Changepoint Detection for Stock Prices

Joyce C. Yang
Department of Mathematics
Harvey Mudd College
jcyang@hmc.edu

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

1 Introduction

In my project I examined changepoints in stock prices. Changepoints are times when the probability distribution of a stochastic process or time series changes.

2 Database

The data was from quandl.com. The database has American stocks from January 1962 to November 2016. It has 3183 companies and fourteen million data points. I examined the daily adjusted closing prices of various stocks and and calculated their daily returns. Stock returns are approximately normal and stock prices are approximately lognormal, as mentioned in Ref.[3]. I also examined the Dow Jones Industrial Average daily return. Figure 1 shows the daily closing price for the Dow Jones Industrial Average for 1985 to 2016.

Figure 2 shows the daily return for the Dow Jones Industrial Average for 1985 to 2016. The stock market crash on Black Monday– October 19 1987– caused the Dow Jones Industrial Average to fall by 22.61 percent.

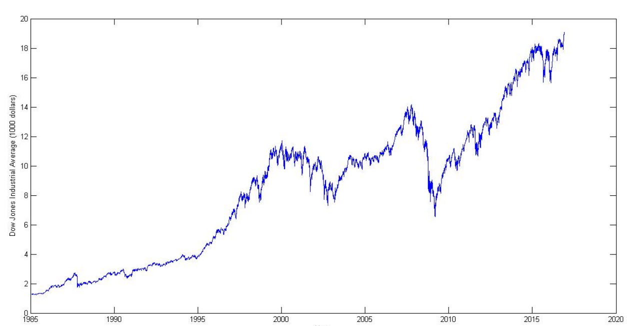

Figure 1: The daily closing price for the Dow Jones Industrial Average for 1985 to 2016

$$Daily\ Return = \frac{Closing\ price\ of\ current\ day}{Closing\ price\ of\ previous\ day} - 1$$

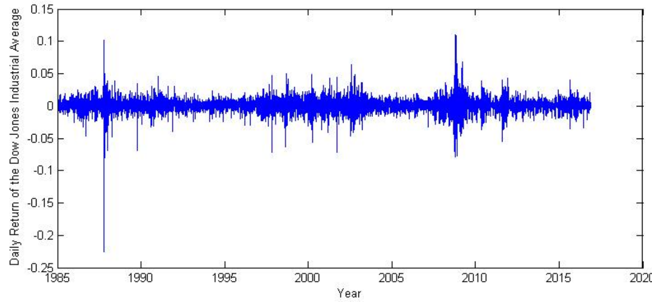

Figure 2: The daily return for the Dow Jones Industrial Average for 1985 to 2016

3 Techniques

I used Bayesian Online Changepoint Detection and Anomaly Detection.

I used Adams and MacKay's (2007) implementation for Bayesian Online Changepoint Detection from Ref. [1] . Ref.[2] used a sampling method to modify this method.

Changepoint detection is the identification of abrupt changes in the generative parameters of sequential data. It is a method that only looks at part of the dataset and updates the changepoints one at a time. I used a prior on the inverse variance that was a gamma distribution. The parameters were $\alpha = 1, \beta = 10^{-4}$. The prior on the returns was a Gaussian distribution with mean 0. The prior on the changepoint interval was exponentially distributed with rate $\lambda = 250$.

For Anomaly Detection I used the Twitter (2015) implementation in Ref.[4] of the generalized Extreme Studentized Deviate, or ESD, of Rosner (1983). See Ref.[5] for a complete description of ESD.

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.

This method 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.

4 Results

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

Our results for the companies for which the most data was available since 2014 are shown in Figure 6. This histogram shows that the number of changepoints is mostly less than 10. The mode is 1, and the time scale is 3 years.

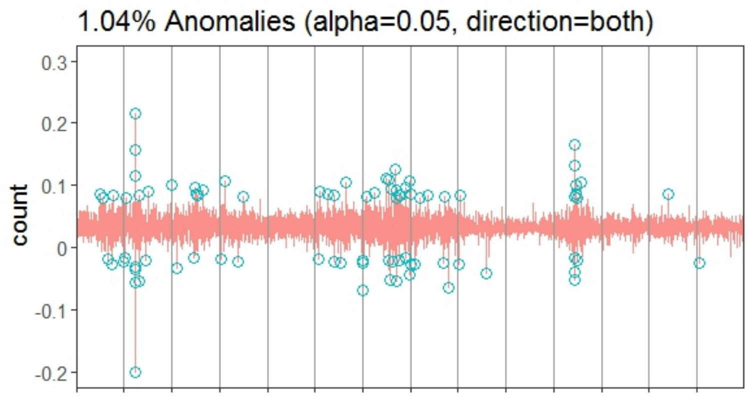

Figure 3: The daily return for KO (Coca-Cola) for 1985 to 2016

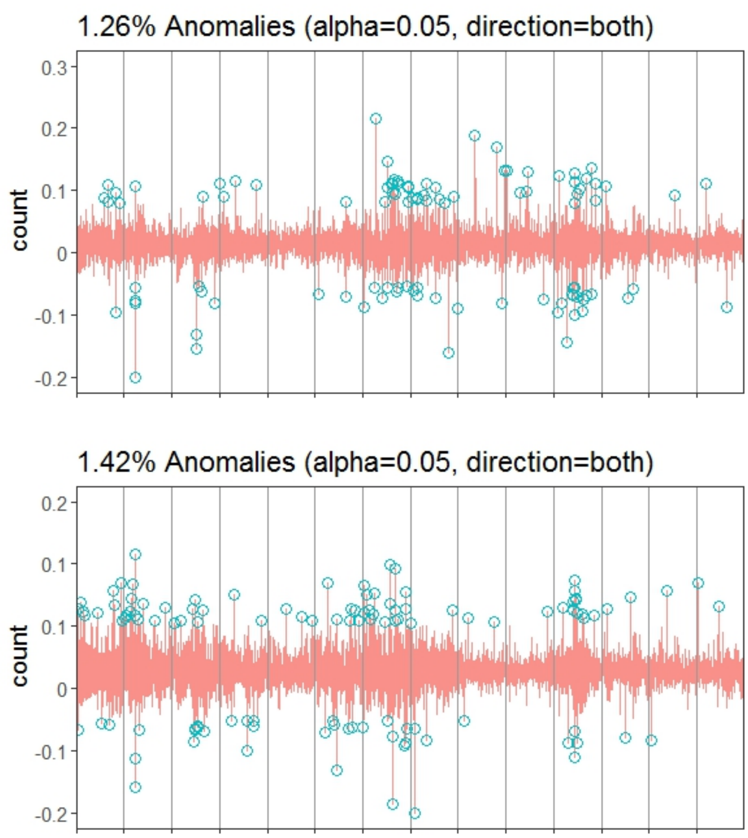

Figure 4: The daily return for HRS (top) and NKE (bottom), from 1985 to 2016

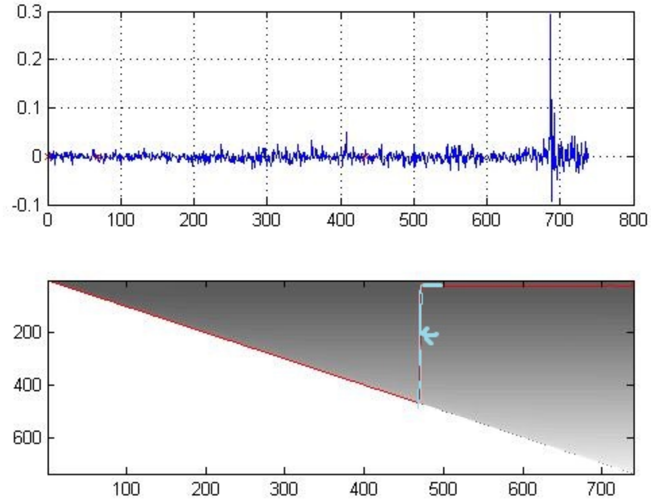

Figure 5: Top: The daily return for the Dow Jones Industrial Average for 1985 to 1987. The Dow fell drastically on black Monday in October 1987. Bottom: Run length distribution of each day. The run length drops to zero at a changepoint, marked by an arrow

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

Table 1: The tickers for which data was available the most days from 1985 to 2016 in alphabetical order, and the corresponding number of changepoints.

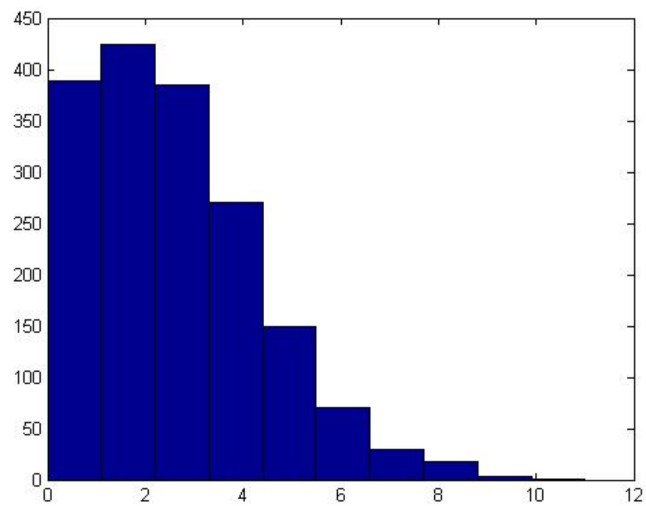

Figure 6: The distribution of the number of changepoints for 1740 stocks from 2014 to 2016

5 Conclusion

I detected changepoints in the stock prices in our database. To see the detailed code, go to github.com/jcyang36/Math189R

6 References

[1] Ryan Prescott Adams and David J.C.MacKay. Bayesian Online Changepoint Detection, arXiv 0710.3742 (2007).

[2] Kelsey Craig Anderson. A Novel Approach to Bayesian Online Changepoint Detection, Computer Science Undergraduate Contributions. Paper 23 (2008).

[3] John Hull. *Options, futures, and other derivatives*. Pearson, 2015.

[4] Arun Kejariwal. Introducing practical and robust anomaly detection in a time series. (2015). https://blog.twitter.com/2015/introducing-practical-and-robust-anomaly-detection-in-a-time-series

[5] Bernard Rosner. Percentage Points for a Generalized ESD Many-Outlier Procedure, Technometrics, 25(2) (1983).