Advanced Time Series Analysis - Paper

Vincent Buekers (r0754046)

20/12/2018

1 Introduction

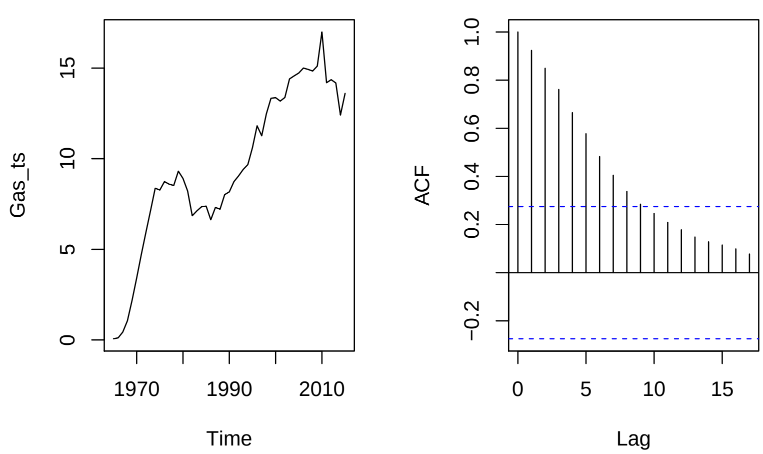
In this paper an analysis is conducted on data regarding four different time series measured between 1965 and 2015 in Belgium, namely: carbon dioxide (CO2) emissions, oil consumption, natural gas consumption and coal consumtion. The goal of the analysis is twofold. On one hand, univeriate modelling will be performed in order to develop a reliable forecast for Belgian gas consumption. On the other hand, multivariate modelling will be used to assess the relationship, if any, between carbon emissions and the consumpion of the aforementioned natural resources. Furthermore, the Box-Jenkins approach to time series modelling will be implemented to structure the modelling process. The data used for this analysis has been retrieved from Quandl (https://www.quandl.com/data/BP-British-Petroleum?keyword=Belgium). Yet, the collection of the data has been done by Britsh Petroleum. The statistical programming language R is used to support the analysis, of which the script has been added in appendix.

2 Univariate Time Series Analysis

Belgian Natural Gas Consumption

The gas consumption data have been measured at the 31st of december of each consequent year, hence this is a time series with an annual frequency. Moreover, the natural gas is measured as millions of tonnes oil equivalent.

***Figure 1**: Plot - Gas & **Figure 2**: Correlogram - Gas*

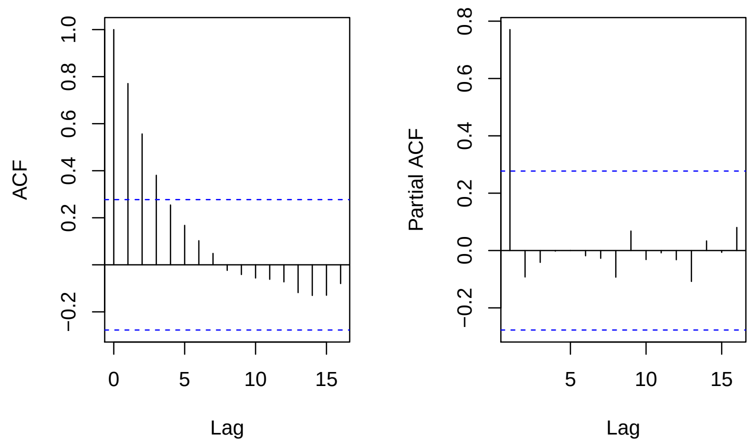


Based on just the graphical assessment of the time series plot (Figure 1), it is quite safe to conclude that this is not a stationary time series. Moreover, there also clearly seems to be an upward trend over the course of the past decades. The correlogram (Figure 2) shows that there are 9 autocorrelations which are significantly different from 0, indicating a strong persistency in the series. An augmented Dickey-Fuller test with trend does not reject there is a stochastic trend in this series ($p = 0.75$). Therefore, also the earlier assumed non-stationarity is indeed the case. Consequently, a difference operator should be applied to the series in levels. In econometrics, it is common practice to also log-transform variables such as the consumption of a natural resource for interpretational purposes. The interpretation of series in log-differences corresponds to the relative increase or percentage-wise growth in gas consumption. Note once again that this is a time series with an annual frequency, hence no seasonal difference operator is applied. Performing once again the Dickey-Fuller test (without trend) on the series in log-differences, the null hypothesis is rejected, thus indicating the series in log-differences is a stationary time series.

2.1 Estimation & Validation

On the basis of the correlogram (Figure 3), the series in log-differences does not look to be generated by a white noise proces. The Ljung-Box statistic confirms this belief (p<<0.05). Since neither the correlogram nor the partial correlogram implode to zero, an ARMA model might be appropriate for the series in log-differences. Specifically an ARMA(3,1) seems plausible since the correlogram (Figure 3) and the partial correlogram (Figure 4) show significant values at lag 1 to 3 and lag 1 respectively.

Figure 3: *Correlogram log(Gas) - I(1) &* ***Figure 4***: *Partial Correlogram log(Gas) - I(1)*



The residual plot and the residual correlorgram do correspond with a model satisfying white noise residuals. Furthermore, the p-value (=0.12) of the Box-Ljung test confirms this visual validation. Hence this is a valid model. For this model, the AR(2) term is not significantly different from 0, as 0 lies within the confidence interval for the corresponding coefficient estimate. The coefficients estimates for the AR(1), AR(3) and the MA(1) term on the other hand, are considered significantly different from 0. The confidence intervals are provided in Table 1 below.

- Confidence interval = [Estimate - 1.96 SE ; Estimate + 1.96 SE]

Table 1: *Confidence intervals and estimates for ARMA(3,1)*

Term	Lower Bound	Estimate	Upper Bound
AR(1)	0.68	1.23	1.77
AR(2)	-0.48	0.28	1.04
AR(3)	-0.96	-0.53	-0.11
MA(1)	-1.26	-0.76	-0.26

Since one can only drop the highest order term from a proposed model, it is not possible to exclude just the insignificant AR(2) term in this case. An ARMA(1,1) could however be an alternative model. When evaluating this ARMA(1,1), both the AR(1) and the MA(1) are considered to be significantly different from zero as provided in table 2 below. Once more, this model still has to be verified since the residuals are supposed to be white noise. Again the residual plot and residual correlogram correspond to white noise residuals, which is again in line with the Ljung-Box test (p=0.93).

***Table 2**: Confidence intervals and estimates for ARMA(1,1)*

Term	Lower Bound	Estimate	Upper Bound
AR(1)	0.86	0.96	1.07
MA(1)	-0.63	-0.33	-0.03

2.2 Model Comparison

The information criteria provided in table 3, namely AIC and SIC, are both lower for the ARMA(1,1) model compared to the ARMA(3,1) model. This is not surprising as the ARMA(3,1) contains two additional terms, for which it is penalised. However, when comparing both models in terms of their mean absolute error, the ARMA(3,1) has a slightly lower measure. Yet, the Diebold-Mariano test is not rejected, so in fact there is no significant difference in forecasting performance (p>.05). As these models perform equally well with respect to forecasting, the ARMA(1,1) is still preferred since it is the more parsimonious model based on the information criteria.

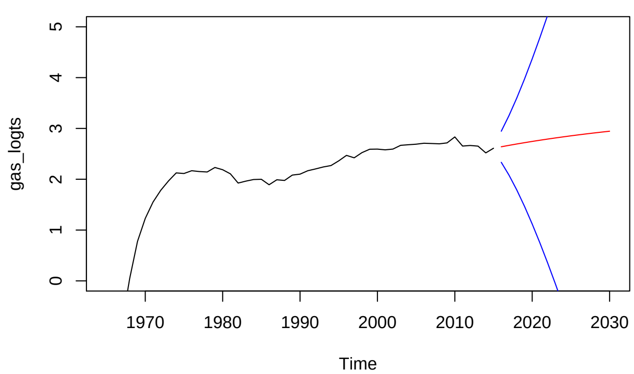
***Table 3**: Model Comparison:*

.	ARIMA(3,1)	ARIMA(1,1)
AIC	-34.93	-35.93
SIC	-25.08	-30.02
MAE	0.07	0.08

2.3 Forecast

The forecasts based on the ARMA(1,1) for the next 15 years are visualised in figure 5, along with their 95% prediction intervals. For the coming years, an upward trend in natural gas consumption is expected on the basis of this model. However, the prediction intervals widen very quickly as the forecast horizon expands. Hence, this model is quite unreliable to forecast far ahead in time.

***Figure 5**: Forecast - ARIMA(3,1,1)*



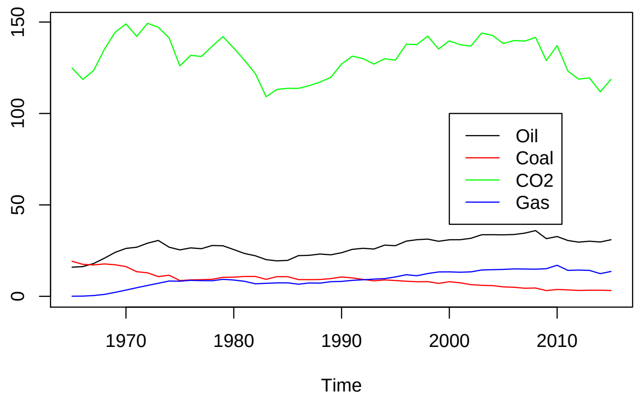
2.4 Conclusion

The Belgian natural gas consumption is stochastic proces characterized mainly by the gas consumption in the preceding year, as well as by an upward moving average. Although rather unreliable, the consumption is expected te keep increasing over the course of the next 15 years.

3 Mulitvariate Time Series Analysis

Now, also the Belgian consumption of the other national resources will be considered, namely of oil and coal. The objective is to be able to explain the variabilty in CO2 as well as to examine a possible cointegrating relationship between CO2 and the natural resource consumptions. Note that British Petroleum has measured CO2 in such a way that it only reflects carbon emissions through the consumption of oil, gas and coal. Oil consumption and CO2 units are both expressed in millions of tonnes, while natural gas and coal consumption are measured as millions of tonnes oil equivalent.

Figure 6: *Multiple Graph*



Unlike the upward trending oil and gas consumption, coal consumption has a decreasing trend over time. With regards to CO2 and the consumption of oil and coal, the unit root test with trend is not rejected for either of these time series. Hence, these are not non-stationary time series with a stochastic trend. Therefore, the difference operator will be applied to oil consumption, coal consumption and CO2 emissions. For the aforementioned interpretational purposes, also the log-transformation will again be used on each of these series. Note that the gas consumption has already been analysed in section 2. After transforming and differencing the three series in levels, they can be considered stationary as the unit root test can be rejected for all three of them ($p << 0.05$). This is important for modelling purposes, as regressing time series on one another tends to lead to spurious regression inference since it involves processes driven by time.

3.1 Linear Regression

The proposed model regresses CO2 on coal, oil and gas consumption on each other (all in log-differences). This corresponds with the following equation:

$$\Delta Log(CO2) = \beta 0 + \beta 1 \Delta log(Oil) + \beta 2 \Delta log(Gas) + \beta 3 \Delta log(Coal)$$

As the residuals are found to be white noise, this can be considered a valid model. Unlike gas consumption ($p=0.785$), both coal and oil consumption are considered statistically significant predictors ($p << 0.05$). They are also reported to be jointly significant by the F-test ($p << 0.05$). Together, they can account for about 80% of the variability in CO2 at a given point in time.

Table 4: *Model summary*

	Estimate	SE	t value	Pr(>t)
Intercept	0.000	0.003	-0.16	0.88
log(Coal)	0.22	0.032	6.80	0.00 ***
log(Gas)	0.004	0.014	0.27	0.79
log(Oil)	0.53	0.061	8.66	0.00 ***

- On average, and all else equal, if coal consumption increases by 1 percentage point, CO2 increases by 0.22 percentage points at a given point in time
- On average, and all else equal, if oil consumption increases by 1 percentage point, CO2 increases by 0.53 percentage points at a given point in time

Yet, it is important to know if there is a cointegrating relationship between CO2, coal, gas and oil consumption. The standard errors of OLS regression need to undergo a Newey-West correction if the series are not cointegrated.

3.2 Cointegration - Johansen Procedure

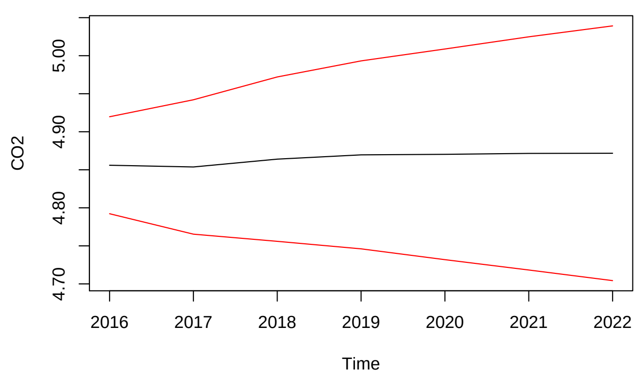
To perfom a Johansen test for cointegration, automated lag selection is used to determine the appropriate amount of lags, based on the SIC. Depending on the maximum amount of lags however, the SIC was either 1 or it produced irregularities such as singular matrices or infinities. The AIC on the other hand, was more consistent and suggested a lag order of 2 with most of the initial maximum lag lengths. From the test procedure based on the trace method, it can be deducted that there are at least 2 cointegration relationships at a 5% significance level. Newey-West corrections need not be applied to the standard errors of the linear regression adressed above.

The following cointegrating relationships are estimated:

$$Log(CO2) - 1.47log(Oil) + 0.73log(Gas) + 0.002log(Coal) - 1.72 = \delta 1$$
$$Log(CO2) - 0.49log(Oil) + 0.01log(Gas) + 0.12log(Coal) - 3.03 = \delta 2$$

Based on these equations, a Vector Error Correction Model (VECM) is estimated. Next, a VAR model is built, for which the associated 7 step-ahead forecasts are illustrated in Figure 7. Slight increases in carbon emissions are expected over the coming 7 years. Like with the natural gas predictions, the predition intervals are once again quickly widening over time.

***Figure 7**: 7-Step ahead forecast for CO2*



3.3 Conclusion

Based on this analysis, there are two long run equilibrium relationships between carbon emissions on one hand and fossil fuel consumption on the other hand. The prediction for carbon emissions in the coming years is stable. A very slight increase is expected however. Compared with the projections of institutions such as the OECD, who predict steep increases in CO2, this is still very a optimistic forecast. Furthermore, there is also a highly impactfull relationship between these variables, regardless of a time-driven dependence. Increases in both coal consumption and oil consumption have been found to go together with increases in carbon emissions. This is consistent with prior findings regarding the harmful nature of fossil fuel consumption on the ecosystem as measured by carbon emissions.