UNC CHARLOTTE

ADVANCED BUSINESS & ECONOMIC FORECASTING

Jasper Peter de Bles

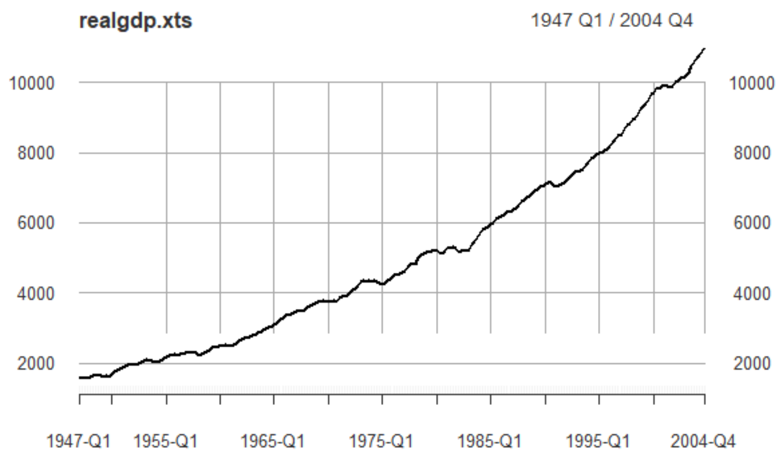
Professor McGregor

Capstone Project

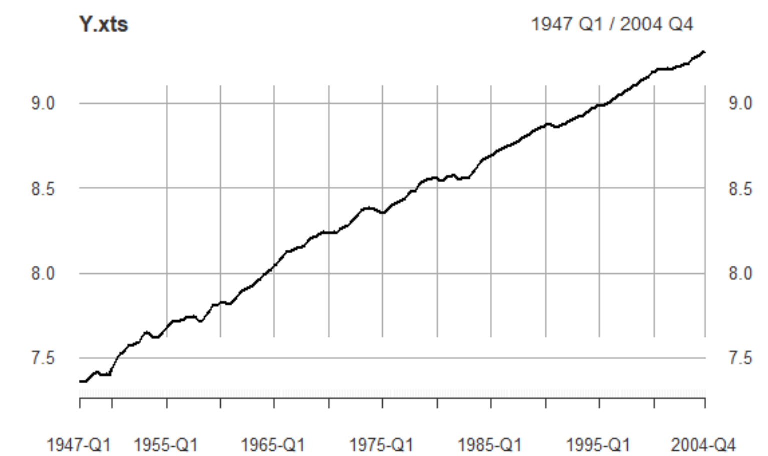
May 1, 2020

1 US GDP

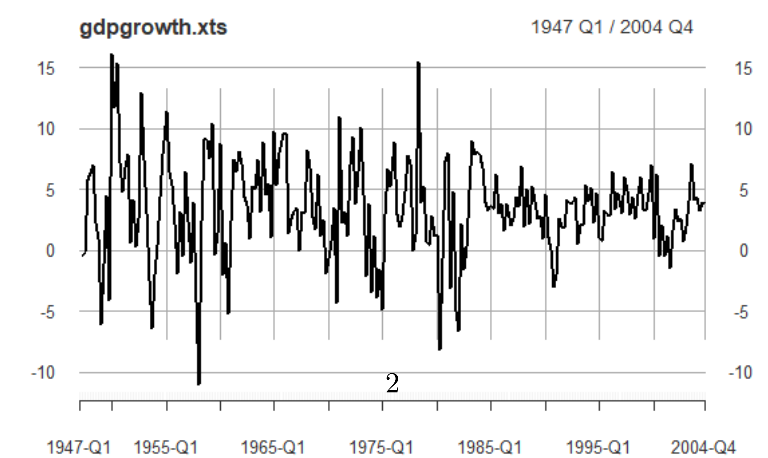
In Figure 1 is plotted how to make the timeseries for US real GDP stationary. First, the natural log is taken because there is an exponential trend in real GDP. Finally, the difference times 400 is taken to get the yearly GDP growth. In Figure 2 is plotted how to make the timeseries for the T-bill rate stationary. One needs to take the difference only to get to the T-bill spread.



(a) Real GDP

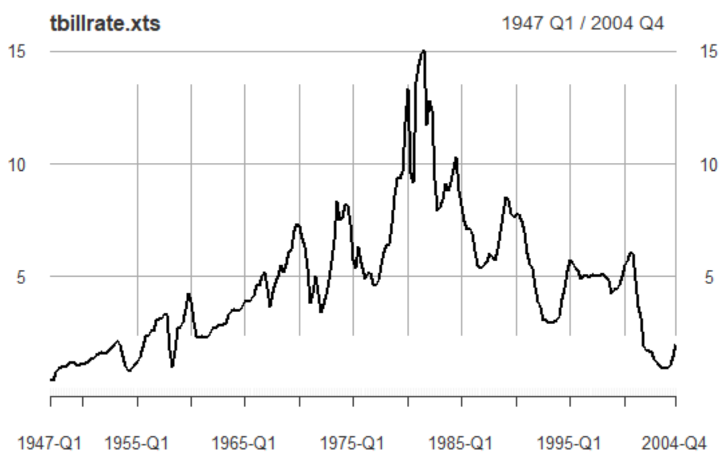


(b) Ln(GDP)

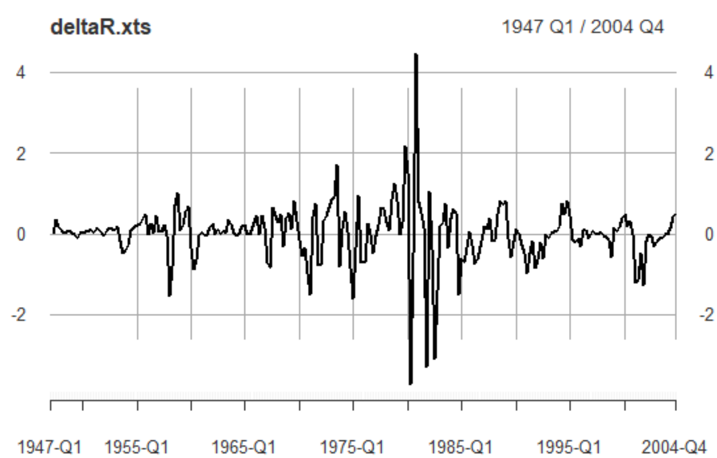


(c) GDP growth

Figure 1: Making the real GDP stationary



(a) T-bill rate



(b) T-bill spread

Figure 2: Making the T-bill rate stationary

1.1 AR(1) model

Now that the data has been made stationary, an autoregression of order one, AR(1), is executed on the sample from 1955q1 to 2004q4. I used the sarima function in *Rstudio*. The results are summarized in Table 1. The equation is:

$$\hat{GDPgrowth_t} = \underset{(0.352)}{3.321} + \underset{(0.068)}{0.295} GDPgrowth_{t-1} \tag{1}$$

Both the constant and the coefficient for the first lag are significant on the one per cent confidence level.

Table 1: Autoregression of order (1)

	Dependent variable:
	xdata
ar1	0.295***
	(0.068)
xmean	3.321***
	(0.352)
Observations	200
Log Likelihood	−535.066
σ^2	12.334
Akaike Inf. Crit.	1,076.132
Note:	*p<0.1; **p<0.05; ***p<0.01

1.2 ADL(1,4) model

Next, an autoregressive distributed lag model, ADL(1,4), is estimated using the dynlm function in *RStudio*. The results can be seen in Table **??**. The equation is:

$$\begin{aligned} GDP\hat{growth}_t = \underset{(0.361)}{2.365} + \underset{(0.069)}{0.317} GDPgrowth_{t-1} + \underset{(0.410)}{0.655} Tspread_{t-1} \\ - \underset{(0.403)}{1.662} Tspread_{t-2} + \underset{(0.404)}{0.342} Tspread_{t-3} - \underset{(0.397)}{1.329} Tspread_{t-4} \end{aligned} \tag{2}$$

Only the first lag of the difference in the interest rate is insignificant. The other coefficients are all significant at the one per cent confidence level.

Table 2: Autoregressive distributed lag model of order (1,4)

	Dependent variable:
	gdpgrowth_ts
L(gdpgrowth_ts)	0.317^{***}
	(0.069)
L(deltaR_ts)	0.655
	(0.410)
L(deltaR_ts, 2)	-1.662^{***}
	(0.403)
L(deltaR_ts, 3)	0.342
	(0.404)
L(deltaR_ts, 4)	-1.329^{***}
	(0.397)
Constant	2.365^{***}
	(0.361)
Observations	195
R^2	0.208
Adjusted R^2	0.187
Residual Std. Error	3.822 (df = 189)
F Statistic	9.924^{***} (df = 5; 189)
Note:	*p<0.1; **p<0.05; ***p<0.01

1.3 Forecast with AR(1)

With the equation obtained in question a, a forecast of the GDP is made. The results are reported in Table3.

Table 3: AR(1) forecast

Year	Observed	Forecast	Error
1990:01	4.593135387	3.62011722	-0.973018166
1990:02	1.022300169	4.67597494	3.65367477
1990:03	0.028048327	3.62257855	3.594530223
1990:04	-3.034988553	3.32927426	6.36426281
1991:01	-2.045663454	2.42567838	4.471341831
1991:02	2.587906219	2.71752928	0.129623062
1991:03	1.925787829	4.08443233	2.158644505
1991:04	1.871833494	3.88910741	2.017273916
1992:01	4.121769434	3.87319088	-0.248578553
1992:02	3.838637207	4.53692198	0.698284776
1992:03	3.905290303	4.45339798	0.548107673
1992:04	4.383250566	4.47306064	0.089810074
1993:01	0.482884488	4.61405892	4.131174429
1993:02	2.021773084	3.46345092	1.44167784
1993:03	2.048762253	3.91742306	1.868660807
1993:04	5.346277539	3.92538486	-1.420892675
1994:01	4.048886549	4.89815187	0.849265325
1994:02	5.182033482	4.51542153	-0.66661195
1994:03	2.235383995	4.84969988	2.614315883
1994:04	4.660070037	3.98043828	-0.679631758
1995:01	1.11018388	4.69572066	3.585536781
1995:02	0.71671583	3.64850424	2.931788415
1995:03	3.246677966	3.53243117	0.285753204
1995:04	2.914934552	4.27877	1.363835448
1996:01	2.810615205	4.18090569	1.370290488
1996:02	6.503181674	4.15013149	-2.353050189
1996:03	3.338989633	5.23943859	1.900448961
1996:04	4.649972512	4.30600194	-0.34397057
1997:01	3.081153845	4.69274189	1.611588046
1997:02	6.032006133	4.22994038	-1.802065749
1997:03	4.94974635	5.10044181	0.150695459
1997:04	2.938901826	4.78117517	1.842273348
1998:01	4.401831241	4.18797604	-0.213855202
1998:02	2.636710325	4.61954022	1.982829891
1998:03	4.580523355	4.09882955	-0.481693809
1998:04	6.029782593	4.67225439	-1.357528203
1999:01	3.38067836	5.09978586	1.719107505
1999:02	3.296986289	4.31830012	1.021313827
1999:03	4.640482217	4.29361096	-0.346871261
1999:04	7.047482452	4.68994225	-2.357540198
2000:01	1.012046991	5.40000732	4.387960333
2000:02	6.234424054	3.61955386	-2.614870191
2000:03	-0.459244634	5.1601551	5.61939973
2000:04	2.07257502	3.18552283	1.112947813
2001:01	-0.489796806	3.93240963	4.422206436
2001:02	1.225388315	3.17650994	1.951121627
2001:03	-1.407697261	3.68248955	5.090186814
2001:04	1.573220909	2.90572931	1.332508399
2002:01	3.356213324	3.78510017	0.428886844
2002:02	2.358570345	4.31108293	1.952512586
2002:03	2.566208932	4.01677825	1.45056932
2002:04	0.734698921	4.07803163	3.343332714
2003:01	1.909424254	3.53773618	1.628311927
2003:02	4.025083056	3.88428016	-0.140802901
2003:03	7.144629785	4.5083995	-2.636230283
2003:04	4.100066771	5.42866579	1.328599015
2004:01	4.391392999	4.5305197	0.139126699
2004:02	3.247357493	4.61646093	1.369103441
2004:03	3.92331882	4.27897046	0.35565164
2004:04	3.776078634	4.47837905	0.702300418

1.4 Forecast with ADL(1,4)

Similarly, with the equation obtained in question b, a forecast of the GDP is made. The results are reported in Table4.

Table 4: ADL(1,4) forecast

Year	Observed	Forecast	Error
1990:01	4.593135387	2.378404238	-2.21473
1990:02	1.022300169	4.185641628	3.163341
1990:03	0.028048327	3.191176894	3.163129
1990:04	-3.034988553	2.50818692	5.543175
1991:01	-2.045663454	1.386556489	3.43222
1991:02	2.587906219	1.817576945	-0.77033
1991:03	1.925787829	4.680879521	2.755092
1991:04	1.871833494	3.941631342	2.069798
1992:01	4.121769434	3.841292398	-0.28048
1992:02	3.838637207	5.191644311	1.353007
1992:03	3.905290303	4.474386815	0.569097
1992:04	4.383250566	4.457485926	0.074235
1993:01	0.482884488	5.523890389	5.041006
1993:02	2.021773084	2.547633273	0.52586
1993:03	2.048762253	3.981506488	1.932744
1993:04	5.346277539	3.007479964	-2.3388
1994:01	4.048886549	4.18443105	0.135545
1994:02	5.182033482	3.678065936	-1.50397
1994:03	2.235383995	4.160557854	1.925174
1994:04	4.660070037	2.146520076	-2.51355
1995:01	1.11018388	3.564620062	2.454436
1995:02	0.71671583	0.86059382	0.143878
1995:03	3.246677966	1.365042238	-1.88164
1995:04	2.914934552	2.564782485	-0.35015
1996:01	2.810615205	2.946630973	0.136016
1996:02	6.503181674	3.32450056	-3.17868
1996:03	3.338989633	5.303107451	1.964118
1996:04	4.649972512	3.353002994	-1.29697
1997:01	3.081153845	4.102365746	1.021212
1997:02	6.032006133	3.502358059	-2.52965
1997:03	4.94974635	3.986985904	-0.96276
1997:04	2.938901826	4.144202913	1.205301
1998:01	4.401831241	3.209708599	-1.19212
1998:02	2.636710325	3.682062763	1.045352
1998:03	4.580523355	3.226389273	-1.35413
1998:04	6.029782593	3.773870413	-2.25591
1999:01	3.38067836	4.180458872	0.799781
1999:02	3.296986289	4.5338916	1.236905
1999:03	4.640482217	3.194704734	-1.44578
1999:04	7.047482452	4.697275073	-2.35021
2000:01	1.012046991	4.341986407	3.329939
2000:02	6.234424054	2.349573286	-3.88485
2000:03	-0.459244634	3.548862465	4.008107
2000:04	2.07257502	1.737078261	-0.3355
2001:01	-0.489796806	1.950479631	2.440276
2001:02	1.225388315	1.270541123	0.045153
2001:03	-1.407697261	3.587090386	4.994788
2001:04	1.573220909	3.109795508	1.536575
2002:01	3.356213324	4.049828738	0.693615
2002:02	2.358570345	6.778120814	4.41955
2002:03	2.566208932	3.632166719	1.065958
2002:04	0.734698921	4.757782741	4.023084
2003:01	1.909424254	2.758121778	0.848698
2003:02	4.025083056	3.353564049	-0.67152
2003:03	7.144629785	3.849590259	-3.29504
2003:04	4.100066771	5.103275462	1.003209
2004:01	4.391392999	4.033690017	-0.3577
2004:02	3.247357493	3.896660471	0.649303
2004:03	3.92331882	3.640843465	-0.28248
2004:04	3.776078634	3.629037636	-0.14704

1.5 RMSFE

The root mean square forecast error is calculated as:

$$RMSFE = \sqrt{\sum_{i=1}^{N} (foreast - observed)^2 / N} \tag{3}$$

For the AR(1) model, the RMSFE for the forecast from 1990 onwards is $\sqrt{\frac{331}{60}} = 2.349$. For the ADL(1,4) model, the RMSFE for the forecast from 1990 onwards is $\sqrt{\frac{310}{60}} = 2.274$. Thus, the RMFSE of the ADL model is slightly smaller than that of the AR model. It is not unexpected that the ADL model is a better predictor than the AR model because four more variables, i.e. the four lags of the Tspread, are included in the model that might help explain the change in GDP growth.

The ADL model might be biased because the T-bill rate is influenced by GDP growth itself.

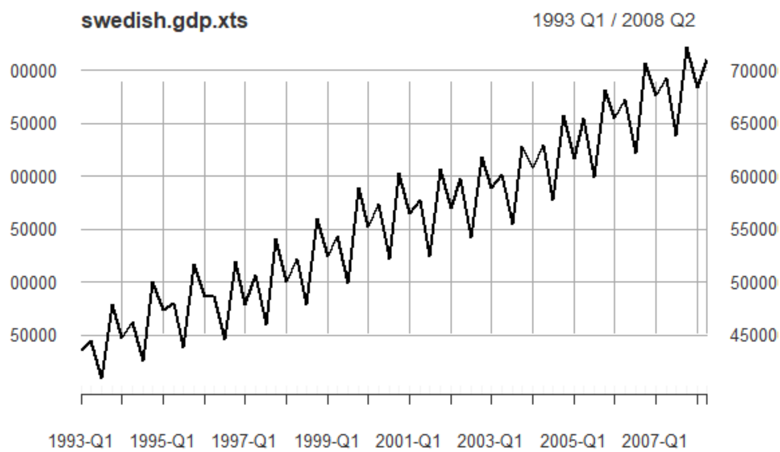


Figure 3: Swedish GDP

2 Swedish GDP

Figure 3 displays the timeseries of the Swedish GDP. Two features can be clearly distinguished. First, the GDP of Sweden is increasing over the years. Second, there is a seasonal pattern with a rise in the first and third quarter of the year, and a drop in the second and fourth quarters. Figure 4 shows the decomposition of the Swedish GDP.

A linear trend model is a model that fits the data into a straight line. Swedish GDP has a positive trend.

$$Data = trend + seasonality + residuals \tag{4}$$

From Figure 6 it is easy to see that including the seasonal dummy variables significantly improves the forecast. The basic linear trend model is just a straight line due to the single variable, the trend. The model with just a trend has an RMSFE of 28447.61233 and the model with the seasonal adjustments only 9285.136434.

The seasonal linear trend model already provides a pretty good fit. A possibility to improve the model even further is to check the model that includes for example an exponential trend or a damped trend. One could also look at moving averages.

Table 5: Results Linear Trend Models Swedish GDP

	Swedish GDP	
	swrgdp	
	Linear Trend Model	Seasonally Adjusted
	(1)	(2)
Constant	$-33{,}563{,}644.00^{***}$	$-31{,}566{,}145.00^{***}$
	(882,059.40)	(723,701.00)
Trend	$17{,}056.17^{***}$	$16{,}054.48^{***}$
	(441.04)	(361.43)
Quarter 2		$13{,}789.38^{***}$
		(2,770.70)
Quarter 3		$-36{,}781.93^{***}$
		(1,611.55)
Quarter 4		$37{,}944.76^{***}$
		(1,864.71)
Observations	62	52
R^2	0.88	0.99
Adjusted R^2	0.88	0.99
Residual Std. Error	28,530.30 (df = 60)	7,290.50 (df = 47)
F Statistic	443.52^{***} (df = 1; 60)	$1{,}081.99^{***}$ (df = 4; 47)
Note:		*p<0.1; **p<0.05; ***p<0.01

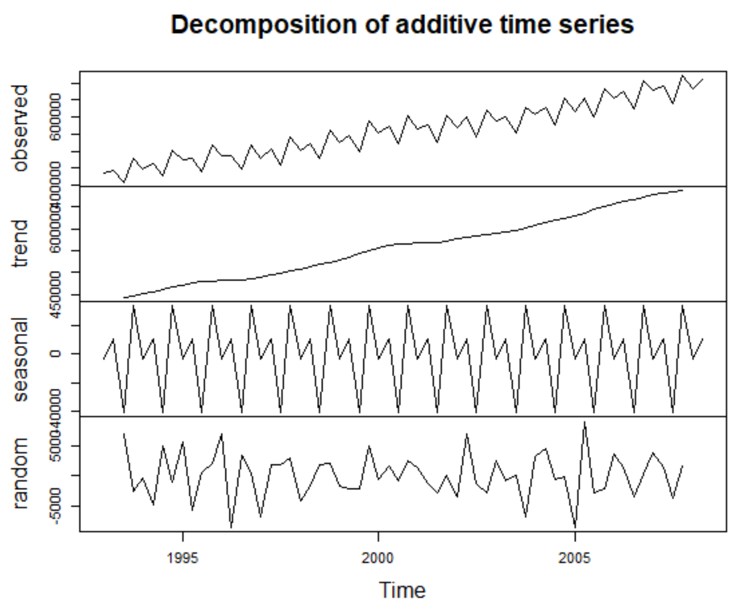
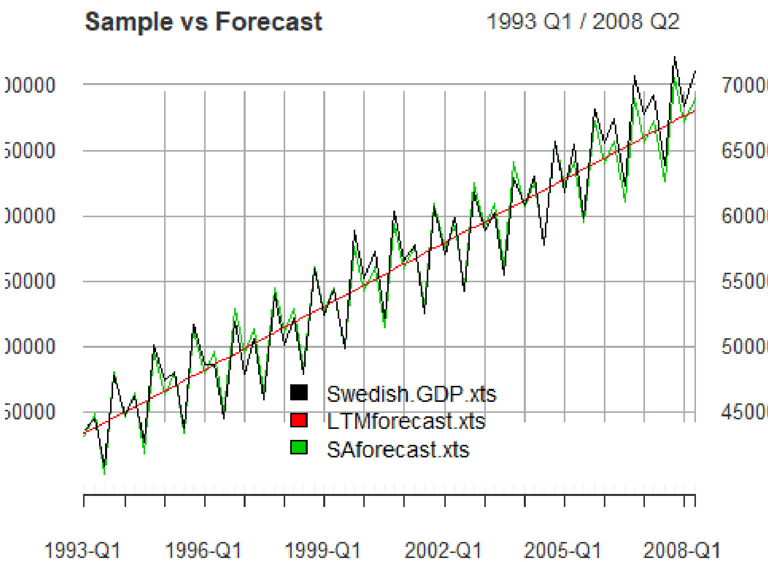
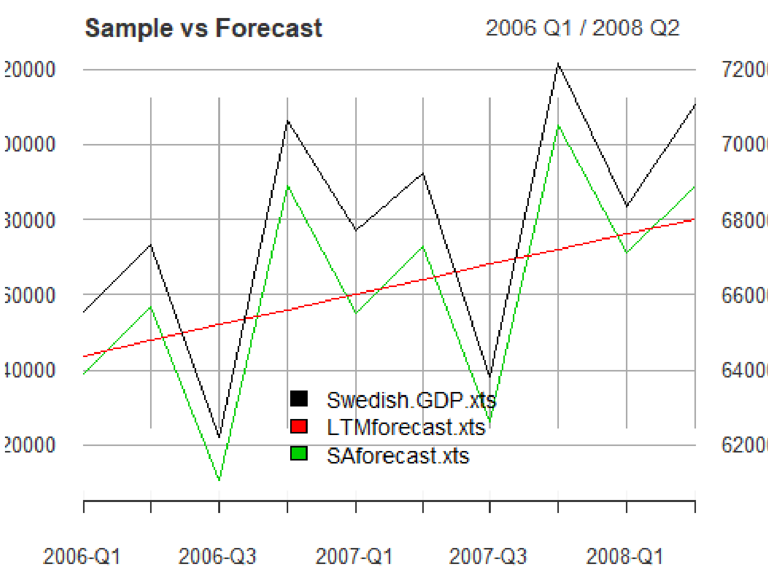


Figure 4: Decomposed factors of Swedish GDP



(a) Entire period



(b) Last two years

Figure 5: Forecast of the Swedish GDP

Date	forecast (1)	forecast (2)
1/1/2006	643808.1	639141.13
4/1/2006	647856.8	656944.13
7/1/2006	651905.5	610386.44
10/1/2006	655954.2	689126.75
1/1/2007	660002.9	655195.61
4/1/2007	664051.6	672998.61
7/1/2007	668100.3	626440.92
10/1/2007	672149	705181.23
1/1/2008	676197.7	671250.09
4/1/2008	680246.4	689053.09

Figure 6: Forecast of the Swedish GDP

Year	Price crude oil	Log	Crude oil growth
1/1/2011	93.06143028	4.53326	51.52736576
2/1/2011	96.68267791	4.571434	45.80932604
3/1/2011	99.27926964	4.597937	31.80303616

Table 6: 3-month forecast crude oil

3 Crude Oil

Figure 7 displays the timeseries for the price of crude oil and how to make it stationary by taking the natural logarithm, take the differences, and finally, since the data is monthly data, multiply by 1200 to make the growth of the oil price a yearly rate.

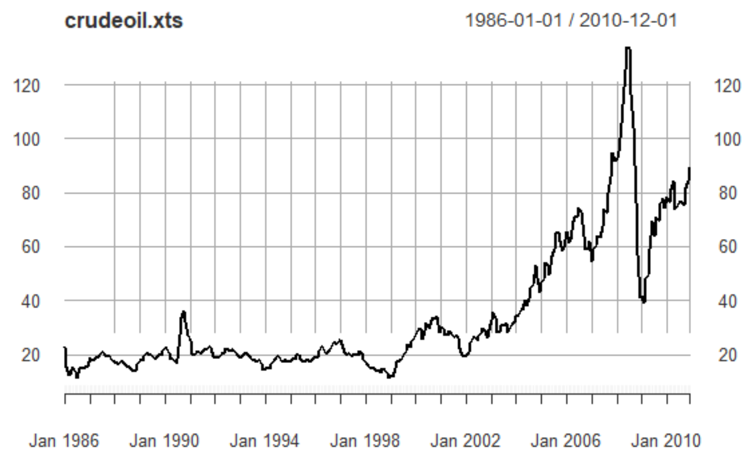
The autocorrelation function and partial correlation function in Figure 8 help to choose the right model. The PACF of the price of crude oil indicates that a maximum of three lags should be used, e.g. AR(1), AR(2), or AR(3).

The AIC and BIC are used in selecting the best model, although the BIC is usually preferred. The lower the values for the AIC and BIC, the better the model. The auto.arima function in r gives for both the ARIMA(3,1,1) model. That means that to predict the price of crude oil, one needs three lags of the price of crude oil itself, and an integrated moving average. Table 8 displays several models and their information criteria. It can be clearly seen that the ARIMA(3,1,1) has the lowest value for both AIC and BIC.

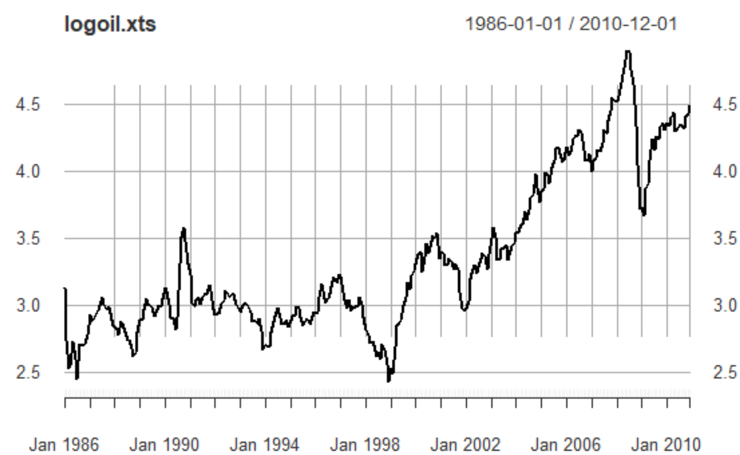
Thus, to summarize, the PACF of the price of crude oil indicates an AR(3) model. However, after taking the auto.arima function, it can be noted that adding an integrated MA(1) is beneficial for lowering the AIC and BIC. The results of the ARIMA(3,1,1) can be seen in Table 7. There is no intercept. The equation is given as:

$$\begin{aligned} \hat{crudeoil}_t = \underset{(0.067)}{1.212} oilgrowth_{t-1} - \underset{(0.092)}{0.102} oilgrowth_{t-2} \\ -\underset{(0.092)}{0.259} oilgrowth_{t-3} - \underset{(0.043)}{0.892} MA \end{aligned} \tag{5}$$

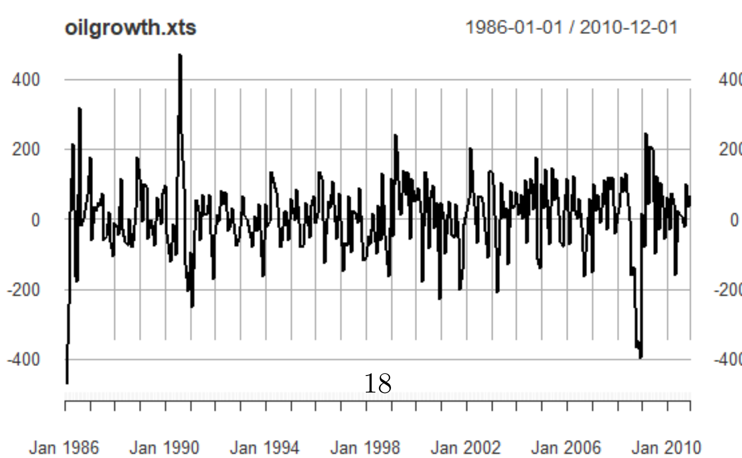
With this equation, the crude oil values for the next three months, January, February, and March 2011, are being forecasted. The results can bee seen in Table 6.



(a) The price of crude oil

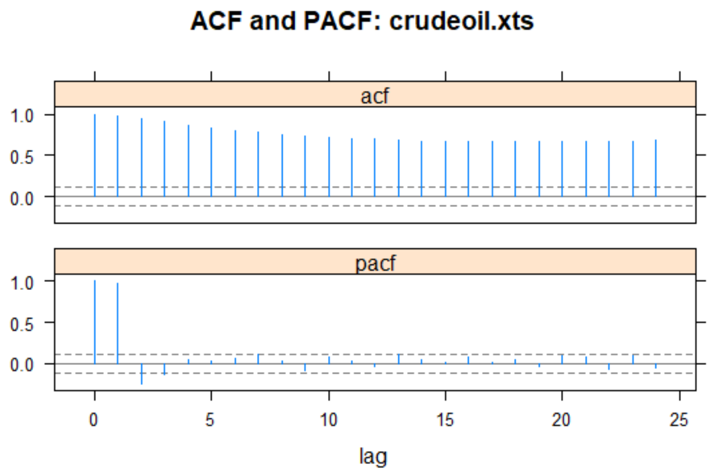


(b) The natural log of the oil price

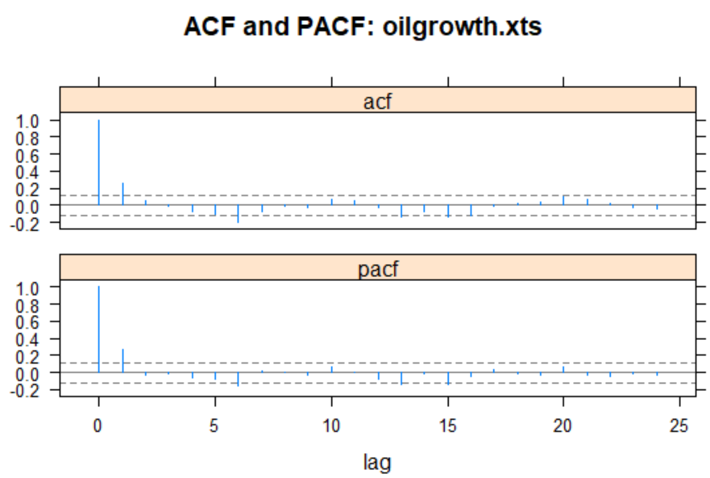


(c) The growth in the oil price

Figure 7: Making the price of oil stationary



(a) Price of crude oil



(b) Crude oil price growth

Figure 8: Autocorrelation functions

	Dependent variable:
	xdata
ar1	1.212***
	(0.067)
ar2	−0.102
	(0.092)
ar3	−0.259***
	(0.057)
ma1	−0.892***
	(0.043)
Observations	299
Log Likelihood	−787.181
σ^2	11.304
Akaike Inf. Crit.	1,584.362
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7: ARIMA(3,1,1)

Model	AIC	BIC
AR(1)	5.58693	5.623968
AR(2)	5.388169	5.437553
AR(3)	5.366704	5.428434
AR(4)	5.363937	5.438013
ARIMA(3,1,1)	5.296663	5.37092

Table 8: Information criteria

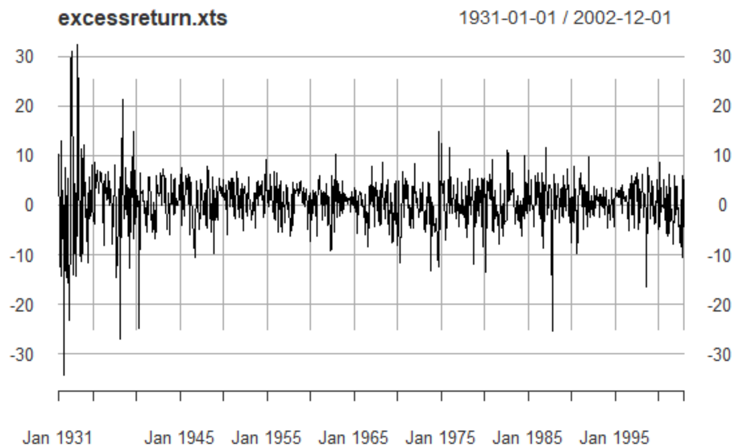


Figure 9: Excess return of the stock

4 Stock & Watson E15.2

Figure 9 plots the timeseries of the excess return of the stock. It is stationary, as also confirmed by the ADF test.

4.1 a

Table 9 shows the results of the three autoregressive models. It should be noted that these models are based on the entire dataset, that is 1932-2002, whereas the example in the book is based on 1960-2002.

4.2 b

The errors are very high. The AR(4) has the lowest RMSFE.

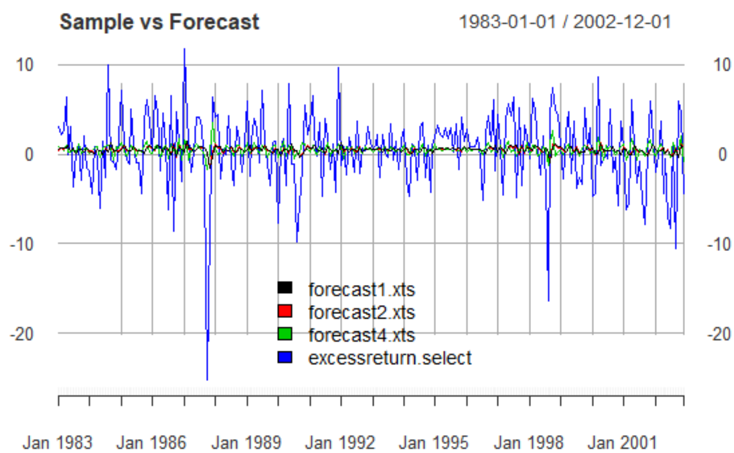
4.3 c

The results from (a) and (b) do not have different conclusions from the book. Except for the constants, none of the coefficients are statistically significant. Furthermore, the values for R^2 and adjusted $R^2$are very low. These kind of results are very normal for stock prices because we simply cannot predict the stockprice. If this was the case all investors would buy or sell the stock at the same time and all make a profit which is not possible because there will allways be winners and losers in the financial market.

Table 9: Autoregressive Models of Monthly Excess Stock Returns

	excess return $_t$		
	AR(1)	AR(2)	AR(4)
	(1)	(2)	(3)
Constant	0.454**	0.452**	0.495**
	(0.180)	(0.186)	(0.206)
excess return $_{t-1}$	0.082	0.082	0.084
	(0.060)	(0.064)	(0.061)
excess return $_{t-2}$		-0.017	-0.003
		(0.059)	(0.054)
excess return $_{t-3}$			-0.102
			(0.067)
excess return $_{t-4}$			0.040
			(0.053)
Observations	863	862	860
R^2	0.007	0.007	0.018
Adjusted R^2	0.006	0.004	0.013
Residual Std. Error	5.362 (df = 861)	5.358 (df = 859)	5.322 (df = 855)
F Statistic	5.864** (df = 1; 861)	2.910* (df = 2; 859)	3.886*** (df = 4; 855)

Note: The sample period is 1932-2002 $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$



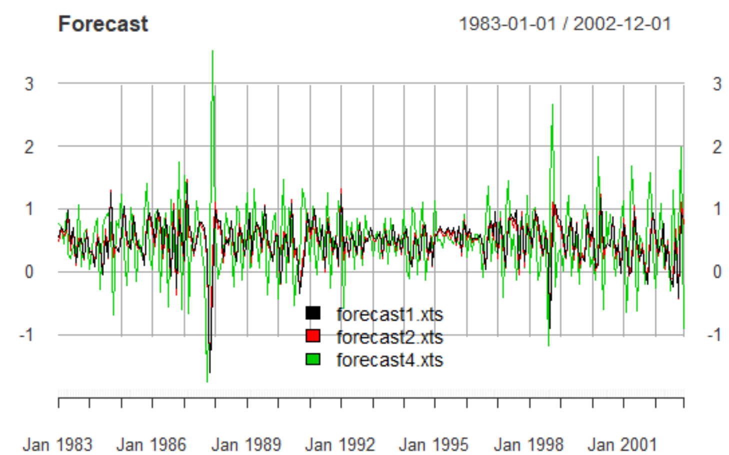


Figure 10: Excess return and AR(p) forecasts

5 R Script

```
# Advanced Business & Economic Forecasting
# Capstone Project
# Jasper P de Bles

# Download packages
suppressMessages(library(xts))
suppressMessages(library(astsa))
suppressMessages(library(stargazer))
suppressMessages(library(dynlm))
suppressMessages(library(sandwich))
suppressMessages(library(lmtest))
suppressMessages(library(HH))
suppressMessages(library(forecast))
suppressMessages(library(aTSA))
suppressMessages(library(quantmod))
suppressMessages(library(tseries))
suppressMessages(library(timeSeries))
suppressMessages(library(basicTrendline))


# Section I
# Import the data
table1 = read.csv("capstone1.csv")
table1$date = as.yearqtr(table1$date, "%Y:%q")
realgdp.xts = as.xts(table1$realgdp, order.by = table1
   $date)
plot.xts(realgdp.xts)

Y.xts = log(realgdp.xts)
plot.xts(Y.xts)

gdpgrowth.xts = 400 * diff(Y.xts)
plot.xts(gdpgrowth.xts)

tbillrate.xts = as.xts(table1$tbillrate, order.by =
   table1$date)
plot.xts(tbillrate.xts)
```

```r
deltaR.xts = diff(tbillrate.xts)
plot.xts(deltaR.xts)

# Question a
gdpgrowth.select = gdpgrowth.xts["1955/2004"]
AR.1 = sarima(gdpgrowth.select, 1, 0, 0)
stargazer(AR.1, type = "latex")

# Question b
gdpgrowth_ts = ts(gdpgrowth.xts,
                 start = c(1955,1),
                 end = c(2004, 4),
                 frequency = 4)

deltaR_ts = ts(deltaR.xts,
                  start = c(1955,1),
                  end = c(2004, 4),
                  frequency = 4)

ADL1.4data = ts.union(gdpgrowth_ts, deltaR_ts)
ADL1.4 = dynlm(gdpgrowth_ts ~ L(gdpgrowth_ts) + L(
   deltaR_ts) + L(deltaR_ts, 2)
                + L(deltaR_ts, 3) + L(deltaR_ts, 4),
                start = c(1955, 1), end = c(2004, 4))
coeftest(ADL1.4, vcov. = sandwich)
stargazer(ADL1.4, type = "latex")



# Sectoin II
# Import the data
table2 = read.csv("capstone2.csv")
table2$date = as.yearqtr(table2$date, "%m/%d/%Y")
swedish.gdp.xts = as.xts(table2$swrgdp, order.by =
   table2$date)
plot.xts(swedish.gdp.xts)
swedish.gdp.xts = swedish.gdp.xts["1993/2005"]

# Linear trend model
```

```
ltm = lm(swrgdp ~ date, data = table2)
plot(ltm$model)
summary(ltm)
stargazer(ltm, type = "text")

swedish.gdp.ts = ts(table2$swrgdp, frequency = 4,
    start = c(1993, 1), end = c(2005,4))
ts.stl = stl(swedish.gdp.ts,"periodic")
ts.sa = seasadj(ts.stl)
plot(swedish.gdp.ts, type = "l")
plot(ts.sa, type = "l")
seasonplot(ts.sa, 4, col = rainbow(4), year.labels =
    TRUE)

# Create dummy variables
DF = data.frame(Date = seq(as.Date("1993-01-01"),
    length.out = 62, by = "3 months"))
dummytable = data.table::dcast(DF, Date ~ paste0("Q",
    lubridate::quarter(DF$Date)), length,
                   value.var = "Date")

table2$Q2 = dummytable$Q2
table2$Q3 = dummytable$Q3
table2$Q4 = dummytable$Q4
table2 = table2[-(53:62),]

# Linear trend model with 3 seasonal dummy variables
ltmseasonal = lm(swrgdp ~ date + Q2 + Q3 + Q4, data =
    table2)
summary(ltmseasonal)
stargazer(ltmseasonal, type = "text")


# HAC
robust.se.1 = list(sqrt(diag(vcovHAC(ltm, type = "HAC1
    ")))),
                  sqrt(diag(vcovHAC(ltmseasonal, type =
                      "HAC1"))))

# Make one table for comparison
```

```
stargazer(ltm,
          ltmseasonal,
          se = robust.se.1,
          type = "latex", intercept.bottom = FALSE,
          column.labels = c("Linear Trend Model", "
             Seasonally Adjusted"),
          title = "Results Linear Trend Models Swedish
             GDP",
          digits = 2,
          dep.var.caption  = "Swedish GDP",
          covariate.labels = c("Constant", "Trend", "
             Quarter 2", "Quarter 3",
                                 "Quarter 4"))

ltm2 = tslm(swedish.gdp.ts ~ trend)
summary(ltm2)
ltmseasonal2 = tslm(swedish.gdp.ts ~ trend + table2$Q2
    + table2$Q3 + table2$Q4)
summary(ltmseasonal2)

decomposedRes <- decompose(swedish.gdp.ts, type="
   additive")
plot (decomposedRes) # see plot below
stlRes <- stl(swedish.gdp.ts, s.window = "periodic")
stlRes

# Forecast

capstone2.forecast = read.csv("capstone2forecast.csv")
capstone2.forecast$date = as.yearqtr(capstone2.
   forecast$date, "%m/%d/%Y")
Swedish.GDP.xts = as.xts(capstone2.forecast$swrgdp,
   order.by = capstone2.forecast$date)
LTMforecast.xts = as.xts(capstone2.forecast$forecast
   ..1., order.by = capstone2.forecast$date)
SAforecast.xts = as.xts(capstone2.forecast$forecast
   ..2., order.by = capstone2.forecast$date)
plot.xts(merge.xts(Swedish.GDP.xts, LTMforecast.xts,
   SAforecast.xts),
```

```
          lwd = 0.1, legend.loc = "bottom", main = "
             Sample vs Forecast")
Swedish.GDP.xts = Swedish.GDP.xts["2006/2008"]
LTMforecast.xts = LTMforecast.xts["2006/2008"]
SAforecast.xts = SAforecast.xts["2006/2008"]


# Section III
# Import the data
table3 = read.csv("capstone3.csv")
table3$date = as.Date(table3$date, "%Y-%m-%d")
crudeoil.xts = as.xts(table3$oilprice, order.by =
   table3$date)
plot.xts(crudeoil.xts)

logoil.xts = log(crudeoil.xts)
plot.xts(logoil.xts)

oilgrowth.xts = 1200 * diff(logoil.xts)
plot.xts(oilgrowth.xts)

oilgrowth.xts = na.omit(oilgrowth.xts)
acf(oilgrowth.xts)

acf.pacf.plot(oilgrowth.xts)
acf.pacf.plot(crudeoil.xts)

ARIMA3.1.1 = sarima(crudeoil.xts, 3, 1, 1, no.constant
    = TRUE)
?sarima
stargazer(ARIMA3.1.1, type = "latex")

auto.arima(crudeoil.xts, ic = "aic")


# Section IV
# Import the data
stockreturn = read.csv("Stock_Returns_1931_2002.csv")
stockreturn$X = 1
```

```
stockreturn$date = as.Date(with(stockreturn, paste(
    time, Month, X, sep="-")), "%Y-%m-%d")
excessreturn.xts = as.xts(stockreturn$ExReturn, order.
    by = stockreturn$date)
plot.xts(excessreturn.xts, lwd = 0.1)

return.ts = as.ts(stockreturn$ExReturn, order.by =
    stockreturn$date)

adf.test(excessreturn.xts)

# Run the AR models
AR.1.stock = dynlm(return.ts ~ L(return.ts))
summary(AR.1.stock)
coeftest(AR.1.stock, vcov. = vcovHAC)

AR.2.stock = dynlm(return.ts ~ L(return.ts) + L(return
    .ts, 2))
summary(AR.2.stock)
coeftest(AR.2.stock, vcov. = vcovHAC)

AR.4.stock = dynlm(return.ts ~ L(return.ts) + L(return
    .ts, 2) +
                        L(return.ts, 3) + L(return.ts, 4)
                            )
summary(AR.4.stock)
coeftest(AR.4.stock, vcov. = vcovHAC)

robust.se = list(sqrt(diag(vcovHAC(AR.1.stock, type =
    "HAC1"))),
                  sqrt(diag(vcovHAC(AR.2.stock, type =
                      "HAC1"))),
                  sqrt(diag(vcovHAC(AR.4.stock, type =
                      "HAC1"))))

# Create a table like in the book
stargazer(AR.1.stock,
          AR.2.stock,
          AR.4.stock,
          se = robust.se,
```

```
          type = "latex", intercept.bottom = FALSE,
          column.labels = c("AR(1)", "AR(2)", "AR(4)")
             ,
          title = "Autoregressive models of monthly
             excess stock returns",
          digits = 3,
          dep.var.caption  = "excess return t",
          covariate.labels = c("Constant", "excess
             return t-1", "excess return t-2",
                              "excess return t-3", "
                                 excess return t-4"))

acf(AR.1.stock$residuals,main="Residuals plot")

# Forecast
forecast.return = read.csv("returnsforecast.csv")
excessreturn.select = excessreturn.xts["1983/"]
forecast.return$X = 1
forecast.return$date = as.Date(with(forecast.return,
   paste(time, Month, X, sep="-")), "%Y-%m-%d")
forecast1.xts = as.xts(forecast.return$AR1, order.by =
     forecast.return$date)
forecast2.xts = as.xts(forecast.return$AR2, order.by =
     forecast.return$date)
forecast4.xts = as.xts(forecast.return$AR4, order.by =
     forecast.return$date)
plot.xts(merge.xts(forecast1.xts, forecast2.xts,
   forecast4.xts, excessreturn.select),
         lwd = 0.1, legend.loc = "bottom", main = "
            Sample vs Forecast")
plot.xts(merge.xts(forecast1.xts, forecast2.xts,
   forecast4.xts),
         lwd = 0.1, legend.loc = "bottom", main = "
            Forecast")
```