

For this exam we will be using the tools we have been working with in modelling the growth of the big four tech stocks (AAPL, AMZN, FB, GOOG). Attached is a dataset of weekly aggregated growth data for each stock between 2014 and February 2019, with the data from 2019 held aside for testing purposes. Values represent decimalized percent growth between the respective date and the first date, for example a value of .5 at date X would represent the stock price is 50% greater at date X than on January 1, 2014.

Below you will find a script with initialization code as well as a script with helper functions that may prove helpful. You are not required to use this code, but it is highly recommended that you use them to ensure consistency of results.

Note: Training data should be used for all questions except when comparing accuracy of forecasts.

Note: after loading the 'TSA' library ACF plots will begin at 1st lag, not 0th. Keep this in mind as you evaluate. Alternatively, you can use the method 'stats::acf(...)' for the typical form of the plot. Either is acceptable.

Exploratory Data Analysis - 12 Points

1. Using relevant graphs of the **first difference**, evaluate the stationarity of each series' difference and comment on the presence of any features of note.
2. For AMZN data fit second order polynomial regression and for FB data fit Splines regression models (both undifferenced); display summaries of both. Calculate Mean Squared Error (MSE) and Mean Absolute Error (MAE) for the fits of each. Can these measures be used to compare relative effectiveness of each model?
3. Graphically evaluate the residuals for both models using appropriate plots.

ARIMA Modeling - 20 Points

1. For the **first difference** of AAPL data produce ACF and PACF plots. Comment if these plots can be used to estimate an ARMA order and if so what the order would be.
2. Use the iterative method with a max order of (4,2,4) and an AIC threshold of 2 to estimate the best ARIMA orders for AAPL and GOOG data. Orders selected must have at least one coefficient (i.e. any order of (0,d,0) should be discarded).
3. For the both models selected, use roots evaluation to describe the nature of the processes selected.
4. For both models, forecast ahead 9 points. Compare forecasts to the test data to calculate Mean Absolute Percentage Error (MAPE) and Precision.
5. For the GOOG data, produce a plot that compares predictions and prediction confidence intervals to actual values in that time period. Comment on what is observed.

Multivariate Modeling - 18 Points

1. Determine what VAR(p) orders would be selected using AIC and HQ of the first differenced training data. Next fit a VAR(2) model on the differenced data. Use appropriate statistical tests to evaluate Normality, Heteroskedasticity and Uncorrelation assumptions of the models' residuals.
2. Forecast ahead 9 points. Again, for AAPL and GOOG, compare forecasts to the differenced test data to calculate Mean Absolute Percentage Error (MAPE) and Precision.
3. Compare your accuracy measures above to those in Question 2.5. Do the results indicate anything of note about the data? (Note: while the frames of the difference differ, MAPE and Precision measures do not depend on scale and they are robust to outliers, although not immune to transforms. Consider this when you compare the models). Can you conclude anything about the data?

Heteroskedasticity Modeling - 20 Points

(Use training data for order selection and model evaluation, all data for rolling forecasts)

1. Starting with the ARMA order (1,1) and using a max possible order of (4,4)x(2,2), use minimum BIC to determine the best ARMA-GARCH order on the undifferenced GOOG data. Your selected model must have at least one coefficient in both components.
2. Print the coefficients of the model with the selected order. Comment on the significance of coefficients and write out the model equation in full. To ensure consistent results, use the "garchFit" method with "trace=false" and extract the coefficients from the resulting model.
3. Perform Ljung-Box tests on the residuals and the squared residuals as well as ARCH and Jarque-Bera tests on the residuals. (Use the Ljung-Box results from the model summary.) Interpret the results.
4. Using the order selected in A, use the rolling forecast method to calculate Mean Absolute Percentage Error (MAPE) and Precision for the last year (52 points) of the training data. (Split the training set into new train and test sets for this.) How do these results compare to the GOOG ARIMA model? Can these metrics be compared? What would you say about the reliability of the results?