Persistent Homology with Stock Prices in Different Sectors

Sierra Cartano

Francis Marion University
Florence, South Carolina
Funded by REAL Grant
Advised by Dr. Ivan Dungan
Overview

• Determine a way to identify the sector* of any random stock
  • Use topological data analysis, specifically, persistent homology.
    • Homology is the mathematics for identifying holes in a shape
    • Persistent homology identifies holes that persist over time.
  • Determine way to identify a random stock by sector with its unique homology
    • Apply to financial data

*Financials, Utilities, Energy, Materials, Industrials, Consumer Discretion, Consumer Staples, Health Care, Information Technology, Telecommunication Services, Real Estate
Example of a Point Cloud
Different Dimensional Holes

• 0-dimensional holes = connected components

• 1-dimensional holes = circles

• 2-dimensional holes = spheres
Example of a Point Cloud
Radius=1
Radius=2
Method of Persistent Homology

- A(-6, -2)
- B(-3.8, -2.2)
- C(-6, 1)
- D(-4, .5)
- E(4, 1)
- F(4.5, -1.5)
- G(11, 1.5)
- H(11.5, -.5)

- Download data into Excel spreadsheet
- Save as text file
- Run through Ripser
Persistence refers to topological features persist over a period of time.
Radius=1, 2, 3
Persistence refers to topological features persist over a period of time.
Circles
BP Prudhoe Bay Royalty Trust

Time Series for BP Prudhoe Bay Royalty Trust

Francis Marion University
BP Prudhoe Bay Royalty Trust (BPT)

\[
SW_{M,t}f(t) = \begin{bmatrix}
  f(t) \\
  f(t + \tau) \\
  \vdots \\
  f(t + M\tau)
\end{bmatrix}
\]

*Blue squares are spheres*
Phillips 66 (PSX)

\[
SW_{M,\tau} f(t) = \begin{bmatrix}
\frac{f(t)}{f(t+\tau)} & \ldots & \frac{f(t+M\tau)}{f(t+M\tau)}
\end{bmatrix}
\]

Birth/Death Plot
BPT vs. PSX Birth/Death Plots

Persistence diagrams can be compared using the bottleneck distance.
Method of Sector Classification

- Companies from energy, financial and technology sectors
- Data gathered from NASDAQ and Yahoo
- 60 companies from each sector
  - 30 training data
  - 30 test data
Ideal Situation

Each point is a persistent diagram from each sector.
Distance between persistence diagrams calculated with bottleneck distance.
Ideal Situation
Output of Results

E P (MIN AVG IQR SSD): [26.67, 3.33, 53.33, 60.00]
F P (MIN AVG IQR SSD): [26.67, 100.00, 23.33, 50.00]
T P (MIN AVG IQR SSD): [36.67, 0.00, 16.67, 3.33]

The program to calculate the bottleneck distance was run 100 times.
Output of Results

E Mean(Min Avg IQRavg SDavg): [26.03, 9.37, 50.60, 33.27]
E SD(Min Avg IQRavg SDavg): [7.28, 24.38, 28.18, 34.80]

F Mean(Min Avg IQRavg SDavg): [34.03, 89.00, 29.17, 61.40]
F SD(Min Avg IQRavg SDavg): [9.02, 28.50, 25.17, 34.70]

T Mean(Min Avg IQRavg SDavg): [28.50, 3.47, 11.57, 5.50]
T SD(Min Avg IQRavg SDavg): [7.81, 10.16, 10.25, 9.64]
Future Work

• What is the correct statistic to describe the results?
• What is the best way to adjust the sliding window function?
  • Is there a way to match up the parameters to produce a valid result?
• Is there a way to include H0, H1 and H2 when running the data?
References

Gunnar Carlsson, February 2005, *Computing Persistent Homology*
Jose A. Perea and John Harer, November 2013, *Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis*
Saba Emrani, Thanos Gentimis, and Hamid Krim, September 2014, *Persistent Homology of Delay Embeddings and its Application to Wheeze Detection*