Changepoint Detection in Stock Prices

Joyce Yang

December 8, 2016

Introduction

Changepoints are times when the probability distribution of a stochastic process or time series changes.

Database

The data was from quandl.com. The database has American stocks for the past 54 years, since 1962. We have 3183 companies and fourteen million data points. I examined the daily adjusted closing prices of various stocks and and calculated their daily returns. I also examined the Dow Jones Industrial Average daily return.
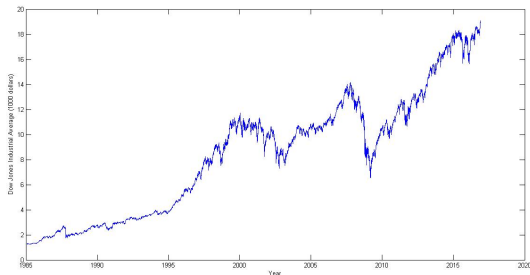
The daily closing price for the Dow Jones Industrial Average for 1985 to 2016

$$Daily\ Return = \frac{Closing\ price\ of\ current\ day}{Closing\ price\ of\ previous\ day} - 1$$
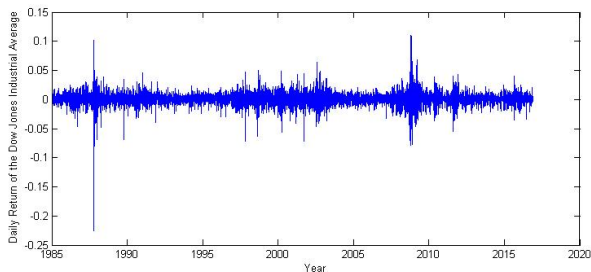
The daily return for the Dow Jones Industrial Average for 1985 to 2016

Techniques

I used Bayesian Online Changepoint Detection and Anomaly Detection.

I used Adams and MacKay's (2007) implementation for Bayesian Online Changepoint Detection.
Changepoint detection is the identification of abrupt changes in the generative parameters of sequential data.

For Anomaly Detection I used the Twitter (2015) implementation of the generalized Extreme Studentized Deviate, or ESD, of Rosner (1983).

ESD takes the datapoint whose distance from the mean is largest, then removes it, then finds the datapoint whose distance from the new mean is largest, and so on.

It uses the absolute value of the difference, which is a robust measure of distance.

Anomalies, or outliers, are the data that are outside a certain percentile range.
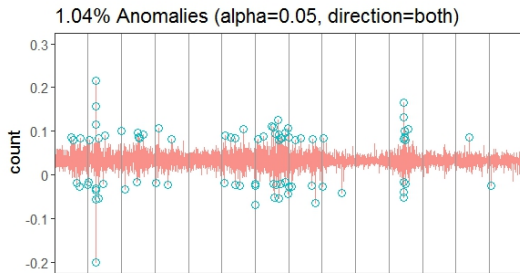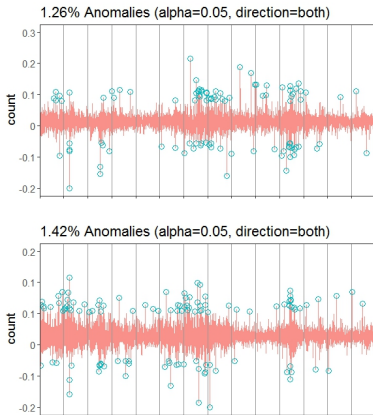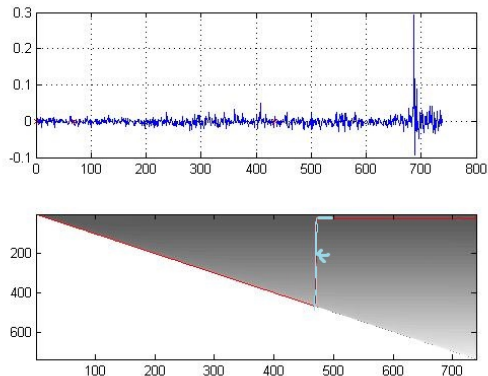
Results

Figure: The daily return for KO (Coca-Cola) for 1985 to 2016

The daily return for HRS (top) and NKE (bottom), from 1985 to 2016

The Bayesian approach finds variations in the generative parameters of the data.

Top: The daily return for the Dow Jones Industrial Average for 1985 to 1987. The Dow fell drastically on black Monday in October 1987. Bottom: Run length distribution of each day. The run length drops to zero at a changepoint, marked by an arrow

AAPL	29
ALCO	39
BOBE	28
GD	30
HES	39
HRS	27
KO	46
MSI	28
NI	36
NKE	24

Table: The tickers for which data was available the most days from 1985 to 2016 in alphabetical order, and the corresponding number of changepoints.
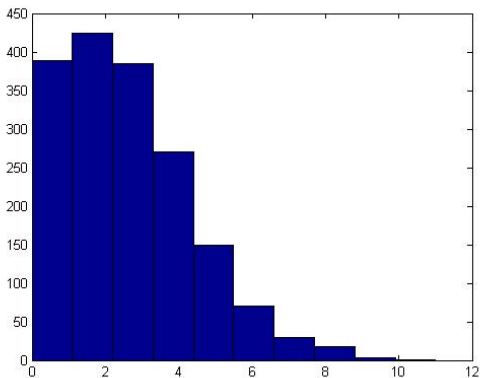
The distribution of the number of changepoints for 1740 stocks from 2014 to 2016

Summary

We detected changepoints in the stock prices in our database using Bayesian techniques.

We also used the frequentist techniques from the Anomaly Detection Twitter package.