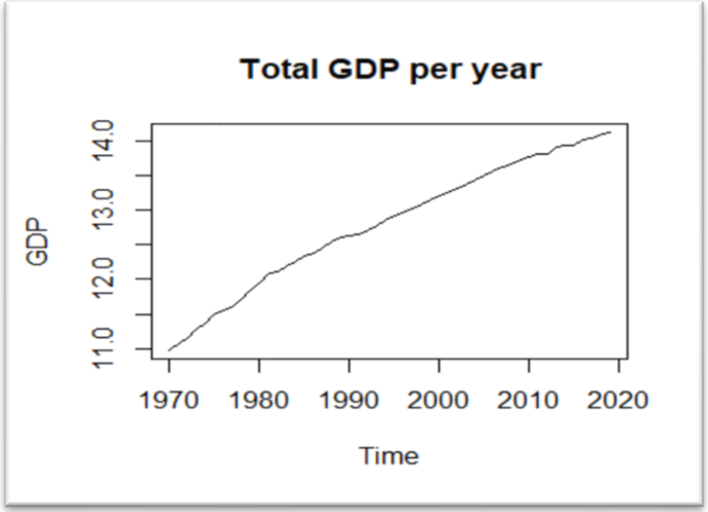
Advanced Time series Analysis

Prerna xxx (r0825809)

prerna.xxx@student.kuleuven.be

Introduction of Univariate Analysis

- The data used shows how the GDP of Australia has developed over 50 years. The dataset starts with the year 1970 and ends at year 2019.
- It is clear that this series has a positive trend, as it increases in the long run.
- R considers GDP to be a typical variable, so we explicitly declare it to be a time series.
- As we deal with GDP value, we construct the time series in log differences. Difference is used to convert time series into stationary. It become stationary after taking log differences.



Data source: https://data.oecd.org/

Univariate analysis

- In the ACF, which we use to determine the number of MA terms, we see a borderline significant lag at lag 1 and a significant lags at lag 2 and 3.
- The PACF, which is used to determine the number of AR terms, shows no significant lags.

Order	AIC	BIC	P value	Validated using acf and pacf	MAE	MAPE
1,1,1	-206.13	-200.4	0.34	Yes	0.019	0.154
1,1,2	-205.19	-197.62	0.99	Yes	0.019	0.155
1,1,3	-204	-194.87	0.85	Yes	0.020	0.159

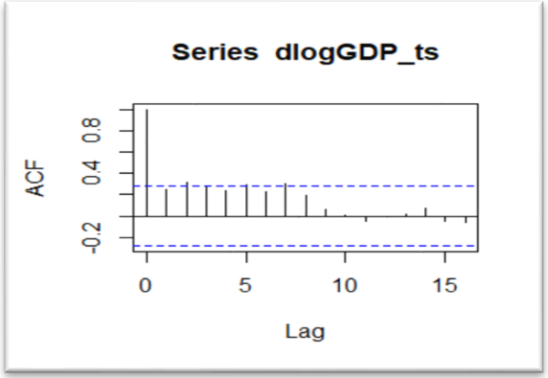


Figure 2: ACF

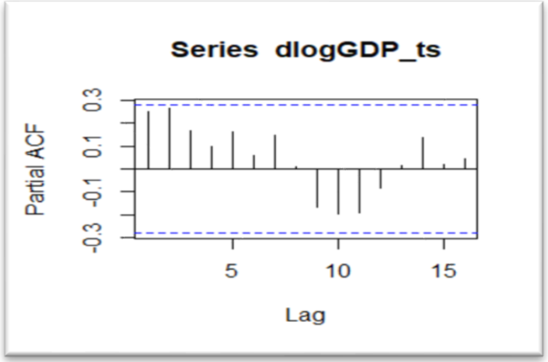


Figure 3: PACF

Forecast and forecast errors

- The result of the 8 step ahead forecast based on the ARIMA (1,1,2) model can be seen in figure 4. Blue lines represent the 95% Confidence Interval.
- Diebold marino test suggests that forecast performance of these 2 models using absolute value loss is not different, for squared value loss it is different for horizon=2.

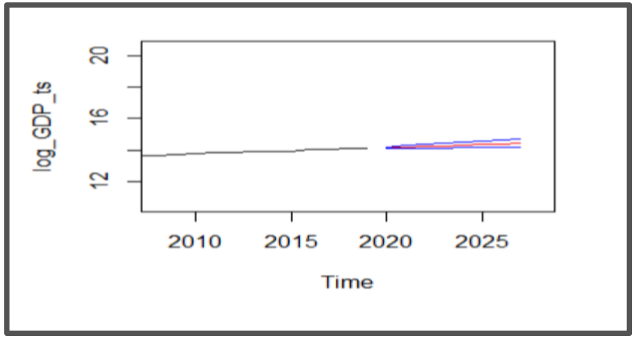
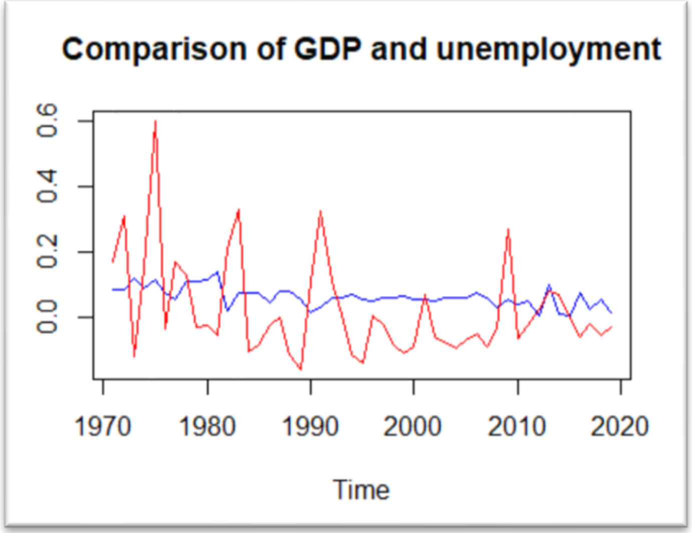


Figure 4: Forecast

	ARIMA(1,1,1) Horizon=1	ARIMA(1,1,2) Horizon=1	ARIMA(1,1,1) Horizon=2	ARIMA(1,1,2) Horizon=2
MAE	0.028	0.028	0.0263	0.0258
MAPE	0.0020	0.026	0.0018	0.02378

Multivariate analysis: Linear regression

- For the multivariate analysis the data on Australia's GDP is complemented with data on Australia's Unemployment data from the same period.
- **We aim to check if GDP is affected by the unemployment data**.
- As these time series are now stationary, we can run a standard regression analysis. We do not get any significant results. We need to pay attention that time is not included in the linear regression model.
- To include the time, we use distributed lag models. Distributed Lag Models provide a way to model for time series data in which we use a regression equation to predict current values of a GDP based on both the current values and the lagged values of unemployment.



Distributed Lag model and Granger causality

- Here a DL(4) model was run. While some of the results initially looked promising, the model could not be validated. This means that using the Q-test, the hypotheses that the residuals are white noise had to be rejected for the model.
- Then we used an Autoregressive DL(3). Using the Q test this model is validated. When we look at the results of the fit, overall p value is less than 0.5, but individual p values of lagged variable is not. Only one variable's individual p value<0.05. This model is also rejected.
- We test for granger causality using ADLM(1). A predictor variable is said to Granger cause a response variable when that predictor variable has incremental predictive power with regards to the response. After doing the test, we get the p value >0.05, so we cannot reject H0 of **no granger causality**.

Cointegration

- We check for cointegration using the Engle-Granger test, which is a residual-based test that uses regression techniques and the Johansen test. The Johansen test is preferred as it is a more powerful and symmetric test. Still, as in this scenario both tests give contradictory results, both are included. As there are only two series, the maximum number of CEs is one.
- **Engle-Granger Test:** The ADF test on the residuals obtained from regressing(Yt ~Xt) gives a test-statistic of −0.62. As we have only one predictor, the critical value at the 10%-level is -3.12. Even at this level we cannot reject the null hypothesis, meaning there is **no evidence for cointegration of these series according to this test.**
- **Johansen Test:** We can see that in this scenario, we get the value of SC to be 2. This is the number of lags that is used in the Trace test and the Max Eigenvalue test, which will show whether or not our two series are cointegrated.

	test	10pct	5pct	1pct
r <= 1 |	9.36	7.52	9.24	12.97
r = 0 |	61.65	17.85	19.96	24.60

Trace

	test	10pct	5pct	1pct
r <= 1 |	9.36	7.52	9.24	12.97
r = 0 |	52.29	13.75	15.67	20.20

Eigen

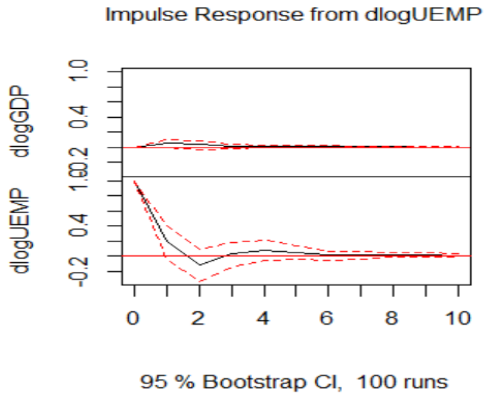
- As per Johansen test, we reject the null hypothesis that the number of Cointegration Equations is equal to 0 at all levels and cannot reject that **there is 1 Cointegration Equation at 1 percent level.**

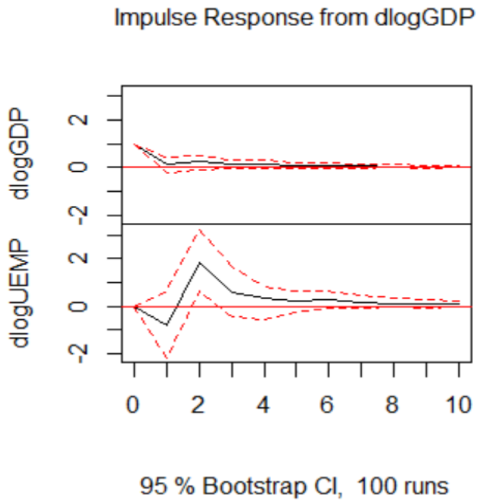
VECM and VAR

- We conclude from the tests that a single Cointegration Equation is present, and that the series are thus cointegrated. As a result, a Vector Error Correction Model is run.
- VECM determine the speed of adjustment towards the equilibrium equations and as such we would expect these elements to have negative values. However, in this scenario the **error correcting terms for both GDP (1.0) and Unemployment (0.90) are positive.**
- This means that neither of these converge to a long-rum equilibrium, but rather they will diverge from a long-run equilibrium. This makes the process inherently instable.
- As the Error Correcting Term is giving an unexpected positive value, we have a look at Vector Autoregression (VAR) models next.
- To help choose a VAR model the function VARselect is used.

Impulse Response Function

- Results of VAR(2) model show that 20% of the variance of dlogGDP can be explained by lagged values of dlogUEMP, and
- Figures on the right represent the impulse response functions based on VAR(2) estimates.
- Given a unitary impulse in dlogUEMP at time t, we observe a significant negative response at time t+2.
- Given a unitary impulse in dlogGDP at time t, we observe pretty much no significant response.





Forecast and Conclusion

- The result of the 8 step ahead forecast based on the VAR model can be seen in figure. Red lines represent the 95% Confidence Interval.

- **Conclusion:**

1. ARIMA(1,1,1) is the best model according to AIC and BIC for univariate modelling.
2. For multivariate analysis we can see from the impulse response function that change in Unemployment can affect GDP for a short time. **We see convergence to the equilibrium after a few time periods in both the impulse response functions, meaning there is no long-term effect.**