All R code in PDF

Sam Ansari, s3765571

4 juni 2020

ALL THE R CODE WE USED

Library

```
library("DataCombine")
library('ggplot2')
library("corrplot")
library("tidyverse")
library("dplyr")
library("openxlsx")
library("tseries")
library('fpp2')     # For forecasting
library('dynlm')    # To estimate ARDL models
library('urca')     # For the Dickey Fuller test
library('corrplot')# For plotting correlation matrices
library('quadprog')# For quadratic optimization
library('forecast')
library('readxl')  # To read Excel files
library('fpp2')     # For forecasting
library('tseries') # To estimate ARMA models
library('dynlm')    # To estimate ARDL models
library('urca')     # For the Dickey Fuller test
library('corrplot')# For plotting correlation matrices
library('quadprog')# For quadratic optimization
library('forecast')# Lots of handy forecasting routines
library('vars')     # VARs
library('zoo')
library('lubridate')
```

Data loading

```
data_1 <- read.xlsx("WEI.xlsx", sheet = 2, detectDates = TRUE)
sp500_newdata <- read.csv("sp500newdata.csv")
CCI       <- read.csv('CCI.csv')
```

Data manipulation

```
sp500_newdata <- sp500_newdata %>%
  mutate(average_open_close = (Open + Close) / 2)

data <- data %>% cbind(sp500data$average_open_close)
```

```
colnames(data)[11] <- "average_open_close"

sp500_52week_change <- PercChange(data = sp500_newdata, Var = "average_open_close", NewVar = "sp500_52we
sp500_52week_change <- sp500_52week_change$sp500_52week_change
sp500_52week_change <- sp500_52week_change[!is.na(sp500_52week_change)]
data_1$sp500_52week_change <- sp500_52week_change
sp_500_52week_diff <- diff(sp500_newdata$average_open_close, lag = 52)
data_1$sp_500_52week_diff <- sp_500_52week_diff
#read CSV file and obtain data from 2007-2020, with values around 0
CCI_data = CCI %>% slice(3:nrow(CCI)) %>% mutate(percentage = Value - 100)
CCI_2007       <- ts(CCI_data[,9],start = 2007,frequency=12)

#Take difference with respect to the value of last year
diff_CCI = diff(CCI_2007, 12)
diff_CCI = ts(as.vector(diff_CCI), start = 2008, frequency = 12)

# Merge low and high freq time series
lowfreq      <- zoo(diff_CCI,time(diff_CCI))
highfreq     <- zoo(WEI,time(WEI))
merged       <- merge(lowfreq,highfreq)

# Approximate the NAs and output at the dates of the WEI
CCIw         <- na.approx(merged$lowfreq, xout = time(WEI),rule=2)
CCIw         <- ts(CCIw,start = 2008,frequency=52)

data_1$CCIw =as.vector(CCIw)

#preparing all time series
WEI_365       <- ts(data_1$WEI, decimal_date(ymd("2008-01-05")), frequency = 365.25/7)
CCIw_365       <- ts(data_1$CCIw, decimal_date(ymd("2008-01-05")), frequency = 365.25/7)
sp500_52week_change_365        <- ts(data_1$sp500_52week_change, decimal_date(ymd("2008-01-05")), freque
sp_500_52week_diff_365       <- ts(data_1$sp_500_52week_diff, decimal_date(ymd("2008-01-05")), frequency
noise<-ts(rnorm(length(CCIw_365))*sqrt(sd((CCIw_365)/100)),decimal_date(ymd("2008-01-05")),frequency=365
CCIn <- CCIw_365+noise
```

Plots we used

```
plot_WEI_SP500_CCI <- ggplot(data = data_1, aes(x = Date)) +
  geom_line(aes(y = WEI, colour = "WEI")) +
  geom_line(aes(y = sp500_52week_change / 10, colour = "S&P500")) +
  geom_line(aes(y = CCIw * 1.5, colour = "CCI")) +
  geom_hline(yintercept = 0, colour = 'black') +
  scale_color_manual("", values = c("WEI" = "green", "S&P500" = "blue", "CCI" = "red")) +
  ggtitle("WEI vs S&P500 52 week % change scaled by 10 vs CCI scaled by 1.5") +
  ylab("WEI, S&P500 and CCI") +
  xlab("Date")
plot_WEI_SP500_CCI

plot_WEI_SP500_CCI_2008_2010 <- ggplot(data = data_1[1:105, ], aes(x = Date)) +
  geom_line(aes(y = WEI, colour = "WEI")) +
  geom_line(aes(y = sp500_52week_change / 10, colour = "S&P500")) +
  geom_line(aes(y = CCIw * 1.5, colour = "CCI")) +
  geom_hline(yintercept = 0, colour = 'black') +
```

```
  scale_color_manual("", values = c("WEI" = "green", "S&P500" = "blue", "CCI" = "red")) +
  ggtitle("WEI vs S&P500 52 week % change scaled by 10 vs CCI scaled by 1.5 during 2008-2010") +
  ylab("WEI, S&P500 and CCI") +
  xlab("Date")
plot_WEI_SP500_CCI_2008_2010

plot_WEI_SP500_CCI_before_covid <- ggplot(data = data_1[560:630,], aes(x = Date)) +
  geom_line(aes(y = WEI, colour = "WEI")) +
  geom_line(aes(y = sp500_52week_change / 10, colour = "S&P500")) +
  geom_line(aes(y = CCIw * 1.5, colour = "CCI")) +
  geom_hline(yintercept = 0, colour = 'black') +
  scale_color_manual("", values = c("WEI" = "green", "S&P500" = "blue", "CCI" = "red")) +
  ggtitle("WEI vs S&P500 52 week % change scaled by 10 vs CCI scaled by 1.5 before covid") +
  ylab("WEI, S&P500 and CCI") +
  xlab("Date")
plot_WEI_SP500_CCI_before_covid

plot_WEI_SP500_CCI_during_covid <- ggplot(data = data_1[630:639, ], aes(x = Date)) +
  geom_line(aes(y = WEI, colour = "WEI")) +
  geom_line(aes(y = sp500_52week_change / 10, colour = "S&P500")) +
  geom_line(aes(y = CCIw * 1.5, colour = "CCI")) +
  geom_hline(yintercept = 0, colour = 'black') +
  scale_color_manual("", values = c("WEI" = "green", "S&P500" = "blue", "CCI" = "red")) +
  ggtitle("WEI vs S&P500 52 week % change scaled by 10 vs CCI scaled by 1.5 during covid") +
  ylab("WEI, S&P500 and CCI") +
  xlab("Date")
plot_WEI_SP500_CCI_during_covid
```

Corrplot

```
data_2 <- data_1
colnames(data_2)[11:13] <- c("S&P500 %change" ,"S&P500 difference", "CCI")
z <- cbind(data_2[4:8], data_2[11:13])
Z <- cor(z)
corrplot(Z, method = "color")
```

(P)ACF plots and Dicky-Fuller tests

```
acf(data_1$WEI)
pacf(data_1$WEI)
```

```
# From the ACF plots we suspected non covariance stationarity, so we conducted some Dicky-Fuller tests.
# For the WEI
pmax    <- floor(12*((length(data_1$WEI)/100)^0.25))
dft     <- ur.df(data_1$WEI,type=c("drift"),lags=pmax,selectlags=c("BIC"))
summdf  <- summary(dft)
print(summdf@test.name)
print(c("Test statistic: ", summdf@teststat[1]))
print(c("Crit. vals", summdf@cval[1,]))

# For the S&P500
pmax    <- floor(12*((length(data_1$sp500_52week_change)/100)^0.25))
```

```
dft     <- ur.df(data_1$sp500_52week_change,type=c("drift"),lags=pmax,selectlags=c("BIC"))
summdf  <- summary(dft)
print(summdf@test.name)
print(c("Test statistic: ", summdf@teststat[1]))
print(c("Crit. vals", summdf@cval[1,]))

#For the CCI
pmax    <- floor(12*((length(data_1$CCIw)/100)^0.25))
dft     <- ur.df(data_1$CCIw,type=c("drift"),lags=pmax,selectlags=c("BIC"))
summdf  <- summary(dft)
print(summdf@test.name)
print(c("Test statistic: ", summdf@teststat[1]))
print(c("Crit. vals", summdf@cval[1,]))
```

BIC and AIC

```
# Information criteria for the WEI.
bic_WEI = matrix(NA,7,6)
aic_WEI = matrix(NA,7,6)
T_est = matrix(NA,7,6)
for (i in seq(2,8)){
  for (j in seq(0,5)){
    fit = Arima(WEI, order = c(i,0,j))
    T_est[i-1,j+1] = length(fit$residuals)
    bic_WEI[i-1,j+1] = fit$bic
    aic_WEI[i-1,j+1] = fit$aic
  }
}
T_est

colnames(bic_WEI) <- c("MA(0)","MA(1)","MA(2)","MA(3)","MA(4)",'MA(5)')
rownames(bic_WEI) <- c('AR(2)',"AR(3)","AR(4)","AR(5)","AR(6)","AR(7)",'AR(8)')
bic_WEI #table 1
min_values_bic= sort(bic_WEI)[1:3]
min_index_bic=c()
for (i in 1:3){
  min_index_bic[i] = which(bic_WEI==min_values_bic[i])
}
min_index_bic
```

```
# Information criteria for the VAR model.
colnames(aic_WEI) <- c("MA(0)","MA(1)","MA(2)","MA(3)","MA(4)",'MA(5)')
rownames(aic_WEI) <- c('AR(2)',"AR(3)","AR(4)","AR(5)","AR(6)","AR(7)",'AR(8)')
aic_WEI #table 2
min_values_aic= sort(aic_WEI)[1:3]
min_index_aic=c()
for (i in 1:3){
  min_index_aic[i] = which(aic_WEI==min_values_aic[i])
}
min_index_aic
```

```
# BIC graph
Y            <- cbind(CCIw_365, sp500_52week_change_365 , WEI_365)
colnames(Y) <- c('CCI', 'SP500', 'WEI')
VARmodel_ic <- VARselect(Y,type=c("const"),lag.max=8)
ic           <- as.data.frame(t(VARmodel_ic$criteria))
ic
ggplot(data=ic, aes(x=seq(1,8),y=`SC(n)`))+geom_line()+ylab("BIC")+xlab("VAR(p)")
ggplot(data=ic, aes(x=seq(1,8),y=`AIC(n)`))+geom_line()+ylab("AIC")+xlab("VAR(p)")
```

WEI residual analysis

```
#residual autocorrelation
fit_1 <- Arima(WEI, order = c(2,0,3)) #figure 4
checkresiduals(fit_1)
```

```
fit_2 <- Arima(WEI, order = c(3,0,0))
checkresiduals(fit_2)
```

```
fit_3 <- Arima(WEI, order = c(2,0,0))
checkresiduals(fit_3)

fit_4 <- Arima(WEI, order = c(5,0,4)) #figure 5
checkresiduals(fit_4)

fit_5 <- Arima(WEI, order = c(6,0,4))
checkresiduals(fit_5)

fit_6 <- Arima(WEI, order = c(7,0,5))
checkresiduals(fit_6)

fit_7 = Arima(WEI, order = c(52,0,3), fixed=c(NA,NA,0,0,0,0,0,
                                              0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
                                              0,0,0,0,0,0,0,0,0,0,0,0,0,NA,NA,NA,NA,NA,NA))
checkresiduals(fit_7)


autoplot(fit_1)
fit_1

autoplot(fit_2)
fit_2

autoplot(fit_3)
fit_3

autoplot(fit_4)
fit_4

autoplot(fit_5)
fit_5

autoplot(fit_6)
fit_6
```

```r
autoplot(fit_7)
fit_7
```

Forecasting

```r
## ARMA forecasts:
# ARMA(2,3)
fit_1 <- Arima(WEI_365, order = c(2,0,3))
fARMA_1 <- forecast(fit_1,h=208)
autoplot(fARMA_1)

# VAR(3)
Y <- cbind(WEI_365, CCIw_365 , sp500_52week_change_365 )
VAR3 <- VAR(Y,p=3,type = c('const'))
fVAR3 <- forecast(VAR3, h=208)
autoplot(fVAR3$forecast$WEI)
VAR3$varresult$WEI$coefficients
```

MSE/MAE

```r
#MSE of the ARMA models
es     <- as.Date("2008/1/5") # Estimation start
fs     <- as.Date("2016/1/2") # First forecast
fe     <- as.Date("2020/03/21")# Final forecast

convert_date <- function(date){
  c(as.numeric(format(date,'%Y')),
    ceiling(as.numeric(format(date,'%W'))))
  # Use %W for weeks and do not divide by 3.
}

dates   <- seq(fs,fe,by="week") # (or "week"...)
n       <- length(dates)                 # number of forecasts
qF      <- convert_date(fs)
qL      <- convert_date(fe)
target  <- window(WEI_365,start=qF,end=qL)

in_out_ARMA = function(hor, p, q){
  fc    <- ts(data=matrix(NA,n,1),start=qF,frequency=365.25/7)
  fce   <- ts(data=matrix(NA,n,1),start=qF,frequency=365.25/7)

  for (i_d in seq(1,n)){
    # Define estimation sample (ends h periods before 1st forecast)
    # Start at the first forecast date,
    # Then move back h+1 quarters back in time
    est   <- seq(dates[i_d],length=hor+1, by = "-1 week")[hor+1]
    # Now define the data we can use to estimate the model
    yest  <- window(WEI_365,end=convert_date(est))
    # Fit the AR models using Arima
    fit            <- Arima(yest,order=c(p,0,q))   #Fit model
    fc[i_d,1]      <- forecast(fit,h=hor)$mean[hor]#Get forecast
    fce[i_d,1]     <- fc[i_d,1]-target[i_d]        #Get forecast error
```

```
  }
  results         <- list()
  results$fc      <- fc
  results$fce     <- fce
  results$target  <- target
  return(results)
}

h_all     <- c(26,52,104)      # Which horizons to consider
lh        <- length(h_all)
mseARMA   <- matrix(NA,lh,3) # Full sample
p = c(2,3,5)
q = c(3,0,4)
parameters = as.data.frame(cbind(p,q))
for (p in 1:3){
  for (i in seq(1,lh)){
    fcARMA             <- in_out_ARMA(h_all[i],parameters[p,1],parameters[p,2])
    mseARMA[i,p]    <- colMeans(fcARMA$fce^2, na.rm = T)
  }
}
rownames(mseARMA)  <- c("26-step","52-step","104-step")
colnames(mseARMA)  <- c('ARMA(2,3)','ARMA(3,0)','ARMA(5,4)')
mseARMA
```

```
# Absolute error/MAE

h_all     <- c(26,52,104)      # Which horizons to consider
lh        <- length(h_all)
abeARMA   <- matrix(NA,lh,3)
p = c(2,3,5)
q = c(3,0,4)
parameters = as.data.frame(cbind(p,q))
for (p in 1:3){
  for (i in seq(1,lh)){
    fcARMA             <- in_out_ARMA(h_all[i],parameters[p,1],parameters[p,2])
    abeARMA[i,p]    <- colMeans(abs(fcARMA$fce), na.rm = T)
  }
}
rownames(abeARMA)  <- c("26-step","52-step","104-step")
colnames(abeARMA)  <- c('ARMA(2,3)','ARMA(3,0)','ARMA(5,4)')

abeARMA
```

IRF

```
Y           <- cbind(sp500_52week_change_365 , CCIw_365  ,  WEI_365)
colnames(Y) <- c('CCI','SP500', 'WEI' )
VARmodel    <- VAR(Y,p=3,type=c("const"))
roots(VARmodel) # computes eigenvalues of companion matrix


irf_WEI <- irf(VARmodel,impulse=c("SP500"),
               response=c("WEI"),ortho=T, n.ahead = 208)
```

```
plot(irf_WEI,plot.type=c("single"))

irf_WEI_CCI <- irf(VARmodel,impulse=c("CCI"),
                   response=c("WEI"),ortho=T, n.ahead = 208)
plot(irf_WEI_CCI,plot.type=c("single"))
```

Combined forecast

```
fit_1 <- Arima(WEI_365, order = c(2,0,3))
fARMA_1 <- forecast(fit_1,h=208)
autoplot(fARMA_1)

Y <- cbind(WEI_365, CCIw_365 , sp500_52week_change_365 )
VAR3 <- VAR(Y,p=3,type = c('const'))
fVAR3 <- forecast(VAR3, h=208)
autoplot(fVAR3$forecast$WEI)
VAR3$varresult$WEI$coefficients

fcombined = matrix(0,length(fARMA_1$mean),6)
for (i in 1:208){
  fcombined[i,2] = 0.5*as.numeric(fVAR3$forecast$WEI_365$mean[i])+0.5*as.numeric(fARMA_1$mean[i])
  fcombined[i,3] = 0.5*as.numeric(fVAR3$forecast$WEI_365$lower[i,1])+0.5*as.numeric(fARMA_1$lower[i,1])
  fcombined[i,4] = 0.5*as.numeric(fVAR3$forecast$WEI_365$lower[i,2])+0.5*as.numeric(fARMA_1$lower[i,2])
  fcombined[i,5] = 0.5*as.numeric(fVAR3$forecast$WEI_365$upper[i,1])+0.5*as.numeric(fARMA_1$upper[i,1])
  fcombined[i,6] = 0.5*as.numeric(fVAR3$forecast$WEI_365$upper[i,2])+0.5*as.numeric(fARMA_1$upper[i,2])
}

combinedForecast_1 = ts( c(as.vector(WEI_365),fcombined[,2]), decimal_date(ymd("2008-01-05")), frequency
combinedForecast_low1 = ts( c(as.vector(WEI_365),fcombined[,3]), decimal_date(ymd("2008-01-05")), freque
combinedForecast_low2 = ts( c(as.vector(WEI_365),fcombined[,4]), decimal_date(ymd("2008-01-05")), freque
combinedForecast_high1 = ts( c(as.vector(WEI_365),fcombined[,5]), decimal_date(ymd("2008-01-05")), frequ
combinedForecast_high2 = ts( c(as.vector(WEI_365),fcombined[,6]), decimal_date(ymd("2008-01-05")), frequ

ts.plot(combinedForecast_low1, combinedForecast_low2, combinedForecast_high1, combinedForecast_high2, co
        col= c('#4842f5','#00b5af','#4842f5', '#00b5af','#000000'), ylab = 'WEI', main = 'Combined Var(3
legend('bottomleft', legend = c('95% low', '80 low', '95% high' ,'80% high','forecast'), col =  c('#484
```

```
# The calculation of the SSR of the combined model.
fcombined2 = matrix(0,636,2)
for (i in 4:639){
  fcombined2[i-3,2] = 0.5*as.numeric(VAR3$varresult$WEI_365$fitted.values[i-3])+0.5*as.numeric(fit_1$fit
}
residuals_combined = c()
for(i  in 4:639){
  residuals_combined[i-3] = as.vector(WEI_365)[i] - fcombined2[i-3,2]
}
SSR_c = sum(residuals_combined^2)
SSR_VAR = sum(as.numeric(VAR3$varresult$WEI_365$residuals)^2)
SSR_ARMA = sum(as.numeric(fit_1$residuals)[4:639]^2)
SSR = matrix(c(SSR_c, SSR_VAR, SSR_ARMA),1,3)
rownames(SSR)  <- c("SSR")
colnames(SSR)  <- c('Combined','VAR(3)','ARMA(2,3)')
SSR
```

This is code we did not use in the paper, but gave different insights

```
# From these plots we suspected non covariance stationarity
acf(data_1$sp500_52week_change)
pacf(data_1$sp500_52week_change)

acf(data_1$CCIw)
pacf(data_1$CCIw)
```

Restricted Var

```
Y           <- cbind(CCIw_365  , sp500_52week_change_365 , WEI_365)
colnames(Y) <- c('CCI', 'SP500', 'WEI')
VARmodel_ic <- VARselect(Y,type=c("const"),lag.max=8)
ic          <- as.data.frame(t(VARmodel_ic$criteria))
ic
ggplot(data=ic, aes(x=seq(1,8),y=`SC(n)`))+geom_line()+ylab("BIC")+xlab("VAR(p)")
ggplot(data=ic, aes(x=seq(1,8),y=`AIC(n)`))+geom_line()+ylab("AIC")+xlab("VAR(p)")
```

```
#restricted VAR
p1        <- 6;
VARr     <- VAR( Y,p=p1,type=c("const"))
nseries  <- 3;
#mones     <- matrix(1,nrow = nseries,ncol=nseries)
#mzero     <- matrix(0,nrow = nseries,ncol=nseries)
vones    <- matrix(1,nrow = nseries,ncol=1)
lag1mat <- matrix(c(1, 1, 1,
                    1, 1, 1,
                    1, 1, 1)
                  ,nrow = nseries,ncol=nseries, byrow = TRUE) # lag matrix cols = cci, sp500 and WEI. R
lag2mat <- matrix(c(0, 0, 0,
                    0, 0, 0,
                    0, 0, 0)
                  ,nrow = nseries,ncol=nseries, byrow = TRUE)
lag3mat <- matrix(c(1, 1, 1,
                    1, 1, 1,
                    1, 1, 1)
                  ,nrow = nseries,ncol=nseries, byrow = TRUE)
lag4mat <- matrix(c(0, 0, 0,
                    0, 0, 0,
                    0, 0, 0)
                  ,nrow = nseries,ncol=nseries, byrow = TRUE)
lag5mat <- matrix(c(1, 1, 1,
                    1, 1, 1,
                    1, 1, 1)
                  ,nrow = nseries,ncol=nseries, byrow = TRUE)
lag6mat <- matrix(c(0, 0, 0,
                    0, 0, 0,
                    0, 0, 0)
                  ,nrow = nseries,ncol=nseries, byrow = TRUE)
lag7mat <- matrix(c(1, 1, 1,
                    1, 1, 1,
                    1, 1, 1)
```

```
                    ,nrow = nseries,ncol=nseries, byrow = TRUE)
lag8mat <- matrix(c(0, 0, 0,
                    0, 0, 0,
                    0, 0, 0)
                    ,nrow = nseries,ncol=nseries, byrow = TRUE)
lag9mat <- matrix(c(1, 1, 1,
                    1, 1, 1,
                    1, 1, 1)
                    ,nrow = nseries,ncol=nseries, byrow = TRUE)
restrict <- matrix(cbind(lag1mat, lag2mat, lag3mat, lag4mat, lag5mat, lag6mat,  vones), nrow = 3, ncol =


VARr     <- restrict(VARr, method = "man", resmat = restrict)
```

```
# Somehow BIC has to be calculated by hand
resid    <- residuals(VARr)
T        <- length(resid[,1])
BIC      <- log(det(t(resid)%*%resid/T)) + (log(T)/T)*sum(restrict)
BIC

fVARr <- forecast(VARr, h=200)
autoplot(fVARr$forecast$WEI)
VARr$varresult$WEI$coefficients
# You can check that now the third lag is omitted by typing
summary(VARr)
roots(VARr)

irf_WEI <- irf(VARr,impulse=c("SP500"),
               response=c("WEI"),ortho=T, n.ahead = 300)
plot(irf_WEI,plot.type=c("single"))

irf_CCI <- irf(VARr,impulse=c("SP500"),
               response=c("CCI"),ortho=T, n.ahead = 300)
plot(irf_CCI,plot.type=c("single"))

irf_WEI_CCI <- irf(VARr,impulse=c("CCI"),
                   response=c("WEI"),ortho=T, n.ahead = 300)
plot(irf_WEI_CCI,plot.type=c("single"))
```

rejected ARMA

```
fit_2 <- Arima(WEI, order = c(5,0,4)) #figure 5
fARMA_2 <- forecast(fit_2,h=208)
autoplot(fARMA_2)

fit_3 = Arima(WEI, order = c(52,0,3), fixed=c(NA,NA,0,0,0,0,0,
                                              0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
                                              0,0,0,0,0,0,0,0,0,0,0,0,0,NA,NA,NA,NA,NA,NA))
fARMA_3 <- forecast(fit_3,h=208)
autoplot(fARMA_3)
```

check ARDL and VAR

```r
#ARDL model
ARDL4 <- dynlm(WEI_365 ~L(WEI_365,(1:4)) +L(sp500_52week_change_365 ,(1:4)) + L(CCIw_365,(1:4)))
summ<-summary(ARDL4)
print(summ$coefficients,digits=1)


Y <- cbind(WEI_365, CCIw_365, sp500_52week_change_365 )
VAR3 <- VAR(Y,p=3,type = c('const'))
corder1  <- order(names(VAR3$varresult$WEI$coefficients))
corder2  <- order(names(summ$coefficients[,1]))
coefVAR  <- cbind(VAR3$varresult$WEI$coefficients[corder1],
                  summ$coefficients[corder2])
colnames(coefVAR)<- c("VAR(4)","ARDL(4,4,4)")

print(coefVAR,digits=3)
```

uin sample ARDL test

```r
forecastARDL <- function(y,X,es,fs,fe,maxARp,hor){
  dates   <- seq(fs,fe,by="week") # (or "week"...)
  n       <- length(dates)                 # number of forecasts
  qF      <- convert_date(fs)
  qL      <- convert_date(fe)
  target  <- window(y,start=qF,end=qL)     # What we are forecasting.

  # Define ts objects where forecasts/forecast errors are saved.
  # (Note that frequency=4 applies to quarterly data!)
  fc    <- ts(data=matrix(NA,n,maxARp),start=qF,frequency=365.25/7)
  fce   <- ts(data=matrix(NA,n,maxARp),start=qF,frequency=365.25/7)

  for (i_d in seq(1,n)){
    # Define estimation sample (ends h periods before 1st forecast)
    # Start at the first forecast date,
    # Then move back h+1 quarters back in time
    est   <- seq(dates[i_d],length=hor+1, by = "-1 week")[hor+1]
    # Now define the data we can use to estimate the model
    Y = cbind(y,X)
    Yest  <- window(Y,end=convert_date(est))
    # Fit the AR models using Arima
    for (j in seq(1,maxARp)){
      fit           <- VAR(Yest,p=j,type=c('const'))   #Fit model
      fc[i_d,j]     <- forecast(fit,h=hor)$forecast$y$mean[hor]#Get forecast
      fce[i_d,j]    <- fc[i_d,j]-target[i_d]        #Get forecast error
    }
  }
  results         <- list()
  results$fc      <- fc
  results$fce     <- fce
  results$target  <- target
  return(results)
}
```

```r
# Get forecasts
X_SP <- cbind(WEI_365,sp500_52week_change_365 )
X_CCI <- cbind(WEI_365,CCIw_365)
fcARDLh1_SP  <- forecastARDL(WEI_365,X_SP,es,fs,fe,maxARp,1)
fcARDLh1_CCI  <- forecastARDL(WEI_365,X_CCI,es,fs,fe,maxARp,1)

# Calculate MSE and compare
mseARDL_SP     <- colMeans(fcARDLh1_SP$fce^2)
mseARDL_CCI     <- colMeans(fcARDLh1_CCI$fce^2)
compare_SP     <- rbind(mseARMA[1,],mseARDL_SP)
compare_CCI     <- rbind(mseARMA[1,],mseARDL_CCI)
rownames(compare_SP) <- c("AR","ARDL")
colnames(compare_SP) <- c("p=1","p=2","p=3","p=4",'p=5','p=6')
rownames(compare_CCI) <- c("AR","ARDL")
colnames(compare_CCI) <- c("p=1","p=2","p=3","p=4",'p=5','p=6')
round(compare_SP,digits=3)
round(compare_CCI,digits=3)
```

leading indicator analysis

```r
plot240 <- ggplot(data = data_1) +
  geom_line(mapping = aes(x = Date, y = sp_500_52week_diff / 100), colour = "red") +
  geom_line(mapping = aes(x = Date, y = WEI), colour = "blue") +
  geom_line(aes(x = Date, y = 0), colour = "black")
plot240

plot239 <- ggplot(data = data_1) +
  geom_line(mapping = aes(x = Date, y = sp500_52week_change / 10), colour = "red") +
  geom_line(mapping = aes(x = Date, y = WEI), colour = "blue") +
  geom_hline(yintercept = 0, colour = 'black') +
  ggtitle("WEI vs S&P500 52 week percentage change") +
  ylab("S&P500 scaled by 10")
plot239
```

```r
correlation <- cor(select(data_1, 4:8, 11:12))
print(correlation)
corrplot(correlation, method = "color", na.remove = TRUE)
```

SP500 zorgen dat je een 52 weken percentage change difference neemt, die plotten tegenover WEI zonder veranderingen.

Oil price, 52 weken percentage change dan verschillen over 52 weken (misschien doet deze stap te weinig voor verlies aan data, dit zelf bekijken)

(vanwege inflatie)

midas modellen

```
SP500change <- PercChange(data = data, Var = "S&P500", NewVar = "SP500change")
SP500change <- SP500change$SP500change
data$SP500change <- SP500change

sp500_perc_change<- PercChange(data = data, Var = "average_open_close", NewVar = "sp500_perc_change")
sp500_perc_change <- sp500_perc_change$sp500_perc_change
data$sp500_perc_change <- sp500_perc_change

data$lnSP500 <- log(data$`S&P500`)
data$lnBB <- log(data$BB)
data$lnM1 <- log(data$M1)
data$lnOil <- log(data$Oil)

WEI_time_series_change <- diff(data$WEI)
WEI_time_series_change <- append(WEI_time_series_change, 0, after = 0)
data <- data %>% cbind(WEI_time_series_change)
```

Plotting several variables against the WEI to identify some correlation

```
plot1 <- ggplot(data = data, mapping = aes(x = BB, y = WEI)) +
  geom_point()
plot1

plot2 <- ggplot(data = data, mapping = aes(x = T10Y3M, y = WEI)) +
  geom_point()
plot2

plot3 <- ggplot(data = data, mapping = aes(x = M1, y = WEI)) +
  geom_point()
plot3

plot4 <- ggplot(data = data) +
  geom_line(mapping = aes(x = Date, y = WEI), color = "blue") +
  geom_line(mapping = aes(x = Date, y = T10Y3M, color = "red"))
plot4
```

```
plot5 <- ggplot(data = data) +
  geom_line(mapping = aes(x = Date, y = WEI), color = "blue", label = "WEI") +
  geom_line(mapping = aes(x = Date, y = sp500_perc_change, color = "red", label = "Change in BB")) +
  geom_line(mapping = aes(x = Date, y = T10Y3M, color = "black", label = "Bond rates")) +
  geom_line(aes(x = Date, y = 0))
plot5

plot6 <- ggplot(data = data) +
  geom_line(aes(x = Date, y = lnOil, color = "red")) +
  geom_line(aes(x = Date, y = lnSP500, color = "blue")) +
  geom_line(aes(x = Date, y = lnBB, color = "black")) +
  geom_line(aes(x = Date, y = lnM1, color = "white"))
plot6

plot7 <- ggplot(data = data, aes(x = Date)) +
  geom_line(aes(y = M1change, color = "Money supply")) +
  geom_line(aes(y =WEIchange, colour = "WEI"))+
  geom_hline(yintercept = 0, color = 'black') + scale_colour_manual("",
                                                                  values = c("Money supply"="blue", "W
  ggtitle("WEI vs Money supply percentage change") +
  ylab("WEI and Money supply percentage change")
plot7

plot8 <- ggplot(data = data, aes(x = Date)) +
  geom_line(aes(y =BBchange, colour = "Bank borrowings")) +
  geom_line(aes(y =WEI, colour = "WEI"))+
  geom_hline(yintercept = 0, color = 'black') + scale_colour_manual("",
                                                                  values = c("Bank borrowings"="blue"
  ggtitle("WEI vs Bank borrowings percentage change") +
  ylab("WEI and the BB percentage percentage change")
plot8

plot9 <- ggplot(data = data) +
  geom_line(aes(x = Date, y = BBchange, color = "darkred")) +
  geom_line(aes(x = Date, y = M1change, color = "lightblue"))
plot9

plot10 <- ggplot(data = data) +
  geom_line(aes(x = Date, y = WEI, color = "Red")) +
  geom_line(aes(x = Date, y = SP500change, color = "Blue"))
plot10

plot11 <- ggplot(data = data) +
  geom_line(aes(x = Date, y = Oil*10, color = "red")) +
  geom_line(aes(x = Date, y = WEI*100, color = "blue"))
plot11

plot12 <- ggplot(data = data) +
  geom_line(aes(x = Date, y = WEI*1000, color = "red")) +
  geom_line(aes(x = Date, y = `S&P500`*10, color = "blue"))
plot12

plot13 <- ggplot(data = data) +
```

```r
  geom_line(mapping = aes(x = Date, y = BBchange, color = "red"))
plot13

plot14 <- ggplot(data = sp500data) +
  geom_line(aes(x = Date, y = average_high_low, group = 1, color = "darkred")) +
  geom_line(aes(x = Date, y = average_open_close, group = 1, color = "lightblue"))
plot14

plot15 <- ggplot(data = data) +
  geom_line(aes(x = Date, y = SP500change, color = "darkred")) +
  geom_line(aes(x = Date, y = sp500_perc_change, color = "lightblue"))
plot15


plot16 <- ggplot(data = data, aes(x= Date)) +
  geom_line(aes(y = sp500_perc_change, colour = "S&P500")) +
  geom_line(aes(y = WEI, colour = "WEI")) +
  geom_hline(yintercept = 0, colour = 'black') + scale_colour_manual("", values = c("S&P500"="blue", "WE
  ggtitle("The WEI vs S&P500 percentage changes") +
  ylab("WEI and S&P500 percentage changes")
plot16

data[(which.min(data$WEIchange) - 1):(which.min(data$WEIchange) + 1), ]
plot(x = data$Date, y = data$WEI, type = "l")
data(length(data$WEIchange - 10))

par(mfcol = c(2,2))
plot(data$BB ,data$WEI)
plot(data$T10Y3M, data$WEI)
plot(data$M1, data$WEI)



plot18 <- ggplot(data = data) +
  geom_line(aes(x = Date, y = WEI_time_series_change, color = "darkred")) +
  geom_line(aes(x = Date, y = BBchange / 10, color = "lightblue")) +
  geom_line(aes(x = Date, y = 0 , color = "green")) +
  ggtitle("Change in WEI vs BB growth rate")
plot18

plot19 <- ggplot(data = data) +
  geom_line(aes(x = Date, y = WEI_time_series_change, color = "darkred")) +
  geom_line(aes(x = Date, y = M1change / 2, color = "lightblue")) +
  geom_line(aes(x = Date, y = 0 , color = "green")) +
  ggtitle("Change in WEI vs M1 growth rate")
plot19

plot20 <- ggplot(data = data[560:630, ]) +
  geom_line(aes(x = Date, y = WEI_time_series_change, color = "darkred")) +
  geom_line(aes(x = Date, y = diff_oil_price / 10, color = "lightblue")) +
  geom_line(aes(x = Date, y = 0 , color = "green")) +
  ggtitle("diff in WEI vs diff in Oil price")
plot20
```

```r
plot17 <- ggplot(data = data[560:628, ]) +
  geom_line(aes(x = Date, y = WEI_time_series_change, color = "darkred")) +
  geom_line(aes(x = Date, y = SP500_diff_change / 50, color = "lightblue")) +
  geom_line(aes(x = Date, y = 0 , color = "green")) +
  ggtitle("Change in WEI vs S&P500 growth rate")
plot17

plottest <- ggplot(data = data[560:628, ]) +
  geom_line(aes(x = Date, y = WEI_time_series_change, color = "darkred")) +
  geom_line(aes(x = Date, y = sp500_perc_change / 5, color = "lightblue")) +
  geom_line(aes(x = Date, y = 0 , color = "green")) +
  ggtitle("Change in WEI vs S&P500 growth rate")
plottest

plothoertje <- ggplot(data = data[500:600, ]) +
  geom_line(aes(x = Date, y = sp500_perc_change, color = "darkred")) +
  geom_line(aes(x = Date, y = SP500_diff_change / 50, color = 'lightblue'))
plothoertje

plot17 <- ggplot(data = data, aes(x = Date)) +
  geom_line(aes(y =WEI_difference100, colour = "WEI difference scaled by 100")) +
  geom_line(aes(y =SP500_time_series_change, colour = "S&P500 difference")) +
  geom_hline(yintercept = 0, colour = 'black') + scale_colour_manual("", values = c("WEI difference scal
                                                                                    "S&P500 difference"=
  ggtitle("Difference within WEI vs S&P500 2008-2020") + ylab("Difference with WEI scaled by 100")
plot21

plot_WEI_SP500 <- ggplot(data = data_1, aes(x = Date)) +
  geom_line(aes(y = WEI, colour = "WEI")) +
  geom_line(aes(y = sp500_52week_change / 10, colour = "S&P500 52 week %change scaled by 10")) +
  geom_hline(yintercept = 0, colour = 'black') +
  scale_color_manual("", values = c("WEI" = "green", "S&P500 52 week %change scaled by 10" = "blue")) +
  ggtitle("WEI vs 52 week % change of S&P500") +
  ylab("WEI and 52 week % change S&P500") +
  xlab("Date")
plot_WEI_SP500

plot_WEI_SP500_2008_2010 <- ggplot(data = data_1[1:105, ], aes(x = Date)) +
  geom_line(aes(y = WEI, colour = "WEI")) +
  geom_line(aes(y = sp500_52week_change / 10, colour = "S&P500 52 week %change scaled by 10")) +
  geom_hline(yintercept = 0, colour = 'black') +
  scale_color_manual("", values = c("WEI" = "green", "S&P500 52 week %change scaled by 10" = "blue")) +
  ggtitle("WEI vs 52 week % change of S&P500 in 2008-2010") +
  ylab("WEI and 52 week % change S&P500") +
  xlab("Date")
plot_WEI_SP500_2008_2010

plot_WEI_SP500_before_covid <- ggplot(data = data_1[560:630, ], aes(x = Date)) +
  geom_line(aes(y = WEI, colour = "WEI")) +
  geom_line(aes(y = sp500_52week_change / 10, colour = "S&P500 52 week %change")) +
  geom_hline(yintercept = 0, colour = 'black') +
  scale_color_manual("", values = c("WEI" = "green", "S&P500 52 week %change" = "blue")) +
  ggtitle("WEI vs 52 week % change of S&P500 before COVID-19") +
```

```
  ylab("WEI and 52 week % change S&P500") +
  xlab("Date")
plot_WEI_SP500_before_covid

plot_WEI_SP500_during_covid <- ggplot(data = data_1[630:639, ], aes(x = Date)) +
  geom_line(aes(y = WEI, colour = "WEI")) +
  geom_line(aes(y = sp500_52week_change / 10, colour = "S&P500 52 week %change")) +
  geom_hline(yintercept = 0, colour = 'black') +
  scale_color_manual("", values = c("WEI" = "green", "S&P500 52 week %change" = "blue")) +
  ggtitle("WEI vs 52 week % change of S&P500 before COVID-19") +
  ylab("WEI and 52 week % change S&P500") +
  xlab("Date")
plot_WEI_SP500_during_covid

plot_WEI_SP500_CCI <- ggplot(data = data_1[1:105, ], aes(x = Date)) +
  geom_line(aes(y = WEI, colour = "WEI")) +
  geom_line(aes(y = sp500_52week_change / 10, colour = "S&P500")) +
  geom_line(aes(y = CCIw * 1.5, colour = "CCI")) +
  geom_hline(yintercept = 0, colour = 'black') +
  scale_color_manual("", values = c("WEI" = "green", "S&P500" = "blue", "CCI" = "red")) +
  ggtitle("WEI vs S&P500 52 week % change scaled by 10 vs CCI scaled by 1.5") +
  ylab("WEI, S&P500 and CCI") +
  xlab("Date")
plot_WEI_SP500_CCI

plot17_2008_2010 <- ggplot(data = data[1:63,], aes(x = Date)) +
  geom_line(aes(y =WEI_difference100, colour = "WEI difference scaled by 100")) +
  geom_line(aes(y =SP500_time_series_change, colour = "S&P500 difference")) +
  geom_hline(yintercept = 0, colour = 'black') + scale_colour_manual("", values = c("WEI difference sca
                                                                                    "S&P500 difference"
  ggtitle("Difference within WEI vs S&P500 2008-2010") + ylab("Difference with WEI scaled by 100")
plot17_2008_2010

plot17_before_covid <- ggplot(data = data[560:630,], aes(x = Date)) +
  geom_line(aes(y =WEI_difference100, colour = "WEI difference scaled by 100")) +
  geom_line(aes(y =SP500_time_series_change, colour = "S&P500 difference")) +
  geom_hline(yintercept = 0, colour = 'black') + scale_colour_manual("", values = c("WEI difference sca
                                                                                    "S&P500 difference"
  ggtitle("Difference within WEI vs S&P500 before COVID-19") + ylab("Difference with WEI scaled by 100"
plot17_before_covid

plot17_during_covid <- ggplot(data = data[630:639,], aes(x = Date)) +
  geom_line(aes(y =WEI_difference100, colour = "WEI difference scaled by 100")) +
  geom_line(aes(y =SP500_time_series_change, colour = "S&P500 difference")) +
  geom_hline(yintercept = 0, colour = 'black') + scale_colour_manual("", values = c("WEI difference sca
                                                                                    "S&P500 difference"
  ggtitle("Difference within WEI vs S&P500 during COVID-2019") + ylab("Difference with WEI scaled by 10
plot17_during_covid
```

Gekloot met time series

```r
BB_time_series = ts(data[, 9], start = c(2008), frequency = 365.25/7)
BB_ts_change = diff(BB_time_series)
data$BB_t
y <- data.frame(rep(c(0), times = 639))

WEI_time_series = ts(data[,4], start= c(2008), frequency = 365.25/7)
WEI_ts_change = diff(WEI_time_series)
time_series_data <- data.frame(WEI_ts_change)
data$WEI_time_series <- WEI_time_series

WEI_time_series_change <- diff(data$WEI)
WEI_time_series_change <- append(WEI_time_series_change, 0, after = 0)
data <- data %>% cbind(WEI_time_series_change)

ggplot(data = data) +
  geom_line(aes(x = Date, y = diff(WEI)))

autoplot(diff(BB))
p1 <- autoplot(diff(WEI))
plot(x = data$Date, y = data$WEI)
plot(x = data$Date, y = diff(data$WEI))

p1$layers +
  geom_line(data = data,
            mapping = aes(x = Date, y = sp500_perc_change),
            inherit.aes = F)

een_variabele_naam <- cbind(diff(WEI), diff(BB_time_series / 100000))
plot.ts(een_variabele_naam, plot.type = "single", col = c("blue", "red"))
```

Correlation matrix and plots

```r
x <- c(cor(data$WEI, data$T10Y3M), cor(data$WEI, data$BB), cor(data$WEI, data$M1))
print(x)
cor(data[4:10])
corrplot(data[4:10], type = "upper", order = "hclust", tl.col = "black", tl.srt = 45)
cor(data[4:11])
cor(data[-1])
cor.test(x = data$WEI, y = data$`S&P500`, method=c("pearson", "kendall", "spearman"))

data_2 <- data_1
colnames(data_2)[12:13] <- c("S&P500 52 week difference", "CCI")
cor_all <- cor(select(data_2, 4:8, 11:13))
corrplot(cor_all, method = "color", na.rm = T)
x <- as.data.frame(cor(data_2[11:13]))
y <- as.data.frame(cor(data_2[4:8]))
z <- as.data.frame(rbind(x, y))
z <- cbind(data_2[4:8], data_2[11:13])
```

```
Z <- cor(z)
corrplot(Z, method = "color")
```

Regressing several variables to identify statsitical significance

```
model1 <- lm(WEI ~ T10Y3M + BBchange, data = data)
summary(model1)
anova(model1)
plot(model1)
summary(data)

model2 <- lm(WEI ~ T10Y3M, data = data)
summary(model2)

model3 <- lm(WEI ~ `S&P500` + Oil + FFR, data = data)
summary(model3)
plot(model3)

model4 <- lm(WEI ~ lnSP500 + Oil + FFR, data = data)
summary(model4)
plot(model4)

model5 <- lm(WEI ~ lnSP500 + lnBB + lnM1 + lnOil, data = data)
summary(model5)


model6 <- lm(WEI ~ lnOil + FFR + T10Y3M, data = data)
summary(model5)
```

Autocorrelation models (acf and pacf)

```
acf(data$sp500_perc_change, na.action = na.pass)
pacf(data$sp500_perc_change, na.action = na.pass)
```