Assignment Tijdreeks 2

Arthur van Roest (505615ar)
Nicky Sonnemans (506125ns)

June 2020

1 Introduction

In this report we cover the second assignment of the course time series analysis. Each sub question is answered in a separate subsection, so 2a is answered in section 2.1, 2b in section 2.2, et cetera.

2 Results

2.1 Key features

(i) **Trend**: Figure 1 shows the original time series of imports. We observe a pronounced trend in the time series over the full sample (1979 Q1 - 2019Q4) looking just at the graph. We can test this hypothesis of a trend on the full sample by using the following regression:

$$imports_t = \alpha + \beta t + \epsilon_t \; for \; t = 1, 2, .., T \tag{1}$$

The regression gives an estimate (obtained by least squares using the complete sample period) of $\beta = 0.223$ *(std. error = 0.004)*, which is significantly different from zero at the 5% significance level (p-value: 0.0000).

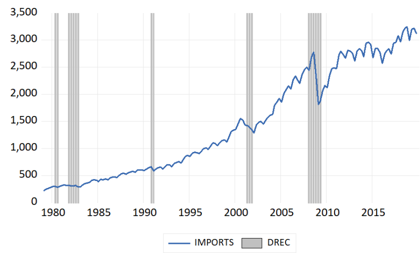
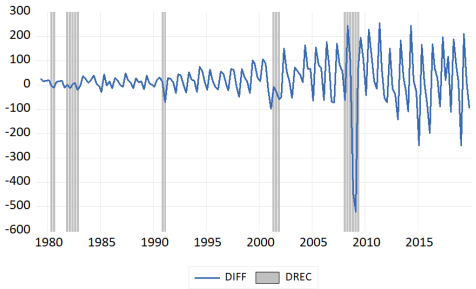


Figure 1: imports of goods and services (1979Q1-2019Q4)

(ii) **Seasonality**: In figure 2 we look at the difference between the observation at time t compared to the observation at time*t-1*, this is denoted as *diff*. The left window of figure 2 contains this time series with the series *DREC* included, we observe a 'saw-tooth' pattern here. In the right window of figure 2 we do not observe remarkable differences between the different quarters. To test whether a seasonal effect exists we regress the difference between the current and previous observation on four dummy variables, each representing a different quarter in the year:

$$diff_t = \mu_1 Q1 + \mu_2 Q_2 + \mu_3 Q_3 + \mu_4 Q_4 + \epsilon_t \; for \; 't = 1, 2, ..., T \tag{2}$$

where Q1,...,Q4 are dummy variables for quarter 1 through 4. A Wald test to test joint significance of all quarterly variables with the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ yields a F-statistic with the value 33.6 (corresponding with a p-value of 0.0000). Therefore we reject the null hypothesis of no seasonality, furthermore since we find an R^2 of 0.39 in this regression this further strengthens the belief of seasonality



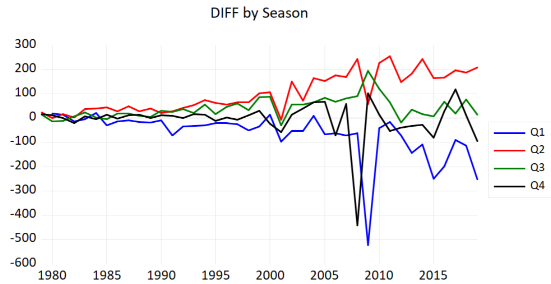


Figure 2: First differences of imports of goods and services (1979Q1-2019Q4)

in the data.
(iii) **Abberant observations**: From the graph in figure 1 we observe some abberant observations during the period of the Great Financial Crisis (2008 Q3 - 2009 Q1). This hypothesis of abberant observations during this time period is strengthened by observing the left window of figure 2, the quarterly difference between import of goods, we notice a big downward spike during this time period.
(iv) **Heteroskedasticity**: In the left window of figure 2 we can not only clearly see an increasing standard deviation over time, but also an extreme downward spike during the period of the Great Financial Crisis (2008 Q3 - 2009 Q1). We start by investigating the sample standard deviation of the period 2000 Q1 - 2019 Q4 compared to 1979Q1 - 1999 Q4, which are respectively 137.4 and 30.4. This is a big difference, hence we can conclude heteroskedasticity.
(v) **Nonlinearity**: From figure 1 it appears that the import of goods experiences declines during recession periods, especially during the Great Financial Crisis. We test this possible non-linearity using the following regression:

$$diff_t = \mu_1 DREC_t + \mu_{2]}(1 - DREC_t) + \epsilon_t \; for \; t = 1, 2, ..., T \tag{3}$$

This regression yields estimates for μ_1 *and* μ_2 of respectively -45.87 (*std. error: 22.66*) and 25.69 (*std. error: 7.99*), with respective p-values of 0.045 and 0.002. We conclude nonlinearity in the data. Furthermore as the R^2 equals 0.05, this further strengthens the idea of nonlinearity in the data.

We found that the time series of import of goods has the following characteristics: Trend, Seasonality, Abberant observations, Heteroskedasticity and Nonlinearity. Specific observations however, the Great Financial Crisis in particular are very influential for these characteristics. Running previously mentioned regressions for seasonality and nonlinearity again, yet excluding the time period of the Great Financial Crisis (2008 Q3 - 2009 Q1). We find for running regression (2), testing for seasonality, an R^2 of 0.48 and a value for the Wald test for joint significance of 47.95 with a corresponding p-value of 0.000. Hence we still conclude seasonality in the data. However, when running regression (3), we find estimates for μ_1 *and* μ_2 of respectively 3.32 (*std. error: 21,02*) and 25.69 (*std. error: 6.76*), with respective p-values of 0.875 and 0.0002. Furthermore as the R^2 equals 0.006, we conclude no nonlinearity in the data excluding the abberant observations of the Great Financial Crisis.

2.2 Trend detection

We take the natural logarithm of the seasonally adjusted import series, this series will be denoted as $logimportssa_t$. Since we wish to determine if the series either contains a determinstic or a stochastic trend, we use the Augmented Dickey-Fuller (ADF) test. As the unconditional mean of the series is not equal to zero, we ought to include an intercept in the ADF test equation. Furthermore we've detected the presence of a trend in the series, which is discussed in the previous section, we ought to include a trend in the ADF test equation.

Using the ADF in Eviews, we find the following results for either the SIC or the AIC to select the lag length.

	AIC	SIC
Lag length	2	1
ADF test-statistic	-1.31	-1.91
ADF prob.	0.88	0.65

Table 1: ADF Unit Root test on sample period 1980 Q1 - 2019 Q4

The outcome on the test does not necessarily depend on how the lag lengths are selected, either SIC or AIC. To investigate how important the period 2009 Q1 - 2019 Q4 is for the above finding, we will run ADF test again on the sample 1980 Q1 - 2008 Q4.

	AIC	SIC
Lag length	2	1
ADF test-statistic	-3.75	-3.88
ADF prob.	0.023	0.016

Table 2: ADF Unit Root test on sample period 1980 Q1 - 2008 Q4

As one can notice, the period 2009 Q1 - 2019 Q4 is very important for the ADF test conclusions of either rejecting or accepting the null-hypothesis. This is because the period excluding 2009 Q1 - 2019 Q4 rejects the presence of a stochastic trend, however the full sample fails to reject the presence of a stochastic trend.

As the ADF test for the full sample (1980 Q1 - 2019 Q4) fails to reject the null-hypothesis, we could further investigate this time series by taking the first differences. This is necessary in order to remove the nonstationarity in the time series due to the presence of a stochastic trend.

2.3 Estimation of AR(3)

We estimate the parameters of the AR(3) model with intercept and deterministic trend for the *log imports* series for the sample period 1980 Q1 - 2019 Q4 using the following regression:

$$\phi_3(L)(logimportssa_t - \mu - \delta t) = \epsilon_t \tag{4}$$

where $\phi_3(L) = 1 - \phi_1 L - \phi_2 L - \phi_3 L$. After estimating this model, the results are shown in the following table. The R^2 equals to 0.999, which implies that the AR(3) model explains a great proportion of the

	Coefficient	Std. Error	Prob.
$\hat{\mu}$	-104.147	23.814	0.000
$\hat{\delta}$	0.000	3.25e^{-5}	0.000
$\hat{\phi}_1$	1.497	0.0786	0.000
$\hat{\phi}_2$	-0.723	0.131	0.000
$\hat{\phi}_3$	0.201	0.079	0.012

Table 3: Estimated parameters of the AR(3) model of log imports

variance in the data. Furthermore according to the t-test $\hat{\phi}_3$ is not significantly different from zero at the 5% significance.

Examining the roots of this AR(3) polynomial, we find the inverse roots using E-views to be $\lambda_{1,3} = 0.267 \pm 0.371i, \lambda_2 = -0.0963$. As the roots lie inside the unit circle, the AR(3) model is stationary.

(i) Autocorrelation properties residuals:

Using the correlogram of the residuals, we find no significant autocorelations or partial autocorellations. We find that the white noise property of $E[\epsilon_s \epsilon_t] = 0$ for $s, t = 1, 2, ..., T$ and $s \neq t$ holds for this model.

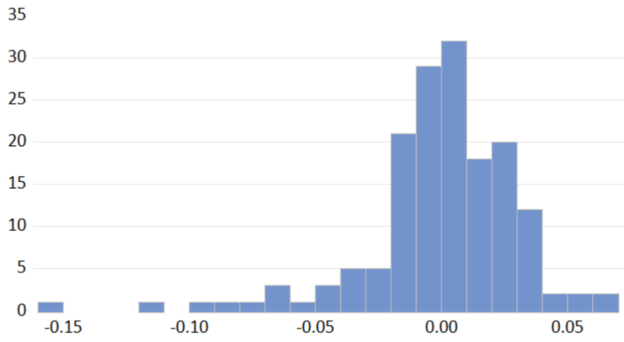
(ii) Autocorrelation properties squared residuals:

Using the corelogram of the squared residuals, we find significant autocorrelations and partial autocorrelations for up to 8 lags, this implies conditional heteroskedasticity. This is confirmed by the White Heteroskedasticity test, which rejects the null hypothesis of homoskedasticity with a p-value of 0.0000. As we find heteroskedasticity in our model, the white noise property of $E[\epsilon_t^2] = \sigma^2$ for $t = 1, 2, ..., T$ does not hold for this model.

(iii) Normality residuals:

The residuals are shown in the figure below. One can observe that the kurtosis differs significantly from 3, which is the kurtosis of the normal distribution. Furthermore the Jarque-Bera null-hypothesis is rejected. Thus we conclude that the residuals are not normally distributed.

We can test if the mean is signicantly different from 0 using the t-test. We find that $t = \frac{-3.03*10^{-11}}{\frac{0.0298}{160}} < 1.645$ (the 5% significance), which implies that we do not reject the null hypothesis and the mean does not differ significantly from zero. This implies that the white noise property of $E[\epsilon_t] = 0$ for $t = 1, 2, ..., T$ holds for this model.



Series: Residuals	
Sample 1980Q1 2019Q4	
Observations 160	
Mean	3.03e-11
Median	0.001847
Maximum	0.067029
Minimum	-0.150810
Std. Dev.	0.029814
Skewness	-1.487611
Kurtosis	7.972041
Jarque-Bera	223.8209
Probability	0.000000

Figure 3: Residuals

2.4 Forecasts

We re-estimate the AR(3) model with a intercept and deterministic trend for log import with the sample period 1980Q1 to 1999Q4:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \delta t + \epsilon_t \tag{5}$$

Where y_t is the logarithm of the seasonally adjusted imports time series at the time t. $t = 0$ at 1980Q1 and increases by 1 every quarter. Running this regression yields the following coefficients: $\alpha = 5.634$, $\phi_1 = 1.036$, $\phi_2 = -0.155$, $\phi_3 = -0.065$ and $\delta = 0.019$. In EViews, we generate point estimates for the forecast period 2000Q1 to 2019Q4. We convert the point estimates to a quarterly growth rate ($\hat{g}_{t+1|t}$), calculated with the following formula:

$$\hat{g}_{t+1|t} = \hat{y}_{t+1|t} - y_t \tag{6}$$

Where $\hat{y}_{t+1|t}$ denotes the one-step ahead forecast in quarter $t + 1$ and y_t is the actual value at time t.

First, we do absolute evaluation of the forecasts. We compare the point estimates for the log imports growth rate to the actual growth rate (g_t) which is calculated by $g_t = y_t - y_{t-1}$, where y_t is the log import at time t.

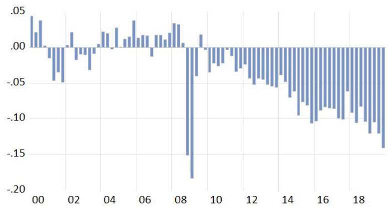


Figure 4: forecast error

We use the forecast error $e_{t+1|t} = g_{t+1} - \hat{g}_{t+1|t}$, that we see in plotted in figure 4, to evaluate the forecast. The average value of the forecast error is -0.035, which is means that the error is not unbiased, which is also visible in figure 4. We also see that the error is negatively biased, with a skewness of -0.558. We measure accuracy with the MSPE which is calculated as follows:

$$\frac{1}{P}\sum_{t=T}^{t=T+P-1}(g_{t+1} - \hat{g}_{t+1|t})^2 \tag{7}$$

Where P is the number of observations and T the start of the forecast period. The value of MSPE equals 0.0036, which is very small, indicating that the forecast has a high level of accuracy. In order to test whether the AR(3) model with intercept and deterministic trend is efficient, we use the Mincer-Zarnowitz regression:

$$g_{t+1} = \beta_0 + \beta_1 \hat{g}_{t+1|t} + \eta_{t+1} \tag{8}$$

Where g_{t+1} is the actual growth rate and $\hat{g}_{t+1|t}$ is the one step ahead forecast of the growth rate. In case the forecast is efficient, the coefficient $\beta_0 = 0$ and $\beta_1 = 1$. The idea is that, for a efficient forecast, it is not possible to predict the forecast error at time $t+1$ using information available at time t. Running the regression yields $\beta_0 = 0.005$ with standard deviation 0.006 and $\beta_1 = 0.111$ with standard deviation 0.114. Testing the null hypothesis $H_0 : \beta_0 = 0$ yields a p-value of 0.428, thus we do not reject the null hypothesis. However, when testing the null hypothesis $H_0 : \beta_1 = 1$ gives p-value 0.000, thus we reject the null hypothesis. We conclude that the forecast is not efficient.

We now compare the forecast using the AR(3) model with intercept and deterministic trend to forecasts made with a AR(2) model with intercept.

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \tag{9}$$

Similarly to the AR(3) model, we create forecasts for the forecast period 2000Q1 to 2019Q4 and we convert these estimates to growth rates. We then calculate the forecast error for the AR(2) model the same way as for the AR(3) model. From now on we will refer to the forecast error of the AR(3) model as $e_{1,t+1|t}$ and we will denote the forecast error of the AR(2) model as $e_{2,t+1|t}$. To perform our analysis we assume that the relevant loss function is the squared forecast error. We calculate the loss differential as follows:

$$d_{t+1} = e^2_{1,t+1|t} - e^2_{2,t+1|t} \tag{10}$$

If both models had the same forecast accuracy we would have $E(d_{t+1}) = 0$, in the figure below we see, on the left, the forecast errors of both models, and on the right we see the loss differential plotted.
To test whether one of the models forecasts more accurately we use the Diebold-Mariano statistic which follows a standard normal distribution and is calculated as follows:

$$DM = \frac{\bar{d}}{\sqrt{V(\hat{d}_{t+1})/P}} \tag{11}$$

The Diebold-Mariano statistic equals 4.940, which means that one of the models is significantly more accurate than the other. Because the average loss differential is positive we know that the forecasts generated with the AR(2) model are more accurate.

The fact that forecasts generated using a model without deterministic trend are superior is in agreement with the result in section 2.2 (2b) where the null hypothesis of the presence of a deterministic trend is rejected.

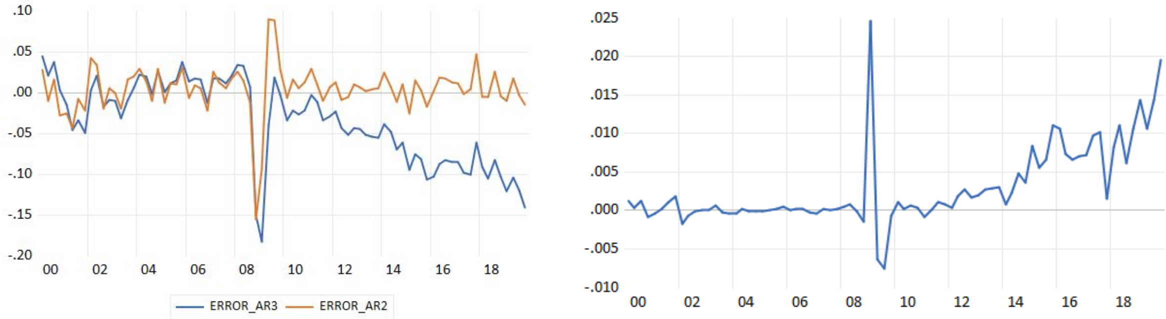


Figure 5: Left: forecast errors for both models, Right: loss differential

2.5 Moving Window

In the section we first repeat the forecast that we made in section 2.4 where we estimate the parameters of the AR(3) model with intercept and deterministic trend that describes the log imports in the period 1980Q1-1999Q4, this model is then used to forecast 2000Q1. Then we re-estimate the parameters using the sample period 1980Q2 to 2000Q1 to forecast the log imports for 2000Q2. We repeat this process until we have an estimate for 2019Q4 with sample period 1999Q4 to 2019Q3. We then convert the forecasts for log imports to growth rates applying the same method we used in section 2.4. The code we used to generate the forecasts can be found in the appendix. In figure 6 we see the forecasted growth rates generated with this method compared to the actual growth rates.

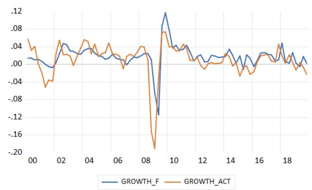


Figure 6: forecast compared to actual growth rate in log imports

We average forecast error is 0.007 with a standard deviation of 0.003, when we apply a t-test we find that the null hypothesis that the average forecast error is 0 is not rejected. Thus, the forecast are unbiased. For efficiency, we test the forecast using the Mincer-Zarnovitz regression.

$$g_{t+1} = \beta_0 + \beta_1 \hat{g}_{t+1|t} + \eta_{t+1} \tag{12}$$

Where g_{t+1} is the actual growth rate and $\hat{g}_{t+1|t}$ is the one step ahead forecast of the growth rate. In case the forecast is efficient, the coefficient $\beta_0 = 0$ and $\beta_1 = 1$. We find that we do not reject the null hypothesis $\beta_0 = 0$ because the t-test yield a p-value of 0.31, we also do not reject the null hypothesis that $\beta_1 = 1$ because the t-test yield a p-value of 0.23. Thus, we conclude that the forecast are efficient.

When comparing the average forecast error in the model in 2.4 to the method used in this section we find that the method used in this section has an average forecast error of -0.007 compared to -0.035 in the model in section 2.4. The results of this forecasting method are are more accurate because they incorporate more recent information which is more relevant to the forecast.

We also compare the AR(3) model with intercept and deterministic trend to an AR(2) model with intercept. The code used to generate this forecast can also be found in the appendix. In the figure 7 we see the forecasts plotted next to the actual growth rates.

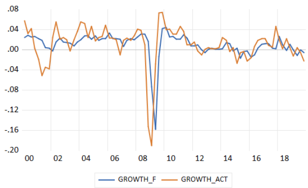


Figure 7: forecast compared to actual growth rate in log imports

The forecast look very similar to the results of the AR(3) model with intercept and deterministic trend. However, when we look at the average forecast error we find 0.001 for the AR(2) model compared to 0.007 for the AR(3) model. We thus conclude that the AR(2) model with intercept is better for forecast the growth rates of log imports using moving windows.

The fact that forecasts generated using a model without deterministic trend are superior is in agreement with the result in section 2.2 (2b) where the null hypothesis of the presence of a deterministic trend is rejected.

3 Appendix

Moving Window

AR(3) with intercept and deterministic trend

```
smpl 1980Q1 2019Q4

scalar mwlength = 80  'Length of moving window
scalar ARp = 3          'order of AR(p) model

scalar nofqtrs = @obsrange-mwlength-ARp  'Number of quarters in the out-of-
sample period

series logimport_f = NA    'Initialize series of one-step ahead point forecasts
series logimport_se = NA 'Initialize series of standard errors

for !i = 1 to ARp
  %ARstr = %ARstr + "logimport(-"+@str(!i) + ") "
next

for !qtr = 1 to nofqtrs    'Loop over out-of-sample period
  smpl @first+!qtr-1+ARp @first+mwlength+!qtr-2+ARp    'Set sample to moving
window
  equation tempAR.ls logimport c trend {%ARstr} 'Estimate AR(p) model

  smpl @first+mwlength+!qtr-1+ARp @first+mwlength+!qtr-1+ARp  'Set sample
to next quarter for forecasting
  tempAR.fit temp_f     'Compute one-step ahead point forecast
  logimport_f = temp_f      'Store point forecast
  logimport_se = tempAR.@se    'Store standard error
next
```

AR(2) with intercept

```
smpl 1980Q1 2019Q4

scalar mwlength = 80  'Length of moving window
scalar ARp = 2          'order of AR(p) model

scalar nofqtrs = @obsrange-mwlength-ARp  'Number of quarters in the out-of-
sample period

series logimport_f = NA    'Initialize series of one-step ahead point forecasts
series logimport_se = NA 'Initialize series of standard errors

for !i = 1 to ARp
  %ARstr = %ARstr + "logimport(-"+@str(!i) + ") "
next

for !qtr = 1 to nofqtrs    'Loop over out-of-sample period
  smpl @first+!qtr-1+ARp @first+mwlength+!qtr-2+ARp    'Set sample to moving
window
  equation tempAR.ls logimport c {%ARstr}    'Estimate AR(p) model

  smpl @first+mwlength+!qtr-1+ARp @first+mwlength+!qtr-1+ARp  'Set sample
to next quarter for forecasting
  tempAR.fit temp_f     'Compute one-step ahead point forecast
  logimport_f = temp_f      'Store point forecast
  logimport_se = tempAR.@se    'Store standard error
next
```