DAT 537: Final Project

Siddhartha Chib

April 21, 2021

Project Instructions

These questions aim to help you gain some hands-on experience in processing and analyzing real marketing and finance data. Project parameters are as follows.

1. The report is due by 23:59 pm on May 5, 2021.
2. The first page of the report have your name followed by the program track (MBA, MSCA, MSF in quantitative finance or MSF in corporate finance etc).
3. BOTH the knitted output file and the Rmd file used to generate the report have to be submitted.
4. The Rmd file should be such that all results are fully reproducible.

Finance Project: Pricing performance of factors

1. Select 10 stocks that you are interested in. Find their yahoo symbols. This website can help you find the symbols that yahoo is using http://investexcel.net/all-yahoo-finance-stock-tickers/. Remember to double check the symbols at yahoo finance.

2. Download monthly premium data for each stock from Jan 2005 to Dec 2018 using the cbw getfinmdat() function. Remember this requires that all 10 stocks you select in step 1 should be available for this time frame.

3. Load the package czzg and use the data(factor12) as given to find the best factor collection by the Chib, Zeng and Zhao (2020) method. Use a student-t distribution for the errors and let nu = 5 in the model scan.

4. Combine the data in factor12 with the data on the 10 stock premiums (this means that at this point you will remove all the rows in factor12 before January 2005).

5. Use the factors from the best model to see how many of the 10 stocks can be priced (for each stock you need to fit two models - one with an intercept and one without, as described in the which factors note).

Marketing Project: Brand complements and substitutes

1. Load the `tuna` data set. There are seven brands in the data set. For each brand, estimate separate independent student-t models where `logsales` for each product is regressed on an intercept, the product's log price and display activity. Use the default training sample prior and use log-marginal likelihoods to find the appropriate-degrees of freedom of the student-t distribution on the grid seq(from = 3,to = 6,length.out = 11).

2. Now estimate a SURE student-t model for the seven brands. Again use marginal likelihoods to find the appropriate degrees of freedom on the grid seq(from = 3,to = 6,length.out = 11).

3. Do you find that the SURE-t model is an improvement over the independent t-models?

4. From the best fitting SURE-t model, what brands appear to be complements and which appear to be substitutes?