Hand-in Assignment 1

1. Introduction

This paper examines the quarterly German GDP for the period from 1991 to the first quarter of 2009, where the real economic impact of the financial crisis was at its worst. Based on a univariate model for this period this paper constructs a forecast describing the expected recovery from 2009 (2) and the development until 2019 (4) and compare this forecast to the actual data. Forecasts of future economic models play an important role in economic policy and are produced by large institutions on a regular basis which shows the importance of the forecasts. This paper predicts the forecast based on the autoregressive (AR) model thus that the model's output depends on its own precious values. To support the AR-model a certain number of lags is implemented. These lags can on the other side make the AR-model prediction less reliable because it may cause the model to convert to the unconditional mean, ignoring significant economics fluctuations. Less information is required for the AR-model which makes it much more reliable useful in short term.

2. Description of data

The dataset contains quarterly data for the real gross domestic product (GDP[1]) in Germany for the period of the first quarter of 1991 to the second quarter of 2022. The time series data are from Reserve Bank of St. Louis. We define the following variables:

$$\log(GDP) = \log\,(GDP_t)$$

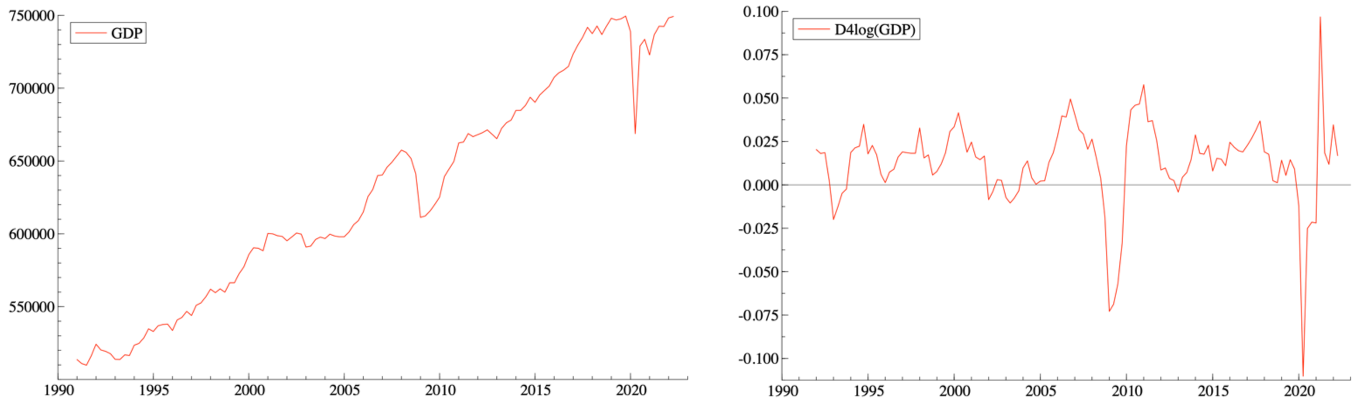
$$D4\log(GDP) = \Delta_4\log(GDP_t) = \log(GDP_t) - \log\,(GDP_{t-4}),$$

where D4log(GDP) is the yearly growth rate in log-percent. To work with the GDP the relative changes are needed and is a achieved by taking the logarithmic transformation of the real GDP. In figure 1 the actual real GDP is plotted which shows a non-stationary trend which is lead to a non-consistent OLS-estimator. For the OLS to obtain consistency both stationarity and weak dependency is necessary. Weak dependence is important because it replaces the assumption of random sampling in implying that the law of large numbers and the central limit theorem hold. Weak dependency implicates that a time series' correlation/covariance asymptotically decreases to zero when the lag distance increases to infinity. To insure that this is fulfilled we will use the first difference transformation

[1] GDP Real Gross Domestic Product, millions of chained 2010 Euros.

(D4log(GDP)). With the first difference we also remove seasonal noise. Because the data are time series data, we cannot assume independent and identical distributed data but with the first differences transformation we achieved stationary data as seen in figure 1.

Figure 1: Actual quarterly GDP and quarterly GDP growth rate for Germany 1991(1)-2022(2) (T = 126).



Further in this paper we will use the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) to examine the best model for forecasting.

3. Economic theory

3.1 Model Specifications

The most suitable specified model for this data would in our case be the autoregressive model (AR) which uses the past forecasts to predict future values. We think that there would exist a seasonal pattern in the German GDP therefore we chose the AR-model. We believe that the AR model fits better than the moving average model (MA) which uses errors from the past forecasts to predict future values and not past forecasts.

The autoregressive (AR) model with p lags, AR(p), is defined by

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_t - p + \epsilon_t \ ,$$

where $e_t \sim i.i.d.(0, \sigma^2)$ and $y_{p-1}, y_{(p-2),\dots y_1, y_0}$ are observed initial values.

The lag-polynomials and its characteristic roots can be used to see if the AR(p) model is stationary.

3.2 Estimator and assumptions

The OLS-estimator would be equal to the Maximum Likelihood Estimator (MLE) if the error term is assumed to be Gaussian-distributed. By estimating the model with the OLS-estimator it leads to a condition of normality and homoscedasticity. The OLS estimator is in our instance given by:

$$\hat{\theta}_i = \frac{\sum_{t=1}^{T} \Delta_4\, y_{t-i} \Delta_4 y_t}{\sum_{t=1}^{T} \Delta_4\, y_{t-i}^2}$$

The stationarity condition is characterized by the characteristic equation for the AR(p) model defined as

$$\theta(z) = 0$$

The p solutions $z_1, \ldots z_p$ are the characteristic roots which contain information on the dynamic properties and may be complex. We now consider the AR(2) case which is our model used to forecast where the characteristic equation is

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 = 0$$

The two roots z_1 and z_2 can be used to factorize the polynomial

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 = (1 - \phi_1 z)(1 - \phi_2 z),$$

where $\phi_1 = z_1^{-1}$ and $\phi_2 = z_2^{-1}$ are the inverse roots.

The polynomial is invertible if $|\phi_1| < 1$ and $|\phi_2| < 1$ or, equivalently if $|z_1| > 1$ and $|z_2| > 1$ meaning that the stationarity condition is fulfilled i.e., that $|\phi_j| < 1, j, \ldots, p$ which leads to the polynomial converging to 0 as $T \to \infty$. That is the AR(p) model is stationary if the sum converges.

To obtain stationarity the data D4log(GDP) is differenced which de-trend the time series data shown in figure 1.

3.3 Model Formulation

In practice, the correct model is unknown. It is the researcher's job to search for regressors, lag-length and test for the necessary assumptions. This suggests a general-to-specific (GETS) principle. GETS is a method used to formulate the most simple and best model. The method removes insignificant lags and replace large residuals/outliers with dummy variables by conducting misspecifications tests. However, we can never prove that a model is correctly specified. But we can conduct theses misspecifications tests to ensure that the main dynamic properties, of interest, are accounted for.

3.4 Misspecifications test

To ensure that the main dynamic properties of the model we must test for no-autocorrelation, no-heteroskedasticity and normality. The no-autocorrelation tests test if the residuals are correlated by using the Breusch-Godfrey LM test. To test the assumption of no-heteroskedasticity we use the auxiliary regression to test if the variance of the residuals changes in the time series. The test is based on the LM statistic. To check the normality, we can plot the residuals in a histogram for the visual inspection. If a residual falls outside the expected interval, it indicates an outlier which form the basis for including an intervention dummy for outliers. The formal test is based on estimated skewness (S) and kurtosis (K).

3.5 Forecast

The ARMA models are suited for forecasting because it only requires very little data and serves as a good benchmark. We want to predict y_{T+k} given all information up to time T. I.e. given the information set

$$I_T = \{y_1, \dots, y_{T-1}, y_T\}$$

The optimal predictor is the conditional expectation where the variance of the forecast is derived as the squared forecast error

$$y_{Y+k} = E(y_{T+k}|I_T)$$

The forecast will converge towards the unconditional mean characterized as when k converges towards infinity which means the forecast is more reliable in the short run.

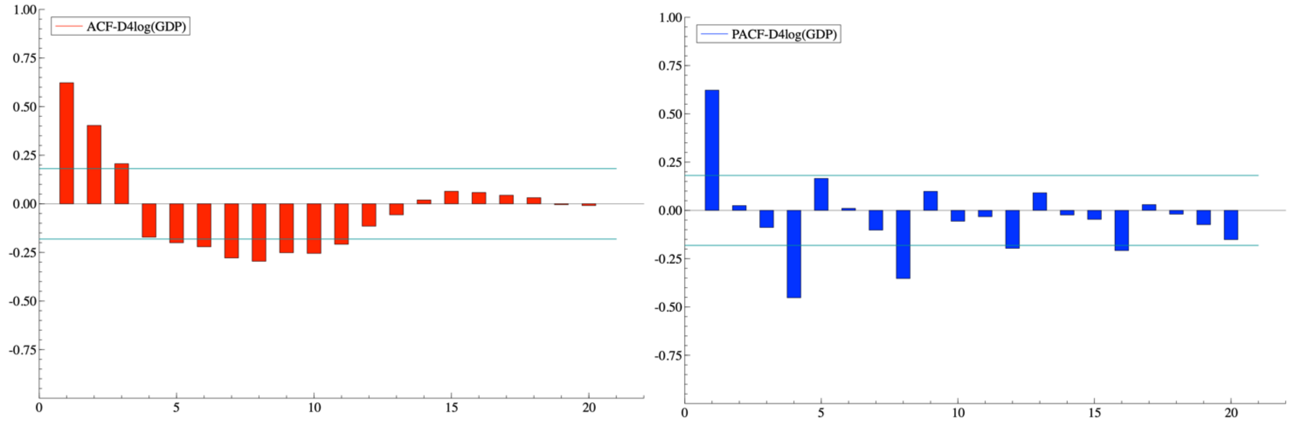
$$E(y_{T+k}|I_T) \to E[y_T] = \mu, \qquad \text{when } k \to \infty$$

4. Empirical analysis

4.1 The model

For the purpose of finding the best model and its associated lag length to use in a forecast of the annual German GDP growth between 2009 (2) and 2019 (4), we construct a ACF and a PACF plot. The plots are made with data from the sample and includes observations from the second quarter of 1993 until and including the first quarter of 2009.

Figure 2: Autocorrelation function (ACF) and partial autocorrelation function (PACF)



We will only analyze the first 5 lags in above figures because we choose our model to consist of 5 lags. First point to note from the plots is the exponential decline of the autocorrelation function (ACF) as is supports the validity of a AR- or ARMA-model and furthermore the rejection of the ARMA-model as the PACF is not exponentially declining. This brings us to the construction of an initial model of 5 lags (1), which is showcased in the appendix. The result shows that the model doesn't pass the heteroscedasticity- and normality-tests, which brings us to the introduction of using dummies for large outliers and making the standard errors robust. Furthermore, the partial autocorrelation function (PACF) is the direct effects of the lags on today's value which is more efficient to look at than the indirect effect in autocorrelation function (ACF). In the PACF graph lag one and four are statistical different from zero because the occur outside of the restriction lines (showcased in the appendix (1) as well). Thus, can lag one and four be used to gain a better and simpler model instead of including all five lags.

The new model is given as model (2) and key numbers can be found the appendix. The model now only contains two lags and one dummy variable derived by GETS method:

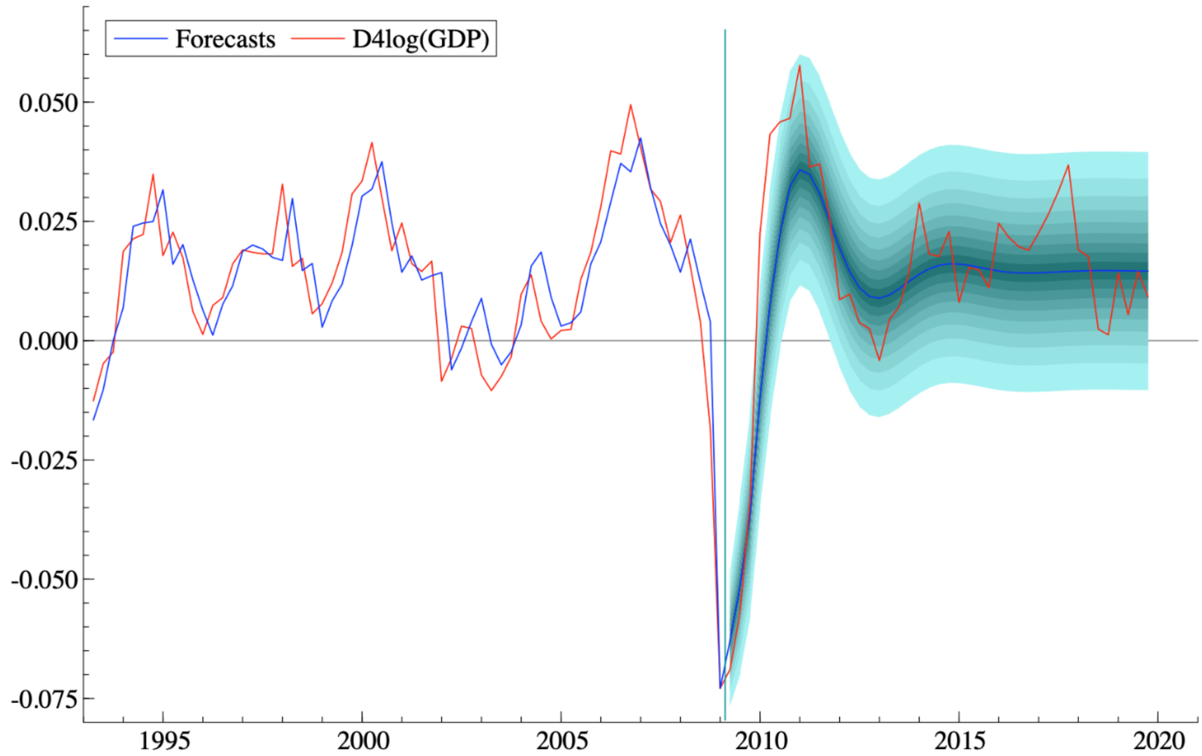
$$AR(2) = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-4} + \phi D_t + \epsilon_t$$

We note that the model passes the heteroscedasticity- and normality-test, as well as it is well specified.

4.1 Forecast

Given the conclusions in the previous section of the report we construct our forecast using the new model (2), which can be seen in figure 3. The figure shows our forecast of the German GDP-recovery from the financial crisis in comparison with the data constructed earlier in the report which was retrieved from the data sample - actual data. We find our model to be more useful for prediction in the short run (as expected) but that the general German GDP-evolution for the period falls within the restriction boundaries of our forecast. In table 7.2 our forecast values are compared with the simple given forecast values.

Figure 3: forecast of the German GDP-growth between 2009 (2) and 2019 (4)



5. Discussion

The most obvious point to make when discussion univariate time models is the oversimplifying of reality. We found in our report and forecast that the AR-model was relatively fitting in comparison with the real data, but that the long-term perspective of the German GDP-growth saw inconsistency. The problem with the univariate time models is that they contribute with little knowledge and fails to acknowledge the real evolution of the business cycles which is based on multiple explanatory variables. To use a univariate model to forecast the GDP-growth of a country is highly questionable (even though it was quite accurate in our case) especially in the long run, as the constructed data is purely based on previous data and therefore converges towards an unconditional mean as time approaches infinity. We can however argue that the univariate time models can be used as a purposeful tool when forecasting the GDP-growth in the short run if you take the right assumptions and condition into account. Solely relying on this type of forecast in any kind of business matter will provide you with inaccurate results as reality is way more complex, but the univariate time models are brilliant in terms of forecasting short-run evolution if the business climates stay true to the nature of previous collected data.

6. Conclusion

Initially we chose to use an AR(5) model as we thought it would fit the quarterly dataset best. In summary this paper argued that an AR(2) model fitted the data best based on the GETS misspecification tests. We removed tree lags and inserted an dummy variable for outlier 2009(1) seen in table 7.1. This allows the conclusion that the univariate time series AR(2) model will be the best suited model for the German GDP for the period until 2009(1) and for forecasting the recovery and later development. This conclusion follows from the fact that the forecast is more reliable in the short run than in the long run because the forecast will converge to the unconditional mean in the long run.

7. Appendix

Table 7.1 Model estimation

	(1)	(2)
Constant	0.00211 (0.961)	0.004809 (3.11)
D4log(GDP)_1	1.115 (7.15)	0.882 (12.9)
D4log(GDP)_2	−0.01514 (−0.0718)	.
D4log(GDP)_3	−0.06644 (−0.318)	.
D4log(GDP)_4	−0.4782 (−2.29)	−0.212 (−2.75)
D4log(GDP)_5	0.2504 (1.54)	.
I:2009(1)	.	−0.05577 (−19.7)
$\hat{\sigma}$	0.0107	0.008382
Log-lik.	202.719	217.281
AIC	-6.147	-6.665
HQ	-6.068	-6.612
SC/BIC	-5.945	-6.530
No autocorr. 1-5	[0.18]	[0.56]
No hetero.	[0.02]	[0.87]
Normality	[0.00]	[0.09]
T	64	64
Sample start	1993(2)	1993(2)
Sample end	2009(1)	2009(1)

Table 7.2 Comparison of forecast values.

	Simple	AR(2)
2009(2)	.015	−.0627604
2009(3)	.015	−.0513844
2009(4)	.015	−.036584
2010(1)	.015	−.0120064
2010(2)	.015	.00752538
2010(3)	.015	.0223402
2010(4)	.015	.0322688
2011(1)	.015	.0358151
2011(2)	.015	.034802
2011(3)	.015	.0307678
2011(4)	.015	.0251048
2012(1)	.015	.0193584
2012(2)	.015	.0145049
2012(3)	.015	.0110796
2012(4)	.015	.00925902
2013(1)	.015	.00887162
2013(2)	.015	.00955887
2013(3)	.015	.0108912
2013(4)	.015	.0124522
2014(1)	.015	.0139112
2014(2)	.015	.0150522
2014(3)	.015	.0157761
2014(4)	.015	.0160837
2015(1)	.015	.0160456
2015(2)	.015	.0157701
2015(3)	.015	.0153737
2015(4)	.015	.0149589
2016(1)	.015	.0146011
2016(2)	.015	.0143439
2016(3)	.015	.0142011
2016(4)	.015	.0141631
2017(1)	.015	.0142055
2017(2)	.015	.0142973
2017(3)	.015	.0144086
2017(4)	.015	.0145149
2018(1)	.015	.0145996
2018(2)	.015	.0146548
2018(3)	.015	.0146799
2018(4)	.015	.0146795
2019(1)	.015	.0146613
2019(2)	.015	.0146334
2019(3)	.015	.0146036
2019(4)	.015	.0145773