Selvstudie_1

Jonathan Strandberg

11 mar 2019

```
## Warning: package 'tseries' was built under R version 3.5.2

## Warning: package 'forecast' was built under R version 3.5.2
```

Selvstudieopgavesæt 1

Nedenstående opgave arbejdes der med ved første selvstudie-kursusgang. Data til opgaven er prisen i dollars pr. tønde (42 gallon) "West-Texas Intermediate Crude Oil (WTI)". Priserne foreligger som månedsdata fra Januar 1986 til Juni 2013, i alt 330 observationer. Datasættet er i filen 'olie.RData' som kan indlæses med funktionen load(). Til identifikationen og estimationen nedenfor bruges data op til december 2012. De sidste 6 måneder, fra januar 2013 til juni 2013 reserveres til out-of-sample forecasting (Hint: Brug funktionen window() til at opdele datasættet)

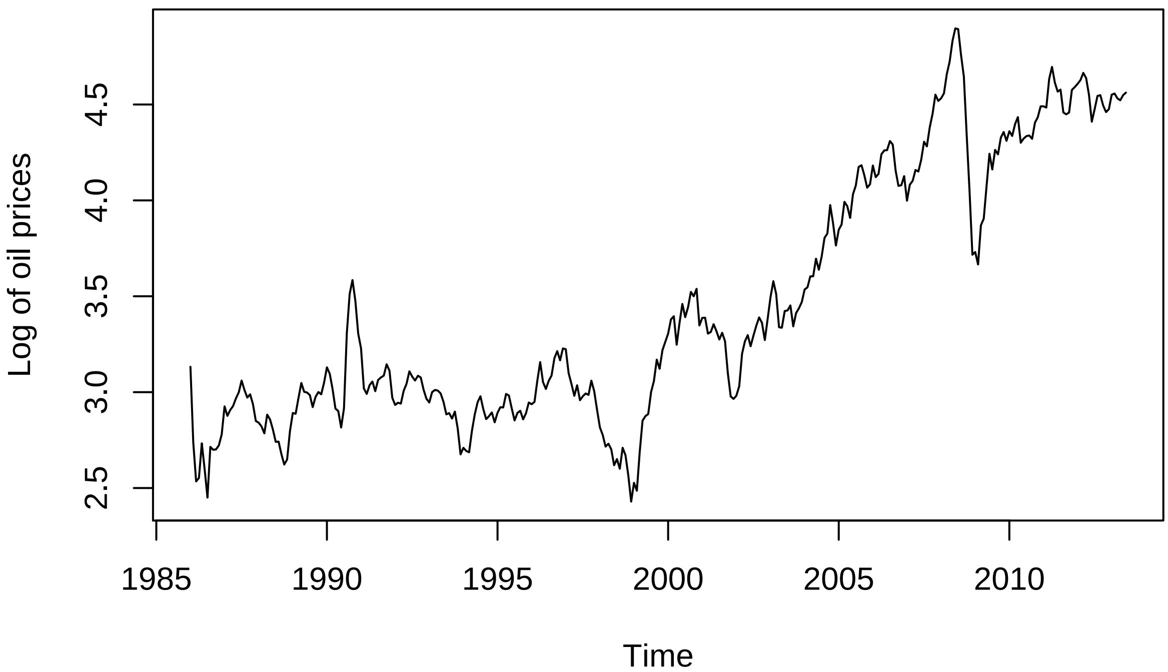
Opgave

Identificér, estimér og prædiktér (forecast) ARMA modeller for logaritmen til olieprisen (her kaldes denne variabel x_t)

1. I det følgende betragtes som nævnt de 324 første observationer, altså data til og med december 2012. Plot og beskriv kort log prisen af olie i perioden. Kommentér på særligt usædvanlige egenskaber ved data. Beregn, plot og fortolk korrelogrmmet for log-olieprisen, samt korrelogrammet for første-differensen af log-olieprisen.

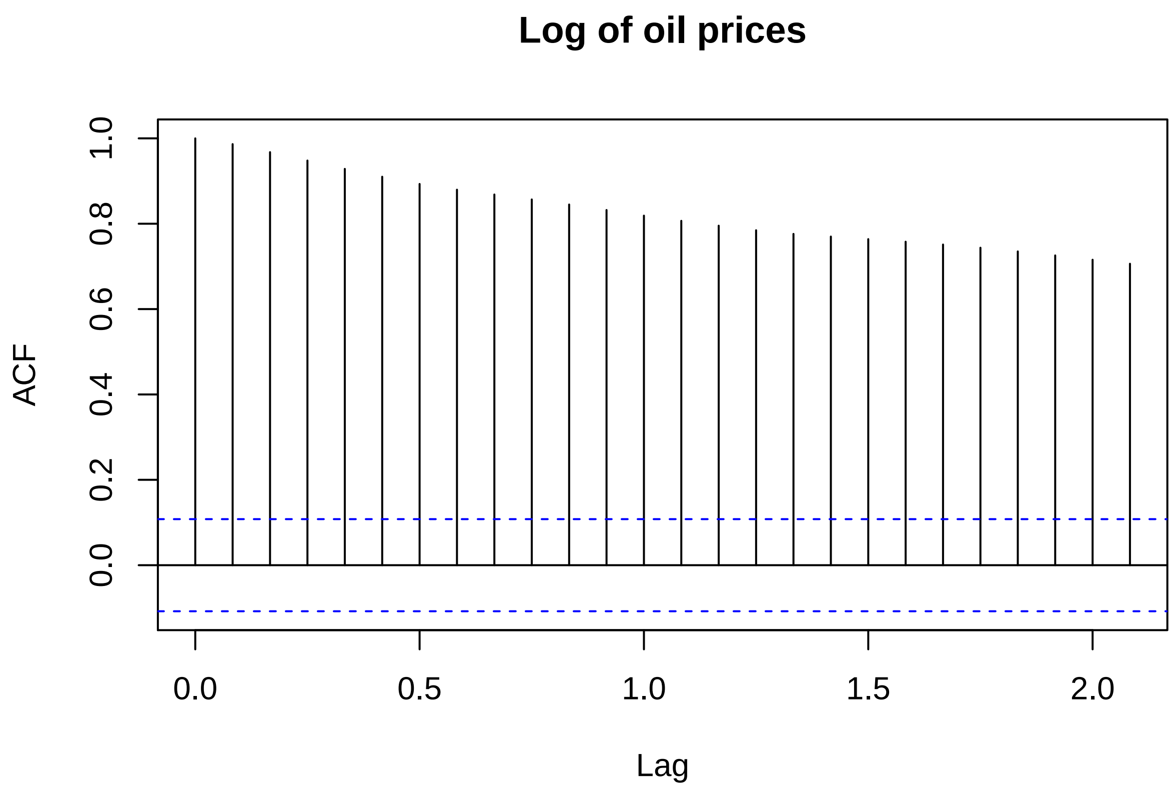
```r
# Indlæser oliedata
load("olie.Rdata")
loil <- log(olie)

# Plotter log-oil
plot(loil , ylab = "Log of oil prices")
```



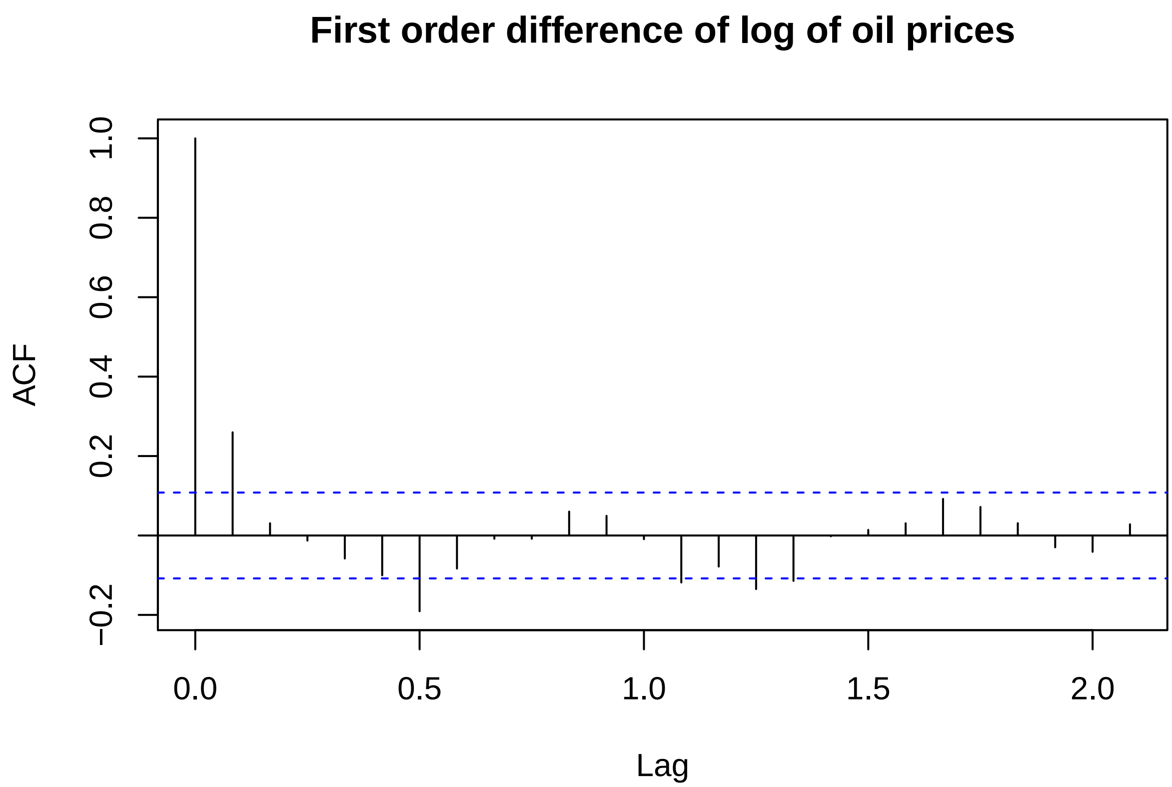
Ses at der forekommer en stærk opadgående trend med spikes

```
# Plotter acf for 'loil
acf(loil , main = "Log of oil prices")
```



Ses at ACF'en ikke aftager eksponentielt, og tidsrækken ser altså ikke stationær ud.

```r
# Plotter acf for første-differensen af 'loil'
acf(diff(loil) , main = "First order difference of log of oil prices")
```



Efter der er blevet taget differens ses det at ACF'en aftager eksponentielt, og tidsrækken ser stationær ud.

2. Estimér en AR(1) model for log-olieprisen på to forskellige måder. Estimér først autoregressionen

$$x_t = \alpha + \phi x_{t-1} + w_t$$

og estimér dernæste den middelværdi-justerede model

$$x_t - \mu = \phi(x_{t-1} - \mu) + w_t$$

Hvad fortæller estimatet på ϕ om "stabiliteten" (stationariteten) af modellen?

```
# Laver x_t og x_{t-1}
start <- c(1986,1)
end   <- c(2012,12)

xins <- window(loil , start = start , end = end)
n <- length(xins)

x      <- xins[2:n]
xlag   <- xins[1:(n-1)]
xdm    <- x - mean(x)
xlagdm <- xlag - mean(xlag)

ar1.coef        <- lm(x ~ xlag)$coefficients
```

```r
ar1.demean.coef <- lm(xdm ~ xlagdm)$coefficients

df <- data.frame(x = ar1.coef , xdemean = ar1.demean.coef)
rownames(df) <- c("Alpha" , "Phi") ; print(df)
```

```
##                x      xdemean
## Alpha 0.02285715 3.459370e-16
## Phi   0.99454695 9.945469e-01
```

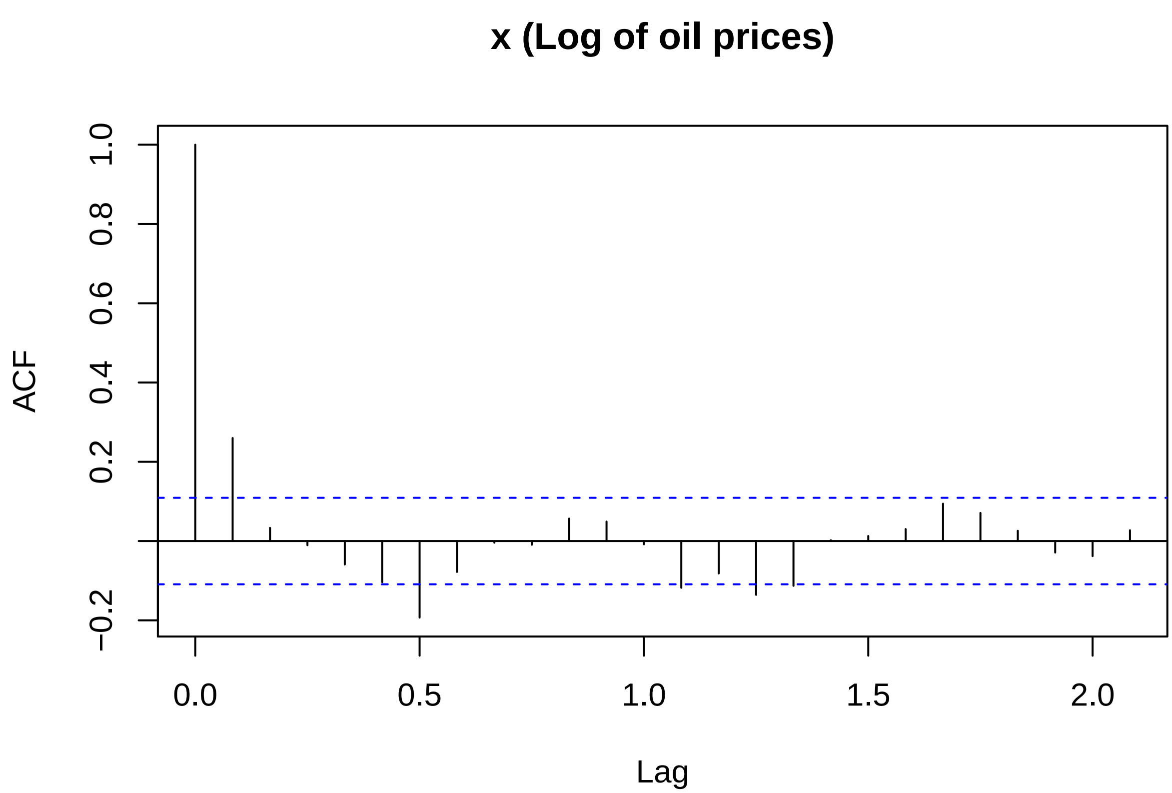
Det ses at når vi har ‘demeanet’ estimeres skæringen til at være tæt på 0

I det resterende sammenlignes to forskellige tranformationer til at gøre tidsrækken stationær, og i kan overveje og vurdére, hvilken transformation der bedst gør tidsrækken stationær.

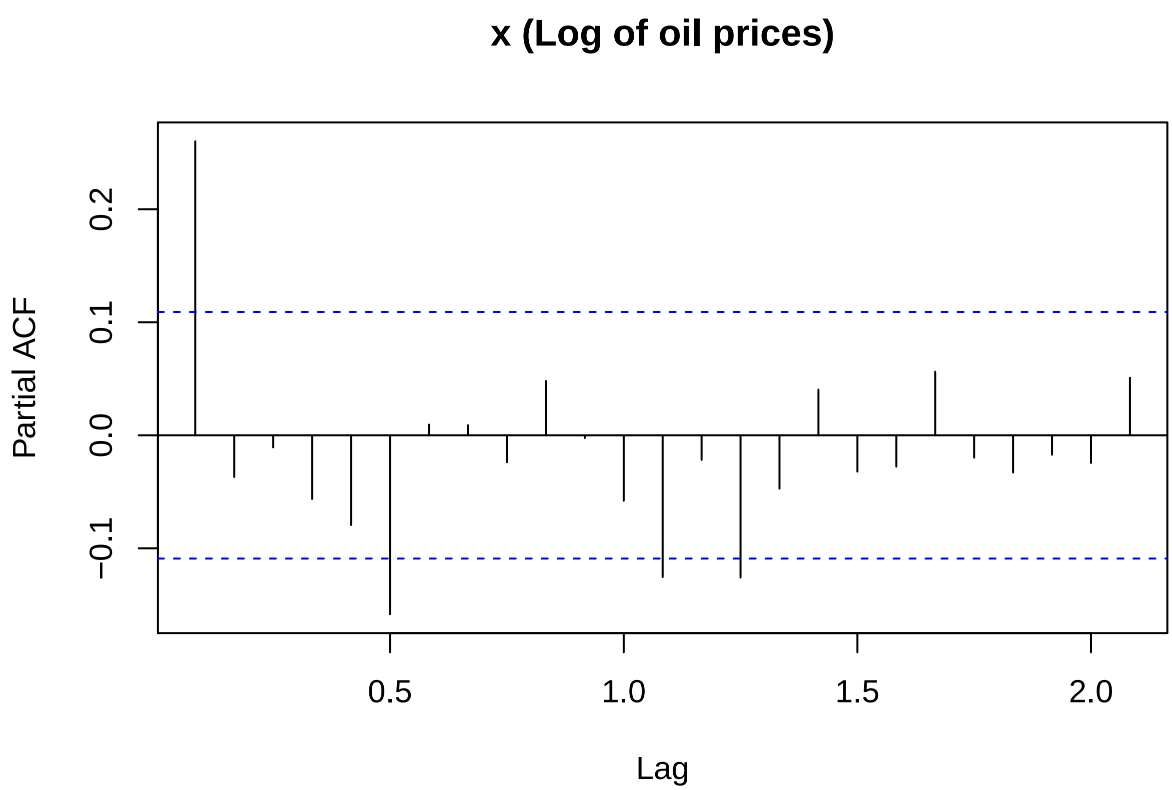
3. Beregn, plot og vurdér ACF og PACF for første-differensen af log-olieprisen

```r
xd    <- diff(xins)
xlagd <- diff(xlag)

# Plotter ACF og PACF for første-differensen af log-olieprisen
acf(xd , main = "x (Log of oil prices)")
```



```r
pacf(xd , main = "x (Log of oil prices)")
```



VURDER

4. *Detrend* log-olieprisen ved at beregne residualerne fra en regression af log-olieprisen på en konstant og en lineær trend. Disse residualer kan vi kalde den *detrendede* log-oliepris, dvs. log-olieprisen når trenden er fjernet

```r
dtx <- lm(xins  ~ time(xins))$residuals
```

5. (1) For den *detrendede* log-oliepris og (2) dernæst for første-differensen af log-olieprisen, beregn AIC (og eventuelt BIC) for alle ARMA(p, q) modeller med $p = 0, 1, 2, 3, 4$ og $q = 0, 1, 2, 3, 4$. Hvilke modeller vælges ud fra disse informationskriterier?

```r
df <- data.frame(q0 = numeric(5) ,
                 q1 = numeric(5) ,
                 q2 = numeric(5) ,
                 q3 = numeric(5) ,
                 q4 = numeric(5))
rownames(df) <- c("p0" , "p1", "p2" , "p3" , "p4")

for(p in 0:4){
  for(q in 0:4){
    df[p + 1, q + 1] <- arima(dtx , order = c(p,0,q))$aic
```

```
  }
}
```

```
## Warning in arima(dtx, order = c(p, 0, q)): possible convergence problem:
## optim gave code = 1
```

```r
print(df)
```

```
##            q0        q1        q2        q3        q4
## p0  185.1862 -170.0345 -368.5780 -500.7160 -545.1432
## p1 -658.5995 -682.9327 -682.2587 -680.9601 -678.9633
## p2 -684.6239 -682.6417 -680.6672 -678.9692 -679.6429
## p3 -682.6398 -680.6555 -679.0795 -684.6729 -675.8146
## p4 -680.6831 -681.6928 -684.8130 -677.6987 -678.2574
```

Ses at $p = 4$ og $q = 2$ giver laveste AIC

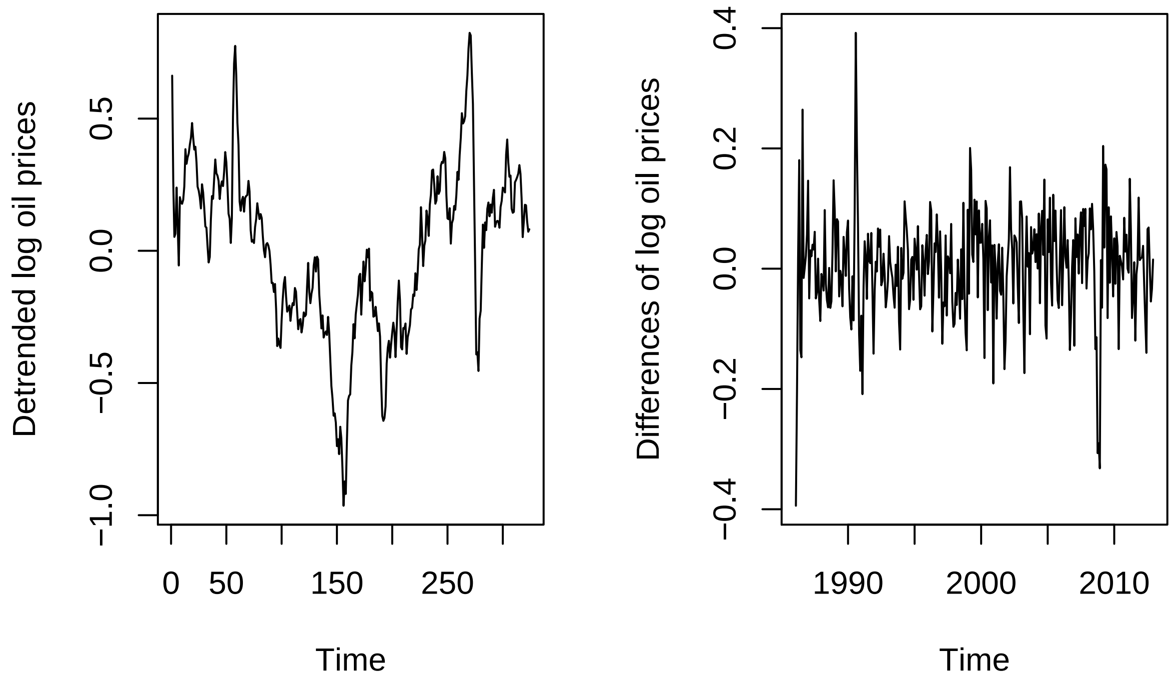
```r
df <- data.frame(q0 = numeric(5) ,
                 q1 = numeric(5) ,
                 q2 = numeric(5) ,
                 q3 = numeric(5) ,
                 q4 = numeric(5))
rownames(df) <- c("p0" , "p1", "p2" , "p3" , "p4")

for(p in 0:4){
  for(q in 0:4){
    df[p + 1, q + 1] <- arima(xd , order = c(p,0,q))$aic
  }
}
print(df)
```

```
##            q0        q1        q2        q3        q4
## p0 -656.1815 -678.0750 -676.6506 -674.7923 -673.2902
## p1 -678.3570 -676.6871 -674.6875 -680.3600 -680.0393
## p2 -676.7100 -674.6873 -672.6876 -680.0251 -669.7590
## p3 -674.7774 -682.0546 -674.8468 -676.7650 -669.8276
## p4 -674.2095 -671.7995 -674.3419 -683.5744 -681.5744
```

Ses at for første ordens differensen giver $p = 4$ og $q = 3$ den laveste AIC

```r
par(mfrow=c(1,2))
plot(dtx , type = "l" , ylab = "Detrended log oil prices" , xlab = "Time")
plot(xd , ylab = "Differences of log oil prices")
```




```
par(mfrow=c(1,1))
```

6. For den ARMA model du vælger for den detrendede log-oliepris: Forklar og beskriv estimationsoutputtet, og hvilken metode der bruges. Foretag passende diagnostiske tjek (modelkontrol)