

## FORECASTING HOMEWORK 1

All data sets are on the course website, <http://www.stern.nyu.edu/~churvich> (click on "Forecasting"). All recent data sets are available in Minitab Portable (.MTP) format, and most are also available in EXCEL (.CSV). The file ReadMe contains a description of the data sets and the other files.

The MTP files (such as GDP.MTP) may be easily read into Minitab using File → Open Worksheet. If you ever need to read label-free text files into Minitab, I recommend the command sequence File → Other Files → Import Special Text. You need to enter column names in which you want the data stored.

1)

A) Plot the GDP series versus time. Note that the data is quarterly, from 1946, Q4 to 2014, Q4, a total of  $n=273$  observations. (When talking about a plot of  $y$  versus  $x$ , it is traditional to give the response variable first, then the predictor variable. So I am asking for GDP on the  $y$ -axis, and time on the  $x$ -axis.) (In Minitab, to make a time series plot, you can use Stat → Time series → Time Series Plot. In the plot menu, I recommend that you click Data View and then un-check the box for Symbols. This removes the big dots that can obscure the fine structure of the plot.) Does GDP seem to grow linearly over time?

B) Plot the log of GDP, in a time series plot. Does log GDP appear to grow linearly over time? If so, then what does this imply about the growth of GDP itself? To create the log of GDP in Minitab, you can use Calculator: Calc → Calculator → Store Result in Variable: LogGDP, Expression:  $\log(\text{GDP})$ .

C) Fit an ordinary linear regression model for LogGDP, using time as the predictor variable. (In Minitab, generate a column containing the predictor variable "time", taking values 1,2,3,...,273. To do this, use Calc → Make patterned data → simple set of numbers.) Based on the output, forecast the LogGDP for the second quarter of the year 2015. Based on the linear regression model, construct a 95% prediction interval to go with the above point forecast. (To get prediction intervals,

after running the regression, use Stat → Regression → Regression → Predict, and enter the  $x$ -value at which you want to predict  $y$ . This  $x$ -value should be 275, since the 273rd point was the final data value, corresponding to the fourth quarter of 2014.) Do you think this interval forecast (prediction interval, denoted by PI in the Minitab output) is valid? Does it seem too wide, or too precise? Explain.

D) On a single plot, superimpose the LogGDP series and the fitted line. (For this, you can use Stat → Regression → Fitted Line Plot.) Does the line fit well?

E) Plot the residuals from the fitted line versus time. (In Minitab, this can be done by selecting "Graphs" in the regression dialog box, and then checking "Residuals versus order"). What potential problems with the linear model are indicated by this plot? Do you think these problems could spoil the validity of the forecast interval?

2) Consider the daily U.S. Dollar/Euro exchange rate, daily, 4 Jan 1999 to 17 Jul 2015 ( $n=4159$ ). The numbers are given in the file Euro.MTP on the course website.

A) Create a time-series plot of Euro. Does a straight-line model seem appropriate?

B) Get point and interval predictions for 17 Jul 2015 using two different methods: First, based on fitting a straight line to observations 1 to 700; Second, based on fitting a straight line to observations 701 to 4158. Did both of the forecast intervals succeed in containing the actual value for 17 Jul 2015? If not, then use what you learned from Problem 1 to give a statistical explanation of what went wrong.