Stat 443. Time Series and Forecasting

Team-based term projects

Reference: https://www.britishecologicalsociety.org/wp-content/uploads/2019/06/ BES-Guide-Reproducible-Code-2019.pdf

This document has good advice on wrangling, workflow and organization of files for team project work. See especially pages 6–10, 14–17.

Aim for a group size of 4 ± 1.

It is best if you can do a project related to one of your interests. Make a post to piazza if you are looking for teammates for a given theme.

- Part of the project grade goes towards whether you have chosen a meaningful (scientific) topic.

- The team could do analysis of several time series of the same format or response variable.

- Compare RMS out-of-sample forecast errors for 1-step forecasts in a holdout set which follows in time from the training set. Also, compare with in-sample RMS prediction errors using 1-step forecasts.

- 3 classes of forecasting methods, based on theory in Stat 443, are to be compared: (a) best exponential smoothing method (allowing for trend and seasonality, if relevant); (b) best model based on Box-Jenkins (ARMA, ARIMA) methodology; (c) best model with external (lagged) predictor variables using ARMAX or regression with possibly forecasted predictors.

- Include baseline methods of persistence (naive) and training set average.

Possible topics for term projects

See sources for data sets available at the course web site.

1. Environmental, climate
(a) any trends in temperature, precipitation, extreme events;
(b) river flows, river heights, flooding possibilities;
(c) trends in pollution levels, concentrations of pollutants, air quality indexes.

2. Economics: what are leading indicators for some macroeconomic series?

3. Prediction/forecasting of sales at a business: exogenous variables used together with previous observations of the response variable.

4. Crime rates: trends, etc.

5. Time series of reported flu counts?

6. Check possibilities at www.kaggle.com for past kaggle competitions (data sets can be obtained after registration).

7. Sports analytics: don't know if there are possibilities.

Forecasting competitions published in *International Journal of Forecasting*.

For example: https://doi.org/10.1016/j.ijforecast.2021.11.013

Makridakis, Spiliotis, Assimakopoulos (2022). M5 accuracy competition: Results, findings, and conclusions. International Journal of Forecasting, v 38, pp 1346–1364.

The datasets were available in kaggle:
https://www.kaggle.com/c/m5-forecasting-accuracy

M5 example: Start Mar 2, 2020; close Jun 30, 2020.

"Note: This is one of the two complementary competitions that together comprise the M5 forecasting challenge. Can you estimate, as precisely as possible, the point forecasts of the unit sales of various products sold in the USA by Walmart? If you are interested in estimating the uncertainty distribution of the realized values of the same series, be sure to check out its companion competition"

The ranking of entries was based on out-of-sample prediction performance:

RMSSE = root mean squared scaled error based on point forecasts up to a forecasting horizon of 28 days.

WRMSSE $= \sum_{i=1}^{42840} w_i RMSSE_i$, w_i is the weight for the ith product.

For Stat 443, the performance measure is different, as it is based on moving 1-step forecasts in the holdout set.

If n is the length of a training set time series, note the distinction between $\widehat{y}_{n+h|n}, h = 1, 2, \ldots$ and $\textcolor{magenta}{\widehat{y}_{n+j|n+j-1}, j = 1, 2, \ldots}$.

For the competition, the teams were required to submit a file with $\widehat{y}_{i,n+h|n}, h = 1, 2, \ldots, 28$ for product i, $i = 1, 2, \ldots$.

For moving 1-step forecasts, code in the form of a function would be needed to evaluate the out-of-sample forecasts.

Summary

The purpose of the project is to make use of the theory and methods presented in Stat 443. Having to use the statistical methods for a project of personal interest can provide motivation for learning.

To do well in the kaggle competitions, typically combination of methods from several different courses are needed.