Advanced Time Series Analysis

Final paper

Daniel IZQUIERDO JUNCAS

Prof. Dr. Cristophe Croux

Academic year 2017-2018

Contents

1 Introduction

This paper consists on a time series analysis of the Trade of Goods in Spain in the period 1990-2017. This dataset is part of the International Financial Statistics dataset of the International Monetary Fund e-library. In the first part, a univariate analysis is conducted on the value of exports in national currency. In the second part, a multivariate analysis is done to relate the value of exports with the value of imports in national currency.

The original dataset also includes the period 1960-1990; however, after taking differences and seasonal differences this period shows a higher variance than the period 1990-2017. The increase from 1990 has been steeper than the one in the period 1960-1990 (Figure A.1). The change is most likely caused by the accession of Spain in the European Economic Community (EEC) in January of 1986. We consider this point as a structural break; so we limit the analysis only to the period 1990-2017.

2 Univariate analysis

2.1 Stationarity

The time series shows a positive trend, increasing over time (Figure 1). Around 2008 there is a noticeable drop in the value of exports, which corresponds with the world financial crisis. This event did not seem to cause a lasting effect so it is not considered as a structural break. The series looks far from stationary. Using the augmented Dickey-Fuller (ADF) test for trend it is checked if the log-transformed series presents a deterministic or a stochastic trend. The maximum lag for the test is given by the rule of thumb $\sqrt{SampleSize} \approx 10$. The test gives a p-value of 0.0671, so although it is a borderline result, at a 0.05 significance level we do not reject the null hypothesis and, thus, the trend is stochastic.

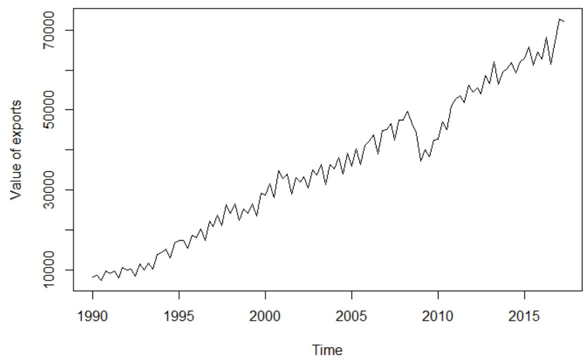


Figure 1: Value of exports

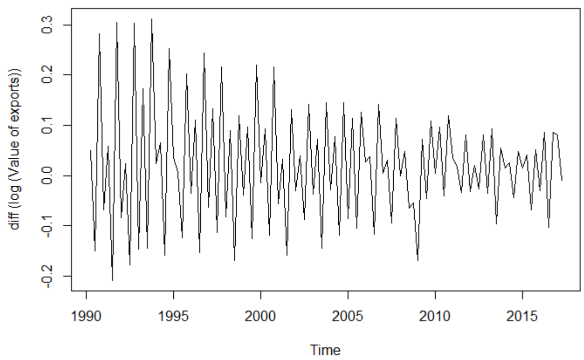


Figure 2: Value of diff(log(exports))

For the previous reason and as we are more interested in the percentage of increase of the value of exports rather than in its absolute value, we apply the log-difference operator to the series (Figure 2). There seems to be pronounced movements from quarter to quarter, which is confirmed with the monthplot in Figure 3 (see Figure A.2 for each year profile). The average value of the exports in Q4 is higher than in the other quarters, thus, there is a seasonal effect. To correct for this, the seasonal difference operator is applied (Figure 4).

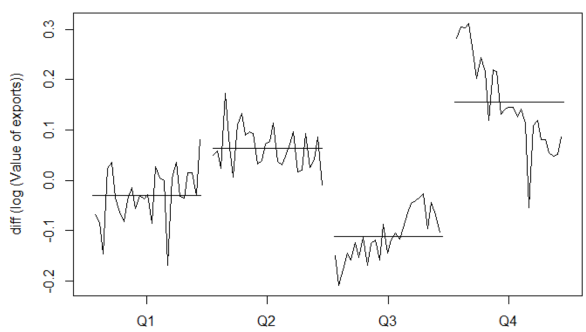


Figure 3: Monthplot

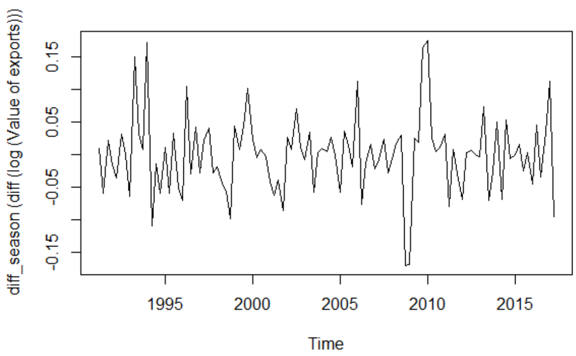
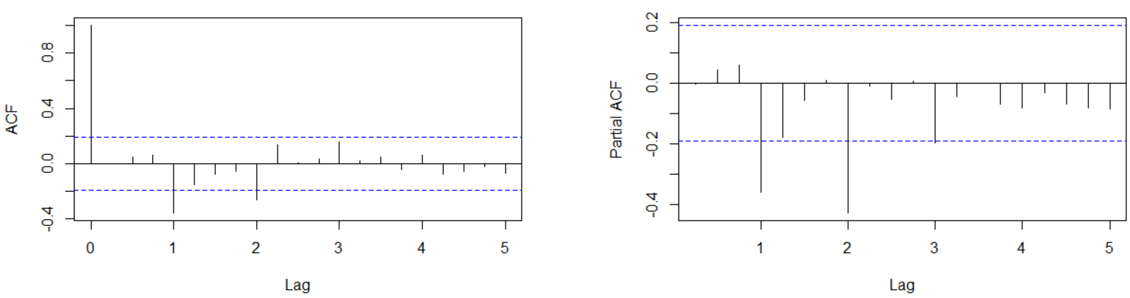


Figure 4: Value of seasonal differenced log(exports)

With the ADF test without trend we formally check if the resulting time series is stationary. The H_0 of the existence of a unit root (hence, no stationary time series) is rejected with a p-value of 7.713e-08 < 0.05, so we conclude that the series is stationary. There does not seem to exist a persistence in volatility.

2.2 Modeling

There are still some significant correlations left at different lags so the series can not be considered white noise. The correlogram (Figure 5) indicates that there are significant correlations at seasonal lag 1 and 2, indicating that the use of the first two seasonal moving average (MA) terms is appropriate. The partial correlogram points to significant correlations at seasonal lags 1, 2 and 3 (barely significant), so the first three seasonal autoregressive (AR) terms are considered. Non-seasonal effects are not observed in neither the correlogram or the partial correlogram.



(a) (b)

Figure 5: (a) Correlogram; (b) Partial correlogram

As a starting point we fit a SARIMA(0,1,0)(0,1,2) and a SARIMA(0,1,0)(3,1,0). For the former, the second MA term is not significant so we fit a SARIMA(0,1,0)(0,1,1) with only one parameter. For the latter, the third AR term is barely significant so we also fit a

SARIMA(0,1,0)(2,1,0). In Table 1 these models and others that combine AR and MA terms are presented with some summary measures.

Model	#Parameters	BIC	MAE	Q-test (p-value)	Validated
SARIMA(0,1,0)(0,1,2)	2	-317.84	0.041296	0.41	Yes
SARIMA(0,1,0)(0,1,1)	1	-322.48	0.049807	0.42	Yes
SARIMA(0,1,0)(3,1,0)	3	-319.91	0.035516	0.88	Yes
SARIMA(0,1,0)(2,1,0)	2	-320.52	0.051063	0.82	Yes
SARIMA(0,1,0)(1,1,1)	2	-317.83	0.036826	0.42	Yes
SARIMA(0,1,0)(2,1,1)	3	-320.85	0.036639	0.90	Yes

Table 1: Model comparison

To validate the models we have to test if the residuals are white noise checking the residual correlograms and using the Ljung-Box Q-statistic. The test H_0 states that the autocorrelations up to lag $k \approx 10$ of the residuals are equal to zero. This null hypothesis is not rejected for any of the models, so their residuals can be considered as white noise.

For the two best models according to the BIC, the estimated coefficients with the corresponding standard errors are displayed in Figure 6. All coefficients are significant at a 0.05 level except for the first seasonal AR term of the SARIMA(0,1,0)(2,1,1) model.

```
Coefficients:
         sma1
      -0.6606
s.e.   0.0627
```

(a)

```
Coefficients:
         sar1     sar2     sma1
      -0.2200  -0.3380  -0.4363
s.e.   0.1466   0.1106   0.1424
```

(b)

Figure 6: (a) SARIMA(0,1,0)(0,1,1) coefficients; (b) SARIMA(0,1,0)(2,1,1) coefficients

2.3 Prediction

Forecast is provided for the SARIMA(0,1,0)(0,1,1) model and also for the SARIMA(0,1,0)(2,1,1) model. The latter presents the second best BIC and as it has 2 more parameters than the former, a better prediction performance can be expected.

To evaluate the model forecasting performance we use out-of-sample criteria, splitting the dataset in an estimation and a validation subsets with a 75-25 proportion respectively. With the h-step ahead forecast errors we compute the root mean squared error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE) for the two mentioned models using h=1 (Table 2). The best prediction performance corresponds to the SARIMA(0,1,0)(2,1,1) model.

Model	MAE	RMSE	MAPE
SARIMA(0,1,0)(0,1,1)	0.049807	0.059527	0.004534
SARIMA(0,1,0)(2,1,1)	0.036639	0.046473	0.003333

Table 2: Prediction measures for the two selected models

Using the Debold-Mariano test it is found that the forecast performance of this model is significantly better than for the other when we take $L(e) = |e|$ as Loss function (p-value = 0.0467). On the other hand, if $L(e) = e^2$ is taken the Debold-Mariano test does not detect a significant difference between the forecasting performance of both models (p-value = 0.0903). In Figure 7, the 8-quarter ahead forecast using the SARIMA(0,1,0)(2,1,1) model is presented. The prediction interval widens over time because the further away we forecast, the more uncertainty there is.

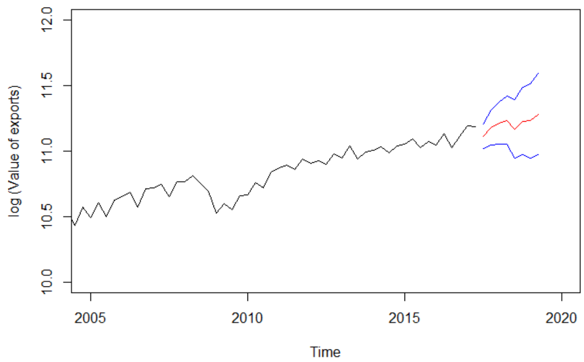


Figure 7: Forecast for the next 8 quarters using SARIMA(0,1,0)(2,1,1) model

3 Multivariate analysis

For the multivariate analysis the value of exports in national currency is related with the value of imports in national currency. The difference between the two defines the country's balance of trade, which influences the economic growth. Normally, one would expect the two to move together, so as to not have too large trade deficit/surplus against the other countries in the long run. Both series follow a similar pattern, although the value of imports is always larger than the value of exports (Figure 8).

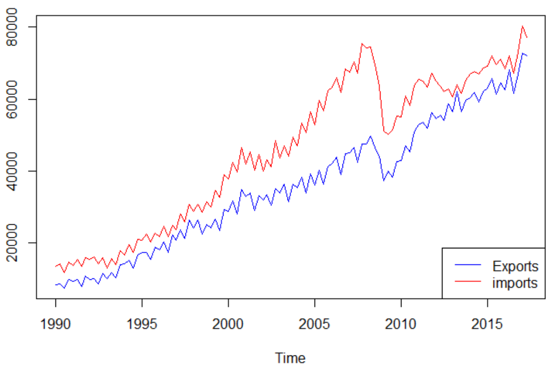


Figure 8: Exports and imports time series

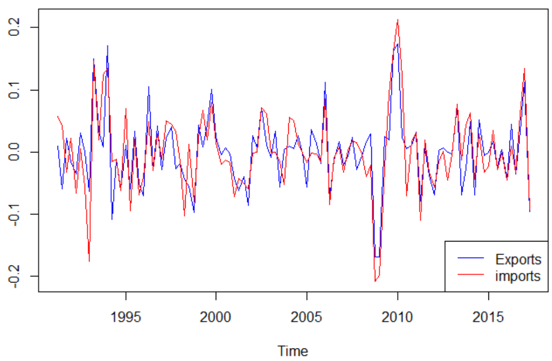
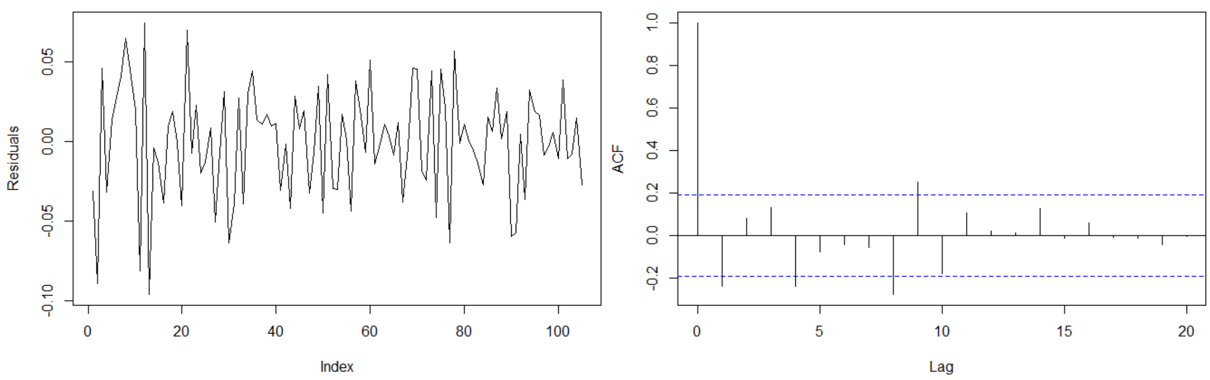


Figure 9: Exports and imports time series in log-differences and seasonal differences

For the reasons exposed in the univariate analysis, in order to make both time series stationary and to eliminate seasonal effects it is necessary to take log-differences and seasonal differences (Figure 9). We check whether log-imports and log-exports are integrated of order 1 (I(1)), that is when the series in differences is stationary although the series in levels is not. With the ADF test we obtain a p-value of 0.007544 for exports and a p-value of 3.461e-05 for imports so both series in differences are stationary at a 0.05 significance level. Note that to reach stationarity it is not necessary to go on seasonal differences.

3.1 Linear regression

As after taking log-differences and seasonal differences both time series are stationary we can fit an ordinary linear regression model. We regress $\Delta\Delta_4$log-exports on $\Delta\Delta_4$log-imports. The F-test indicates that there is a regression relation (p-value < 2e-16) with R^2 = 0.6671.



(a) (b)

Figure 10: Linear regression. (a) Residual plot; (b) Residual correlogram

To validate the model we need to check if the residuals are white noise (Figure 10). The correlogram shows that the correlations at lags 1, 4, 8 and 9 are significant. The Q-test leads to the rejection of the null hypothesis with a p-value of 6.877e-05, so we conclude that the residuals are not white noise and the regression model is not valid.

3.2 Autoregressive distributed lag model

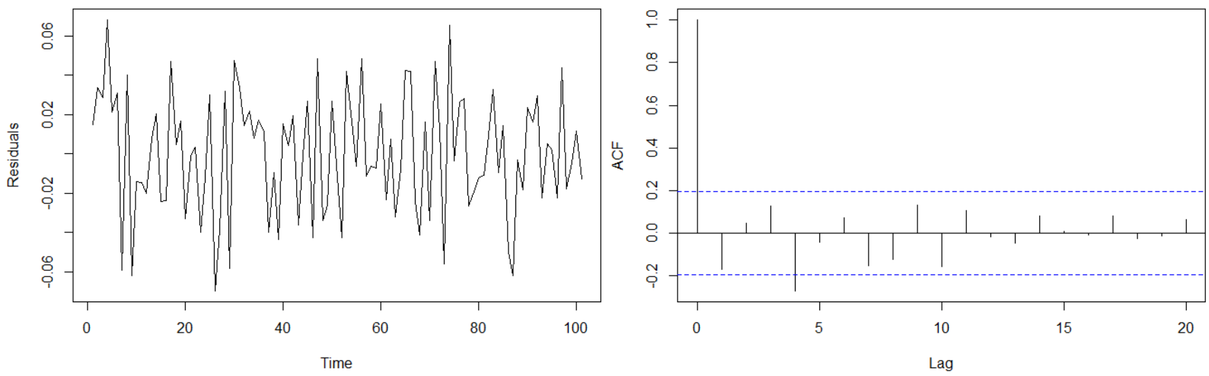
3.2.1 Distributed lag model

Next, we regress $\Delta\Delta_4$log-exports on $\Delta\Delta_4$log-imports not only including the current value of imports in the model but also earlier values of this variable. As in the univariate analysis it was found that the year effects were important, a distributed lag (DL) model of order 4 is applied. That is a model that includes as covariates the value of imports and also the last 4 values of it:

$$Y_t = c + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \beta_4 X_{t-3} + \beta_5 X_{t-4} + \varepsilon_t \tag{1}$$

where Y_t denotes the $\Delta\Delta_4$log-exports time series and X_t the $\Delta\Delta_4$log-imports.

We check if the residuals are white noise in order to validate the model (Figure 11). Only the correlation at lag 4 (after one year) is significant. The Q-test (p-value = 0.0126) leads to the rejection of the null hypothesis at a 0.05 level but not at a 0.01 level.



(a) (b)

Figure 11: DL(4). (a) Residual plot; (b) Residual correlogram

3.2.2 Autoregressive distributed lag model

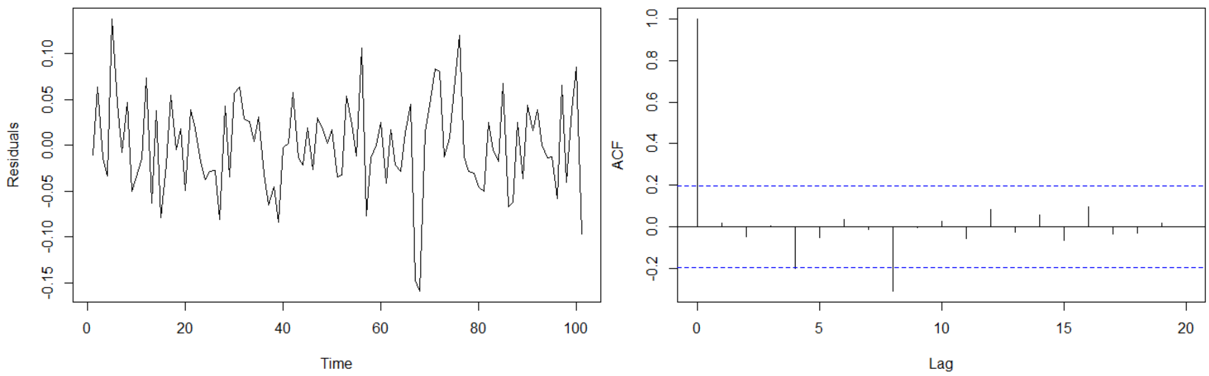
As the DL(4) model is not completely validated we also include earlier values of exports as covariates. We fit an autoregressive distributed lag (ADL) model where we include up to the forth lagged value of imports and exports. In addition, we drop X_t from the model to make it suitable for prediction.

$$\begin{aligned} Y_t =& \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} + \alpha_4 Y_{t-4} \\ &+ \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + \varepsilon_t \end{aligned} \tag{2}$$

The residual correlogram of this model shows a significant correlation at lag 8 (Figure 12). However, the p-value of the Q-test is 0.1051 so we do not reject the null hypothesis that the residuals are white noise and, thus, the model is validated. The overall F-statistic has p-value = 0.0002 < 0.05, so there is a significant regression relation. R^2 = 0.2715, so 27 % of the variance of $\Delta\Delta_4$log-exports is explained by the covariates in the model (note that including the not lagged value of imports would increase the R^2 substantially). Only the forth lagged value of $\Delta\Delta_4$log-imports is significant with p-value = 0.00029 < 0.05 and β_4 = -0.5279. If this covariate is increased by one unit, $\Delta\Delta_4$log-exports decreases by approximately 0.5 units.

3.2.3 Granger causality

A predictor variable is said to Granger cause a response variable when that covariate provides incremental predictive power for predicting the response. To test for Granger Causality we compare the previous fitted ADL(4) model with a model that only includes the exports autoregressive terms without the lagged imports terms. So, taking (2) we test:



(a) (b)

Figure 12: ADL(4). (a) Residual plot; (b) Residual correlogram

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

With a p-value = 0.0054 < 0.05 in the F-test we reject the H_0 of no Granger Causality. $\Delta\Delta_4$log-imports has incremental explanatory power in predicting $\Delta\Delta_4$log-exports.

3.3 Vector autoregressive model

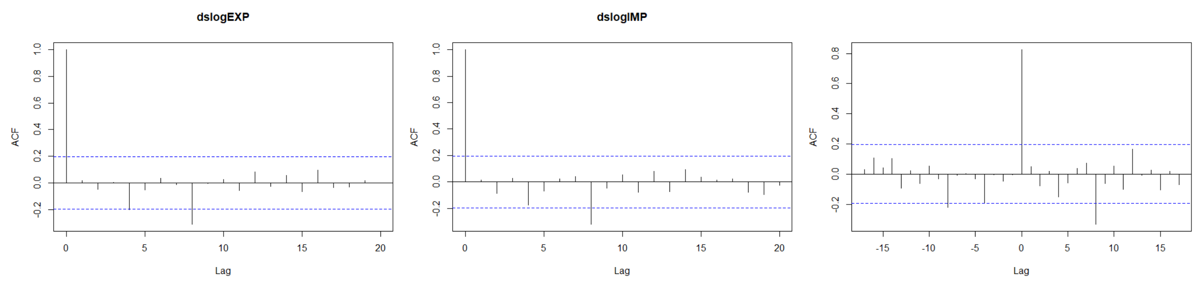
Until this point we have considered the exports as the response and the imports as the predictor. This choice was arbitrary; an analysis taking imports as the response and exports as the predictor could have also been conducted. A vector autoregressive (VAR) model is a generalization of ADL. VAR can be used to model any number of stationary time series in a symmetric approach to clarify the mutual influence among these series and provide simultaneous forecasts. Each variable have an equation that explains its evolution based on its own lags and the lags of the other variable.

We use automatic lag selection to find the best order specifying 10 as the maximum number of lags. The selected model by the BIC is VAR(1) with BIC = -12.24. The results given by the estimated VAR(1) model are not satisfactory as neither of the regression F-statistics are significant. Sometimes BIC tends to oversimplify the models so in this case we also fit VAR(4), the second best model according to the BIC with BIC = -12.17. Another possibility would be to fit VAR(8) model which is preferred by the AIC. VAR(4) model is defined as:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} a_{11}^1 & a_{12}^1 \\ a_{21}^1 & a_{22}^1 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \ldots + \begin{pmatrix} a_{11}^4 & a_{12}^4 \\ a_{21}^4 & a_{22}^4 \end{pmatrix} \begin{pmatrix} x_{t-4} \\ y_{t-4} \end{pmatrix} + \begin{pmatrix} u_{x,t} \\ u_{y,t} \end{pmatrix} \tag{3}$$

where x_t denotes the $\Delta\Delta_4$log-exports time series and y_t the $\Delta\Delta_4$log-imports. Note that the first equation of (3) is the same expression as (2).

To validate the model we check if the residuals $\vec{u}_t$ are multivariate white noise plotting the residual correlograms for each time series and the cross-correlogram (Figure 13). The only significant residual correlations are found in lag 8; apart from that, the residuals look like white noise. We test for no correlation at leads and lags between the $\vec{u}_t$ components with the Breusch-Godfrey test and we get a p-value = 0.03959 < 0.05 leading to the rejection of the H_0 of no correlation. Nonetheless, the result of the test is borderline.



(a) (b) (c)

Figure 13: VAR(4).residual correlograms. (a) dslogEXP; (b) dslogIMP; (c) cross-correlogram

In Figure 14 the ordinary least squares (OLS) estimated coefficients are presented. The results for the $\Delta\Delta_4$log-exports (dslogEXP) equation were already discussed in the ADL(4) section so here we only interpret the second equation. The regression with $\Delta\Delta_4$log-imports (dslogIMP) as the response is significant with p-value = 5.1e-05 < 0.05. R^2 = 0.30, so 30% of the variance of dslogIMP is explained by the lagged observations of dslogEXP and dslogIMP at lag 4. For this model there are two significant predictors: the forth lag of dslogIMP and the forth lag of dslogEXP.

```
Estimation results for equation dslogEXP:
=========================================
dslogEXP = dslogEXP.l1 + dslogIMP.l1 + dslogEXP.l2 + dslogIMP.l
 dslogEXP.l3 + dslogIMP.l3 + dslogEXP.l4 + dslogIMP.l4 + const

             Estimate Std. Error t value Pr(>|t|)
dslogEXP.l1 -1.540e-01  1.745e-01  -0.882 0.379863
dslogIMP.l1  1.584e-01  1.571e-01   1.008 0.316096
dslogEXP.l2 -2.784e-02  1.734e-01  -0.161 0.872795
dslogIMP.l2  5.944e-02  1.542e-01   0.386 0.700735
dslogEXP.l3  4.814e-02  1.682e-01   0.286 0.775348
dslogIMP.l3  7.932e-02  1.510e-01   0.525 0.600691
dslogEXP.l4  9.397e-02  1.618e-01   0.581 0.562867
dslogIMP.l4 -5.279e-01  1.402e-01  -3.764 0.000294 ***
const       -4.742e-05  5.354e-03  -0.009 0.992953
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.05377 on 92 degrees of freedom
Multiple R-Squared: 0.2715,     Adjusted R-squared: 0.2081
F-statistic: 4.286 on 8 and 92 DF,  p-value: 0.0002065
```

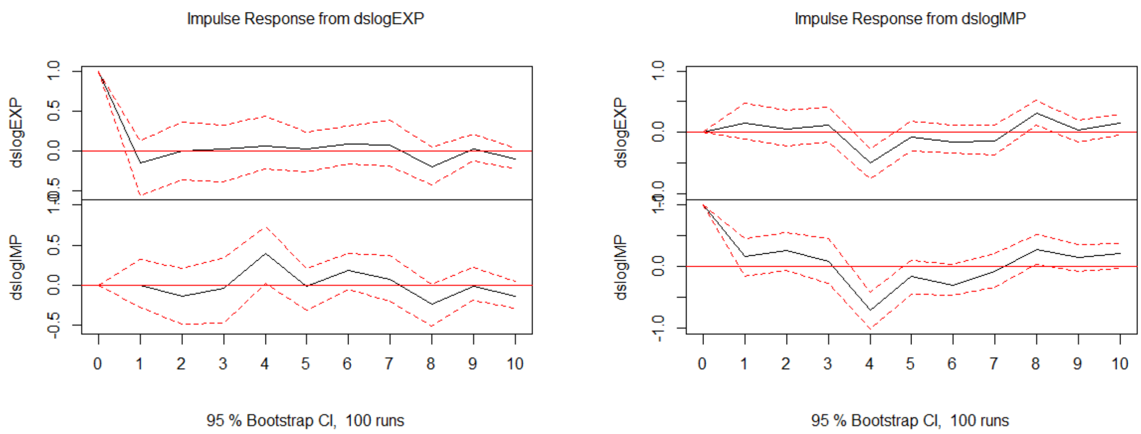
(a)

```
Estimation results for equation dslogIMP:
=========================================
dslogIMP = dslogEXP.l1 + dslogIMP.l1 + dslogEXP.l2 + dslogIMP.l2 +
 dslogEXP.l3 + dslogIMP.l3 + dslogEXP.l4 + dslogIMP.l4 + const

             Estimate Std. Error t value Pr(>|t|)
dslogEXP.l1 -0.0028140  0.1930507  -0.015   0.9884
dslogIMP.l1  0.1623781  0.1737909   0.934   0.3526
dslogEXP.l2 -0.1436483  0.1918106  -0.749   0.4558
dslogIMP.l2  0.2359568  0.1705520   1.383   0.1699
dslogEXP.l3 -0.0391156  0.1860486  -0.210   0.8339
dslogIMP.l3  0.0187851  0.1670834   0.112   0.9107
dslogEXP.l4  0.4318911  0.1790321   2.412   0.0178 *
dslogIMP.l4 -0.7736118  0.1551498  -4.986 2.89e-06 ***
const       -0.0002939  0.0059230  -0.050   0.9605
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.05949 on 92 degrees of freedom
Multiple R-Squared: 0.2969,     Adjusted R-squared: 0.2357
F-statistic: 4.856 on 8 and 92 DF,  p-value: 5.171e-05
```

(b)

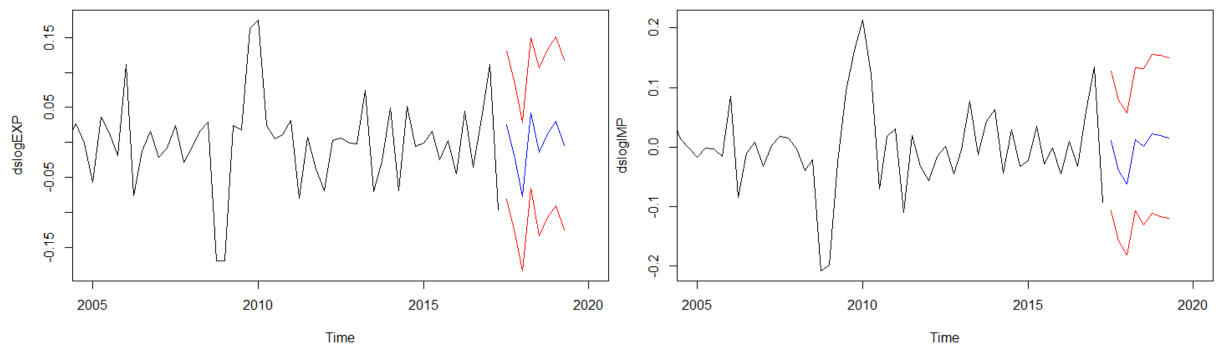
Figure 14: Results of VAR model. (a) dslogEXP; (b) dslogIMP

We use a non-orthogonal impulse response function and calculate 95% confidence intervals using bootstrap (Figure 15). When an impulse of one unit is given to dslogEXP there is no significant effect on the next lags of dslogEXP and there is a barely significant positive effect at lag 4 of dslogIMP and a barely significant negative effect at lag 8. When an impulse of one unit is given to dslogIMP there is a significant negative effect at lag 4 of both time series and a barely significant positive effect at lag 8 of both series. Again, we observe a clear seasonal (yearly) behavior. Both series tend to the equilibrium after some lags, so there is no long-term effect. Figure 16 presents the 8-step ahead prediction given by the VAR(4) model with its 95% confidence intervals for both time series.



(a) (b)

Figure 15: Impulse response function from shocks of (a) dslogEXP; (b) dslogIMP



(a) (b)

Figure 16: Forecast for the next 8 quarters using VAR(4) model. (a) dslogEXP; (b) dslogIMP

3.4 Cointegration

Two I(1) time series are cointegrated if there exists a linear combination of the two that is integrated of order zero (stationary). To fit an ADL and a VAR model we used the time series after taking non-seasonal and seasonal differences. If a cointegrating equation is found it is possible to work with the original log-transformed time series with an error correcting model (ECM). We use the Engle-Granger test and the Johansen test to check for cointegration of the log-transformed series.

3.4.1 Engle-Granger test

The Engle-Granger test is based on testing the stationarity of the residuals of the time series regression model. The Engle-Granger test is not symmetric, and it gives us a test statistic of -2.26 when we regress log-exports on log-imports and a test statistic of -2.15 when we regress log-imports on log-exports. This results are larger than -3.41, the critical value for this test with one explanatory variable at the 0.05 level, so we do not reject the null hypothesis in neither of both cases. Hence, the series are not cointegrated.

3.4.2 Johansen test

One of the advantages of the Johansen test is that it is symmetric in both time series. The optimal lag length needed for the Johansen test equation is obtained running a VAR automatic selection in levels (for log-exports and log-imports). VAR(5) is chosen by the BIC.

	test	10pct	5pct	1pct
r <= 1	3.24	7.52	9.24	12.97
r = 0	16.48	17.85	19.96	24.60

(a)

	test	10pct	5pct	1pct
r <= 1	3.24	7.52	9.24	12.97
r = 0	13.24	13.75	15.67	20.20

(b)

Figure 17: Johansen test results. (a) trace test statistic; (b) maximum eigenvalue test statistic

For r = 0, the test statistic is smaller than the critical value even at 10% significance level (Figure 17). There is no cointegrating relation between log-imports and log-exports. Thus, ECM can't be used in this case.

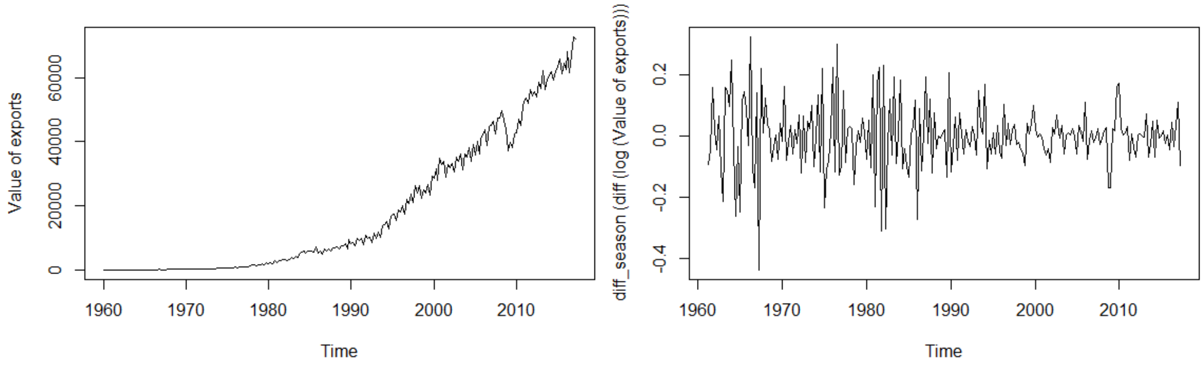
4 Conclusion

In the first part of the paper, the value of exports of Spain in the 1990-2017 period was analyzed. Taking non-seasonal and seasonal differences of the log-transformed series, stationarity was reached. The best model for this series according to in sample and out of sample criteria is SARIMA(0,1,0)(2,1,1). This model has a good prediction performance and was used to do a forecast for the following 8 quarters.

In the second part, the exports were related with the imports of the same period of time. First, a linear regression was fitted; however, it could not be validated. Next, models that include lagged terms were fitted and ADL(4) was validated and interpreted. We generalized the ADL(4) result with a VAR(4), which is symmetric for both time series and we presented the effect of an impulse and a forecast. Finally, it was shown that no cointegrating equation exists with the Engle-Granger and Johansen tests.

As it was expected, the imports and exports are related, and the use of past values of them allows to make valid predictions. It is also remarkable that an impulse on imports has a more noticeable effect (for both series) than an impulse on exports. The relation between this time series does not imply that one causes the other, it is very likely that a third latent variable is affecting the values of exports and imports.

A Appendix: Figures



(a) (b)

Figure A.1: Value of exports for the period 1960-2017 (a) Original time series; (b) Seasonal differences log-time series

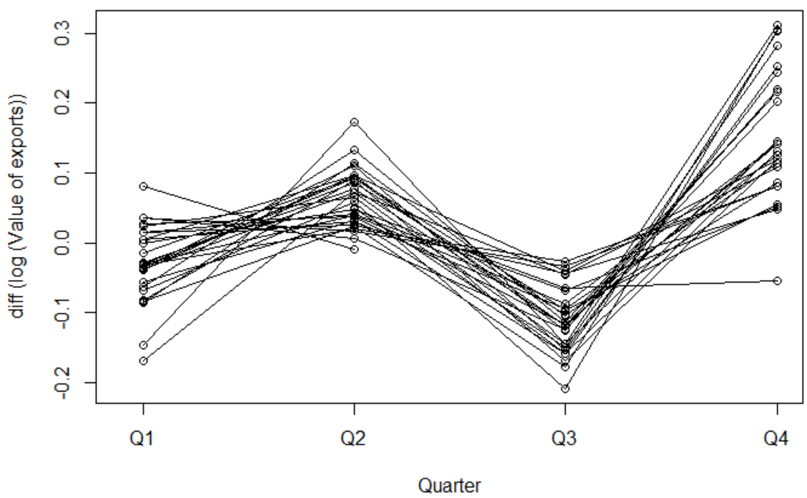


Figure A.2: Seasonplot

B Appendix: R script

```
rm(list=ls())

#libraries
library(readxl)
library(CADFtest)
library(forecast)
library(vars)
library(urca)
library(xtable)

#Import data
Trade<- read_excel("C:/Users/Daniel/Desktop/Trade_of_Goods.xlsx")

attach(Trade)

####UNIVARIATE ANALYSIS####

#Declare time series
exp_ts<-ts(`Goods, Value of Exports, National Currency`,frequency=4,start=c
    (1960,1))
exp_ts
ts.plot(exp_ts,ylab="Value of exports")

#log-differences
logexp_ts<-log(exp_ts)
ts.plot(logexp_ts)
dlogexp_ts<-diff(log(exp_ts))
ts.plot(dlogexp_ts,ylab="diff (log (Value of exports))")
acf(dlogexp_ts)

#correct for seasonality
monthplot(dlogexp_ts, ylab="diff (log (Value of exports))")
dslogexp_ts<-diff(diff(log(exp_ts)),lag=4)
ts.plot(dslogexp_ts,ylab="diff_season (diff (log (Value of exports)))")

#Seems like a structural break: subset 1990-2017

detach(Trade)
Trade<-Trade[121:230,]
attach(Trade)

exp_ts<-ts(`Goods, Value of Exports, National Currency`,frequency=4,start=c
    (1990,1))
ts.plot(exp_ts,ylab="Value of exports")

logexp_ts<-log(exp_ts)
ts.plot(logexp_ts)
#test if deterministic or stochastic trend
max.lag<-round(sqrt(length(exp_ts))) #10
CADFtest(logexp_ts, type= "trend", criterion= "BIC", max.lag.y=max.lag)

dlogexp_ts<-diff(log(exp_ts))
ts.plot(dlogexp_ts,ylab="diff (log (Value of exports))")
```

```
#correct for seasonality
monthplot(dlogexp_ts,ylab="diff (log (Value of exports))")
seasonplot(dlogexp_ts,ylab="diff (log (Value of exports))",main="")
dslogexp_ts<-diff(diff(log(exp_ts)),lag=4)
ts.plot(dslogexp_ts,ylab="diff_season (diff (log (Value of exports)))")

#unit root test
max.lag<-round(sqrt(length(dslogexp_ts))) #10
CADFtest(dslogexp_ts, type= "drift", criterion= "BIC", max.lag.y=max.lag)

#correlograms
acf(dslogexp_ts,main="")
pacf(dslogexp_ts,main="")


#SARIMA modeling
max.lag<-round(sqrt(length(logexp_ts)))

#MA model
fit_ma<-arima(logexp_ts,order=c(0,1,0),seasonal=list(order=c(0,1,1)))
fit_ma
acf(fit_ma$residuals)
pacf(fit_ma$residuals)
Box.test(fit_ma$residuals,lag=max.lag,type="Ljung-Box")
BIC(fit_ma)

#AR model
fit_ar<-arima(logexp_ts,order=c(0,1,0),seasonal=list(order=c(3,1,0)))
fit_ar
acf(fit_ar$residuals)
pacf(fit_ar$residuals)
Box.test(fit_ar$residuals,lag=max.lag,type="Ljung-Box")
BIC(fit_ar)

#ARMA model
fit_arma<-arima(logexp_ts,order=c(0,1,0),seasonal=list(order=c(2,1,1)))
fit_arma
acf(fit_arma$residuals)
pacf(fit_arma$residuals)
Box.test(fit_arma$residuals,lag=max.lag,type="Ljung-Box")
BIC(fit_arma)


#Prediction

#ARMA model

myforecastARMA<-predict(fit_arma,n.ahead=8); myforecastARMA
expected<-myforecastARMA$pred; expected

#CI
lower<-myforecastARMA$pred-qnorm(0.975)*myforecastARMA$se
upper<-myforecastARMA$pred+qnorm(0.975)*myforecastARMA$se

ts.plot(logexp_ts,xlim=c(2005,2020),ylim=c(10,12),ylab="log (Value of exports)")
lines(expected,col="red")
```

```
lines(lower,col="blue")
lines(upper,col="blue")

#MA model

myforecastMA<-predict(fit_ma,n.ahead=8)
expected<-myforecastMA$pred
lower<-myforecastMA$pred-qnorm(0.975)*myforecastMA$se
upper<-myforecastMA$pred+qnorm(0.975)*myforecastMA$se
cbind(lower,expected,upper)
ts.plot(logexp_ts,xlim=c(2005,2020),ylim=c(10,12))
lines(expected,col="red")
lines(lower,col="blue")
lines(upper,col="blue")

#Compare forecasts: MAE and RMSE of best 2 BIC models
y<-logexp_ts
S=round(0.75*length(y))
h=1
error1.h<-c() #SARIMA(0,1,0)(0,1,1)
for (i in S:(length(y)-h))
{
  mymodel.sub<-arima(y[1:i], order = c(0,1,0),seasonal=c(0,1,1))
  predict.h<-predict(mymodel.sub,n.ahead=h)$pred[h]
  error1.h<-c(error1.h,y[i+h]-predict.h)
}
error2.h<-c() #SARMA(0,1,0)(2,1,1)
for (i in S:(length(y)-h))
{
  mymodel.sub<-arima(y[1:i], order = c(0,1,0),seasonal=c(2,1,1))
  predict.h<-predict(mymodel.sub,n.ahead=h)$pred[h]
  error2.h<-c(error2.h,y[i+h]-predict.h)
}
#MAE
MAE1 <- mean(abs(error1.h)); MAE1 #SARIMA(0,1,0)(0,1,1)
MAE2 <- mean(abs(error2.h)); MAE2 #SARMA(0,1,0)(2,1,1)

#RMSE
sqrt(mean(error1.h^2))
sqrt(mean(error2.h^2))

#MAPE
rerror.h<-c()
for (i in S:(length(y)-h))
{
  mymodel.sub<-arima(y[1:i], order = c(0,1,0),seasonal=c(2,1,1)) #choose model
  predict.h<-predict(mymodel.sub,n.ahead=h)$pred[h]
  rerror.h<-c(rerror.h, (y[i+h]-predict.h)/y[i+h])
}
MAPE <- mean(abs(rerror.h)); MAPE

#Debold-Mariano test
dm.test(error1.h,error2.h,h=h,power=1)
dm.test(error1.h,error2.h,h=h,power = 2)
```

```
#####MULTIVARIATE ANALYSIS####

#import TS
imp_ts<-ts('Goods, Value of Imports, CIF, National Currency',frequency=4,start=c
    (1990,1))
ts.plot(exp_ts,imp_ts,col=c("blue","red"))
legend("bottomright",legend=c("Exports","imports"),col=c("blue","red"),lty = c
    (1,1))

#log-differences
logimp_ts<-log(imp_ts)
ts.plot(logexp_ts,logimp_ts,col=c("blue","red"))
dlogimp_ts<-diff(log(imp_ts))
ts.plot(dlogexp_ts,dlogimp_ts,col=c("blue","red"))

#correct for seasonality
monthplot(dlogimp_ts)
dslogimp_ts<-diff(diff(log(imp_ts)),lag=4)
ts.plot(dslogexp_ts,dslogimp_ts,col=c("blue","red"))
legend("bottomright",legend=c("Exports","imports"),col=c("blue","red"),lty = c
    (1,1))

#unit root test for imports
max.lag<-round(sqrt(length(dslogimp_ts))) #10
CADFtest(dslogimp_ts, type= "drift", criterion= "BIC", max.lag.y=max.lag)

#test I(1)
CADFtest(dlogexp_ts, type= "drift", criterion= "BIC", max.lag.y=max.lag)
CADFtest(dlogimp_ts, type= "drift", criterion= "BIC", max.lag.y=max.lag)


#Linear regression

fit<-lm(dslogexp_ts~dslogimp_ts)
summary(fit)
xtable(fit)
ts.plot(fit$residuals,ylab="Residuals",xlab="Index")
acf(fit$residuals,main="")
pacf(fit$residuals)
Box.test(fit$residuals, lag = max.lag, type = "Ljung-Box")


#DL(4)

lag <- 4
n <- length(dslogexp_ts)
dslogexp.0 <- dslogexp_ts[(lag+1):n]
dslogimp.0 <- dslogimp_ts[(lag+1):n]
dslogimp.1 <- dslogimp_ts[lag:(n-1)]
dslogimp.2 <- dslogimp_ts[(lag-1):(n-2)]
dslogimp.3 <- dslogimp_ts[(lag-2):(n-3)]
dslogimp.4 <- dslogimp_ts[(lag-3):(n-4)]
fit_dlm <- lm(dslogexp.0 ~ dslogimp.0+dslogimp.1+dslogimp.2+dslogimp.3+dslogimp
    .4)
summary(fit_dlm)
```

```
ts.plot(fit_dlm$residuals,ylab="Residuals")
acf(fit_dlm$residuals,main="")
Box.test(fit_dlm$residuals, lag = max.lag, type = "Ljung-Box")


#ADL(4)

dslogexp.1 <- dslogexp_ts[lag:(n-1)]
dslogexp.2 <- dslogexp_ts[(lag-1):(n-2)]
dslogexp.3 <- dslogexp_ts[(lag-2):(n-3)]
dslogexp.4 <- dslogexp_ts[(lag-3):(n-4)]
fit_adlm <- lm(dslogexp.0 ~ dslogexp.1+dslogexp.2+dslogexp.3+dslogexp.4
               +dslogimp.1+dslogimp.2+dslogimp.3+dslogimp.4)

summary(fit_adlm)
ts.plot(fit_adlm$residuals,ylab="Residuals")
acf(fit_adlm$residuals,main="")
Box.test(fit_adlm$residuals, lag = max.lag, type = "Ljung-Box")


#Granger-Causality
fit_adlm_nox <- lm(dslogexp.0 ~ dslogexp.1+dslogexp.2+dslogexp.3+dslogexp.4)
anova(fit_adlm,fit_adlm_nox)
xtable(anova(fit_adlm,fit_adlm_nox))


#VAR

dslogTrade<-data.frame(dslogexp_ts,dslogimp_ts)
names(dslogTrade)<-c("dslogEXP","dslogIMP")
attach(dslogTrade)

VARselect(dslogTrade,lag.max=10,type="const")

#VAR(1)
fit_var1<-VAR(dslogTrade,type="const",p=1)
summary(fit_var1)

#VAR(4)
fit_var4<-VAR(dslogTrade,type="const",p=4)
summary(fit_var4)
var4_residuals<-resid(fit_var4)
acf(var4_residuals[,1],main="dslogEXP")
acf(var4_residuals[,2],main="dslogIMP")
ccf(var4_residuals[,1],var4_residuals[,2],main="")

test_bg<- serial.test(fit_var4, lags.pt=10, type = "BG"); test_bg

#impulse response function
irf_var4<-irf(fit_var4,ortho=FALSE,boot=TRUE)
plot(irf_var4)

#VAR prediction
myforecastVAR<-predict(fit_var4,n.ahead=8)
## dslogEXP
dslogEXP_ts_forecast<-ts(myforecastVAR$fcst$dslogEXP[,1], frequency=4,start=c
```

```
    (2017,3))
dslogEXP_ts_lower<-ts(myforecastVAR$fcst$dslogEXP[,2],frequency=4,start=c(2017,3)
    )
dslogEXP_ts_upper<-ts(myforecastVAR$fcst$dslogEXP[,3],frequency=4,start=c(2017,3)
    )
ts.plot(dslogEXP,dslogEXP_ts_forecast,dslogEXP_ts_lower,dslogEXP_ts_upper,col=c("
    black","blue","red","red"),
        xlim=c(2005,2020),ylab="dslogEXP")
## dslogIMP
dslogIMP_forecast<-ts(myforecastVAR$fcst$dslogIMP[,1],frequency=4,start=c(2017,3)
    )
dslogIMP_lower<-ts(myforecastVAR$fcst$dslogIMP[,2],frequency=4,start=c(2017,3))
dslogIMP_upper<-ts(myforecastVAR$fcst$dslogIMP[,3],frequency=4,start=c(2017,3))
ts.plot(dslogIMP,dslogIMP_forecast,dslogIMP_lower,dslogIMP_upper,col=c("black","
    blue","red","red"),
        xlim=c(2005,2020),ylab="dslogIMP")


#Cointegration: Engle-Granger
coint1 <- lm(logexp_ts ~ logimp_ts)
summary(coint1)
plot.ts(coint1$res)
CADFtest(coint1$res, type="drift", criterion="BIC", max.lag.y=10)
coint2 <- lm(logimp_ts ~ logexp_ts)
summary(coint2)
plot.ts(coint2$res)
CADFtest(coint2$res, type="drift", criterion="BIC", max.lag.y=10)

#Cointegration: johansen test
logTrade<-data.frame(logexp_ts,logimp_ts)
names(logTrade)<-c("logEXP","logIMP")

VARselect(logTrade,lag.max=10,type="const",season = 4)

trace_test<-ca.jo(logTrade,type="trace",K=5,ecdet="const",spec="transitory",
    season=4)
summary(trace_test)
maxeigen_test<-ca.jo(logTrade,type="eigen",K=5,ecdet="const",spec="transitory",
    season=4)
summary(maxeigen_test)


#Persistence in volatility?
library(fGarch)
acf((fit_ma$residuals)^2)
acf((fit_ar$residuals)^2)
acf((fit_arma$residuals)^2)
```

AFDELING
Straat nr bus 0000
3000 LEUVEN, BELGIË
tel. + 32 16 00 00 00
fax + 32 16 00 00 00
www.kuleuven.be