Stats 3540 Project
Due on or before April 4, 2008

Reminder: The projects can be done individually, or in groups of two students.

For your project, you are required to find a time series dataset and conduct a thorough analysis, based on methods discussed in class and in your textbook. This handout contains some websites where lots of time series data are available.

Your written project should be split up into different sections, which contain complete paragraphs and sentences. Your report can be typed or handwritten. I will also expect you to submit any relevant computer output, as described below. This means that I will not want to see things Minitab (or Stata, etc.) did that you did not want it to do. In particular, please **do not print your entire dataset** and submit it with your report.

Your report will be evaluated based on clarity of presentation (including grammar and spelling), application of correct statistical methods, interpretation of results and creativity.

A general outline of the format is given below. Note that not all sections will be of equal length.

Format of Report

In preparing your project, break your report up into the following sections, and use these sections titles to aid in making your project clear and easy to follow:

- **Introduction**: Include a description of data, and what you will try to do in your analysis.
- **Exploratory Analysis**: Include things like plots, tables, summary statistics. Not all of these may be needed or applicable. Some of the plots that could be useful will be covered a bit later in the course.
- **Analysis and Modelling**: This will include discussion on the potential models for your time series and estimation for those models. This could include using time series regression models, decomposition methods, exponential smoothing methods or ARIMA models (which we will cover in the last few weeks of class). You should include any inference (hypothesis testing) that is appropriate.

  For the model(s) you select, you should use some model diagnostics (residual plots, etc.) if possible, to see if your model assumptions are satisfied. You should also use your model to demonstrate its forecasting ability.
- **Conclusions**: Brief summary of findings.
- **Appendix**: **Relevant** computer output.

Finding Data

- `www.library.mun.ca/qeii/govdocs/govdocs.php#statistics`

  and click on the link for **CANSIM II**. This site contains tons of socio-economic and demographic time series data.

  WARNING: **CANSIM II** is easiest to access if you're accessing the internet on campus at MUN. Just click on the link and you get access. If you want to access it from home, you have to follow the directions that are given when you click on the **MUN Only** link beside the **CANSIM II** link.

- `www-personal.buseco.monash.edu.au/~hyndman/TSDL/`

  This site contains about 500 time series from many different fields. This site also has some links to other time series collections and resources on the internet.

- `dss.ucar.edu/catalogs/free.html`

  A large collection of atmospheric and oceanographic datasets. However, a lot of the datasets are really big, or measure variables that may not be familiar to you.

You are welcome to find a time series from a different source than the ones listed above.

When selecting a time series, please make sure that it is **interesting**: it has a trend and/or a seasonal component, etc.