Kellogg School of Management
FINC 585-3: Asset Pricing
Spring 2024

Prof: Torben G. Andersen & Zhengyang Jiang

Assignment 3

Problem 1.

For this question, we are going to use the dataset `stocks_pca.zip` available on CANVAS. This dataset contains the daily price of the current constituents of the S&P 500 index from 2004 to 2023 available on CRSP. Stocks are identifies by their unique PERMNO. Using this data, calculate daily log-returns and drop stocks that contain any missing values in the sample.

Additionally, we are going to use the Fama-French 5 factors + Momentum (FF) dataset available on CANVAS under the `ffdaily.csv` file. This dataset contains the daily log-returns of the Fama-French 5 factors + Momentum from 2004 to 2023.

A) We start by making some simple Principal Component Analysis (PCA). Extract the first principal component of the log-returns. How does it look like on the time series? How much of the variance is explained by the first principal component?

B) Now extend this to up to 8 factors. How much does the explained variable increase with the addition of new factors? Using the Bayes Information Criteria, how many factors should we choose? How does your choice compare with using just the explained variance? Explain.

C) Pick the model with the greatest number of factors. Using rotation, how do the principal components correlate with the FF factors?

D) We now turn to using the Canonical Correlation Analysis (CCA) to analyze the relationship between the FF factors and the PCA factors. Using 1 to 5 components, calculate the canonical correlation between these two sets of factors. How does it change with the number of components?

E) Now obtain the canonical variables by rotating the PCA and FF factors according to the canonical loadings. How much correlation is there between the canonical variables?

References