CSE 6242 Data and Visual Analytics
December 2019
Project Final Report
Peter Butler, Daeil Cha, Yewon Kim, Hyun Jae Park

Introduction - Motivation

We believe the economic value of a company is reflected in its stock price in the long run and as such, traditional fundamental stock research provides the motivation for our project. Pure quantitative investing fails to address fundamental business performance, to utilize deep economic reasoning, and to provide interpretable results. Our method addresses these deficiencies in quantitative stock investing. Our final deliverable is an interactive visualization (see page 5) of a stock trading strategy that utilizes both fundamental and quantitative research methods.

Problem Definition

Our objective is to utilize fundamental research (the primary investigation of a company's business performance) to select stocks systematically. We utilized clustering techniques in the realm of stock investing; our method is novel in that it requires domain expertise in stock analysis. We then visualized the performance and nature of an investment strategy that utilizes our methods.

Literature Survey 1: Current Practice and Our Improvements

Fundamental investors investigate individual stocks through deep research of the underlying company's business performance [1], to discover the value of a stock relative to its price [2]. This approach is the motivation for our project, and we will enhance it with quantitative and visualization techniques.

Quantitative investing selects stocks via their numerical characteristics ("factors") [3]. Below we describe quantitative strategy development and testing; while straightforward to implement, neither requires nor reveals business insight. We addressed this by embedding "fundamental" (business) data into our process. This is critical as it will generate interpretable results to the end user.

- Development: Analysts propose a general idea about the behavior of stock prices [4] by assuming factors influence future prices. There is no business basis for this, so they cannot readily explain the strategy's performance.
- Backtesting: Analysts train a model with a portion of data, then apply it to a test set [5]. They then analyze the return stream generated [6] by calculating Sharpe

ratio (% return per risk taken) [7], or with Monte Carlo analysis [8], which randomizes a return stream. While backtests can be prone to biases and overfitting, analysts can make adjustments to avoid these pitfalls [9].

Recent quantitative techniques use fundamental data, including news articles [10, 11], company annual reports [12], press releases [13], company public filings [14], and management conference calls [15]. However, these methods do not require domain expertise, nor are their results interpretable.

Literature Survey 2: Novelty of Our Approach

We grouped stocks via business model and valuation, and our strategy buys stocks that decline the most relative to their peers. No "one-size-fits-all" metric can group companies that have different business models, debt levels, growth rates, tax rates, and so on. It is critical to group companies based on economic performance by considering metrics like:

- Stock price / earnings per share ("P/E")
- Total enterprise value / EBITDA ("TEV/EBITDA")
- Stock price / book value per share ("P/B")
- Dividend per share / stock price ("Dividend yield")
- Stock price / cash flow per share ("P/CF")

Prior research has clustered stocks [16] with techniques that we may consider, but these still do not embed business information. Our approach to grouping stock time-series data is novel. We consider the following methods:

1. <u>Distance Measures:</u> Euclidean distance, dynamic time warping [17].
2. <u>Clustering Algorithms</u>: K-means [18] and self-organizing maps [19, 20].

Proposed Method

<u>Intuition</u>

Assuming that a company's stock price over time will tend to that company's economic value, we expect to see great returns if we can find a method of identifying 'undervalued' stocks at given points in time, buying them, and reselling them when they rise in value. This is the essence of our approach and the main challenge is identifying which stocks are 'undervalued' and to see if our assumption holds over time. To find undervalued stocks, we employ two novel ideas:

1. Embed fundamental business reasoning into a systematic stock selection process

2. Utilize unsupervised learning clustering algorithms to find similar patterns

Historically, quantitative analysts typically look at only stock price data to make buying and selling decisions. We are not concerned solely with a company's stock price, but rather with its stock price relative to its business performance. This is where we embed fundamental business reasoning, using domain expertise to determine a relevant performance metric for each company; and instead of using just the stock price data, we look at the time series of stock price-to-business performance for each company. We then employ unsupervised clustering algorithms to determine which companies have a similar behavior over time in terms of their price-to-performance ratio.

Companies that are traded on the stock market can belong to one of many sectors, and the most relevant business performance metrics for each sector vary. Thus, we use the most relevant performance metric for each company by sector to determine the price-to-performance ratio over time. We then cluster the stocks within each sector and compare each company's price-to-performance ratio to that of its neighbors within the same cluster. Intuitively, if the price-to-performance ratio for a given company is low relative to its neighbors, we deem that company's stock price to be undervalued and assume that it will rise in the next few months. Thus, our goal is to identify and buy these undervalued stocks so as to sell them in a few months time.

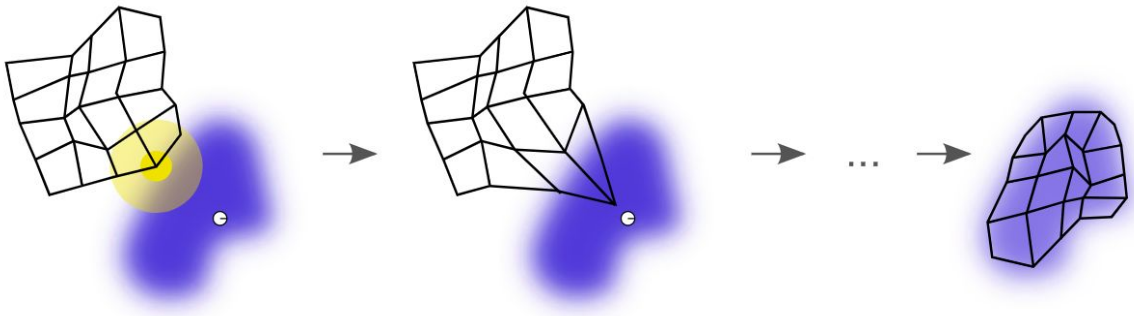
Description of Specific Methods Used

Dynamic Time Warp (DTW): This is an algorithm used to find the ideal lineup of two time series. It has the advantage of correcting for similarities that may be at different ticks. We experimented with this method but found it much too time intensive for our purposes. DTW scales with $O(n^2)$, making it a very slow task given the lengths of our time series.

Euclidean Distance: This is the most standard distance metric used for clustering. It ran quickly and gave good results.

K-Means Clustering: This is one of the most commonly used unsupervised clustering algorithms. Given the initial number of cluster centers as in input, the algorithms uses successive iterations of expectation-maximization to determine a locally optimal positioning of the cluster centers. This method is a heuristic, so it is not guaranteed to give the globally optimal positioning, but it generally performs well. Data points can then be clustered by the closest center using any distance metric.

Self-Organizing Map (SOM): This is a neural network algorithm that can be used to produce a two-dimensional representation of data, and to cluster. Given an initial neuron grid, the grid is iteratively updated to fit the data as shown below. The data can then be assigned cluster labels based on which neuron they are closest to.



<u>User Interface</u>

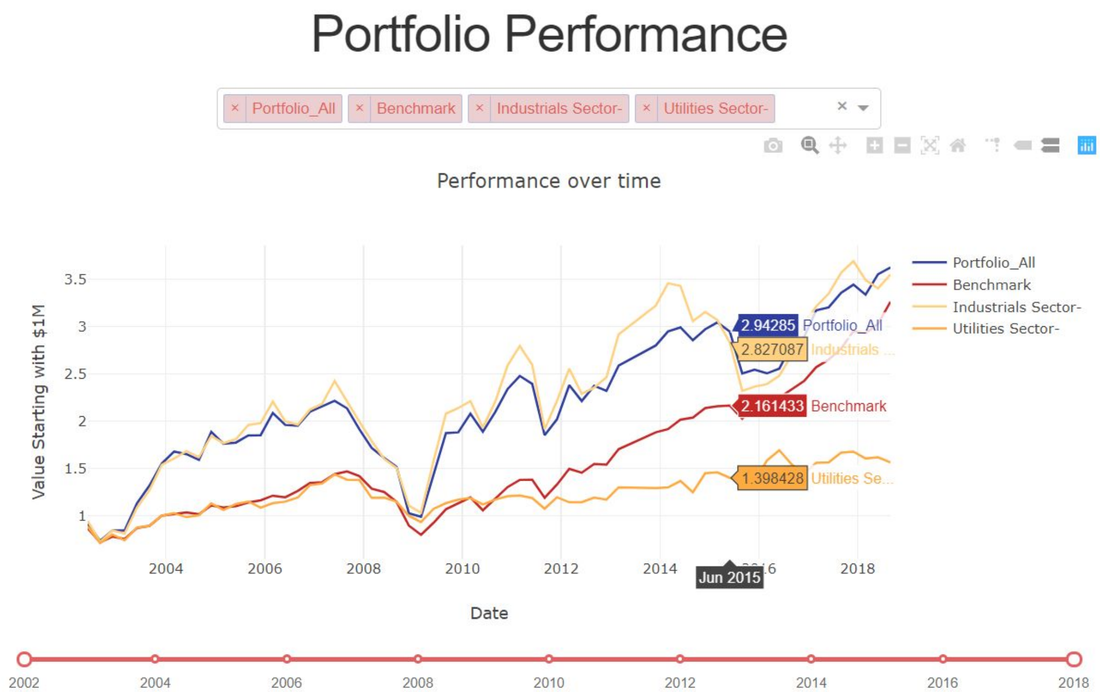
We created an interactive dashboard using the Python toolkit Dash, which is built on the popular visualization platform Plotly. We visualized the performance of an investment strategy that utilizes our methods, and compared it to the returns of a benchmark (the S&P 500 index).

We also visualized the nature of our strategy; our user interface plots the performance of our portfolio by sector over time. This gives the user a sense of dislocations in the broad stock market and business ecosystem. This may also further spark investment ideas for a discretionary investor.

We utilized line charts as it is the most basic and effective way to show the trend of our performance. The X-axis of our graph is the date ranging from 1993 to 2018 and Y-axis shows the monetary value starting from $1M.

The user interface has three main interactive features.

1. Dropdown to select which line charts to appear on the graph. We selected a color scheme so that the two most important lines (total portfolio & benchmark) are highlighted.
2. Mouseover effect to depict number values at a specific time.
3. Timeline Slider so that the user can zoom into a period of time.



Design of Experiments

Description of testbed: List of questions your experiments are designed to answer

1. How should we measure the effectiveness of our grouping techniques?
2. Does our technique(s) work as a long-only stock investment strategy?
3. What is the risk/return of our trading strategy?
4. How robust is the strategy? Does its effectiveness diminish over time?
5. What adjustments can we make to our techniques to keep return high?
6. What parameters of this approach can be optimized? (number of clusters, etc)
7. What is the nature of the 'portfolio' generated by our strategy? Define in terms of industry, size, profitability, valuation levels, etc. What does this tell us about the idiosyncratic, fundamental performance of businesses that trade at 'cheap' levels?
8. Look at significant historical market events (booms, busts, geopolitical conflict, etc). Do these events impact our investment strategy?

Data Collection, Cleaning, and Wrangling

Our approach required stock valuation data for the universe of U.S. stocks, typically in the form of:

(Stock price) / (Estimate of financial performance)

We downloaded data from Wharton Research Data Services (WRDS), a comprehensive data service that provides access to several databases. The data included :

- Performance estimates (1GB) - from IBES
- Historical quarterly financial data (250MB) - from Compustat
- Stock metadata (100MB) - from IBES, Compustat, CRSP
- Daily stock and index prices (3.5GB) - from CRSP

All stock and index prices were adjusted for dividends and stock splits. Nearly all data (except the metadata) were time series and ranged from 1970 to 2019. Data from ~16,000 stocks were collected. While the required data was readily available, we still had to perform comprehensive cleaning and wrangling. That is, while a stock's price and earnings estimate data were downloaded, we had to construct price/estimate ratio data manually, and each sector required its own specific estimate metric. Generally, we performed this with the following logic:

(Price at date T) / (Earnings estimate available prior to and up to date T)

Further, there were some earnings estimates that were too old (or 'stale') to use for a price; we eliminated estimates that were older than each price by 12 months.

This was a challenging, meticulous process.

Design of experiments:

1. Separate dataset by sector ('gsector' code, 11 in total). This step is essential in accounting for the fundamental business characteristics of stocks.
2. Separate sector datasets by a regular time interval. Separate each of these 11 datasets into increments of 3 months (or "quarter").
3. Clean each dataset. Within each of these quarters, remove missing data and normalize the rest. We found that if data was missing, it was either at the start or the end of our training period, meaning that either the stock has not been on the market for the full training period, or that stock was taken off the market before the end of the training period. In either of these cases, we do not want to consider these stocks, so we just removed any stocks that had missing entries.
4. Group the stocks. For each sector dataset, group stocks that behave most similarly to each other based on the past 12 months of data. Here, we tried several values for the number of clusters 'k.'
5. Identify cheap stocks. At the end of each calendar quarter, identify stocks (across **all** sectors) that are trading below their normal relationship to their respective peers. We tried different cut-off values for the lowest percent of stocks to buy.

6. Backtest trading strategy. Assess the investment performance of a trading strategy that simulates our techniques. The trading strategy is as follows:
    a. Rebalance every three months.
    b. Apply a 1% penalty for every round-trip trade to account for slippage and trade execution costs [22]. We believe this estimate to be on the high-end of costs compared to historical observations [23], but we utilized it to be conservative in our methods.
    c. Assume equal weighting of stock purchases at each iteration.
    d. Measure the performance of the portfolio every three months and compare it to a benchmark.

Evaluation and Insights

Overall, we observed that the k-means algorithm provided the best results for our trading strategy. Further, we found that using 12 clusters per sector with a 'cheapness' cutoff of 10% performed the best in the backtest. From inception to date (1993 to 2018), our strategy produced a 15x return on investment, compared to an 11x return for the S&P 500 index. We observe that our strategy underperformed the benchmark during periods of market decline, while it recovered strongly as the market recovered. The intuition for this is based on our strategy as a 'reversion' approach; that is, our approach presumes that stocks that decline relative to their peers will return to their normal levels. This is problematic if those stocks do not 'revert,' and instead continue to decline. This is precisely what happens when the broader market experiences a prolonged decline; our strategy significantly underperforms during these periods.

1. From March 2002 to September 2002, our portfolio saw a drawdown of 25.3% (versus the S&P 500's decline of 13.8% in the same period).
2. From June 2007 to March 2009, our portfolio saw a drawdown of 55.3% (versus the S&P 500's decline of 44.6%).
3. From March 2014 to September 2015, our portfolio saw a drawdown of 17.8% (versus the S&P 500's decline of 6.1%)

By contrast, our portfolio's returns are strong when the broader market is stable.

1. From September 2002 to June 2008, our portfolio saw a return of 202.2% (versus the S&P 500's rise of 100.5% in the same period).
2. From March 2009 to March 2011, our portfolio saw a return of 150.5% (versus the S&P 500's rise of 73.2%).

Insights and Interpretable Results

To explain our results in terms of business characteristics (another novelty of our approach), we embedded fundamental data into our backtest. The data included were:

1. Sector. This is the broader business vertical that a company operates in.
2. Market capitalization (size of the company).
3. Leverage ratio (debt to assets). This measures how much debt a company owes. While some debt can help a company grow, too much debt can create financial risk during poor business performance.
4. Liquidity ratio (current assets divided by current liabilities). This represents how easily a company can pay its short term liabilities.
5. Operating income growth.
6. Operating margin (operating income divided by sales)
7. Sales growth.

As such, we can determine some potential sources of our strategy's performance. For instance, if we investigate our underperformance from June 2007 to March 2009, this period was the onset and aftermath of the financial crisis of 2008. *Prior to the crisis* (which we define as starting on September 2008), our strategy bought stocks of companies that:

- Were large relative to other companies in their sector
- Had high levels of debt (relative to assets) compared to their sector
- Saw high earnings growth compared to their sector
- Saw high operating margins compared to their sector

During and after the crisis, our strategy purchased stocks of companies that:

- Were small relative to other companies in their sector
- Had somewhat lower levels of debt (relative to assets) compared to their sector
- Saw slower earnings growth compared to their sector
- Saw low operating margins compared to their sector

And once the crisis occurred, Our strategy also bought more companies in the industrial sector at the expense of companies in the financial sector. One interpretation of these results is that prior to the crisis, the market assumed the large, highly indebted companies were most at risk -- and therefore those stocks underperformed. But once the public realized the crisis was systemic, the market punished the smaller companies that had mediocre business performance. Another interpretation is that small, struggling companies with less resources and cost flexibility suffered more in the crisis, and thus their stocks tumbled. This business disruption was severe and long lasting, and those stocks continued to underperform (thus hurting our performance).

For detailed visualizations of our portfolio in terms of business characteristics, see the Appendix.

Conclusions and Discussion

The main takeaway from our project is that <u>when paired with domain expertise, k-means is an effective method to cluster stock time-series data</u>. The self-organizing map had good performance but not as strong as that of k-means. One area of further investigation is exploring other clustering methods like affinity propagation and DBSCAN. We could also consider ways to enhance the strategy's performance by restricting our investment to certain industries or companies with specific business characteristics. We found that the majority of our outperformance was driven by the communications, healthcare, information technology, and materials sectors. Further, because our strategy underperforms when the broader market declines consistently, we could implement risk management procedures to manage underperformance. For instance, we could freeze trading activity when if we believe the market will be weak for an extended period of time.

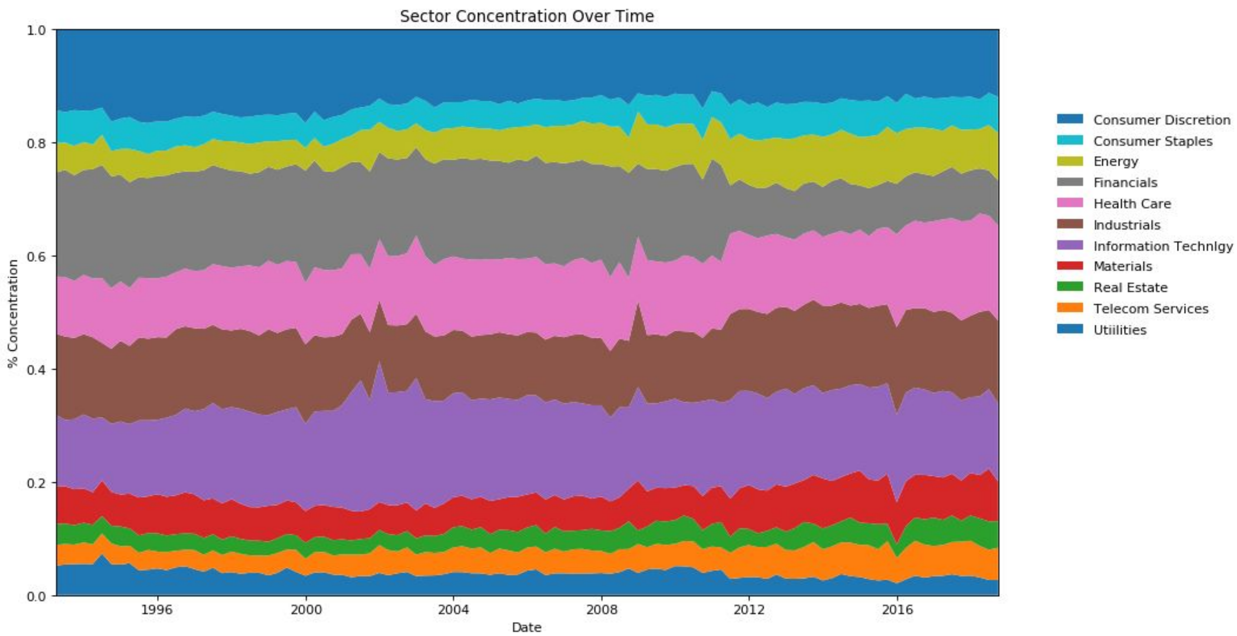
References

1. Klarman, Seth A. "Speculators and Unsuccessful Investors," in *Margin of Safety: Risk-Averse Value Investing Strategies for the Thoughtful Investor*, 1-55. New York, NY: HarperBusiness, 1991.
2. Graham, Benjamin, Warren E. Buffett, and Jason Zweig. "Security Analysis for the Lay Investor: General Approach," in *The Intelligent Investor: The Definitive Book on Value Investing*, 280-309. HarperCollins, 2015.
3. Fama, Eugene F., and Kenneth R. French. "Common risk factors in the returns on stocks and bonds." *Journal of financial economics* 33, no. 1 (1993): 3-56.
4. Pardo, Robert. "Trading System Development," in *Design, Testing, and Optimization of Trading Systems*, 6-28. Vol. 2. John Wiley & Sons, 1992.

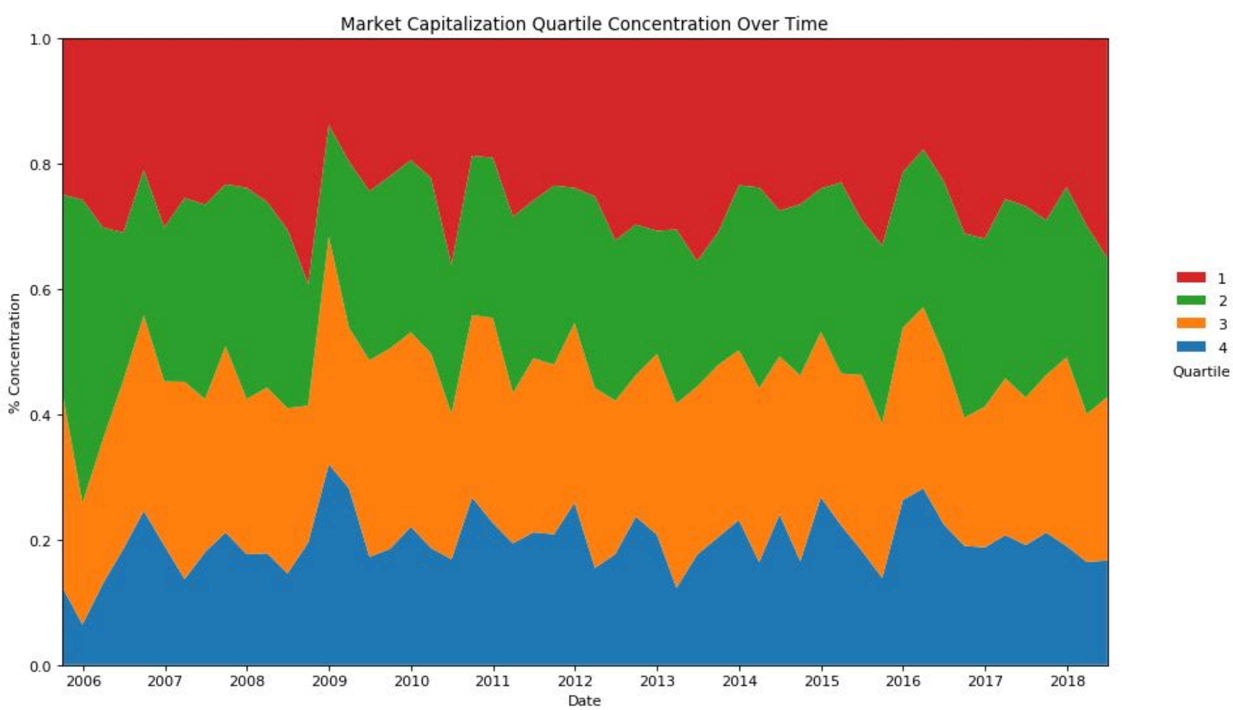
5. Davey, Kevin J. “Testing and Evaluating a Trading System,” in *Building Winning Algorithmic Trading Systems: A Trader's Journey from Data Mining to Monte Carlo Simulation to Live Trading*, 43-52. John Wiley & Sons, 2014.
6. Harvey, Campbell R., and Yan Liu. "Evaluating trading strategies." *The Journal of Portfolio Management* 40, no. 5 (2014): 108-118.
7. Sharpe, William F. "The Sharpe Ratio." *Journal of Portfolio Management* 21, no. 1 (1994): 49-58.
8. Davey, Kevin J. “Testing and Evaluating a Trading System,” in *Building Winning Algorithmic Trading Systems: A Trader's Journey from Data Mining to Monte Carlo Simulation to Live Trading*, 61-70. John Wiley & Sons, 2014.
9. Bailey, David H., and Marcos López de Prado. "The deflated Sharpe ratio: correcting for selection bias, backtest overfitting, and non-normality." *The Journal of Portfolio Management* 40, no. 5 (2014): 94-107.
10. Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy. "More than words: Quantifying language to measure firms' fundamentals." *The Journal of Finance* 63, no. 3 (2008): 1437-1467.
11. Hillert, Alexander, Heiko Jacobs, and Sebastian Müller. "Media makes momentum." *The Review of Financial Studies* 27, no. 12 (2014): 3467-3501.
12. Li, Feng. "Annual report readability, current earnings, and earnings persistence." *Journal of Accounting and Economics* 45, no. 2-3 (2008): 221-247.
13. Henry, Elaine. "Are investors influenced by how earnings press releases are written?." *The Journal of Business Communication* (1973) 45, no. 4 (2008): 363-407.
14. Feldman, Ronen, Suresh Govindaraj, Joshua Livnat, and Benjamin Segal. "Management’s tone change, post earnings announcement drift and accruals." *Review of Accounting Studies* 15, no. 4 (2010): 915-953.
15. Brockman, Paul, Xu Li, and S. McKay Price. "Differences in conference call tones: Managers vs. analysts." Financial Analysts Journal 71, no. 4 (2015): 24-42.
16. Gavrilov, Martin, Dragomir Anguelov, Piotr Indyk, and Rajeev Motwani. "Mining the stock market: Which measure is best." In *Proc. of the 6th ACM SIGKDD*, pp. 487-496. 2000.
17. Salvador, Stan, and Philip Chan. "Toward accurate dynamic time warping in linear time and space." *Intelligent Data Analysis* 11, no. 5 (2007): 561-580.
18. Rashidi, Parviz, M. Analoui, and Javad Azizmi. "Modified k-means algorithm for clustering stock market companies." In *1st Iran Data Mining Conference*, pp. 201-21. 2007.
19. Kohonen, Teuvo. "The self-organizing map." *Proceedings of the IEEE* 78, no. 9 (1990): 1464-1480.

20. Fu, Tak-chung, Fu-lai Chung, Vincent Ng, and Robert Luk. "Pattern discovery from stock time series using self-organizing maps." In *Workshop Notes of KDD2001 Workshop on Temporal Data Mining*, pp. 26-29. 2001.
21. Damodaran, Aswath. "Valuing Financial Service Firms" in *Investment valuation: Tools and techniques for determining the value of any asset. Vol. 666*, 581-610. John Wiley & Sons, 2012.
22. Chan, Louis KC, and Josef Lakonishok. "Institutional equity trading costs: NYSE versus Nasdaq." *The Journal of Finance* 52, no. 2 (1997): 713-735.
23. Robert, Engle, Ferstenberg Robert, and Russell Jeffrey. "Measuring and modeling execution cost and risk." **The Journal of Portfolio Management** 38, no. 2 (2012): 14-28.

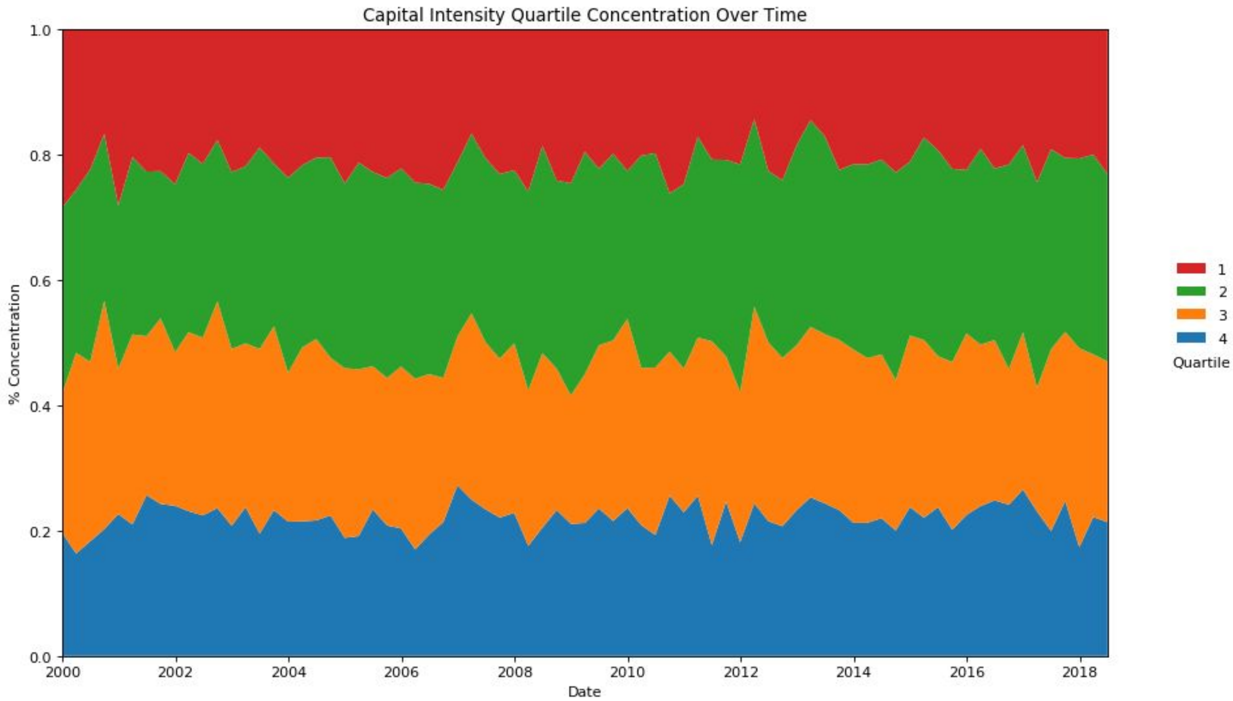
Appendix



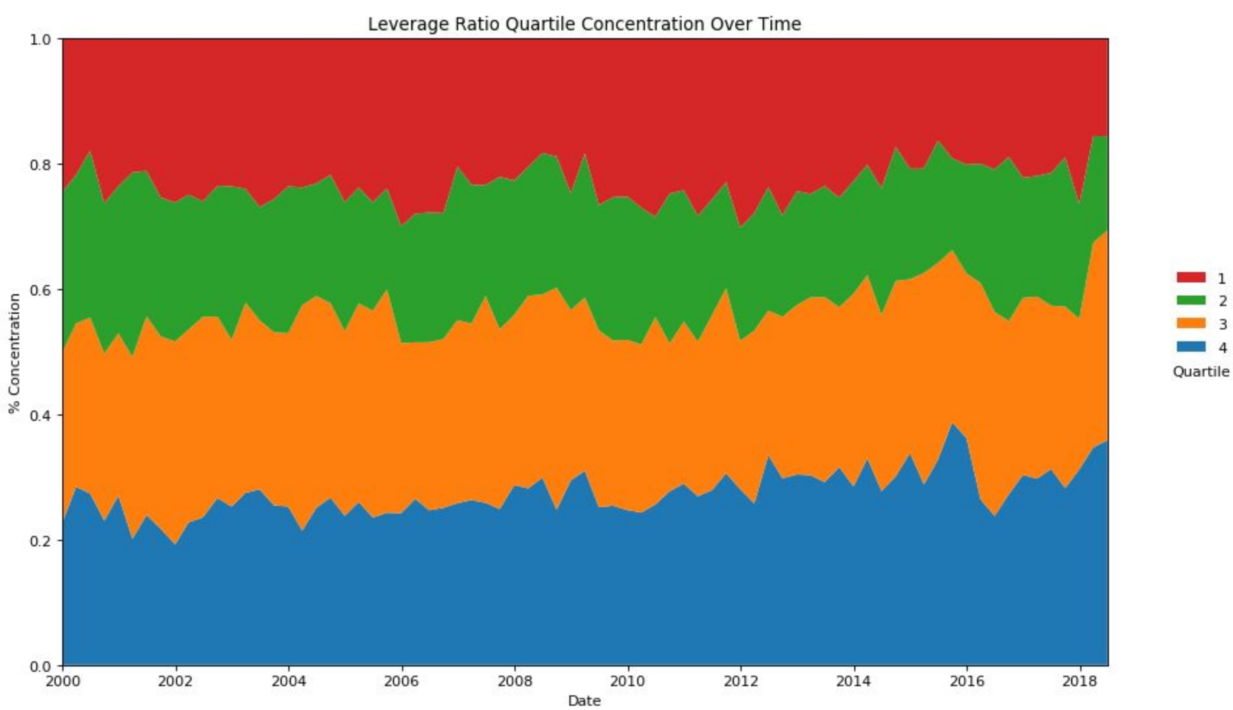
Below, 1st quartile denotes stocks that have the largest market capitalizations relative to their respective sector.



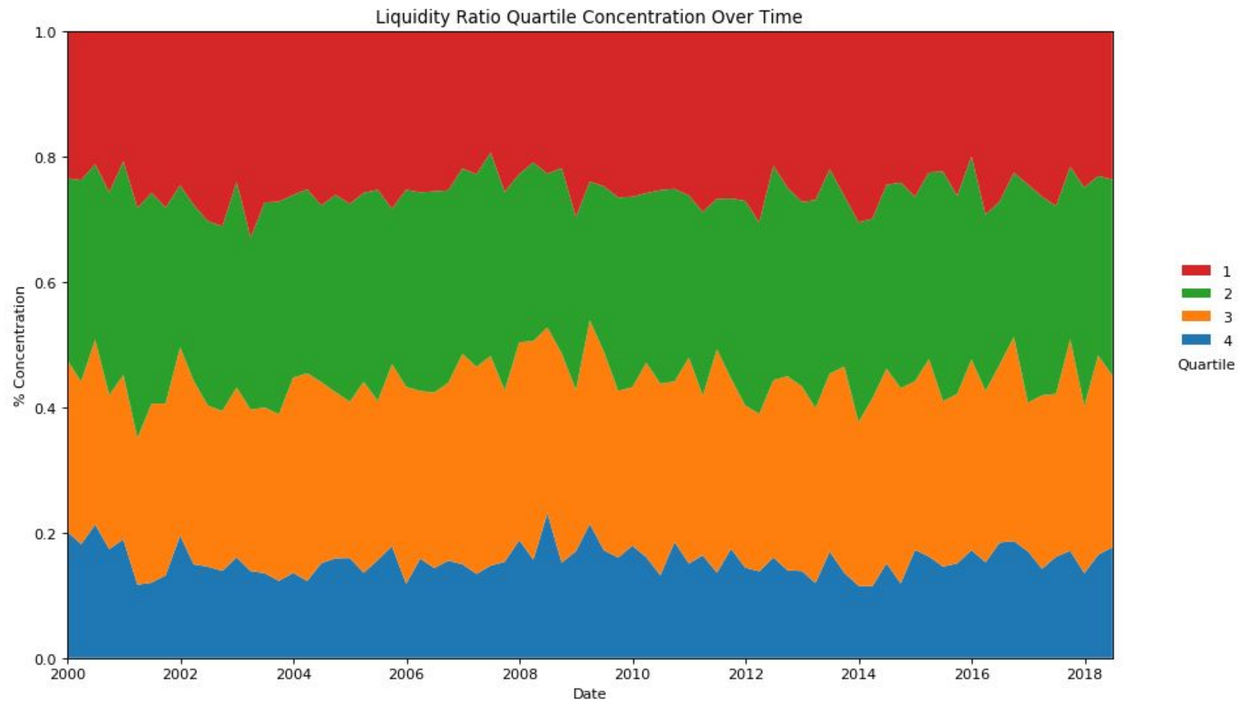
Below, 1st quartile denotes stocks that have the lowest capital intensity (capital expenditures / sales) relative to their respective sector.



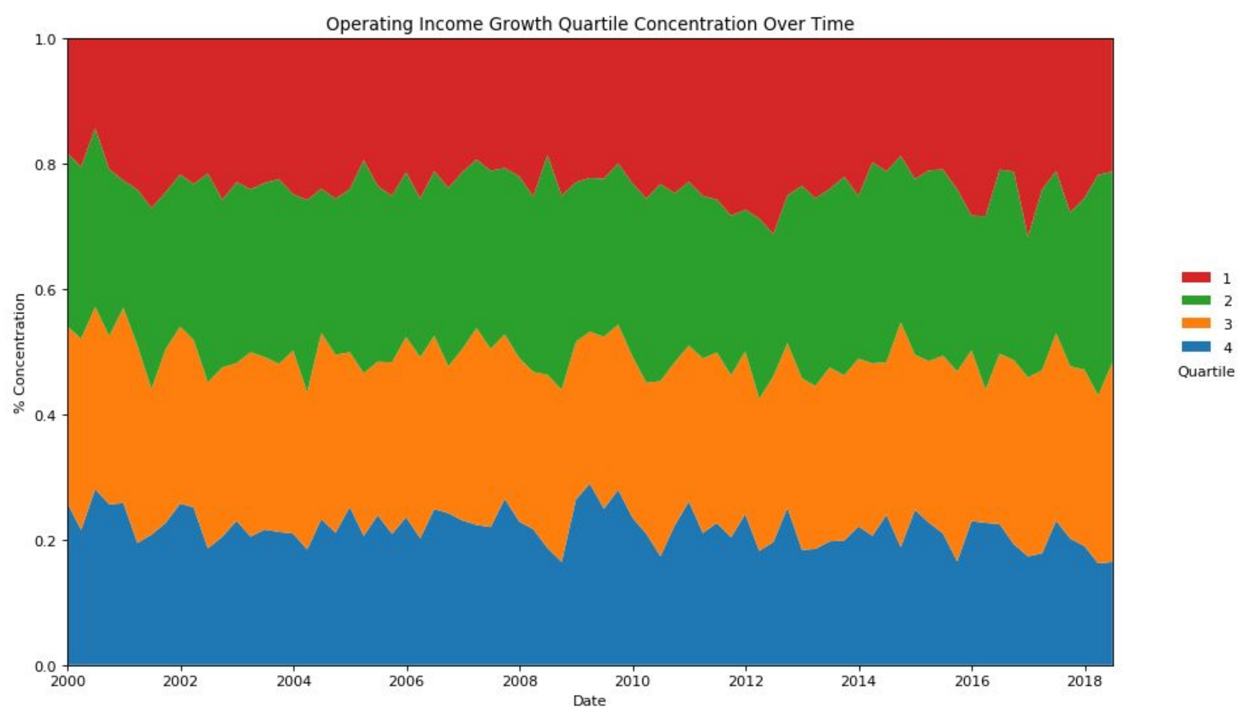
Below, 1st quartile denotes stocks that have the lowest leverage ratio (debt / assets) relative to their respective sector.



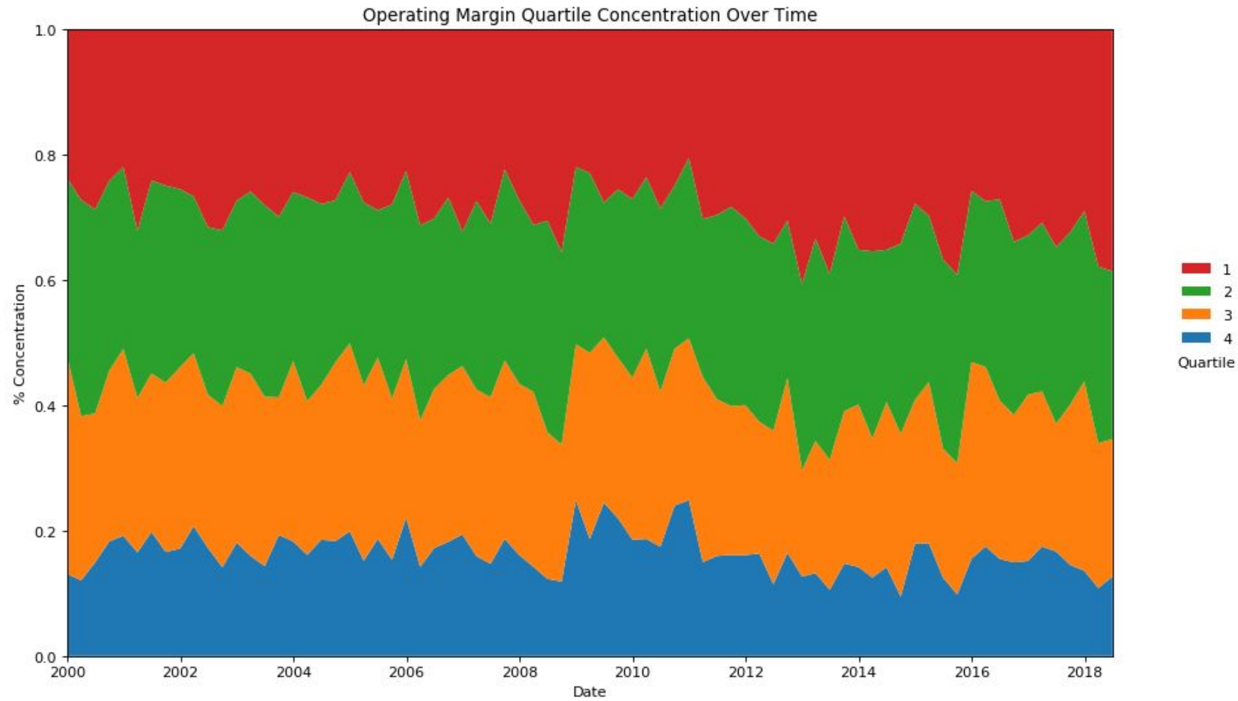
Below, 1st quartile denotes stocks that have the highest liquidity ratio (current assets / current liabilities) relative to their respective sector.



Below, 1st quartile denotes stocks that have the highest quarterly operating income growth (year-over-year) relative to their respective sector.



Below, 1st quartile denotes stocks that have the highest quarterly operating margin (operating income / sales) relative to their respective sector.



Below, 1st quartile denotes stocks that have the highest quarterly sales growth (year-over-year) relative to their respective sector.