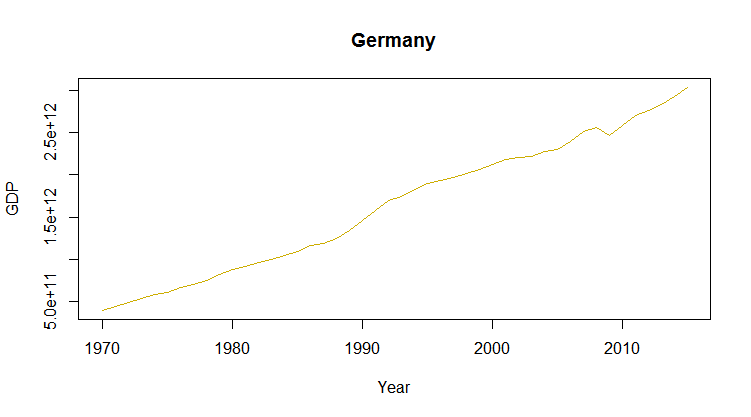
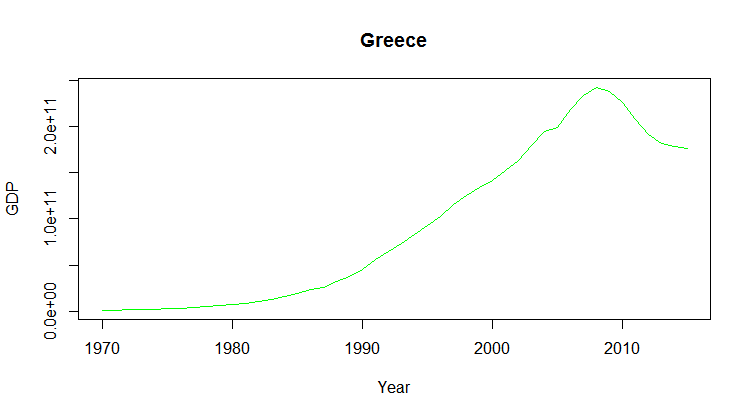
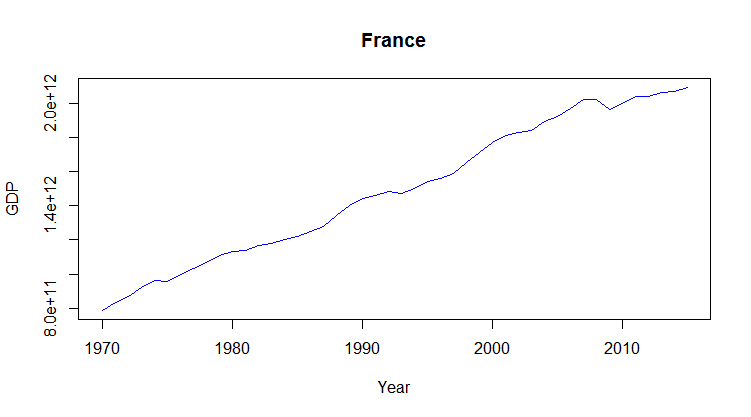
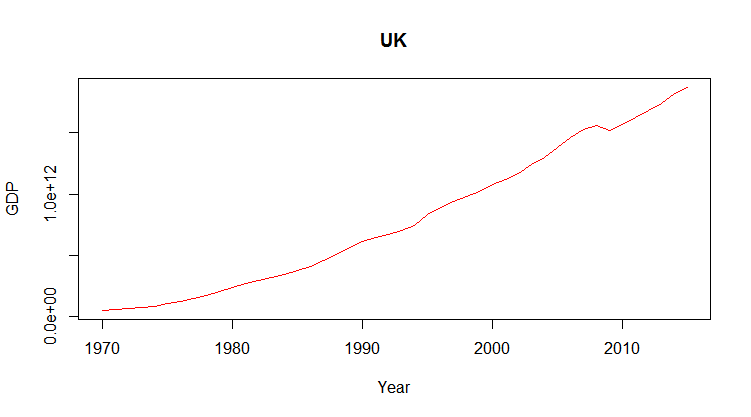
Comparing Accuracy of Naïve, ARIMA, and OECD Forecasts of European Countries from 2006 to 2015

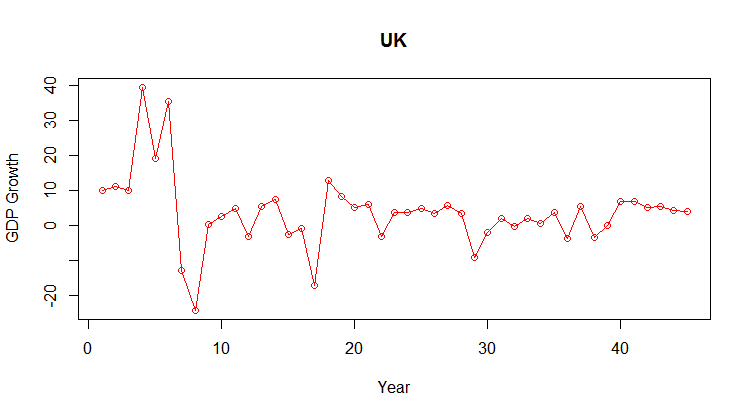
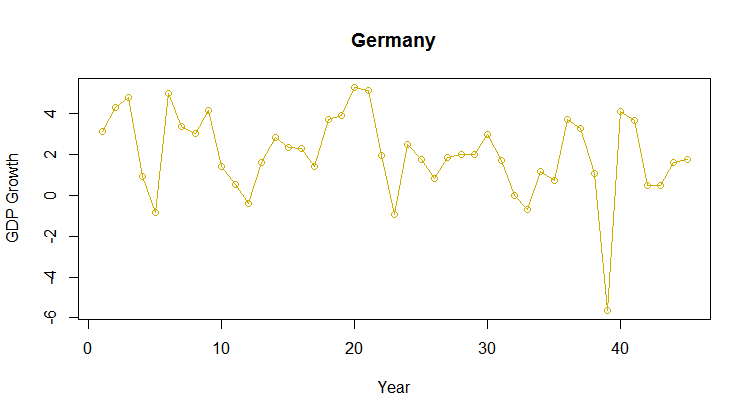
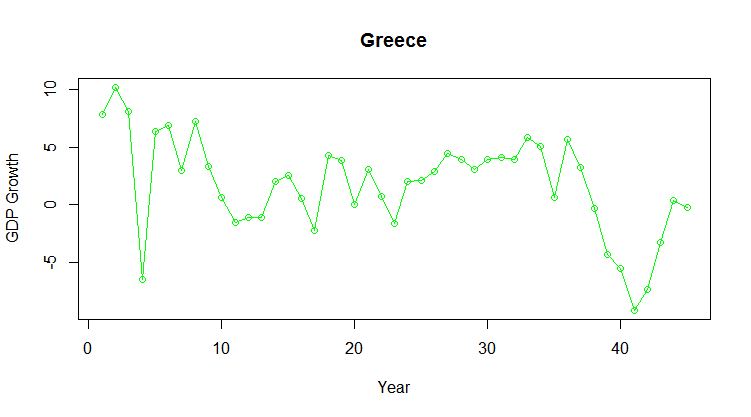
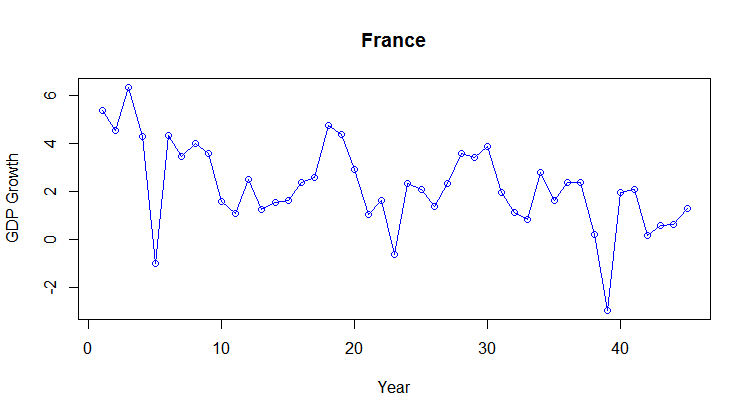
GDP growth is a commonly forecasted statistic in economics. Forecasters predict quarterly and yearly data on numerous countries, including those in the Group of 20, the Organization for Economic Co-operation and Development (OECD), the BRICS, and the European Union. However, it is also one of the most difficult variables to forecast, due to a variety of reasons. This is evidenced by the constantly changing leading indicators used by forecasters to predict recessions and expansions.[[1]](#endnote-1) Even the most basic components of GDP (consumption, investment, government spending, and net exports) are separable into finer and harder to predict pieces. However, approximations of the Wold decomposition may still be useful in forecasting GDP growth. This paper will compare the ability of ARIMA forecasts to predict GDP growth, evaluating accuracy in comparison to naïve forecasts and OECD, 1-year ahead forecasts for Greece, Germany, France, and the UK over the period 2006 to 2015.[[2]](#endnote-2)

Figures 1-4 plot World Bank measures of GDP[[3]](#endnote-3) of the four nations in question. Looking at this we can see a significant turning point for all four nations occurring in or around 2009 where GDP peaks and then declines. Some nations recover faster than others (Greece does not appear to recover at all) before returning to the previous trend. However, GDP is an absolute measure in this case. Looking at GDP growth in figures 5-8, we can see the relative impacts of the turning point on the four economies. Germany and France suffer a large hit in period 39 before returning to comparatively similar GDP growth, whereas GDP growth in Greece becomes negative and remains below zero despite increases after the trough occurs. From the first data visualizations, we can see that some series may be more easily forecasted by an ARIMA forecast, e.g. Greece’s GDP growth does not appear stationary, given the large decline in periods 38 to 45.

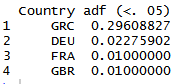




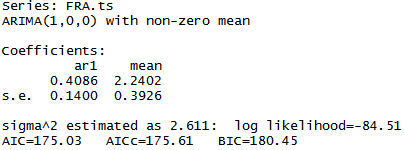
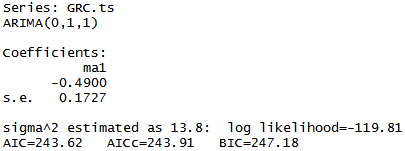
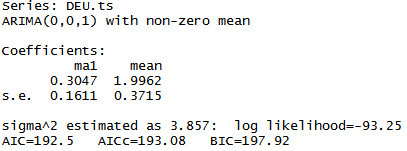
Figures 1-4 shown above, left to right depicting World Bank measures of GDP[[4]](#footnote-1)

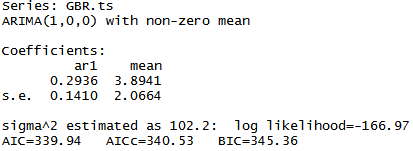


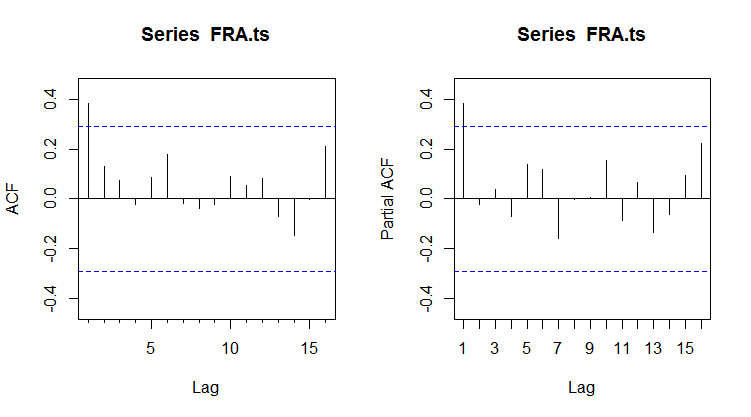
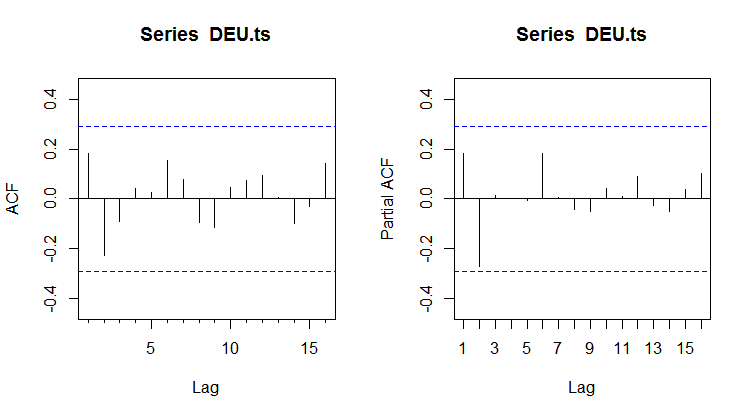
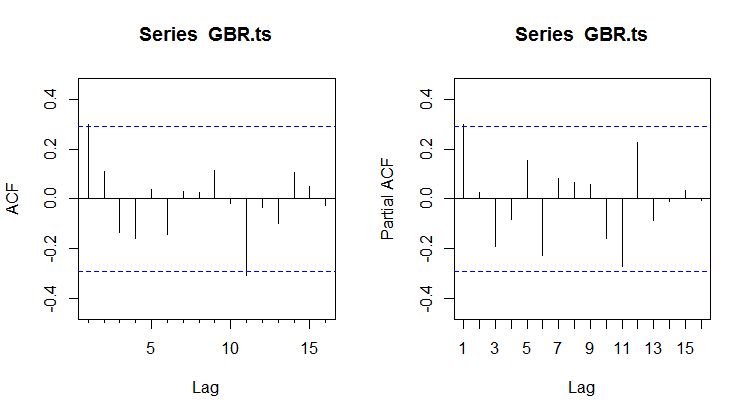
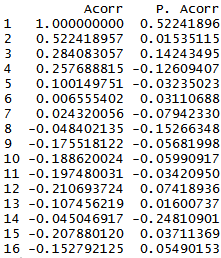
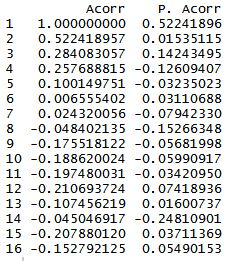
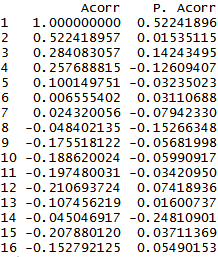
Figures 5-8 shown above, left to right depicting World Bank measures of GDP growth

By using the Augmented Dickey–Fuller (ADF) Test we can determine if our times series are covariance stationary. The alternative hypothesis of the ADF test is that our time series does not have a unit root. Our data will be covariance stationary if the mean and covariance structure are stable over time. Looking at the results (table. 1 on the left) of our ADF tests on our four time series, we see that only in the case of Greece’s GDP growth can we not reject the null hypothesis that a unit root exists. Nonetheless, we proceed with the creation of our forecasts.

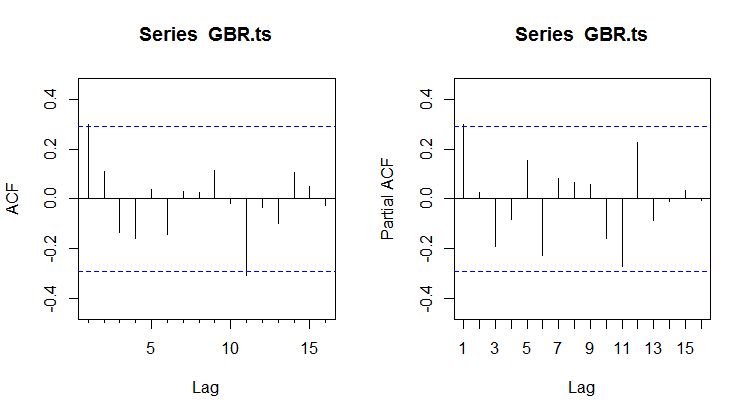
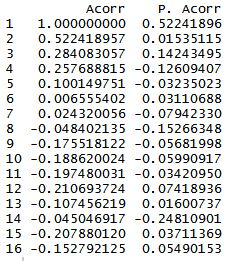
Observing our autocorrelations and partial autocorrelations (figures 9-12 and tables 2-5), we see that Greece and Germany are likely to be MA forecasts, autocorrelations are close to zero and partial autocorrelations oscillate while reducing in absolute value, while France and the UK exhibit the qualities of an AR forecast, autocorrelations reduce to zero and partial autocorrelations are not significantly different from zero beyond the first period. Using the auto ARIMA function in R, we generate the following four forecasts. The auto ARIMA function creates 72 forecasts with various magnitudes of AR and MA roots from zero to five as well as the possibility of seasonally differencing the data. It then compares the AIC and BIS of each forecast and selects the forecast with the best test statistics.











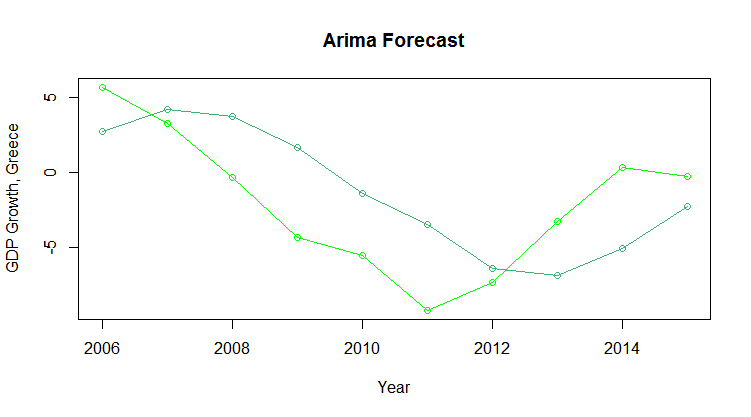
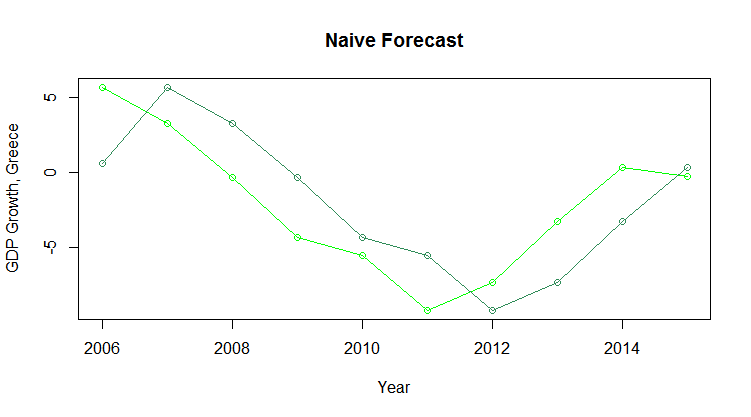
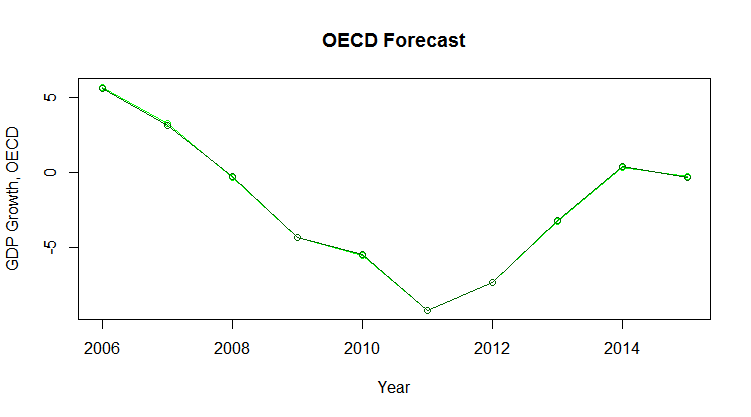


After creating a naïve forecast for each time series and recording OECD forecasts, we can begin evaluating our forecasts. First, we create 1-step ahead in-sample values of the naïve and ARIMA forecasts for our specified period. For each country, we compare the three forecasts graphically and by examining the Root Mean Square Errors (RMSEs). Additionally, we perform Diebold-Mariano tests to determine if our forecasts’ errors are significantly different, check the forecasts for bias, regress the predicted values on actual values as an additional accuracy check, and test for the presence of autocorrelations in our errors with a Durbin-Watson statistic.

RMSE =

RMSE is an evaluator of the accuracy of a forecast that represents the sample standard deviation of the residuals of a forecast. Given that RMSE is scale-dependent, it is useful in evaluating multiple forecasts of the same time series by comparing it to the time series’ range. A worse forecast will have a higher RMSE. However, in some cases differences between RMSEs of two forecasts may be irrelevant. This is when a Diebold-Mariano test is useful. The Diebold-Mariano test used in this paper is a two-sided test of the null-hypothesis that two forecasts have the same accuracy according to a comparison of a squared error loss function. The Diebold-Mariano uses the standard p-value threshold of .05 to reject the null hypothesis. A Diebold-Mariano test will useful when comparing two forecasts with similar RMSEs. Regardless, it is conducted on all three forecasts of each time series of GDP growth.

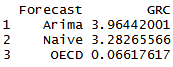
Bias =

Another useful measurement of forecast accuracy is bias. Bias is a simple measure of the arithmetic mean of the residuals of a forecast. With this statistic, we can determine if a particular forecast tends to overestimate or underestimate a time series. Similar to RMSE, bias is scale dependent and useful in context of the range of a time series. Poor forecasts will have a relatively large bias in absolute terms.

As a final accuracy check, we regress the predicted values of each forecast on the actual values of GDP growth.

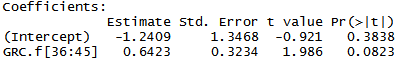
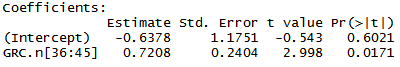
In an ideal forecast and . Lastly, we use the Durbin-Watson statistic to test for autocorrelation in our errors. The Durbin-Watson statistic ranges from zero to four; anything less than two indicates autocorrelation in the residuals of the forecast.

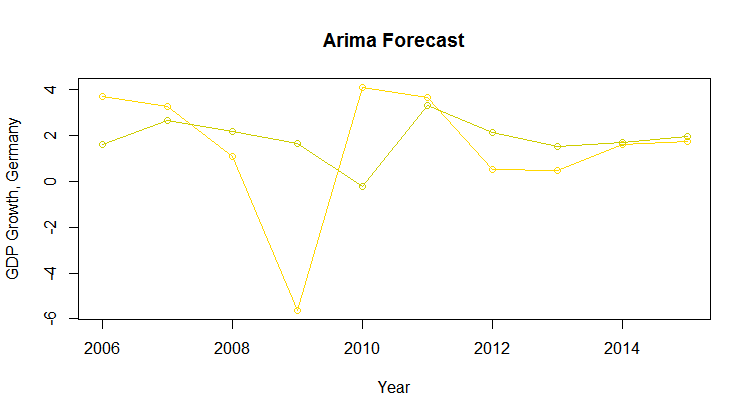
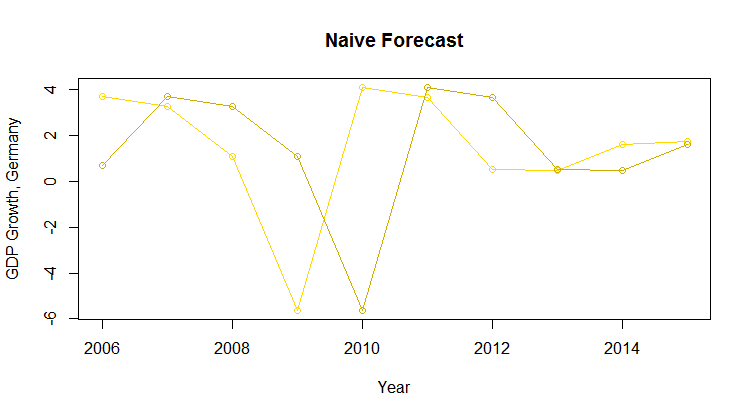
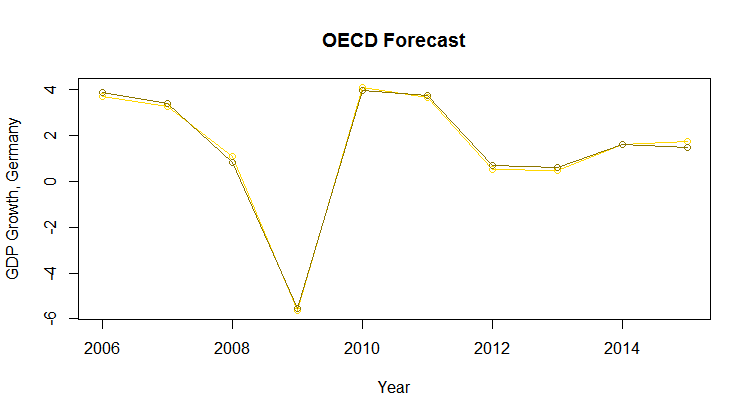
**Greece**

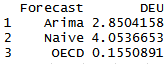
Observing figures 13-15 of the fitted values on the right, we can visually observe that the OECD forecast is far more accurate than our naïve and ARIMA forecasts. Checking the RMSEs of the forecasts (Table 6, bottom left), we see that the OECD forecast of Greece’s GDP growth is significantly more accurate, with a RMSE near zero and 49 times smaller than the next largest RMSE. Running DM tests to compare the OECD forecast to our ARIMA and naïve forecast generates p values of 0.0036 and 0.0020, respectively. Rejecting the null hypothesis that these forecasts have the same accuracy adds weight to the conclusion that the OECD forecast is more accurate.

Comparing the other two forecasts, it appears that the best available ARIMA performs no better than a naïve forecast for this period. A DM test comparing our naïve and ARIMA forecasts returns a p-value of 0.2334, so we cannot reject the null-hypothesis. Evaluating the bias in our forecast errors (Table 7, above), we see a tendency by both forecasts to overestimate GDP growth. Additionally, the mean of the ARIMA forecast’s errors is nine times greater than the naïve forecast’s average. In terms of GDP growth, which ranges between -9.1 and 5.7, a systematic error of .778 can have a significant impact on our forecasts.

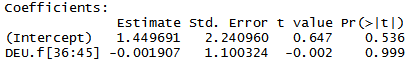
The next statistic test for accuracy is a regression of our time series on our forecasted values. The outputs of the three regressions are shown below. Only the OECD forecast fits the ideal regression of a near zero intercept and a coefficient of one on the independent variable. Lastly, we check the residuals of our forecasts using the Durbin-Watson Statistic (Table 8, above). All three test statistics are far below two, indicating autocorrelation in our residuals. These low values indicate potentially forecastable errors that future forecasts could incorporate.

Of our three forecasts, an ARIMA performs the worst on numerous test statistics, and is not significantly different from a naïve forecast over the same period. Despite this, all three of Greece’s GDP forecasts could be improved with additional information incorporated into the forecasts.

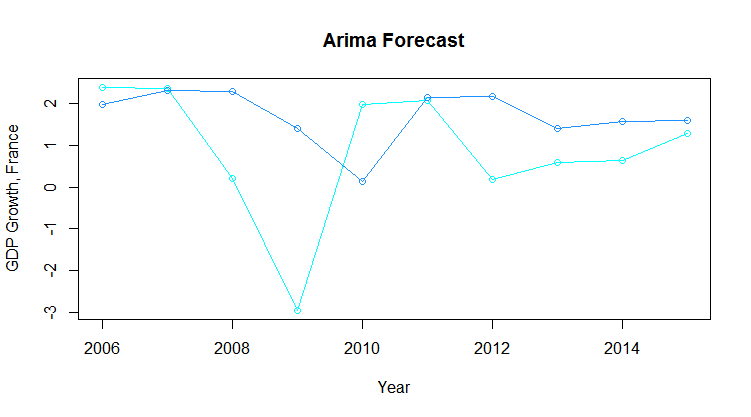
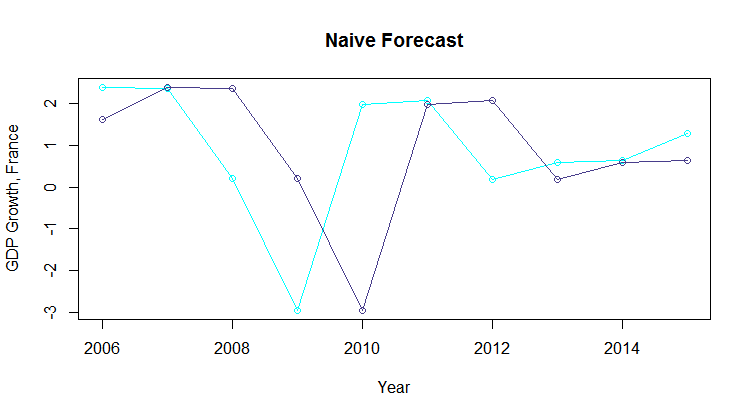
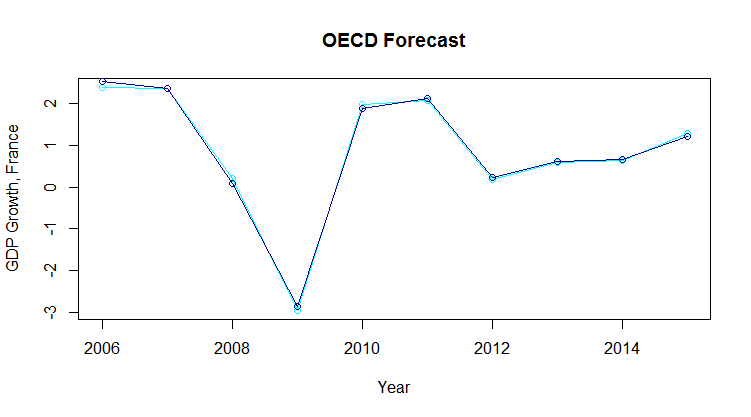
******Germany**

The plots of the three forecasts (Figures 16-18, right), show similar results to the forecasts of Greece’s GDP growth; the OECD forecast errors are small, confirmed by much smaller RMSEs (Table 9, bottom left) The OECD forecast root mean square errors are .15, 18 times smaller than the statistic for the ARIMA forecast and 26 times smaller compared to the naïve forecast. However, DM tests fail to reject the null hypothesis that the OECD forecast is significantly different from the ARIMA (p value: 0.1601) or the naïve (0.1231) forecast. Additionally, we also fail to reject the null hypothesis that the ARIMA and naïve forecast are significantly different from one another (0.3018). Comparing biases in our forecast errors (Table 10, bottom left), the OECD forecast has a much smaller bias than either the ARIMA or the naïve forecast. The systematic errors of the ARIMA in this case are large enough to cause systematic overestimation of Germany’s GDP growth, given a range of -5.6 to 4.07. Bias in the naïve forecast is smaller, although still underestimates GDP growth significantly. The OECD forecast displays very little bias in context of the scope of the time series.

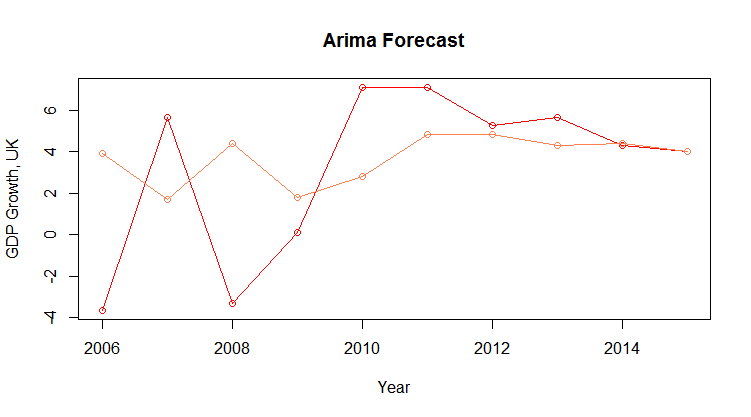
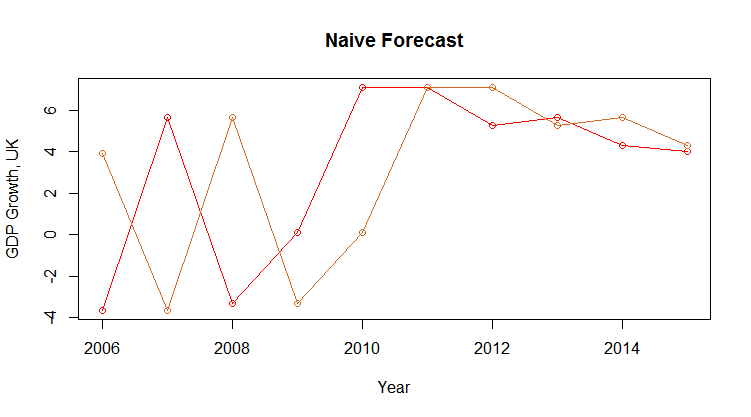
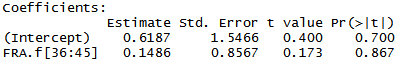
After creating the regressions of our time series on our predicted values, shown in the tables below, we see additional evidence that the OECD forecast is highly accurate; the intercept is near zero and the prediction coefficient is one one-hundred-thousandths greater than one. The ARIMA and naïve forecasts have near zero coefficients on their prediction values and intercepts above one. Observing the DW statistics (Table 11, right), we see that again all three forecasts have autocorrelated errors and could be improved in future forecasts.

Similar to the forecasts of the Greece’s GDP growth, the OECD forecast of Germany’s GDP growth is far superior according to several metrics, although we cannot reject the null hypothesis that it is significantly different the two comparable forecasts. Similarly, the ARIMA forecast has smaller errors than the naïve, but it is not significantly different and has a greater bias than the naïve forecast. Due to very low DW statistics, all three forecasts could be improved.

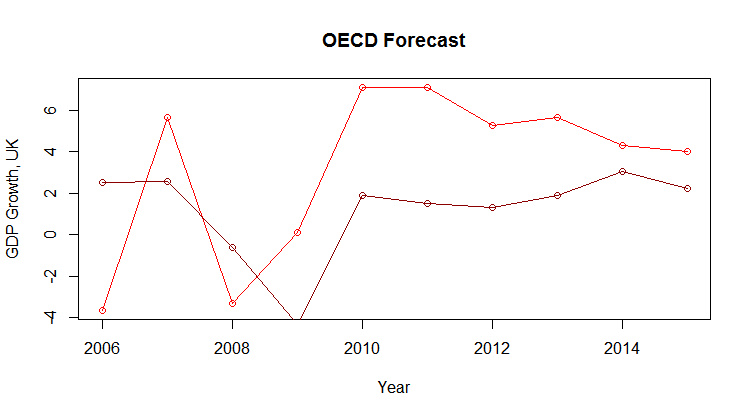
**France**

In the plots of France’s GDP growth forecasts (Figures 19-21, next page), the OECD forecast errors appear small in comparison to the ARIMA and naïve forecasts. Comparing the RMSEs (Table 12, bottom right), the ARIMA and naïve forecasts have similar errors, which are 24 and 28 times larger, respectively. Despite this, none of the forecasts is significantly different from each other. Comparing the OECD forecast to the ARIMA and naïve with a DM test generates p-values of .1096 and .1061, too high to reject the null hypothesis. A DM test of the ARIMA and naïve forecasts results in a p-value of .3018,also too high to reject the null hypothesis. The biases of the three forecasts (Table 13, bottom left) show that the ARIMA forecast has the most bias (.8237), which is large enough in a range of -2.9413 to 2.3749 to significantly overestimate forecasts of France’s GDP growth. The naïve and OECD forecasts, however, have much smaller biases; the OECD forecast in particular has a very small overestimation of France’s GDP growth (about 99 times smaller than the bias in the ARIMA forecast).

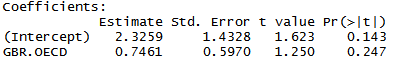
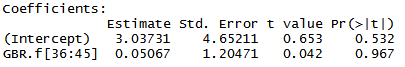
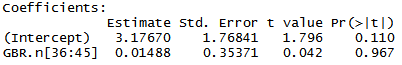
The regressions indicate that the ARIMA and naïve forecasts are very poor. The intercepts are not zero and estimated predictor coefficients are not near one. The OECD forecast regression (similar to our results for Greece and Germany) is very close to ideal, despite an intercept of -.012. When considering the DW statistics (Table 14, bottom right), there is evidence of autocorrelation in the errors of all three forecasts; none of the DW statistics for the forecasts is above two.

Of the three forecasts, the OECD is again the most accurate on all metrics, despite not being significantly different from either the ARIMA or the naïve forecast. The ARIMA is a better forecast than the naïve, although presents a much larger bias in its errors. All three forecasts are improvable, as their errors present significant autocorrelation.

**United Kingdom**

For the forecasts of the UK’s GDP growth, none of the three forecasts appears particularly accurate (Figures 22-24, left). Even the OECD forecast is off, having missed the recovery of GDP growth significantly and forecasting growth too low in nearly every period afterwards. Looking at the RMSEs (Table 15, below), the ARIMA performs best, although by a very slim margin. Similar to the forecasts of Germany and France’s GDP growth, the results of the DM tests indicate that none of the forecasts is significantly different from each other (all three tests result in p-values above .11). The naïve forecast displays the least bias (Table 16, right), and both the ARIMA and OECD forecasts have significant bias, considering that the UK’s GDP growth ranges from -3.6 to 7.1. Specifically, The OECD forecast considerably underestimates the UK’s GDP growth over this period.

Using regressions to evaluate the forecasts indicates that none of the three forecasts is particularly accurate. All three regressions output high intercepts with predictor coefficients near zero, likely due to a very scattered plot of predicted and actual values. Lastly, the Durbin-Watson tests provide a similar output to the forecasts of the previous three nation’s GDP. All three have very low DW statistics, indicating the potential for improvement due to autocorrelated errors.

In the case of the UK, the most accurate forecast is the ARIMA forecast, despite not being significantly different from a naïve forecast and exhibiting a greater bias. This is the only case observed in this paper where the OECD forecast performed the worst, as evidenced by a very high RMSE and considerably larger bias. However, none of these forecasts is ideal, and can incorporate additional information to improve accuracy.

**Predicting the Recession**

Forecasts are notoriously bad at predicting major turning points like those shown in figures 5-8. This is because forecasters rarely predict recessions under any circumstances.[[5]](#endnote-4) Evaluating our forecasts, we see that the naïve and ARIMA forecasts are especially bad at foreseeing short term declines, and the quick recovery of GDP growth in some nations (such as France and Germany) damages their accuracy a second time. This is likely because we are using simple forecasting methods that only incorporate the most recent value in the time series as a form of either smoothing or averaging the data for an optimal forecast. Models that are more sophisticated use a variety of leading and concurrent indicators to predict recessions. However, quantitative testing shows that these are not significantly better than ARIMA models beyond a horizon of three or four quarters (Fildes, Stekler, 2002).

Surprisingly, the OECD managed to forecast the recession very accurately in three of the four countries studied in this paper as well as GDP growth during the recovery period. In Greece, the prediction error in the trough year (2010) was only .0202, for Germany, the error was 0.0514 in the year of lowest GDP growth (2009), and for France, .0789 in the similar period (2009). In Germany and Greece, the error in this year was less than the RMSE of that particular OECD forecast over the ten-year period. This is considerably accurate forecasting, given that the OECD not only predicted negative GDP growth, but also predicted the trough year’s GDP growth with very little error across three separate countries. However, this may be a poor sampling of the OECD’s ability to forecast recessions. The European credit crisis affected the entire Eurozone,[[6]](#endnote-5) which contains 19 countries, as well as nine additional nations in the European Union that do not use the common currency. A broader test of the OECD’s forecasts of this period may reveal a poor overall track record.

**Summary**

ARIMA forecasts are poor at forecasting a negative turning point in GDP growth as well as the periods immediately following a recession. Not once in this analysis did the best ARIMA forecast perform significantly better than a naïve forecast according to a Diebold-Mariano test and sometimes performed worse according to Root Mean Square Error measurements. However, OECD forecasts performed significantly better than both ARIMA and naïve forecasts in three of four cases, despite not being significantly different according to a Diebold-Mariano test, except in the case of Greece where the OECD forecast performed significantly well over the course of a long recessionary period. Interestingly, the OECD was also able to predict recessions in three countries with relatively little error, perhaps they know something not everyone else does.

1. *Business Cycle Indicators Handbook*. New York: Conference Board, 2001. Web. [↑](#endnote-ref-1)
2. OECD (2017), Real GDP forecast (indicator). doi: 10.1787/1f84150b-en (Accessed on 03 May 2017) [↑](#endnote-ref-2)
3. World Bank. GDP growth (annual %). Web. 03 May 2017. [↑](#endnote-ref-3)
4. A note on identification. Greece may be referred to by GRC, Germany by DEU, France by FRA, and the UK by GBR. [↑](#footnote-ref-1)
5. Fildes, Robert and Stekler, H.O., The State of Macroeconomic Forecasting. Journal of Macroeconomics, Vol. 24, No. 4, November 2002. Available at SSRN: https://ssrn.com/abstract=342240 [↑](#endnote-ref-4)
6. "Timeline: The Unfolding Eurozone Crisis." *BBC News*. BBC, 13 June 2012. Web. 03 May 2017. [↑](#endnote-ref-5)