STA5069Z Topic Ideas

Topic Idea 1: Introduction

In the financial world a yield curve is a line that plots yield, or interest rates of bonds that have equal credit quality but differing maturity dates. The slope and shape of the yield curve can predict future interest rate changes and economic activity. Volatile markets and economic environments can significantly distort the shape and smoothness of yield curve movements and hence bond prices.

Topic: This study aims to identify the main underlying movements present in the South African bond market and to further determine the dominant factors that are responsible for driving South African bond price movements.

PCA one of the techniques used to project high-dimensional data into a lower-dimensional subspace by creating a reduced set of linear transformations of the input variables will be used to analyse the major drivers of South African financial bond market prices. Once complete one will be able to tell, among the multiple factors (e.g. duration, convexity or volatility) which main factors mainly quantify the movement of yield curves and hence interest rates in the South African market.

Exploratory Data Analysis

The dataset **SA_bonds_1994_daily** looks at the SA bond market data from the period 1994-2017. It looks at the bonds namely ALBI, 1 to 3 years bonds, 3 to 7 years, 7 to 12 years, over 12 years and GOVI. The variables are as follows:

- Annualised Volatility Close
- TRI (Total Return Index) Average Yield
- Interest Yield
- Duration
- Closing Price
- Convexity
- Total Return YTD (Year to Date)
- Total Return Index
- Return YTD
- BP (Basis Point) Spread
- All in Price
- Clean Price

The data sourced from *UCT Open Data Portal – ZivaHub* , dataset titled **Daily Sampled South African Market Index Data 1994-2017**, the author is Prof Tim Gebbie.

Topic Idea 2: Introduction

Clustering analysis will be performed on the Breast Cancer Wisconsin (Original) dataset to determine the breast cancer variables that can be clustered or grouped together i.e. we wish to partition the 9 variables into K distinct groups. The method that will be used is the K-means clustering algorithm .

Exploratory Data Analysis

The Breast Cancer Wisconsin data describes the characteristics of the cell nuclei present in a breast cancer image. The data set has 9 variables and 569 different observations. Ten valued variables are computed for each cell nucleus and are as follows **clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli and mitoses** .

The data will be obtained from UC Irvine Machine Learning Repository and is titled Breast Cancer Wisconsin (Original)