

The Econometrics of Predictability

Individual Assignment

June 6, 2014

Submission

Important Note: This assignment is due **Friday 9th week 12:00 (noon)** at *Exam Schools*. By proctor imposed rules, we have individual, formal work, and so unfortunately this requires submission at exam schools.

Assignment

This assignment ask you to examine some of the features of high-dimensional modeling strategies.

1. Download data from the course website. The data zip contains a file called StockWatson2012.mat, which is the Stock & Watson dataset used in the course. The data files also contains the raw excel file used to generate the data set. The variables included are:
 - (a) factors - The factors transformed according to transform. Note that the actual factors used to estimate the PC use include code 1.
 - (b) factors_raw - The entire data set only not transformed
 - (c) names - Short name (code) for each series
 - (d) include - Numeric vector indicating whether to include the data in the principal components (1) or not (2)
 - (e) long_descr - Long Description of Variable
 - (f) short_descr - Short description
 - (g) mldate - Dates in MATLAB data forms
 - (h) date - Dates in the form a T by 2 matrix with [yyyy qq]
 - (i) transform - Numeric variable indicating transformation type:
 - i. 1 Level
 - ii. 2 First Difference
 - iii. 3 Second Difference
 - iv. 4 Log-Level
 - v. 5 Log-First-Difference
 - vi. 6 Log-Second-Difference
 - (j) category - Numerical category data. Values like 5.01, 5.02,... indicate variables from category 5, and .01, .02 are counts.

2. Compare the forecasting performance of (pure) DFMs, PLS, 3PRF, and RRRR-Spectral for the three components of the yield curve, level, slope and curvature. You will need to download data to construct L , S and C from Fred using GS1, GS5 and GS10 (1-year, 5-year and 10-year yields). Use quarterly data, which you will need to aggregate using either (a) quarter end and (b) average during the quarter. You should focus on the basic methods covered in the lectures here. Use 50% of the sample and perform the evaluation out-of-sample. Note to use RRRR-spectral, you should *jointly* forecast all 3 series since there are differences gains when forecasting a single series.
3. Add in AR(4) (up to, can be smaller, but need explain choice) to the previous models, and consider a pure AR as well. Do own values improve out-of-sample fits?
4. Use the hard thresholding method to select variables, and repeat 2/3 using the variables that survive thresholding. Be sure to implement this using only in-sample data.

MATLAB

`pca` - principal components analysis from the MFE Toolbox. The PCs are the second output and are T by k .

`eig` - eigenvalue and eigenvector calculation, using $[V,D]=\text{eig}(S)$ where V are the eigenvectors and is a diagonal matrix of eigenvalues. $[V,D]=\text{eig}(A,B)$ computes generalized eigenvectors and eigenvalues.

MATLAB's statistics toolbox contains an implementation of PLS although it differs from the description in the slides. Both PLS and 3PRF are very simple to implement.