Katholieke Universiteit Leuven (KUL)

Advanced Time Series Analysis (D0M63A)

Project

Name: Amalesh Vemula
Student ID: r0691839

Professor:
Christophe Croux

Introduction:

The data used shows how the GDP(in M.£) of UK has developed on a quarterly basis from 1988 until the 3rd quarter of 2018 and CPI (consumer price index) of UK was observed on a quarterly basis from 1988 until the 3rd quarter of 2018 (fig.1). The data of GDP of UK and CPI of UK is taken from Office for National Statistics UK website. Dataset consists of 123 observations.

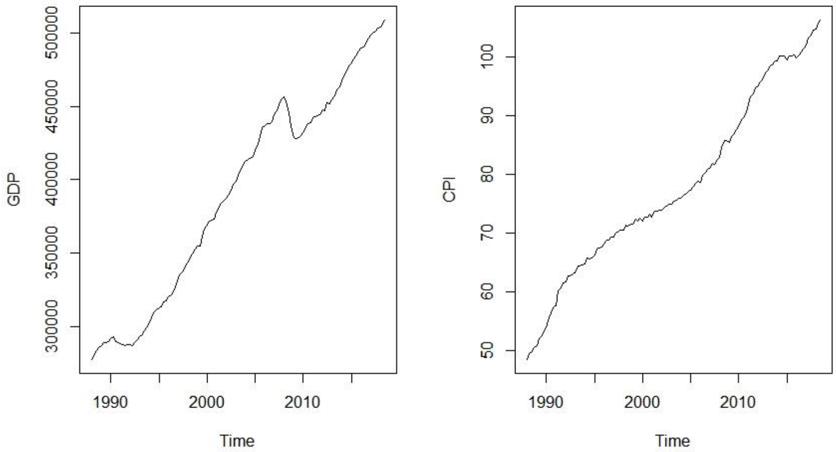


Figure 1: Time series plot of GDP and CPI over a period of 1987-2018

Univariate Analysis:

Firstly, we checked if linear regression of GDP time series on the index of observation and quarter data to see if it's fitted well. From the summary of the model, we see the high coefficient of determination of 0.968 (adjusted R-squared) which means that the variance of 96.8% is explained by the regressors and none of the quarters make a significant effect on the value of GDP. From the residual plots, we see model violates homoscedasticity. We confirm that the model is not valid.

Fig.1 clearly shows that the GDP has a positive trend in the long run as a whole. The time series looks not a stationary series and to confirm we run augmented Dickey-Fuller test. The ADF test result has the p-value of 0.872. This indicates that we cannot reject the null-hypothesis and series is non-stationary time series. Then we've gone for the transformations to check if stationary can be achieved. Log-transformed the GDP then checked for the ADF test still we cannot reject the null hypothesis (p-value of 0.818), so non-stationary. We have then used difference operator to remove the trend and to check for the seasonality in the data. From the figure x, we can see that the there is no seasonality. Then performed augmented DF test and we reject the null hypothesis as the p-value is 3.335e-05. Thus, we conclude that the time series is stationary, and take the assumption that the time series doesn't have significant seasonality.

Then performed autocorrelation function and partial autocorrelation function of the stationary series. We can see that there are some significant correlations. From the ACF, we use to determine the number of MA terms and we can see that there are significant lags until three lags (fig.2). From the PACF, we use to determine the number of AR terms, which shows that most of the correlations are not significant and we normally ignore the last one to keep our model more parsimonious.

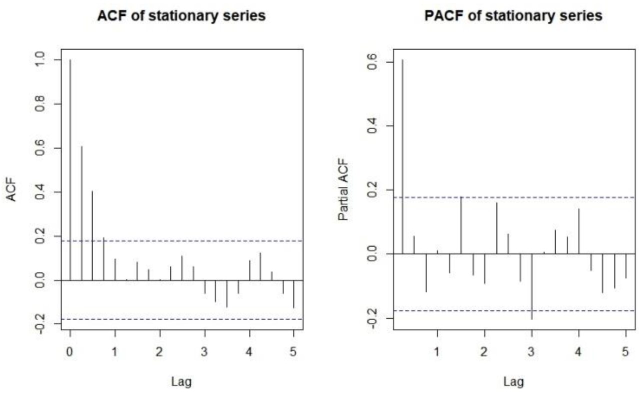


Figure 2

From the insights that we have received from the ACF and PACF plots of the stationary series we proceeded by an ARMA(0,1) model and checked the validity by Q-test for residuals, the Q-test rejects the hypothesis that the residuals are white noise. Hence, we have checked for the ARMA(0,2) and ARMA(0,3) model where the Q-test for the residuals has been validated but from the PACF plots, they indicated seasonal effects by quarter (fig. appendix). Hence, We considered to model by using 1st and 2nd SMA order (seasonal moving average). From table 1, we can see that the models 2,3,4,5 are all valid models but considering that the seasonal effects, we consider the SARIMA model. Model 5 looks better model having the less BIC value, and will be used for the forecast.

#	ARIMA	AIC	BIC	P-value	validation
1	(0,1,1)	-896.67	-891.04	0	No
2	(0,1,2)	-925.29	-916.85	0.25	yes
3	(0,1,3)	-934.49	-923.24	0.35	yes
4	(0,1,2)(0,0,1)	-928.47	-917.21	0.401	yes
5	(0,1,3)(0,0,1)	-947.32	-933.25	0.788	yes

Table 1: SARIMA and ARIMA Models overview

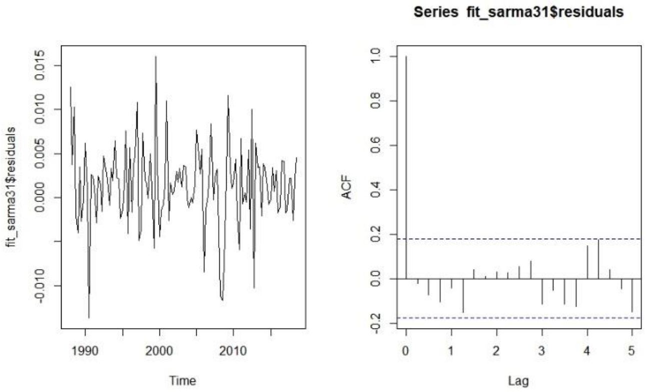


Figure 3: Plot of the residuals and ACF for the final fitted model

Forecasting:

The result of the forecast for the SARIMA (0,1,3)(0,0,1) model can be seen in fig.4. We can see the predictions in (Appendix Output 1) which shows that the GDP fluctuates around £511,095m. The residuals by the forecasts were then summarized by the Root mean square error (RMSE) and Mean absolute error (MAE). Table 2 shows the values for the different forecasting error measurements, which are very small and is chosen as the optimal one for the forecast.

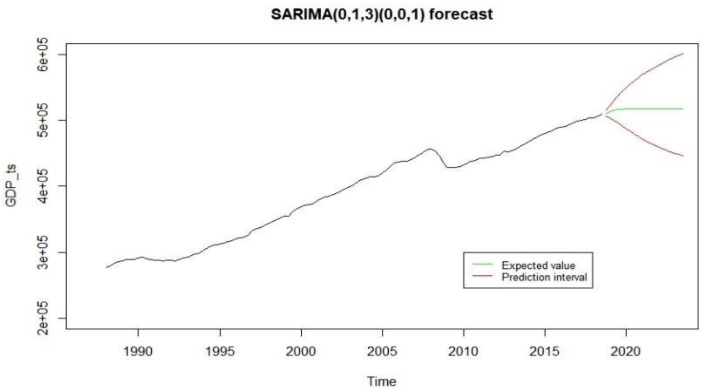


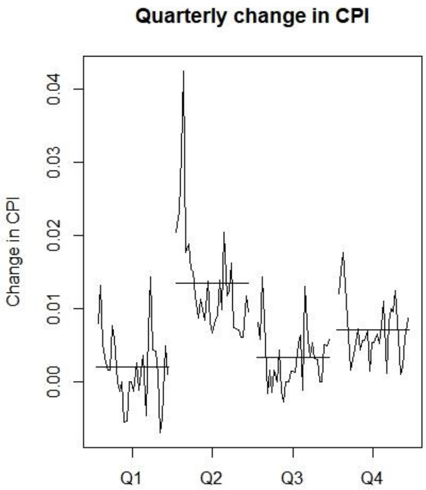
Fig. 4: Plot of the Stepahead forecast of SARMA(0,1,3)(0,0,1)

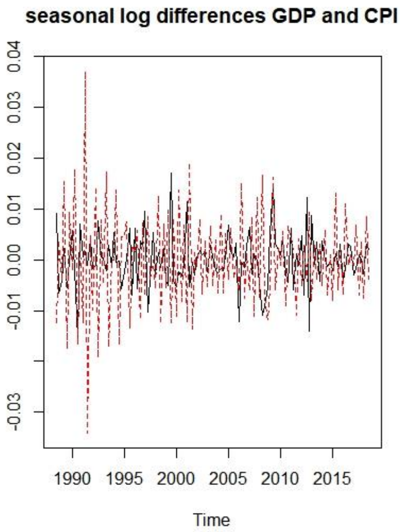
Measure	Value
RMSE	0.004606549
MAE	0.003720946

Table 2: Forecast errors for the pseudo-out-of-sample expanding window approach

Multivariate Analysis:

Firstly, we've seen that the CPI (consumer price index) is not stationary and there is a need for transformation and then checked for being stationary then decided to go with differences and from the quarterly plot, we can see that there is seasonality in quarter 2. Then transform the result to the seasonal differences which are stationary by using the unit root test.





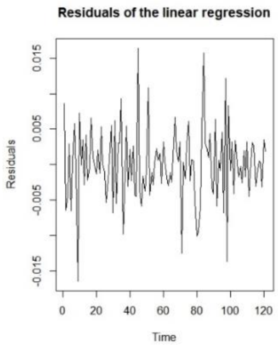
Linear model:

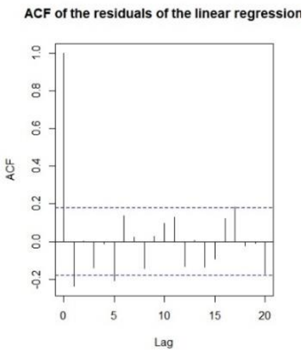
Firstly, We checked for linear regression analysis on the stationary series taking seasonal differences of the logCPI explanatory variable for seasonal differences of logGDP. The results are as follows

	Estimate	Std. Error	t value	Pr(>|t|)
(intercept)	0	0.0005	-0.01	0.991
dslogCPI	-0.0521	0.0476	-1.1	0.2756

Table 3: Summary of the Linear model

From the linear model shows that the effect is not significant and the Adjusted R-squared indicates that CPI alone explains the variance of GDP of only 0.1%. From fig. 6 we can see that the residuals plot which looks stationary and there are some significant autocorrelations but from the Q-test for the residuals, confirms that the residuals are not white noise (p-value<<0). This means that the linear regression model cannot be validated.





Distributed Lag Models and Autoregressive Distributed Lag Models:

Considering that the CPI, is a predictor for GDP but also GDP might also be influenced by the lagged values of CPI. Considering to model the data by using the distributed lag models. DL(1), DL(2) models were run to check for validity but none of both models are validated using the Q-test(Table 4), all the residuals of the models are not white noise. Thus confirming that the models are not valid.

Autoregressive Distributed Lag-models(ADL) are further considered. In these models, we also take lagged values of the GDP or the dependent variable also taken into consideration for predicting the current values of GDP.

Models of ADL(1) and ADL(2) are both validated (Table 4) by checking with the Ljung-Box test showing that the residuals are white noise. Although the ADL(2) model explains only 5 percent of variance deduced from the Adjusted R-squared value. We choose ADL(2) model could be the better choice to explain the variation of GDP compared to ADL(1) as there are no significant correlations for the residuals.

Model	Sign. lag in regression?	Sign. residuals?	Q-test p-value	White noise
DL(1)	No	>5 lags	2.40E-02	No
DL(2)	No	>5 lags	1.10E-02	No
ADL(1)	Lag 1	<5 lags	0.07	Yes
ADL(2)	Lag1 and 2	None	0.062	Yes

Table 4: Overview of (Autoregressive) Distributed lag models.

Granger Causality:

From the ADL models, we can check for the Granger causality. When the predictor variable has incremental predictive power with regards to the response then the predictor variable is said to be granger cause a response variable. To test for Granger causality two models that are considered are the model having both the autoregressive lags and lags of the predictor variable, in our case we can consider ADL(2) and the model having normal regression model.

We check for the null hypothesis stating that there is no Granger causality.

Analysis of Variance Table gives that the p-value is 0.3567 (appendix fig A). This shows that there is no evidence for Granger causality from CPI to GDP.

```
Analysis of Variance Table

Model 1: dslogGDP.0 ~ dslogGDP.1 + dslogGDP.2 + dslogCPI.0 + dslogCPI.1 +
    dslogCPI.2
Model 2: dslogGDP.0 ~ dslogGDP.1
  Res.Df       RSS Df    Sum of Sq      F Pr(>F)
1    113 0.0028531
2    117 0.0029649 -4 -0.00011183 1.1073 0.3567
```

Fig. 5: Output of ANOVA

Cointegration: As both series are integrated, we can check for cointegration. We can check for cointegration using the Engle-Grander test and Johansen test.

Engle-Granger Test:

This is a residual based test that uses regression techniques. Now we test GDP and CPI for cointegration using Engle-Granger test. The series is shown to be the order of integration to be one. As there are only two series, the maximum number of CEs is one. As the Engle-Granger is not symmetric one we need to regress GDP on CPI and regress CPI on GDP. ADF test on the residuals for regressing on GDP on CPI gave a p-value of 0.26 and for regressing, CPI on GDP is given a p-value of 0.604. As the null hypothesis is such that there is no cointegration. Therefore, we cannot reject the null hypothesis and there is no evidence for cointegration by Engle-granger test.

Johansen Test:

Johansen test is a symmetric test and the test equation is as follows

$$\Delta \vec{y}_t = \Pi \vec{y}_{t-1} + A_1 \Delta \vec{y}_{t-1} + \ldots + A_{p-1} \Delta \vec{y}_{t-(p-1)} + \vec{u}_t .$$

Such that H0: rank(π)=0 → no cointegration, Ha: rank(π)>0 → At least one cointegration equation. The maximum number of cointegration equation is equal to k-1 where k is a number of time series. So a number of cointegration equations in our case is between 1 and k-1 and Vector error correcting model to be used. VAR(p) model is fitted to Yt and we use VARselect to do this procedure by taking the season as 4 as it's a 4quarterly data. The output is given as

AIC(n)	HQ(n)	SC(n)	FPE(n)
5	2	2	5

We will be using SC(BIC) as a criterion, and the lowest value of SC is 2. This is the number of lags to be used in the Trace test and max eigenvalue test, which can give clarity to whether the two series are cointegrated or not.

.	test	10pct	5pct	1pct
r <= 1 |	4.60	7.52	9.24	12.97
r = 0 |	53.44	17.85	19.96	24.60

Output for Trace test

	test	10pct	5pct	1pct
r <= 1 |	4.60	7.52	9.24	12.97
r = 0 |	48.84	13.75	15.67	20.20

Output for Max Eigen test

From the trace test output, we will be rejecting the null hypothesis that the number of cointegration equations is equal to 0 at all the levels and its same as for the max eigen test. We fail to reject the null hypothesis that the number of cointegration equations equal to 1 at all levels in both the trace test and max eigen test. Therefore, we can conclude that there is a single cointegration equation and thus the series are cointegrated. As Johansen test is more powerful and symmetric, we consider the result from Johansen test even though there is a contrast in results between the Engle-Granger test and Johansen test.

Vector Error Correction Model:

VECM model considers cointegration relationship by adding error corrections to the multivariate model such that the model shows short run and a long run relationship between series that are cointegrated. From the output of VECM(1) model gives that the error correcting terms for GDP is -0.00165 and CPI is -0.00192 (Appendix Output 2). As both the error terms are negative which means that they converge to a long-run equilibrium.

Forecast:

From the 8-step-ahead forecast of logarithms of GDP and CPI based on the VECM(1) model. We first retransform the VECM to a VAR specification in levels and the resultant, red lines show the prediction intervals and the green line shows a forecast of the GDP and CPI. Values of the predictions are provided in (Appendix Output 3). The output shows that the GDP fluctuate around £510,456 m and CPI 106.93. Due to the increase in uncertainty, there is a widening of intervals over time. The forecast of the log GDP and log CPU are shown in the fig.6.

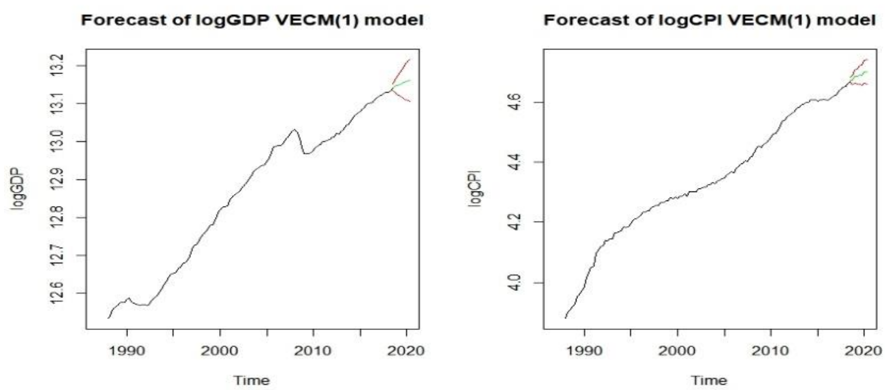


Fig.6. 6-step-ahead forecast of log(GDP) and log(CPI) based on VECM(1)

Conclusion:

In this paper both a univariate analysis on GDP of UK and multivariate analysis on GDP and CPI of UK. In the univariate analysis, the first differences in GDP resulted in the stationary time series and were fitted by a SARIMA (0,1,3)(0,0,1) model. This model was validated and proved forecasting values with an error of less than 1%. For the multivariate analysis, GDP and CPI were proved to be having an order of integration 1, and several models were fitted on these time series. The series turned out to have no cointegration according to Engle-Granger test but contradicted results were given by Johansen test then considered the latter as its more powerful. VECM model showed that the series GDP and CPI would be reaching a long-run equilibrium.

Appendix :

Outputs:

```
$`pred`
         Qtr1     Qtr2     Qtr3     Qtr4
2018                            13.14431
2019 13.14853 13.15202 13.15396 13.15514
2020 13.15672 13.15758 13.15758 13.15758
2021 13.15758 13.15758 13.15758 13.15758
2022 13.15758 13.15758 13.15758 13.15758
2023 13.15758 13.15758 13.15758

$se
            Qtr1        Qtr2        Qtr3        Qtr4
2018                                     0.004759968
2019 0.009256155 0.014609536 0.020074169 0.025362503
2020 0.030344327 0.035222731 0.039917597 0.044115625
2021 0.047947494 0.051495009 0.054813409 0.057942071
2022 0.060910241 0.063740342 0.066450019 0.069053449
2023 0.071562229 0.073985989 0.076332826
```

Output 1: Predictions of logGDP

```
Call:
lm(formula = substitute(form1), data = data.mat)

Coefficients:
            logGDP.d    logCPI.d
ect1        -0.0016506  -0.0019254
sd1          0.0024621  -0.0065047
sd2         -0.0005406   0.0067799
sd3          0.0041886  -0.0076087
logGDP.dl1   0.5455579  -0.1269107
logCPI.dl1  -0.2836165   0.3642297


$beta
                ect1
logGDP.l1   1.000000
logCPI.l1   2.075046
constant  -24.298642
```

Output 2: VECM summary

```
$`logGDP`
         fcst    lower    upper          CI
[1,] 13.14306 13.13446 13.15167 0.008605062
[2,] 13.14615 13.12992 13.16237 0.016223198
[3,] 13.14817 13.12447 13.17188 0.023703763
[4,] 13.15118 13.12036 13.18200 0.030822229
[5,] 13.15344 13.11595 13.19093 0.037488556
[6,] 13.15667 13.11298 13.20037 0.043691924
[7,] 13.15888 13.10942 13.20834 0.049459083
[8,] 13.16200 13.10717 13.21683 0.054831377

$logCPI
         fcst    lower    upper          CI
[1,] 4.672229 4.663258 4.681200 0.008970927
[2,] 4.672209 4.656945 4.687473 0.015263898
[3,] 4.683230 4.662510 4.703949 0.020719286
[4,] 4.683969 4.658402 4.709537 0.025567273
[5,] 4.688440 4.658497 4.718384 0.029943442
[6,] 4.687838 4.653902 4.721775 0.033936359
[7,] 4.698545 4.660935 4.736155 0.037609888
[8,] 4.699066 4.658053 4.740079 0.041013243
```

Output 3: VECM stepahead forecast prediction intervals