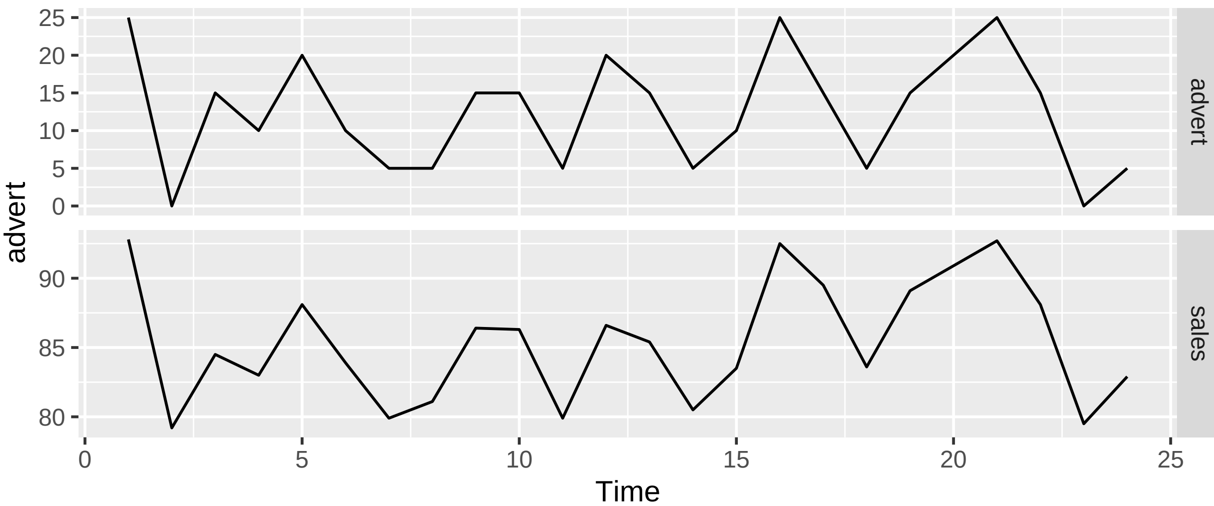
The forecasts are almost identical!

Exercise 4 (R)

Consider monthly sales and advertising data for an automotive parts company (data set `advert`).

a. Plot the data using `autoplot`. Why is it useful to set `facets=TRUE`?

```
autoplot(advert, facets=TRUE)
```



Because the data are on different scales, they should not be plotted in the same panel. Setting `facets=TRUE` puts them in different panels.

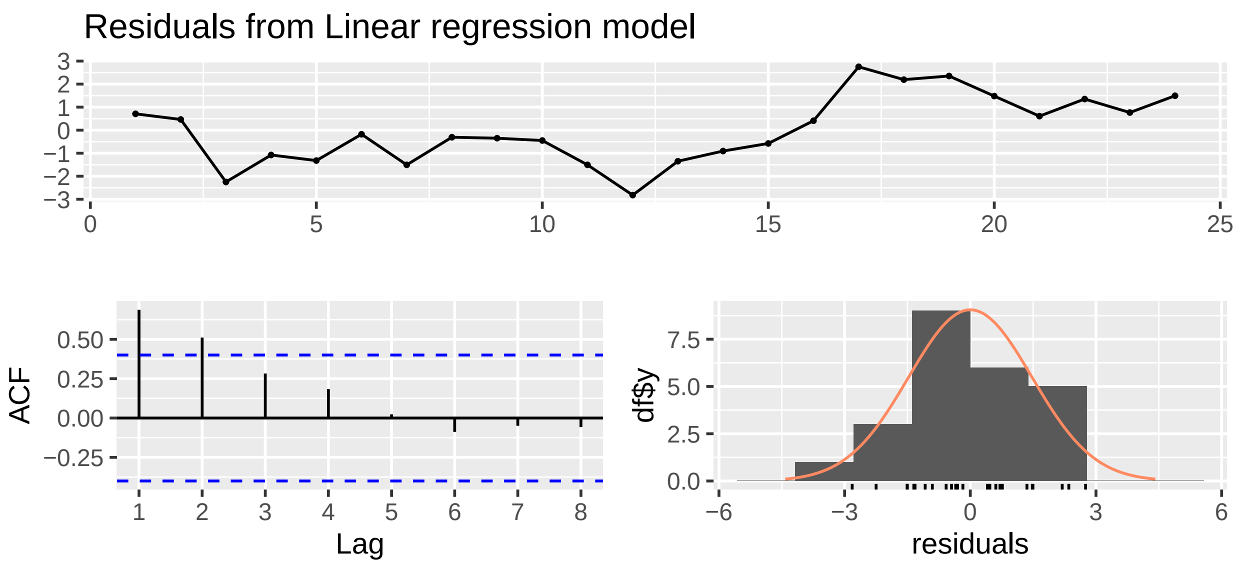
b. Fit a standard regression model $y_t = a + bx_t + \eta_t$ where y_t denotes sales and x_t denotes advertising using the `tslm()` function.

```
(fit <- tslm(sales ~ advert, data=advert))
```

```
##
## Call:
## tslm(formula = sales ~ advert, data = advert)
##
## Coefficients:
## (Intercept)       advert
##      78.734        0.534
```

c. Show that the residuals have significant autocorrelation.

```
checkresiduals(fit)
```




```
##
##  Breusch-Godfrey test for serial correlation of order up to 5
##
## data:  Residuals from Linear regression model
## LM test = 12.5, df = 5, p-value = 0.029
```

d. What difference does it make if you use the `Arima` function instead:

```
Arima(advert[,'sales'], xreg=advert[,'advert'], order=c(0,0,0))
```

```
## Series: advert[, "sales"]
## Regression with ARIMA(0,0,0) errors
##
## Coefficients:
##       intercept   xreg
##          78.734  0.534
## s.e.      0.572  0.039
##
## sigma^2 = 2.27:  log likelihood = -42.83
## AIC=91.66   AICc=92.86   BIC=95.2
```

The model is identical.

e. Refit the model using `auto.arima()`. How much difference does the error model make to the estimated parameters? What ARIMA model for the errors is selected?

```
(fit <- auto.arima(advert[,'sales'], xreg=advert[,'advert']))
```

```
## Series: advert[, "sales"]
## Regression with ARIMA(0,1,0) errors
##
## Coefficients:
```

```
##        xreg
##       0.506
## s.e.  0.021
##
## sigma^2 = 1.2:  log likelihood = -34.22
## AIC=72.45   AICc=73.05   BIC=74.72
```

There is first order differencing, so the intercept disappears. The advert coefficient has changed a little. The error model is ARIMA(0,1,0).

f. Check the residuals of the fitted model.

```
checkresiduals(fit)
```


```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(0,1,0) errors
## Q* = 1.56, df = 5, p-value = 0.91
##
## Model df: 0.   Total lags used: 5
```

All good.

g. Assuming the advertising budget for the next six months is exactly 10 units per month, produce sales forecasts with prediction intervals for the next six months.

```
(fc <- forecast(fit, xreg=matrix(rep(10,6))))
```

```
##    Point Forecast  Lo 80  Hi 80  Lo 95  Hi 95
## 25         85.432 84.028 86.836 83.284 87.579
## 26         85.432 83.446 87.418 82.395 88.469
```

```
## 27          85.432 83.000 87.864 81.712 89.151
## 28          85.432 82.623 88.240 81.137 89.727
## 29          85.432 82.292 88.572 80.630 90.234
## 30          85.432 81.992 88.871 80.171 90.692
```

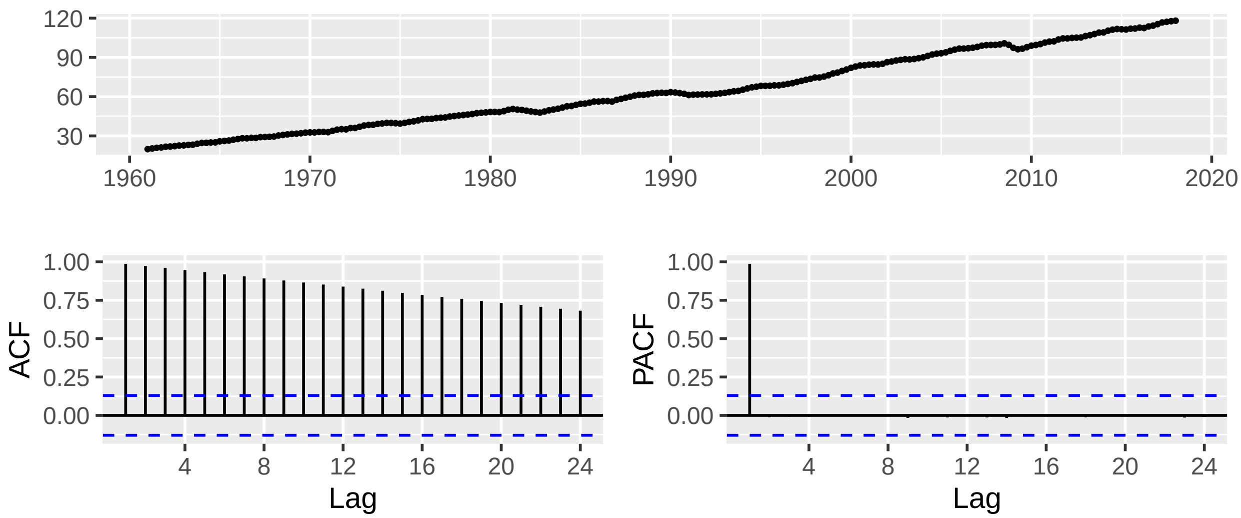
```
autoplot(fc) + xlab("Month") + ylab("Sales")
```

Exercise 5 (R)

The file `NAEXKP01CAQ661S.csv` contains the series of quarterly real gross domestic product (RGDP) for Canada for the quarters 1961:Q1 to 2018:Q1, measured in millions of 2010 Canadian dollars and seasonally adjusted.

a. Use R to plot the series, the ACF, and PACF. Does the series appear to be stationary?

```
# read canada quarterly real gdp data
data <- read.table("NAEXKP01CAQ661S.csv", sep=",", header=TRUE)
cangdp <- ts(data$NAEXKP01CAQ661S,start=1961,frequency=4)
ggtsdisplay(cangdp)
```



The time plot shows the series exhibits a time trend, which is a typical indication of non-stationarity. ACF does not drop quickly to zero, moreover the value r_1 is large and positive (almost 1 in this case). PACF value r_1 is almost 1. All these are signs of a non-stationary time series.

b. Here we will fit a *trend stationary model* for the sample 1961Q1 to 2009Q4. Using the AIC, find the AR model *with* time trend that adequately describes RGDP. Make sure your model includes a trend component. Motivate the steps that you take.

```
smpl1 <- window(cangdp,end=c(2009,4))
smpl2 <- window(cangdp,start=c(2010,1),end=c(2017,4))
trend <- seq_along(smpl1)
AIC(Arima(smpl1, order=c(1,0,0), xreg=trend))
```

```
## [1] 276.88
```

```
AIC(Arima(smpl1, order=c(2,0,0), xreg=trend))
```

```
## [1] 225.16
```

```r
AIC(Arima(smpl1, order=c(3,0,0), xreg=trend))
```

```
## [1] 226.59
```

```r
AIC(Arima(smpl1, order=c(4,0,0), xreg=trend))
```

```
## [1] 228.47
```

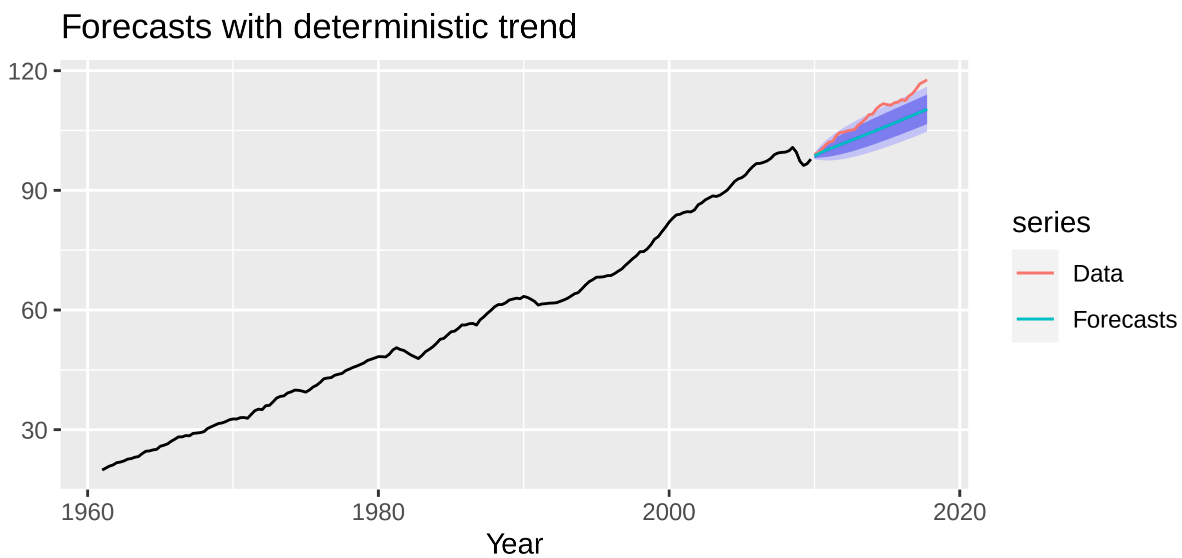
An AR(2) with deterministic time trend is preferred by AIC.

```r
(fit1 <- Arima(smpl1, order=c(2,0,0), xreg=trend))
```

```
## Series: smpl1
## Regression with ARIMA(2,0,0) errors
##
## Coefficients:
##          ar1     ar2  intercept   xreg
##        1.472  -0.491     17.440  0.405
## s.e.   0.062   0.062      2.382  0.020
##
## sigma^2 = 0.175:  log likelihood = -107.58
## AIC=225.16   AICc=225.48   BIC=241.55
```

c. Using this model, compute and plot the quarterly forecasts of real GDP for the next 8 years (32 quarters) along with the prediction intervals for the forecasts and the actual value of real GDP. Interpret your results.

```r
fore1 <- forecast(fit1, h=32, xreg=length(smpl1) + 1:32)
autoplot(fore1) +
  autolayer(smpl2, series="Data") +
  autolayer(fore1$mean, series="Forecasts") +
  ggtitle("Forecasts with deterministic trend") +
  xlab("Year") + ylab("")
```



d. Here we will fit a *difference stationary model* for the sample 1961Q1 to 2009Q4. Using the AIC, find the AR model that adequately describes the change RGDP. Make sure your model uses $d = 1$ and includes a drift. Motivate the steps that you take.

```
AIC(Arima(smpl1, order=c(1,1,0), include.drift=TRUE))
```

```
## [1] 220.09
```

```
AIC(Arima(smpl1, order=c(2,1,0), include.drift=TRUE))
```

```
## [1] 221.14
```

```
AIC(Arima(smpl1, order=c(3,1,0), include.drift=TRUE))
```

```
## [1] 223.12
```

```
AIC(Arima(smpl1, order=c(4,1,0), include.drift=TRUE))
```

```
## [1] 224.98
```

An ARIMA(1,1,0) is preferred by AIC.

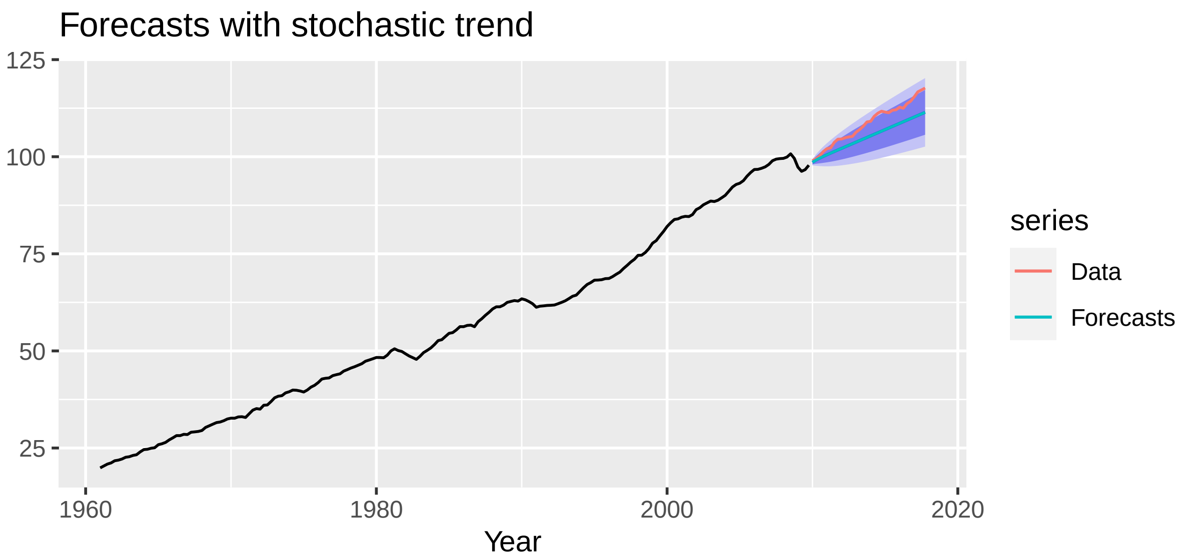
```
(fit2 <- Arima(smpl1, order=c(1,1,0), include.drift=TRUE))
```

```
## Series: smpl1
## ARIMA(1,1,0) with drift
##
## Coefficients:
##          ar1   drift
##        0.484   0.404
## s.e.   0.063   0.058
##
## sigma^2 = 0.177:  log likelihood = -107.04
```

```
## AIC=220.09   AICc=220.21   BIC=229.91
```

e. Using this model, compute and plot the quarterly forecasts of real GDP for the next 8 years (32 quarters) along with the prediction intervals for the forecasts and the actual value of real GDP. Interpret your results.

```
fore2 <- forecast(fit2, h=32)
autoplot(fore2) +
  autolayer(smpl2, series="Data") +
  autolayer(fore2$mean, series="Forecasts") +
  ggtitle("Forecasts with stochastic trend") +
  xlab("Year") + ylab("")
```



f. Compare your results from (c) and (e).

Point forecasts are very similar, prediction intervals are quite different. A deterministic trend implies intervals that may be too narrow.