

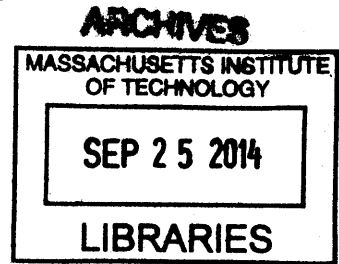
# LEARNING CONNECTIONS IN FINANCIAL TIME SERIES

by

Gartheeban Ganeshapillai

S.M., Massachusetts Institute of Technology (2011)

B.Sc., University of Moratuwa (2009)



Submitted to the Department of Electrical Engineering and Computer Science

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2014

© Massachusetts Institute of Technology 2014. All rights reserved.

## Signature redacted

Author .....  
Department of Electrical Engineering and Computer Science  
August 26, 2014

Certified by. **Signature redacted** .....  
John V. Guttag  
Professor, Electrical Engineering and Computer Science  
Thesis Supervisor

**Signature redacted**  
Accepted by .....  
/ UU Leslie. A. Kolodziejski  
Chair of the Committee on Graduate Students



# LEARNING CONNECTIONS IN FINANCIAL TIME SERIES

by

Gartheeban Ganeshapillai

Submitted to the Department of Electrical Engineering and Computer Science  
on August 26, 2014, in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

## Abstract

Much of modern financial theory is based upon the assumption that a portfolio containing a diversified set of equities can be used to control risk while achieving a good rate of return. The basic idea is to choose equities that have high expected returns, but are unlikely to move together. Identifying a portfolio of equities that remain well diversified over a future investment period is difficult.

In our work, we investigate how to use machine learning techniques and data mining to learn cross-sectional patterns that can be used to design diversified portfolios. Specifically, we model the connections among equities from different perspectives, and propose three different methods that capture the connections in different time scales. Using the “correlation” structure learned using our models, we show how to build selective but well-diversified portfolios. We show that these portfolios perform well on out of sample data in terms of minimizing risk and achieving high returns.

We provide a method to address the shortcomings of correlation in capturing events such as large losses (*tail risk*). Portfolios constructed using our method significantly reduce tail risk without sacrificing overall returns. We show that our method reduces the worst day performance from  $-15\%$  to  $-9\%$  and increases the Sharpe ratio from 0.63 to 0.71. We also provide a method to model the relationship between the equity return that is unexplained by the market return (excess return) and the amount of sentiment in news releases that hasn't been already reflected in the price of equities (excess sentiment). We show that a portfolio built using this method generates an annualized return of  $34\%$  over a 10-year time period. In comparison, the S&P 500 index generated  $5\%$  return in the same time period.

Thesis Supervisor: John V. Guttag

Title: Professor, Electrical Engineering and Computer Science



## Acknowledgments

To quote Jeffrey Archer, “It is often spur-of-the-moment decisions, sometimes made by others, that can change our whole lives.” I didn’t apply to work with John Guttag nor did I know about his research when he offered to take me into his research group. I accepted that offer, an exercise in which I had little choice, and I was driven by the urge to pursue my doctoral studies at MIT. In hindsight, for a decision made with very little information, it turned out to be the best decision I could have made. I couldn’t have asked for a better advisor.

Over the last five years, John has been an amazing teacher, mentor, colleague, friend, and family. He simply loves what he is doing, and that makes all the difference. I will cherish the rapport I developed with him forever.

None of the work presented in this thesis would have been possible without his inspiration and guidance. His positive influence on my life, growth, and personality extends beyond the realm of research. It has been an absolute dream to have worked with John. I am looking forward to have this pleasure so much more in the years to come.

I have been fortunate to have an amazing set of people in my thesis committee. The advice and guidance of Prof. Andrew Lo had a direct impact on this work from the very beginning. I have to also thank Prof. Tommi Jaakkola and Dr. Bradley Betts for serving on my thesis committee.

This thesis would not have been possible without the help of many amazing people I worked with at MIT. In my research group, I had the fortune of working with Prof. Collin Stultz, Zeeshan Syed, Ali Shoeb, Eugene Shih, Al Kharbouch, Jenna Wiens, Anima Singh, Guha Balakrishnan, Joel Brooks, Yun Liu, Jen Gong, Amy Zhao, Jessica Liu, and Mitchell Kates. Numerous conversations, ideas, and feedback I got from them during our group meetings and other discussions have been invaluable. I owe a great deal to Dorothy Curitis, Sheila Marian, Janet Fischer and many others at CSAIL, especially those from the 9th floor of the STATA center, Prof. Sam Madden, Eugene Wu, and Alvin Cheung.

I am indebted to many at the University of Moratuwa, and Saman Amarasinghe of MIT. Their support and guidance enabled me to come to MIT for my PhD.

It would be impossible to mention all of my other friends at MIT, but I thank Satchuthananthavale Branavan, Oshani Seneviratne, Sachithra Hemachandra, Narmada Herath, Charith Mendis, Savithru Jayasinghe, Prashan Wanigasekara, and Wardah Inam, in particular, for being part of my life during my stay at MIT. I mention John Cheng separately for I owe a great deal to him.

Finally, I dedicate my efforts to my family, especially my parents, and my grandparents, who have made many sacrifices to allow me to be where I am today. Finally, I owe a great debt to my lovely wife, without whose consistent support, I wouldn't have remained far from home for this many years, let alone finish my PhD.

# Contents

<b>1</b>	<b>INTRODUCTION</b>	<b>15</b>
1.1	Motivation . . . . .	16
1.2	Background . . . . .	17
1.3	Proposed Approaches . . . . .	18
1.4	Contributions . . . . .	19
1.5	Organization of Thesis . . . . .	22
<b>2</b>	<b>BACKGROUND</b>	<b>23</b>
2.1	Finance . . . . .	23
2.1.1	Reward - Risk . . . . .	24
2.1.2	Portfolio . . . . .	25
2.2	Data . . . . .	30
2.2.1	Return Data . . . . .	30
2.2.2	News Data . . . . .	31
<b>3</b>	<b>LEARNING CONNECTIONS IN TERMS OF LARGE LOSSES</b>	<b>37</b>
3.1	Introduction . . . . .	37
3.2	Related Work . . . . .	39
3.2.1	Correlation and Partial Correlation . . . . .	39
3.2.2	Understanding Connections on Extreme Returns . . . . .	40
3.3	Method . . . . .	41
3.3.1	Problem Formulation . . . . .	41
3.3.2	Model . . . . .	41

3.3.3	Similarity to Other Methods . . . . .	43
3.3.4	Connectedness Matrix . . . . .	44
3.4	Evaluation . . . . .	46
3.4.1	Data . . . . .	47
3.4.2	Top-K Ranking . . . . .	47
3.4.3	Portfolio Construction . . . . .	53
3.5	Application Considerations . . . . .	64
3.6	Summary . . . . .	66
<b>4</b>	<b>LEARNING THE RELATIONSHIP BETWEEN EQUITY RETURNS AND NEWS RELEASES</b>	<b>69</b>
4.1	Introduction . . . . .	70
4.2	Related Work . . . . .	71
4.2.1	Efficient Market Hypothesis and Events . . . . .	71
4.2.2	Behavioral Finance and Data Mining . . . . .	71
4.2.3	Sentiment Mining Applied to Finance . . . . .	72
4.3	Method . . . . .	73
4.3.1	Problem Formulation . . . . .	73
4.3.2	Modeling Excess Return . . . . .	74
4.3.3	Variance Decomposition . . . . .	77
4.4	Experiments . . . . .	78
4.4.1	Data . . . . .	79
4.4.2	Methods . . . . .	79
4.4.3	Top-k Ranking . . . . .	80
4.4.4	Portfolio Construction . . . . .	84
4.5	Application Considerations . . . . .	87
4.6	Interday returns . . . . .	89
4.7	Summary . . . . .	90
<b>5</b>	<b>LEVERAGING NETWORK STRUCTURE IN FINANCIAL TIME SERIES</b>	<b>93</b>
5.1	Introduction . . . . .	93



5.2	Related Work . . . . .	96
5.3	Financial Filtered Networks . . . . .	97
5.3.1	Methods . . . . .	97
5.3.2	Empirical Observations . . . . .	99
5.4	Separating Peripheral and Central Regions . . . . .	102
5.4.1	Methods . . . . .	102
5.4.2	Empirical Observations . . . . .	104
5.4.3	Portfolio Returns . . . . .	107
5.5	Graph Clustering . . . . .	111
5.5.1	Methods . . . . .	111
5.5.2	Empirical Observations . . . . .	114
5.5.3	Portfolio Returns . . . . .	116
5.6	Graph Sparsification . . . . .	116
5.6.1	Method . . . . .	117
5.6.2	Portfolio Returns . . . . .	117
5.7	Application Considerations . . . . .	117
5.8	Summary . . . . .	118
<b>6</b>	<b>CONCLUSION AND FUTURE WORK</b>	<b>121</b>
6.1	Future Directions . . . . .	123



# List of Figures

2-1	Efficient frontier and the optimal portfolio choices . . . . .	25
2-2	The Sharpe ratio Vs. Information ratio. . . . .	29
2-3	Snapshot of the relevant fields in Newscope dataset. . . . .	32
2-4	Historical distribution of the number of feeds in Newscope dataset . .	33
2-5	Historical distribution of the number of equities covered in Newscope dataset . . . . .	34
2-6	Historical distribution of the number of feeds across sectors in News- cope dataset . . . . .	35
2-7	Positive and negative sentiments over time . . . . .	36
2-8	Hourly variation of positive and negative sentiments . . . . .	36
3-1	Comparing daily returns of Bank of America (BAC) with that of S&P500 Index (SPX) . . . . .	51
3-2	Factor model for Bank of America . . . . .	52
3-3	Return characteristics for energy sector . . . . .	57
3-4	Portfolio holdings for energy sector . . . . .	58
3-5	Temporal evolution of connections . . . . .	60
4-1	Timeline of return calculation for intraday returns . . . . .	75
4-2	Mean average precision at different cutoffs . . . . .	82
4-3	Mean average precision at different event levels . . . . .	82
4-4	Explained variance . . . . .	83
4-5	Return characteristics of the <i>LONG-Weighted</i> portfolio . . . . .	86

4-6	The annualized returns and information ratios for different transaction costs . . . . .	88
4-7	The annualized returns and information ratios for holding sizes . . . .	88
4-8	Timeline of return calculation for interday returns . . . . .	90
4-9	Interday return characteristics under different lags . . . . .	90
5-1	A financial filtered network . . . . .	94
5-2	Example clustering of a financial filtered network . . . . .	94
5-3	Minimal spanning tree network generated by different linkages for 2001-2002 . . . . .	100
5-4	The network topologies of correlation and $FAC_{10}$ during the 2008 financial crisis . . . . .	101
5-5	The 10 most central and the 10 most peripheral equities as identified for each year from 2003 to 2012 by different topological measures . .	104
5-6	Composition of the 10 most central equities and the 10 most peripheral equities . . . . .	105
5-7	Sector composition of the 10 most central equities and the 10 most peripheral equities . . . . .	106
5-8	Minimal spanning tree network generated with $FAC_{10}$ as linkage metric for 2007-2008 . . . . .	108
5-9	The characteristics of the equities in the 10 most central and the 10 most peripheral equities in the network . . . . .	109
5-10	Mean and standard deviation of daily returns of the 10 most central equities and the 10 most peripheral equities . . . . .	110
5-11	Histogram of cluster sizes for clusters identified by spectral clustering	113
5-12	Sector composition of the clusters for years 2003-2012 . . . . .	115

# List of Tables

3.1	Connectedness matrix construction . . . . .	45
3.2	MAP scores for different methods. . . . .	49
3.3	MAP scores for different sizes of unknown set $B$ . . . . .	50
3.4	Characteristics of portfolio returns . . . . .	56
3.5	Market wide portfolios. . . . .	59
3.6	Portfolio returns . . . . .	61
3.7	Impact of survivorship bias on Market wide portfolios: 2001-2012. . .	63
3.8	Market wide portfolios: 2001-2012. . . . .	63
3.9	Market wide portfolios: 2013. . . . .	64
4.1	MAP scores across sectors for events based on returns. . . . .	81
4.2	Portfolio characteristics. . . . .	85
5.1	Correlation matrix of topological measures for year 2012 . . . . .	103
5.2	Central/peripheral portfolios . . . . .	111
5.3	Cluster connectivity . . . . .	114
5.4	Cluster index portfolios . . . . .	116
5.5	Cluster index portfolios with graph sparsification . . . . .	118



# Chapter 1

## INTRODUCTION

Much of modern financial theory is based upon the assumption that a portfolio containing a diversified set of equities can be used to control risk while achieving a good rate of return. Identifying a portfolio of equities that remain well diversified over a future investment period is difficult. The main reason is that achieving diversification benefits depends upon estimating the “correlation” structure among asset returns.

In this dissertation, we present novel ways to leverage data to learn cross-sectional connections in multivariate financial time series. By cross-sectional connections, we mean the connections across different variables in a multivariate time series, as opposed to connections on each time series across different time periods. We show how these connections can be used to construct a selective but well-diversified portfolio of equities. We also show that these portfolios perform well on out of sample data in terms of minimizing risk and achieving high returns.

We particularly explore three aspects of the data-driven modeling in time series: predicting large price changes in response to news, modeling connections between equities focusing on the co-occurrences of large losses, and using network structure to model connections that span multiple equities, such as clusters and hierarchies in the network. Our research uses techniques drawn from machine learning and data mining coupled with the understanding of the underlying financial economics.

## 1.1 Motivation

When investing in an asset, an investor is mainly concerned with the return, and, because it is impossible to know the future return of the asset, with the risk of not receiving the expected return. Furthermore, since people prefer to avoid risk, more risky equities typically have higher returns. Much of modern financial theory is based upon the assumption that a portfolio containing a diversified set of equities can be used to control risk while achieving a good rate of return.

Diversification is one of two general techniques for reducing investment risk (the other is hedging). Diversification reduces the risk associated with an investment because the likelihood that a portfolio of equities will perform poorly during an investment period is smaller than the likelihood of a single equity going down.

Given the advantages of diversification, maximum diversification, also known as “buying the market portfolio,” has been widely recommended [86, 45]. Indeed for many investors an index fund is the best way to invest in equities. However, this is inappropriate in a world in which risk preference plays an important role. Some investors, for example, hedge fund managers, expect high returns, and in exchange, expect to bear corresponding risks. Hence, selective as opposed to full diversification is important [40, 94, 42].

A simple, ancient, well-known, and still useful method of diversification is allocating wealth equally across asset classes. In an Aramaic text from the 4th Century, Rabbi Isaac bar Aha proposed: “*A third in land, A third in merchandise, A third at hand*” [27, 87]. Centuries later, mathematical analysis of portfolio diversification within an asset class began with the seminal work of Markowitz in 1952, which gives the optimal rule for allocating wealth across risky assets in a setting where investors care only about the mean and variance of a portfolio’s return.

Over the next six decades, there have been several methods proposed to provide an improved diversification of equities [7, 50, 71, 43, 97, 38, 27]. The underlying problem common across these methods is that the statistical moments estimated using historical returns perform poorly on out of sample data. It is because the estimates



of the statistical averages are so noisy that the portfolio design choices often produce little difference in the outcome [43]. DeMiguel, et al (2009) [27] provides a study on comparing these improved diversification methods, and concludes that none of the portfolios evaluated performed consistently better than an equal weighted portfolio. Furthermore, citing the evaluation results, it recommends the cross-sectional characteristics of assets as a direction worth pursuing [27].

Identifying a portfolio of equities that remains well diversified over a future investment period is difficult. The main reason is that the diversification benefits depend on the “correlation” structure among equity returns [53], and the structure is multifaceted, and always evolving. In our work, we investigate using machine learning techniques and data mining to learn cross-sectional patterns across equity returns, and ways of using these patterns to build selective but well-diversified portfolios that perform well on out of sample data. Specifically, we model connections from different perspectives, and propose applications that capture the connections in different time scales. For example, we consider the task of diversifying against *tail risk*, the risk of improbable but potentially catastrophic negative return events, and the task of learning the influence of news on the returns of equities.

## 1.2 Background

In this section, we provide a short background to portfolio construction and evaluation. A detailed discussion is deferred to Chapter 2.

To reduce risk, investors seek a set of equities that have high expected return and low risk. This process is called *portfolio construction*. Since there is a tradeoff between risk and expected return, portfolio design usually starts by the investor choosing a desired level of expected return (or a risk tolerance). For a given desired expected return  $r_e$ , in the absence of any side information, the *minimum variance portfolio (MVP)*, a portfolio that has the lowest possible risk for its expected level of return, is the optimal portfolio.

The following are few of the criteria that we use to evaluate the performance of

the portfolios.

- Largest loss in a day,
- Annualized return,
- Sharpe ratio, a measure of risk-adjusted return [85].

We also consider the following benchmark portfolios in evaluating the performance of our approach.

- Equal weighted portfolio, where the portfolio is rebalanced to equal weights. This portfolio incurs large turnovers. Therefore, in practice it is hard to implement. While this portfolio has been shown to possess good theoretical performance, i.e., extremely low volatility with high returns, large turnover will negate most of the returns via transaction costs.
- SPX, the Standard & Poor 500 index.

### 1.3 Proposed Approaches

We discuss three areas of complimentary applications for leveraging data to learn previously unknown cross-sectional patterns in financial time series. These methods share the common goal of constructing selective but well-diversified portfolios that perform well.

1. **Learning connections in financial time series:** A problem with building a minimal variance portfolio is that the optimization uses a covariance matrix based on correlation, which gives equal weight to positive and negative returns and to small and large returns. We believe that this is inappropriate in a world in which risk preference plays an increasingly important role. We present a machine learning-based method to build a connectedness matrix to address the shortcomings of correlation in capturing events such as large losses.

## 2. **Learning the relationship between equity returns and news releases:**

A problem with using news releases to predict future returns is that often the information in the news has already been incorporated into the pricing of the equity. We present a machine learning-based method to model the relationship between the equity return that is unexplained by the market return (excess return) and the amount of sentiment in the news releases that hasn't been already reflected in the price of equities (excess sentiment).

## 3. **Leveraging network structures in financial time series:**

We provide methods to leverage network structure to model connections that span multiple equities, such as clusters and hierarchies in a network of financial time series. We discuss how the topological properties of the network can be used to obtain an improved segmentation of equities as opposed to the traditional sector-based grouping, thus improving the portfolio diversification process.

We show how these methods can be used separately or together in building portfolios with desirable characteristics such as greater annualized returns or smaller largest daily loss. In financial time series analysis, patterns can be learned in four time scales: long-term (fundamental based), mid-term (technical and macro-based), short-term (event based), and ultra-short-term (based on inefficiencies in statistical arbitraging). The proposed methods capture the cross-sectional dynamics in the short-term (Chapter 4), mid-term (Chapter 3), and long-term (Chapter 5).

## 1.4 Contributions

We briefly discuss some of the major contributions of our work. A detailed discussion is deferred to subsequent parts of the thesis.

- **Reducing tail risk:** We propose an alternative to the usual approach of using a correlation matrix to represent relationships among equities. Instead, we use what we refer to as a connectedness matrix. This differs from a correlation matrix in two important ways:

1. Traditional correlation matrices do not account for interactions among neighbors. Specifically, correlation is calculated between  $i$  and  $j$  independently of other neighbors. Therefore, these methods may end up incorporating information provided by neighbors multiple times. Our method uses supervised learning to discover the connections between two entities while discounting the influence of others.
2. Extreme returns occur rarely and therefore play a minor role in a traditional correlation matrix. The connectedness matrix focuses on extreme returns without ignoring the non-extreme returns.

To build the connectedness matrix, we model returns using three factors: active return, market sensitivity, and the connectedness of equities. We formulate the problem of estimating one return in terms of other returns as a recursive regression problem, and provide a method to solve it using unconstrained least squares optimization. The method ensures that the resulting connectedness matrix is positive semi-definite.

We show that this matrix can be used to build portfolios that not only “beat the market,” but also outperform optimal (i.e., minimum variance) portfolios over a 13-year period (2001 - 2012). More importantly, portfolios constructed using our method significantly reduced the tail risk without sacrificing overall returns. Compared to the traditional covariance approach, an MVP portfolio built with connectedness matrix reduces the worst day performance from  $-15\%$  to  $-9\%$ , and increases the Sharpe ratio from 0.63 to 0.71. In comparison, the corresponding values for the equal weighed portfolio are  $-11\%$  and 0.61, and the corresponding values for SPX are  $-9\%$  and 0.1.

- **Explaining news sentiment:** We propose a novel alternative to the usual approach of directly learning the relationship between the sentiment in the news releases and returns.
  - First, we learn the expected sentiment, i.e., the sentiment that has already

been reflected in the prices and discount this from current sentiment scores.

- We then model the returns on the excess sentiment instead of the raw sentiment score.

We demonstrate the utility of this learned relationship in constructing a portfolio that generate high risk-adjusted returns as well as our method's ability to model the variance attributable to news (news risk) better. We show that while using news directly yields an annualized return of 22% over a 10-year period (2003 - 2012), our proposed way of handling the past boosts the annualized return to 34% over the same period. In comparison, an equal weighted portfolio generates 6% annualized return, and SPX generates 5% annualized return over the same period.

- **Grouping of equities:** One common way to diversify a portfolio is holding equities from different sectors (e.g., financial and technological stocks). We propose a novel alternative to the traditional sector-based grouping by leveraging network structure to model connections that span multiple equities.
  - The underlying idea is that equities differently positioned within a financial network exhibit different patterns of behavior, and therefore selecting equities from different regions of the network leads to a well-diversified portfolio.
  - We investigate the use of centrality and peripherality measures to find two distinct regions of the network.
  - We discuss how graph clustering could be used to find regions that are sparsely connected with each other, and build a portfolio based on these clusters.

We compare the use of different linkage metrics, and different topological properties of a financial network to obtain an improved segmentation of equities.

We demonstrate the utility of this approach in constructing a diversified portfolio that performs well on out of sample data. For instance, our way of grouping

equities produces an annualized return of 22% for a 10-year period (2003 - 2012), compared to 10% with the sector grouping. SPX generates 5% annualized return over the same period.

## 1.5 Organization of Thesis

The remainder of this thesis is organized as follows.

- Chapter 2 presents background on the financial jargon and the datasets.
- Chapter 3 presents a machine learning-based method to build a connectedness matrix that addresses the shortcomings of correlation in capturing events such as large losses, and demonstrates the utility of this learned relationship in constructing a portfolio containing assets that are less likely to have correlated large losses in the future.
- Chapter 4 presents a machine learning-based method to model the relationship between the equity return that is unexplained by the market return (excess return) and the amount of sentiment in the news releases that hasn't been already reflected in the price of equities. It demonstrates the utility of the method in constructing a portfolio that effectively makes use of the firm specific news releases.
- Chapter 5 presents a set of network analysis techniques for multivariate time series to recover a sparse graph of connected entities, and discusses the application of this method in obtaining improved segmentation of equities as opposed to the traditional sector-based grouping, thus improving the portfolio diversification process.
- Chapter 6 concludes the thesis with a summary of main points and directions for future work.

# Chapter 2

## BACKGROUND

In this chapter, we start with a review of the finance background for our work. First, we provide a discussion on equities, in particular, on equities. Next, we discuss building a portfolio of equities and metrics for evaluating the performance of the portfolio. We use these metrics for evaluating and comparing our methods against benchmark portfolios later in this thesis. Those readers already knowledgeable about these topics may want to skip to Section 2.2, which provides a description of the data used in our work. This chapter borrows heavily from the discussion of these subjects in the literature [66, 29, 40, 33, 19, 60].

### 2.1 Finance

The area of finance is broad and well studied. It covers a numerous topics, but essentially all are related to money and the associated temporal evolution of reward and risk profile in terms of money. The fundamental objective is to seek ways to invest money in (financial) equities, and the objective is achieved by building a portfolio of equities.

There are many types of financial equities available for investment, including equities, bonds, commodities, and the derivatives of these fundamental equities. They are collectively called financial instruments. While the characteristics of investment vary across different equity classes, the nature of risk-return equilibrium is common

for them. In our work, we focus on publicly traded equities, often called stocks.

### 2.1.1 Reward - Risk

The return of an equity is the earning resulting from investing in the equity. Returns can be considered under different time periods (e.g., daily, monthly, quarterly, or yearly) and can include different sources (dividends or capital gains). If one time period is a day, and the equity is purchased at the closing price  $p_t$  of day  $t$  and sold at the closing price  $p_{t+1}$  of day  $t + 1$ , the daily return  $r_t$  for day  $t$  is given as

$$r_t = \frac{p_{t+1} - p_t}{p_t} + \frac{d_t}{p_t} \quad (2.1)$$

Here,  $d_t$  is the dividend given on day  $t$ .

Typically, the equity is held over several consecutive time periods  $\{1, 2, \dots, T\}$ , and the returns form a time series  $\mathbf{r} = \{r_1, r_2, \dots, r_T\}$ . A return series  $\mathbf{r}$  can be characterized by the following

1. Arithmetic mean return  $\bar{r} = \frac{1}{T} \sum_{t=1}^T r_t$
2. Cumulative return  $R_T = \prod_{t=1}^T (1 + r_t) - 1 = \frac{P_{T+1} - P_1}{P_1}$
3. Standard deviation  $\sigma_r = \sqrt{\frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^2}$

Return is probably the most important factor that the investors consider. They want to make as much as the return as early as possible. Arithmetic mean gives the average change in the capital over the entire time period. Cumulative return gives the over all change in the capital by the end of the time period.

While the past returns of an equity are known, the future return of an equity is unknown. Therefore, when an investment is made, there is a risk associated with the expected future return. This *risk* is defined as the uncertainty of the future return. As in the case of the future return, the uncertainty of the future risk is unknown. However, this is usually approximated by the standard deviation of the past returns  $\sigma_r$ . This is the risk of whether the actual future return is higher, lower, or equal to the expected future return.



## 2.1.2 Portfolio

If two equities offer the same gain, most investors will pick the equity with the lowest risk for investing. Therefore, high-risk equities need to offer higher return rates than low-risk equities to attract an investor. To reduce risk, investors seek a set of equities that have high expected return and low risk. This process is called *portfolio construction*. It is widely accepted that in designing a portfolio of equities there is a tradeoff to be made between risk and return. The root cause of the tradeoff is that more volatile equities typically have higher returns. Much of modern financial theory is based upon the assumption that a portfolio containing a diversified set of equities can be used to control risk while achieving a good rate of return.

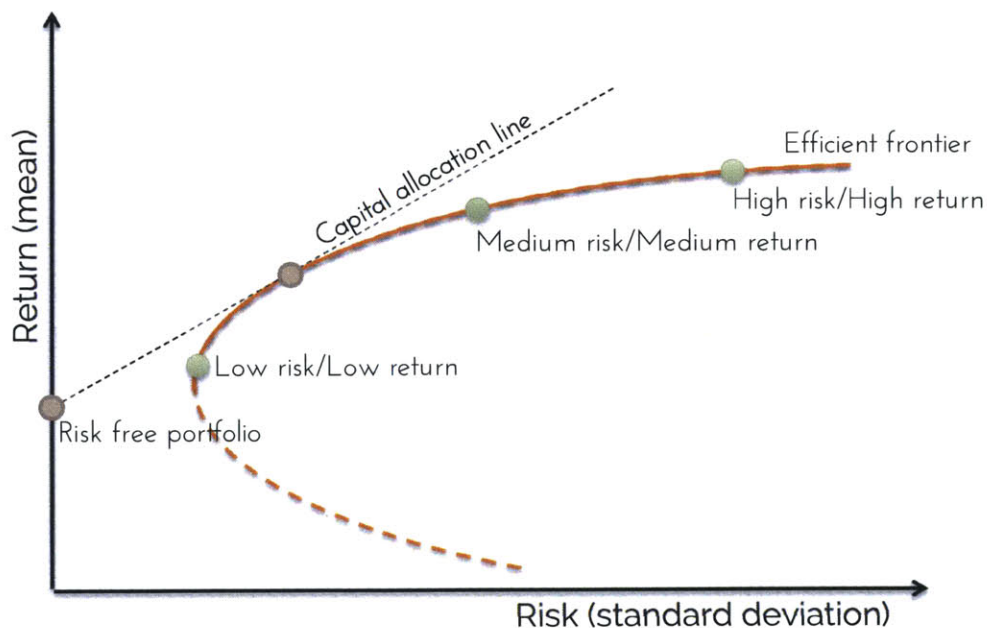


Figure 2-1: Efficient frontier (orange line) and the optimal portfolio choices (green dots) along the efficient frontier. Dashed line shows the possible mathematical solutions to Equation 4.8, but as portfolios are meaningless. Portfolios that lie below the efficient frontier are sub-optimal, because they do not provide enough return for the level of risk. Portfolios that lie above the efficient frontier are theoretically impossible.

Since there is a tradeoff between risk and expected return, portfolio design usually starts by the investor choosing a desired level of expected return (or a risk tolerance).

For a given desired expected return  $r_e$ , in the absence of any side information, the minimum variance portfolio (MVP), a portfolio that has the lowest possible risk for its expected level of return, is the optimal portfolio. Frontier of the set of optimal portfolios is called the efficient frontier (Figure 2-1) [62]. Portfolios that lie below the efficient frontier are sub-optimal, because they do not provide enough return for the level of risk.

For the MVP, portfolio weights  $\omega$  are derived by solving the optimization problem:

$$\begin{aligned} \min_{\omega} \quad & \frac{1}{2} \omega^T C \omega & (2.2) \\ \text{subject to} \quad & \sum_{j=1}^{m+1} \bar{r}_j \omega_j \geq r_e \\ & \sum_{j=1}^{m+1} \omega_j = 1; 0 \leq \omega_j \leq 1, j = 1, \dots, m + 1 \\ & f(x) = x & (2.3) \end{aligned}$$

Here,  $C$  is the covariance matrix of returns, and  $\bar{r}_j$  is the expected return of equity  $j$ .

In this formulation, the goal is to minimize the portfolio risk given an acceptable level of expected return. Of course, not all investors have the same goals, and this formulation can be rewritten to satisfy the following objectives:

- Minimize the portfolio risk.
- Maximize the return.
- Maximize the return given an acceptable level of risk.
- Maximize the risk-adjusted return.

There are many assumptions both explicitly and implicitly made by this model. Here, we assume fixed capital (no leverage), perfect liquidity, no market friction, no short positions, fractional amounts of equities could be bought, and no transaction costs.

Similar to that for an equity, portfolio risk  $\sigma_{r_p}$  is defined as the standard deviation of the portfolio return. For the portfolio weights  $\omega = \{\omega_1, \dots, \omega_m\}$ , and correlation matrix  $\rho = \{\rho_{ij} | 1 \leq i \leq m, 1 \leq j \leq m\}$  it is given by

$$\sigma_{r_p} = [\omega^T C \omega]^{1/2} \quad (2.4)$$

$$= \left[ \sum_{i=1}^m \omega_i^2 \cdot r_i^2 + 2 \sum_{i=1}^m \sum_{j>i}^m \omega_i^2 \cdot r_i^2 \cdot \omega_j^2 \cdot r_j^2 \cdot \rho_{ij} \right]^{1/2} \quad (2.5)$$

In Equation 2.5, we see that the portfolio risk is made of two components: risk of individual equities (*specific risk*), and risk common to all equities (*systematic risk* also known as *market risk*). Specific risk can be reduced away by finding equities that are uncorrelated, i.e., unlikely to move in tandem. This is called *diversification*. In contrast, in a portfolio of equities, systemic risk cannot be diversified away, except for short selling.

### Short selling

Typically an investment strategy involves attempting to buy the equity (or any other financial instrument) when the price is low and selling it when the price is high. The practice of buying equity (in the hopes of selling at a higher price) is called taking *long* position in the equity. The opposite, selling equities (or any other financial instruments) that are not currently owned when the price is high, and subsequently buying them when the price is low is called short selling. The practice is called taking a *short* position in the equity. In the interim period, the investor is required to borrow the equities to satisfy the short sale, and generally there is a fee involved which might also incorporate any dividends the borrower gets in that period.

Similar to how long positions allow an investor to bet on the upward trend, short positions allow an investor to bet on the downward trend. Short selling is typically used in *hedging*, a way of minimizing the risk of an investment via a more complex set of transactions. Hedging can be used to offset the potential losses/gains that may be incurred in a companion investment. For example, the undiversifiable risk (market risk) of a long only portfolio (like the one we discussed above) can be hedged away

by short selling the appropriate amount of the S&P 500 index.

Of course, there is no free lunch: short selling involves additional risks as well. After taking a short position, if the price of a stock rises significantly, the investor is forced to close their positions. This involves buying the shares, which will in turn increase the demand for the stock, resulting in further rise in the price, causing more short positions to be closed. This positive feedback effect is called a *short squeeze*. Similarly, the stock might be *hard to borrow*, resulting in significantly high fees.

In our work, we mainly use long only portfolios (except where stated otherwise). When appropriate, we discuss how the systemic risk of our portfolios can be hedged away by taking a companion short position.

## Portfolio Evaluation

A portfolio is evaluated by its performance measured from different perspectives. Different investors have different goals in investing. Thus, the equities selected by distinct investors differ. The most important features of an equity are its return and its risk. Some investors, for example, hedge fund managers, expect high returns, and in exchange, expect to bear corresponding risks.

Therefore it is worth looking at different tools that have been used to characterize the return and the risk of an equity. Some of these measures focus purely on different types of risk. For instance, largest loss in a day (worst day performance) is a measure of *tail risk*, the risk of improbable but potentially catastrophic loss. Whereas, maximum drawdown measures losses sustained over a time period. Some of these measures try to balance the risk with return. For instance, the Sharpe ratio measures the risk adjusted return.

In our work, for a set of daily portfolio returns  $\{r_t | 1 \leq t \leq T\}$ , we use the following criteria to evaluate the performance of the portfolios:

- Largest loss in a day: Given by  $\min(r_t)$ .
- The expected shortfall (also known as CVaR) at 5% level gives the expected return on the portfolio in the worst 5% of the cases [81].

- Max drawdown: Maximum drawdown is the largest peak-to-subsequent-trough decline in cumulative return. It is given by  $M = \min_{1 \leq i \leq j \leq T} \prod_{t=i}^j (r_t + 1) - 1$ .
- Annualized return: Cumulative return  $R_T$  from day 1 to day  $T$  is given by  $R_T = \prod_{t=1}^T (r_t + 1) - 1$ . Annualized total return is  $(R_T + 1)^{1/(T/252)} - 1$ .
- Sharpe ratio: The Sharpe ratio measures the risk-adjusted return [85]. It is given by  $S = E(r - r_f) / \sqrt{\text{var}(r)}$  where  $r_f$  is the risk free return (assumed to be 1% annually). We quote annualized Sharpe ratio (i.e.,  $S\sqrt{252}$ ) in our results.
- Information ratio: Information ratio measures the excess return for additional risks taken [85]. It is given by  $I = E(r - r_\Lambda) / \sqrt{\text{var}(r - r_\Lambda)}$ , where  $r$  is the daily return of the portfolio and  $r_\Lambda$  is the reference return (return on the S&P500 index). We quote annualized information ratio (i.e.,  $I\sqrt{252}$ ) in our results. A positive information ratio implies that excess return is greater than the additional risk taken.

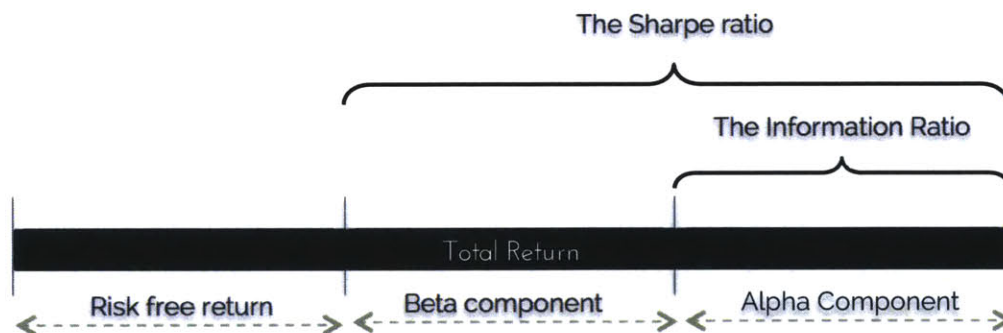


Figure 2-2: The Sharpe ratio Vs. Information ratio.

The Sharpe ratio and information ratio are both indicators of risk-adjusted returns, and are ratios of mean returns to standard deviations of some flavor. Total returns can be considered the sum of the risk free rate, the beta (i.e., the return for taking on market risk), and the alpha (excess return). The Sharpe ratio is the measure of risk-adjusted return of beta plus alpha components, where as information ratio is the measure of risk-adjusted return of alpha component. Figure 2-2 illustrates

this. While the Sharpe ratio is widely used, we believe for our work information ratio is more appropriate.

## 2.2 Data

### 2.2.1 Return Data

In our work, we mainly use equity daily return data from CRSP<sup>1</sup>.

We examine all 369 companies that were in the S&P500 continuously from 2000 to 2012. This time period contains two major financial crises (2001 and 2008). The set of companies are from ten sectors: consumer discretionary, consumer staples, energy, financials, health care, industrials, information technology, materials, telecommunications services, and utilities.

There is a survivorship bias inherent in this dataset. For example, companies that entered or left the S&P 500 between 2001 and 2012 are not part of our dataset, and these companies are likely to be more volatile than the ones included in our dataset. Therefore, the results of our method (and benchmark methods) will be overstated. However, in our evaluations we focus on the relative improvement of our method compared to the benchmark methods. Because our methods are specifically designed to deal with adverse events, we believe that the survivorship bias leads to an underestimation of the relative performance of our methods. We discuss this in Chapter 3.

The data represent a multivariate time series  $\mathbf{R} = \{r_{t,j}\}; 1 \leq t \leq T, 1 \leq j \leq m$ , made of daily returns of  $m$  stocks. We use indexing  $t$  for days, and  $j$  for equities. Daily return of an equity  $j$  on day  $T$  is given by  $r_{T,j} = (p_{T,j} - p_{T-1,j})/p_{T-1,j}$ , where  $p_{T,j}$  and  $p_{T-1,j}$  are closing prices of stock  $j$  on day  $T$  and  $T - 1$ .

---

<sup>1</sup>CRSP, Center for Research in Security Prices. Graduate School of Business, The University of Chicago (2004). Used with permission. All rights reserved. [www.crsp.uchicago.edu](http://www.crsp.uchicago.edu)

## 2.2.2 News Data

We use the sentiment data from Thomson Reuters, called Thomson Reuters NewsScope dataset. On firm specific news releases collected from more than 60 reporting sources, Thomson Reuters uses a proprietary natural language processing based algorithm to score the news feeds, and constructs a database of sentiment scores (positive or negative). Thus, Thomson Reuters NewsScope database assigns a “hard” number to a “soft” concept of news sentiments.

The algorithm using natural language processing techniques analyses news provided by more than 60 reporting sources (news agencies, magazines etc) and compares with an expert built dictionary of about 20,000 words to identify relevant content. The outcome produced by the algorithm is a set of various (quantitative) indices based on the presence and the position of the words in the text. Two main indices computed by the algorithm are the relevance of a news item for the company mentioned, and the sentiment (positive or negative) of the news reported.

### Contents

The dataset contain various types of news feeds: *alerts*, *first-take articles*, *follow-ups*, *append*, and *overwrites*. We only look at the feeds that are *alerts* or *first-take articles* so that we could focus only on the “fresh” news. Figure 2-3 shows the snapshot of relevant fields in the database. *Relevance* measures how targeted a news story is to an equity. For example, a earnings announcement of a company would have high relevance for that company’s stock. *Sentiment* measures the positive and negative tone of the article using natural language processing algorithms. An earnings announcement with disappointing results would have high negative sentiment. *Take sequence* indicates the step of the evolution of a news story. A news story evolves in multiple steps, starting with an alert, followed by one or more updates as news breaks, and may continue into full length articles. They may also be amendments and further updates.

stock_ric	idn_time	relevance	sent_pos	sent_neg	item_type	take_seqno	bcast_text
OCR	08 JAN 2005 00:12:14.249	1	0.771771	0.079536	ARTICLE	1	Omnicare Extends Tender Offer for NeighborCare until February 4,
IGT	08 JAN 2005 00:47:30.606	1	0.536143	0.0902621	ARTICLE	1	Wheel of Fortune(R) Dollars Pays Texas Tourist \$3.8 Million <IGT.N
KR	08 JAN 2005 02:18:32.338	0.408248	0.226677	0.158741	ARTICLE	0	
GE	24 JAN 2005 03:17:15.333	1	0.8303	0.0417572	ARTICLE	1	GE Commercial Finance, Business Property's Tom Pretty Named Ler
IBM	24 JAN 2005 04:04:11.675	1	0.361991	0.483157	ARTICLE	1	UPDATE 1--Lenovo shares jump on report of IBM deal hurdle
SO	24 JAN 2005 11:01:35.807	0.385337	0.722993	0.116785	ARTICLE	1	Georgia Power Selects DYMEC for Substation WAN Deployment Proj
PPC	24 JAN 2005 11:30:16.443	1	0.175775	0.0141482	ALERT	2	PILGRIM'S PRIDE CORP <PPC.N> Q1 SHR \$0.73
ETN	24 JAN 2005 11:57:02.257	1	0.850421	0.0310171	ALERT	1	EATON REPORTS FOURTH QUARTER NET INCOME OF \$1.16 PER SH
EPD	24 JAN 2005 12:00:36.309	0.333333	0.464351	0.192042	ARTICLE	1	Ei Paso Corporation Sells Gathering System, Processing Plant for \$7
AV	24 JAN 2005 12:30:22.124	1	0.837355	0.0378435	ARTICLE	1	TeleTech and Avaya Deliver New Hosted Services to Help Compani

Figure 2-3: Snapshot of the relevant fields in Newscope dataset.

## Observations

We look at the feeds corresponding to the 369 companies that were in the S&P 500 from 2005 to 2012. We observe that the daily number of feeds for this set of equities vary from as low as 0 to as high as 5708 feeds in a day (Figure 2-4). It exhibits, daily, weekly, and seasonal patterns with few or no feeds on weekends, and increased activity during the period of earnings announcements. Over all, the number of feeds increases with time. However, the distribution of news across sectors vary with the macroeconomic conditions (Figure 2-6). We can also see the increase in the coverage by observing the number of stocks receiving at least one news release daily (Figure 2-5). Out of 369 companies we considered, the data cover 310 of them. In Figure 2-7 and Figure 2-8, we look at the variations in the market sentiment over the time, and how its changes corresponding to hour of the day. We observe that the number of news releases peak in the middle of the day, with another spike just after the closing of the market. In contrast, there is very little activity before the opening of the market.



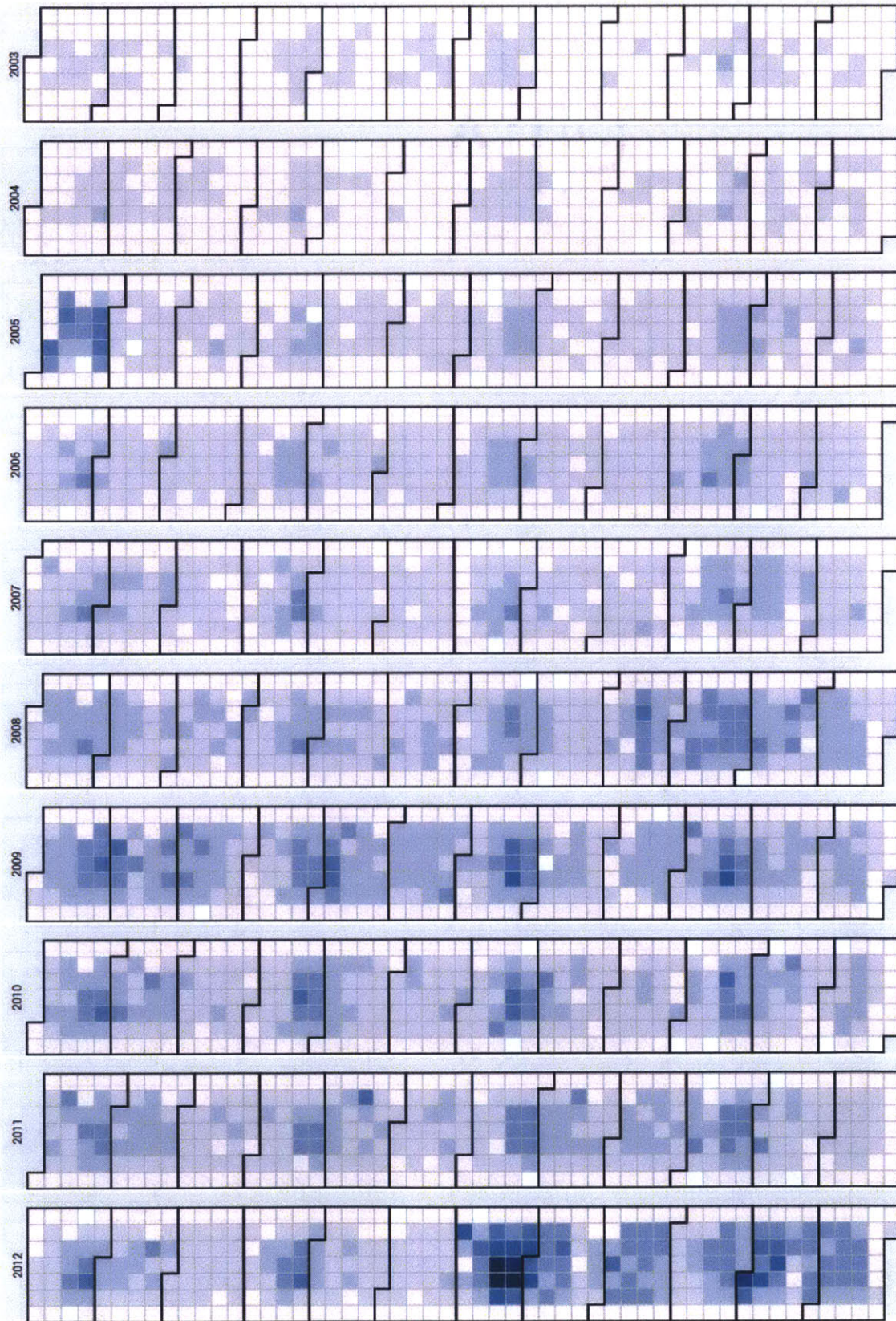


Figure 2-4: Historical distribution of the number of feeds in Newscope dataset. They vary from 0 (particularly on most of the weekends) to 5708.

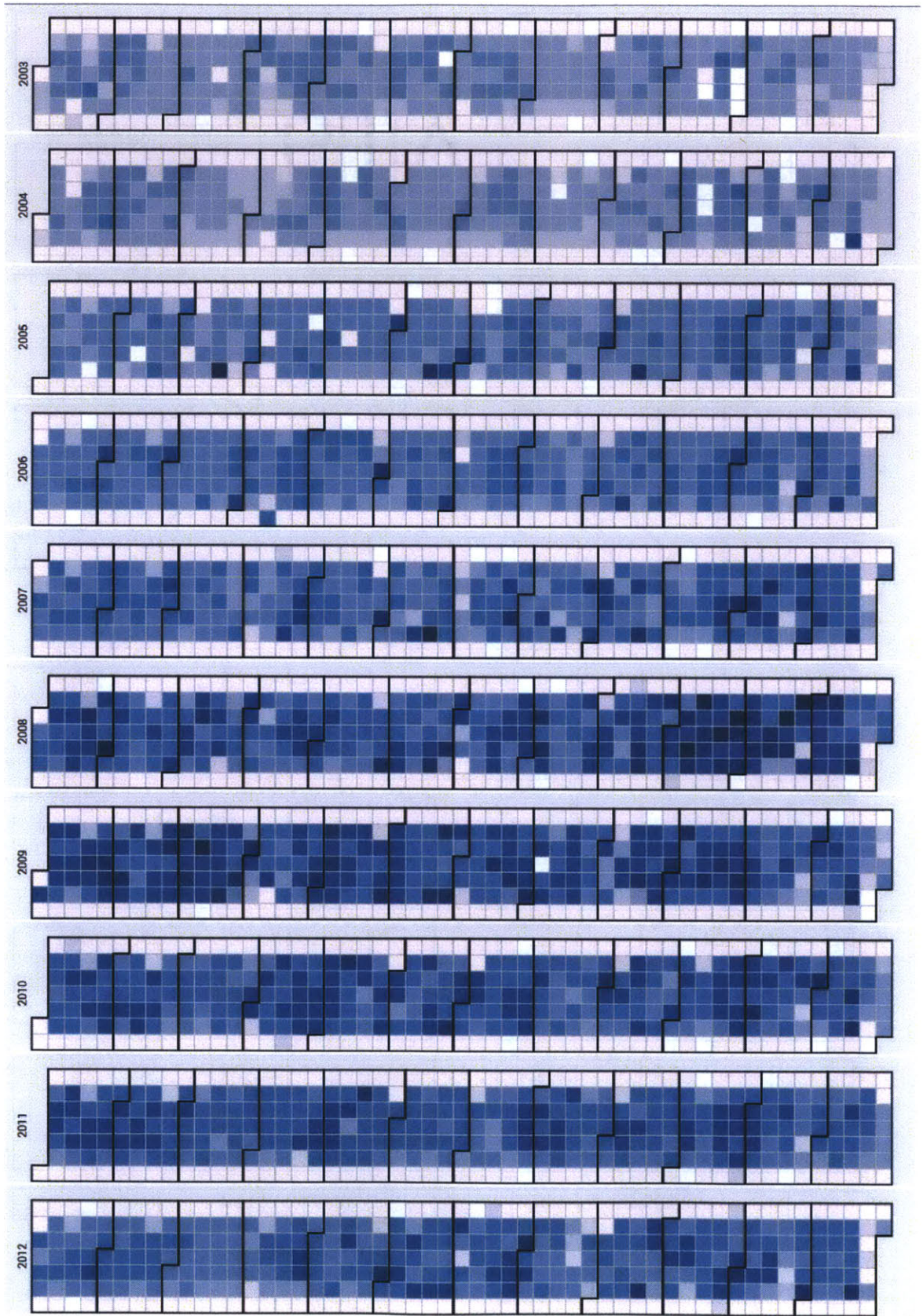


Figure 2-5: Historical distribution of the number of equities covered in Newscope dataset. They range from 0 to 310.

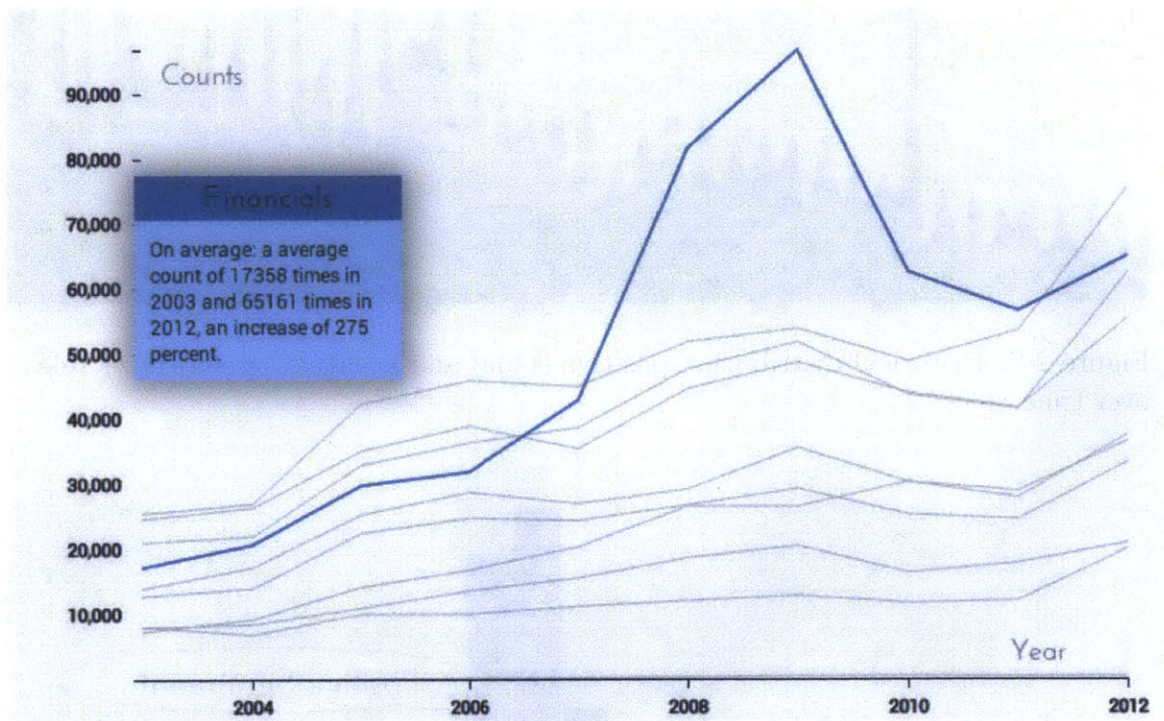


Figure 2-6: Historical distribution of the number of feeds across sectors in Newscope dataset. Over all the coverage increased by three fold in the last decade. Note that during the 2008 crisis, the financial sector received significantly more news feeds than other sectors.

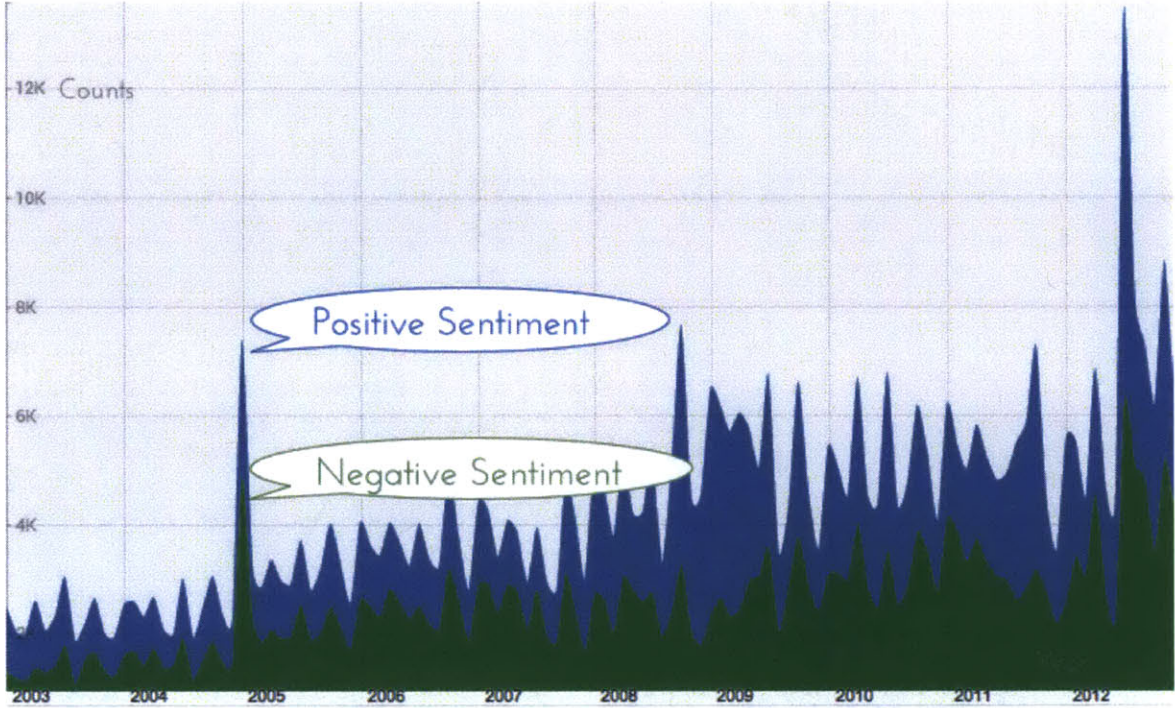


Figure 2-7: Historical distribution positive (blue) and negative (green) news releases over time.

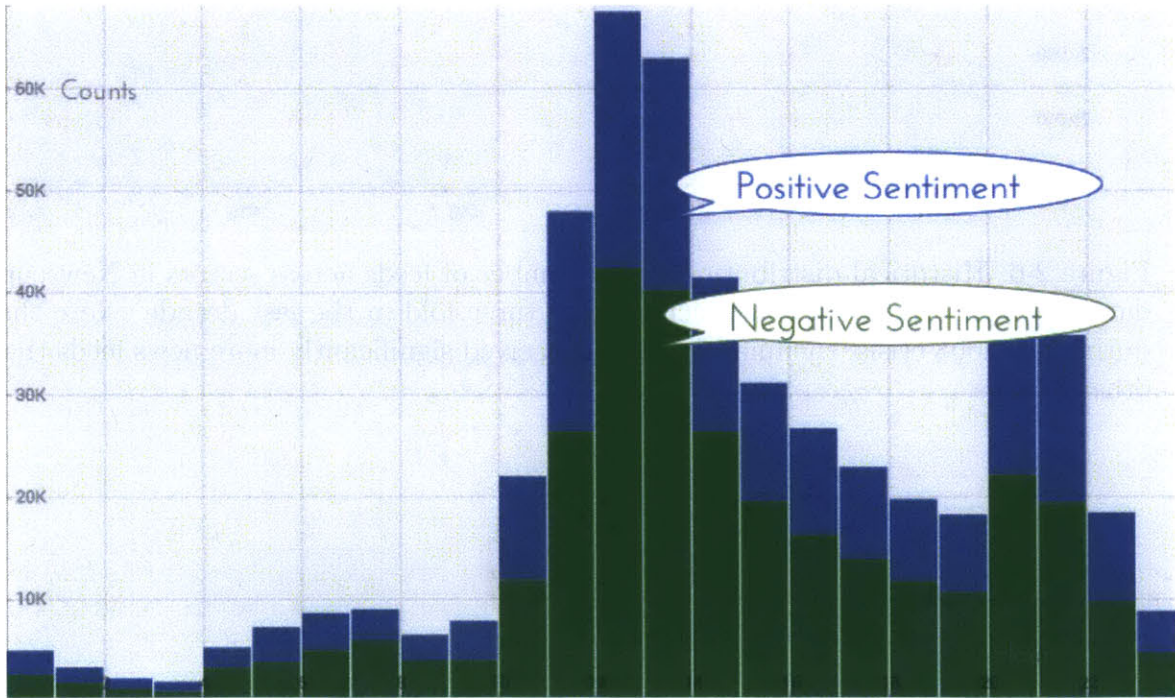


Figure 2-8: Hourly distribution of the news releases scored by their positive (blue) and negative (green) sentiments.

# Chapter 3

## LEARNING CONNECTIONS IN TERMS OF LARGE LOSSES

In this chapter, we consider the task of learning the connections in financial time series in terms of large losses. We describe a machine learning-based method to build an alternative to a correlation matrix, called a connectedness matrix, that addresses the shortcomings of correlation in capturing events such as large losses. Our method uses an unconstrained optimization to learn this matrix, while ensuring that the resulting matrix is positive semi-definite. We further present and demonstrate the potential real world utility of such a method in constructing portfolios. We show that the portfolios built with our method not only “beat the market,” but also outperform optimal (i.e., minimum variance) portfolios.

### 3.1 Introduction

It is widely accepted that in designing a portfolio of equities there is a tradeoff to be made between risk and return. The root cause of the tradeoff is that more volatile equities typically have higher expected returns. Much of modern financial theory is based upon the assumption that a portfolio containing a diversified set of equities can be used to control risk while achieving a good rate of return. The basic idea is to choose equities that have high expected returns, but are unlikely to move down in

tandem.

When building a diversified portfolio investors often begin by choosing a minimum desired expected return as the independent variable. They then formulate portfolio design as an optimization problem with return as a constraint and variance minimization as the objective function. Central to this optimization is the estimation of a covariance matrix for the daily returns of the equities in the portfolio. If the investor cares only about the mean and the variance of the portfolio returns, such formulation gives the optimal portfolio. This is the idea behind modern portfolio theory.

Different investors have different goals. Some investors, for example, hedge fund managers, expect high returns, and in exchange, expect to bear corresponding risks. For such investors, it is critical to control for *tail risk*, the risk of improbable but potentially catastrophic negative return events [5, 36]. A problem with the previously mentioned approach is that the covariance matrix uses correlation, which gives equal weight to positive and negative returns and to small and large returns. This is inappropriate in a world where risk preference plays an increasingly important role. Hence, selective as opposed to full diversification is gaining popularity [40, 94]. There have been works demonstrating the importance of addressing the improbable but potentially catastrophic negative return events in designing portfolios [1, 64]. Learning the connectedness between equities in terms of large losses and exploiting this knowledge in portfolio construction is what this chapter is about.

We present a machine learning-based method to build a connectedness matrix to address the shortcomings of correlation in capturing large losses, which we refer to as events. Our method uses an unconstrained optimization to learn this matrix, while ensuring that the resulting matrix is positive semi-definite. We further present and demonstrate the potential real world utility of a method that constructs portfolios using the learned relationships. We show that this matrix can be used to build portfolios that not only “beat the market,” but also outperform traditional “optimal” (i.e., minimum variance) portfolios.

## 3.2 Related Work

We begin by discussing methods that have been used to study various correlation measures among returns. We then move on to discuss work specific to understanding the connections of extreme returns.

### 3.2.1 Correlation and Partial Correlation

If one knows the correlation of equity  $e$  with all other equities, one can estimate the expected return of  $e$  as a weighted average over known returns of other equities.

Correlation measures give equal weight to small and large returns, and therefore the differential impact of large returns (both positive and negative) may be hidden.

To address this, researchers have proposed conditional correlations to focus on certain segments such as returns outside a standard deviation [92]. However, it has been shown that conditional correlation of multivariate normal returns will always be less than the true correlation. For example, if two time series of zero mean and unit standard deviation have correlation 0.5, semi-correlation of only positive returns drops to 0.33, for returns larger than one standard deviation correlation drops to 0.25, and for returns larger than two standard deviation correlation drops to 0.19. Conditional correlation goes to zero for extreme returns. This effect also exists when a GARCH model generates the returns [57]. Therefore, conditional correlation is likely to understate the connectedness between equities based on extreme returns.

Longin (1999) provides a formal statistical method, based on extreme value theory, to model the correlation of large returns [57]. First, the authors model the tails of marginal distributions using generalized Pareto distribution (GPD) [17]. Then, they learn the dependence structure between two univariate distributions of extreme values. Thus, they express joint distribution in terms of univariate marginal distribution functions. This dependence structure between the variables is known as a *copula*. There has been research done on extreme value distributions [63, 26] and extreme value correlation measures [58, 77]. Semi-parametric models have since been proposed to address the inflexibilities of such parametric models [11].

A downside of these methods is that the linkage is learned between two time series independently of the rest of the time series in the multivariate time series.

Partial correlation measures the degree of association between two time series while discounting the influence of other time series in the multivariate time series. Thus, it removes the transitive connections. It is calculated by fitting a regression model for each of these two time series on the rest. The correlation between the residuals of these regression models gives the partial correlation [47]. But, partial correlation doesn't distinguish extreme values.

### 3.2.2 Understanding Connections on Extreme Returns

Correlation between stocks has traditionally been used when measuring co-movements of prices, and discovering contagion in financial markets [80, 5]. Researchers have used partial correlations to build correlation-based networks. These networks are then used to identify the dominant stocks that drive the correlations present among stocks [48].

Bae (2003) distinguishes extreme returns in establishing the linkages between financial time series [5]. It captures the transmission of financial shocks to answer questions such as how likely is it that two Latin American countries will have extreme returns on a day given that two countries in Asia have extreme returns on that or the preceding day. There has been extensive research on multivariate extreme values [24, 73]. Chen (2007) provides a method to model the temporal sequence associations for rare events [20]. Arnold (2007) examines algorithms that, loosely speaking, fall under the category of graphical Granger methods, such as model selection methodologies, Lasso, sparsification, and structured learning to predict the co-occurrences of events in time series [4].

There have been works on using constraints on downside risk and value-at-risk based optimizations to minimize the effect of event risks in portfolio constructions [8, 44]. Following these works, Poon (2004) presents a framework based on joint-tail distribution. It shows that the extremal dependence structure on the joint-tail distribution helps build a better portfolio selection and risk assessment on five major indexes [77].



## 3.3 Method

In this section, we first formally define our task and describe the evaluation methods. Thereafter, we describe an algorithm for learning the connections, and detail how it could be used to build connectedness matrix.

### 3.3.1 Problem Formulation

We formulate the learning problem as given a set of equities  $A$  on a given day, some of which had events and some of which didn't, which equities in a disjoint set  $B$ , are mostly to experience an event on the same day as those in  $A$ . In our model, we update the weights daily and predict the returns for the following day. We rank the equities in  $B$  using the predicted returns.

It may seem that this is useless, because by the time we have the returns for equities in  $A$ , we would already know the returns for equities in  $B$ . However, the goal of this phase is not to learn to predict events, but to learn historical relationships among equities. This learned relationship will then be used to construct a portfolio containing assets that are less likely to have correlated events in the future.

We evaluate the accuracy of our method in producing a list of equities ordered by their likelihoods of having large losses, given information about the behavior of other equities. We compare this list against the true occurrences of events using mean average precision (MAP) scores.

Later, we present and demonstrate the potential real world utility of using the learned relationships in constructing portfolios. The performance of portfolios constructed using our methods are then compared to the performance of portfolios constructed using conventional approaches, including traditional correlation-matrix based methods.

### 3.3.2 Model

We use a factor model to describe the daily return of each equity in terms of the equity's active return, market sensitivity, and the daily returns of other equities in the

sector (e.g., financial or energy) to which that equity belongs. We learn these factors using a recursive regression. We train the regression model on the historical data using regularized least squares and estimate the parameters using gradient descent. In contrast to methods that quantify the connectedness between equities using pairwise relationships, our method accounts for interactions with all other equities in the set. Since extreme events are rare, we use all of the historical data rather than just events. We use a cost function that differentially weights returns of different magnitudes. Thus we provide a model-independent approach to prioritize the connections on large losses.

If the closing prices of equity  $j$  on day  $T$  and  $T - 1$  are  $p_{T,j}$  and  $p_{T-1,j}$ , the return for equity  $j$  on day  $T$  is given by  $r_{T,j} = (p_{T,j} - p_{T-1,j})/p_{T-1,j}$ . On day  $T + 1$ , we are given historical daily returns for  $m$  equities in a  $T \times m$  matrix  $\mathbf{R} = \{r_{t,j}\}; 1 \leq t \leq T, 1 \leq j \leq m$ . We use indexing  $t$  for days, and  $j, k$  for equities. When  $r_{t,j} < -0.1$  (a 10% loss), we say that equity  $j$  had an event on day  $t$ .

We assume that daily returns (rows of  $\mathbf{R}$ ) are independent. This is because, while daily returns are generally believed to be heteroskedastic [98], we focus only on large returns that are rare. We use regularization to tackle over fitting. The regularization parameter  $\lambda$  is determined by cross validation.

Factor model representation of returns is common in finance and econometrics [57, 49]. We model the return of equity  $k$  on day  $t$  by

$$\hat{r}_{t,k} = \underbrace{a_k + b_k r_{t,\Lambda}}_{d_{t,k}} + \sum_{j=1:m;j \neq k} w_{j,k}(r_{t,j} - d_{t,j}) \quad (3.1)$$

In this model, we explicitly learn the factors for equity  $k$ : the equity's active return  $a_k$ , the equity's sensitivity to the market  $b_k$ , and the equity's connectedness with other equities  $w_{j,k}$ , where  $1 \leq j \leq m; j \neq k$ . We represent the market return with the S&P500 index return ( $r_{t,\Lambda}$ ) that averages the returns of all the equities on a given day.

We use least squares minimization to estimate the weights. We find that better performance is achieved, when we capture the differential impact of certain values by

weighting with a cost function  $f(x)$ . We discuss the choice of cost function in Section 3.4.3.

$$\min_{a^*, b^*, w^*} \sum_{\substack{t=1:T \\ k=1:m}} f(r_{t,k})(r_{t,k} - \hat{r}_{t,k})^2 \quad (3.2)$$

We can efficiently compute model parameters ( $\theta = \{(a_k, b_k, w_{j,k} | 1 \leq k \leq m, 1 \leq j \leq m, j \neq k\}$ ) by estimating the inner products. However, estimating the weights directly on the observed data is prone to overfitting [9]. Therefore, we learn the parameters by solving the following regularized least squares problem:

$$\min_{a^*, b^*, w^*} \sum_{\substack{t=1:T \\ k=1:m}} f(r_{t,k})(r_{t,k} - \hat{r}_{t,k})^2 + \lambda(a_k^2 + b_k^2 + |w|^2) \quad (3.3)$$

We use gradient descent to minimize the regularized square errors. For each  $r_{t,k} \in \mathbf{R}$ , we update the parameters by:

$$\begin{aligned} a_k &\leftarrow a_k + \eta(e_{t,k} - \lambda \cdot a_k) \\ b_k &\leftarrow b_k + \eta(e_{t,k} \cdot r_{t,\Lambda} - \lambda \cdot b_k) \\ w_{j,k} &\leftarrow w_{j,k} + \eta(e_{t,k}(r_{t,j} - d_{t,j}) - \lambda \cdot w_{j,k}) \quad \forall j \neq k \end{aligned}$$

Here,  $\eta$  is the learning rate that is dynamically adjusted using line search, and  $e_{t,k} \stackrel{def}{=} f(r_{t,k})(r_{t,k} - \hat{r}_{t,k})$ . We use the last 500 days in the historical data to train our model. We iterate 100 times for the initial estimate of the parameters. The model is updated daily to make predictions for the next day. Since this new training set differs from the previous one for only two days, convergence is achieved within a few iterations.

### 3.3.3 Similarity to Other Methods

In our model (Equation 3.1), we represent the relationship between the returns of equities after discounting their interactions with the market. In the factor model, we simultaneously fit the regression model and learn the correlation weights. Thus, the interpolation weights  $w_{j,k}$  resemble partial correlation estimates. Further, regularization is employed in our model to reduce the likelihood of spurious estimates.

For a matrix  $X$ , a column vector  $Y$ , and a regression problem expressed as  $Xw = Y$ , an explicit solution, denoted by  $\hat{w}$  is given by:  $\hat{w} = (X^T X)^{-1} X^T Y$ . If the variables are mean adjusted,  $\Sigma$  is the covariance of  $X$ , and  $C$  is the covariance between  $Y$  and each column of  $X$ , it can be rewritten as,  $\hat{w} = \Sigma^{-1} C$ . In Equation 3.1, we can observe the similarity between these variables ( $X$  and  $Y$ ) and adjusted returns, i.e.,  $y_t \approx (r_{t,k} - d_{t,k})$ , and  $x_{t,j} \approx (r_{t,j} - d_{t,j})$ . The connectedness matrix  $G$ , built from the interpolation weights, models the pairwise connection between the adjusted returns of two equities while discounting the connectedness among all other equities.

According to Peng (2009), for a covariance matrix  $\Sigma$  of random variables  $(y_1, \dots, y_p)$  and corresponding concentration matrix  $\Sigma^{-1} = \{\sigma^{ij}\}_{p \times p}$ , partial correlation between  $y_i$  and  $y_j$  is given by  $\rho^{ij} = \frac{-\sigma^{ij}}{\sqrt{\sigma^{ii}\sigma^{jj}}}$ . Then, it can be shown that for  $1 \leq i \leq p$ ,  $y_i = \sum_{j \neq i} \beta_{ij} y_j + \epsilon_i$ , such that  $\epsilon_i$  is uncorrelated with  $y_i$  if and only if  $\beta_{ij} = \rho^{ij} \sqrt{\frac{\sigma^{jj}}{\sigma^{ii}}}$ . Then, under the loss function  $\frac{1}{2} (\sum_{i=1}^p \sigma^{ii} \|Y_i - \sum_{i \neq j} \beta_{ij} Y_j\|^2) + \lambda \|\rho\|_2$  and  $\beta_{ij} = \rho^{ij} \frac{\sigma^{jj}}{\sigma^{ii}}$  we can obtain regularized estimates for partial correlation [74]. Then,  $\rho^{ij} = \text{sign}(\beta_{ij}) \sqrt{\beta_{ij} \beta_{ji}} = \rho^{ji}$ , and therefore  $\beta_{ij}$  converges to  $\rho^{ij}$ . Regression weights  $\beta_{ij}$  resemble the regression weights of our model.

A matrix  $G_{obs}$  that represents an observed network, whose diagonal is set to zero and whose non-diagonal entries represent the weighted links in the network, can be written in terms of direct network paths in  $G_{dir}$  and indirect effects  $G_{dir}^2, G_{dir}^3, \dots$  by  $G_{obs} = G_{dir} + G_{dir}^2 + \dots = G_{dir} (I - G_{dir})^{-1}$ . The network deconvolution allows us to remove the transitive effects and recover the direct paths from the observed network by  $G_{dir} = G_{obs} (I + G_{obs})^{-1}$  [34]. We note the similarity between  $(1 + G_{obs})^{-1}$  in this approach and  $(\lambda \cdot I + \Sigma)^{-1}$  in our approach.

### 3.3.4 Connectedness Matrix

In portfolio construction, connectedness between equities is used to find less correlated assets. Generally, a covariance matrix  $C$  is used to find the optimal diversification, i.e., minimum variance portfolio. In contrast, we use the connectedness matrix  $G$  in place of the correlation estimates that make up  $C$ . Here,  $G$  is learned using our factor

Table 3.1: Connectedness matrix construction

	1	...	$j$	...	$m$
1	1	...	$w_{1,j}$	...	$w_{1,m}$
...	...	...	...	...	...
$j$	$w_{j,1}$	...	1	...	$w_{j,m}$
...	...	...	...	...	...
$m$	$w_{m,1}$	...	$w_{m,j}$	...	1

model. We assume that the portfolio is built with  $m$  equities that belong to a sector.

We build the connectedness matrix  $G \in \mathfrak{R}^{m \times m}$  from the interpolation weights of the factor model. A direct way of doing this from the weights learned via our factor model is given in Table 3.1.

Such a matrix should be positive semi-definite to be used in portfolio optimization involving quadratic programming. Since any positive semi-definite matrix  $G$  can be decomposed into  $PP^T$ , where  $P \in \mathfrak{R}^{m \times m}$ , we reformulate Equations 3.1 and 3.3 as:

$$\hat{r}_{t,k} = a_k + \underbrace{\sum_v P_{k,v} P_{\Lambda,v} r_{t,\Lambda}}_{d_{t,k}} + \sum_{\substack{j \neq k \\ v}} P_{k,v} P_{j,v} (r_{t,j} - d_{t,j}) \quad (3.4a)$$

$$\min_{a^*, P^*} \sum_{\substack{t=1:T \\ k=1:m}} f(r_{t,k}) (r_{t,k} - \hat{r}_{t,k})^2 + \lambda (a_k^2 + |P|^2) \quad (3.4b)$$

We begin the training by initializing  $P$  to  $\tilde{P}$ , where  $\tilde{P}$  is the Cholesky factorization of the covariance matrix  $C$ , i.e.,  $C = \tilde{P}\tilde{P}^T$ . We compute the covariance matrix  $C$  on historical data. For each  $r_{t,k} \in \mathbf{R}$ , we update  $P$  and  $a_k$  by moving against the gradient. For  $1 \leq j \leq m; j \neq k$ , and  $1 \leq v \leq m$ , the updates are:

$$\begin{aligned}
\tau_{t,j} &= r_{t,j} - d_{t,j} \\
P_{k,v} &\leftarrow P_{k,v} + \eta(e_{t,k}(P_{\Lambda,v} \cdot r_{t,\Lambda} + \sum_{j \neq k} P_{j,v} \tau_{t,j}) - \lambda \cdot P_{k,v}) \\
P_{j,v} &\leftarrow P_{j,v} + \eta(e_{t,k} \cdot P_{k,v} \tau_{t,j} - \lambda \cdot P_{j,v}) \\
P_{\Lambda,v} &\leftarrow P_{\Lambda,v} + \eta(e_{t,k} \cdot P_{k,v} \cdot r_{t,\Lambda} - \lambda \cdot P_{\Lambda,v}) \\
\alpha_k &\leftarrow \alpha_k + \eta(e_{t,k} - \lambda \cdot \alpha_k)
\end{aligned}$$

### Comparison to Covariance Matrix

In portfolio construction, the minimum variance formulation is typically used to find the optimal portfolio (in terms of mean and variance of portfolio returns). The resulting portfolio provides the minimum expected variance of returns such that the expected mean return is at least the desired return. This formulation uses covariance estimates in the optimization function. In contrast, the constituents of the connectedness matrix are related to partial correlation estimates, which in turn have interpretation in terms of inverse covariance estimates [47]. The inverse covariance estimates capture the pairwise dependencies, or more precisely, the direct relationships between two equities. In contrast, covariance estimates capture both direct and indirect (i.e., transitive connections). We hypothesize that the direct estimates will be less noisy and more representative of the future than the indirect estimates (or the combination of them). Thus, we expect a portfolio constructed using the direct estimates (connectedness matrix) to perform better than the one using the combination of both direct and indirect estimates (covariance matrix).

## 3.4 Evaluation

We evaluate our method in two ways. First, we evaluate its accuracy in producing a list of equities ordered by their likelihoods of having large losses, given information

about the behavior of other equities. We then present and demonstrate the potential real world utility of a method that constructs portfolios using the learned relationships. The performance of portfolios constructed using our methods are compared to the performance of portfolios constructed using conventional approaches, including traditional correlation-matrix based methods.

### 3.4.1 Data

We use daily return data from CRSP<sup>1</sup>. We examine all 369 companies that were in the S&P500 from 2000 to 2012. This time period contains two major financial crises (2001 and 2008). The set of companies are from ten sectors: consumer discretionary, consumer staples, energy, financials, health care, industrials, information technology, materials, telecommunications services, and utilities.

### 3.4.2 Top-K Ranking

Given all returns for days 1 to  $T$  and returns on day  $T + 1$  for equities in  $A$ , we predict which equities from  $B$  will have events (losses greater than 10% on day  $T + 1$ ). We produce an ordered list of equities from  $B$ , ranked by their likelihoods of having events on day  $T + 1$  based on their predicted returns  $\hat{r}_{T+1}$ .

We use cost function  $f(x)$  to capture the differential impact of certain ranges of returns. The flexibility in choosing the cost functions allows us to optimize different aspects of the problem. Although arbitrary, for this experiment, we defined the event as  $r < -0.1$ . The predictive model will have the maximum ambiguity of whether an event occurred or not around the boundary (i.e., at  $\hat{r} = 0.1$ ). Therefore, in order to achieve higher accuracy in predicting returns around  $r = -0.1$ , we use  $f(x) = e^{-(x-r_\theta)^2/0.01}$ ;  $r_\theta = -0.1$  in Equation 3.2. This helps us achieve high MAP scores in producing the ordered list of equities that are likely to have events at  $-10\%$  level. We discuss the choice of cost function in Section 3.4.3.

---

<sup>1</sup>CRSP, Center for Research in Security Prices. Graduate School of Business, The University of Chicago (2004). Used with permission. All rights reserved. [www.crsp.uchicago.edu](http://www.crsp.uchicago.edu)

## Evaluation

We use mean average precision (MAP) to evaluate the correctness of the ordered list. Average precision (AP) is a popular measure that is used to evaluate an ordered list while taking into account both recall and precision [100]. MAP is the mean of average precision across a test set. For an ordered list of top- $k$  items, MAP and AP are given by:

$$AP(k) = \sum_{j=1}^k p(j)\Delta r(j) \quad (3.5)$$

$$MAP(k) = \sum_{i=1}^n AP_i(k)/n \quad (3.6)$$

Here,  $n$  is the size of the test set,  $p(j)$  is the precision at cut-off  $j$ , and  $\Delta r(j)$  is the change in the recall from  $j - 1$  to  $j$ . We produce a top-10 list, and evaluate with  $MAP(10)$ .

## Benchmark Methods

We compare the performance of our method with the following alternatives.

- Correlation(CR): Standard implementation.
- Partial correlation(PCR): Standard implementation.
- Extreme value correlation (EVCR): First, we apply a GARCH model to remove serial dependencies, if there are any. Then, we fit a univariate generalized Pareto distribution with tail fraction 0.1 on the innovations of the GARCH model. Finally, we model the multivariate distribution with a t-Copula, and learn the linear correlations on extreme values [21].

## Experimental Results

Since diversification inevitably involves using equities from multiple sectors, we focused on the question of which equities are connected within sectors. Our problem



formulation is that given a set of equities  $A$  on a given day, some of which had events and some of which didn't, which equities in a disjoint set  $B$ , are mostly to experience an event on the same day as those in  $A$ . We randomly select 20% of the companies in each sector for set  $B$ , and use the rest for set  $A$ . For each day, we train our factor model and learn the weights on last 500 days of historical data (consist of both  $A$  and  $B$ ), and produce an ordered list of equities from  $B$  according to their likelihoods of having an event on that day. We run our experiments from 2000 to 2012. We evaluate our methods only on days that had at least two events. Three sectors had less than 5 such days in the decade considered, and therefore are excluded in the experiments. Across all the sectors, there are 539 days that had two events out of 3019 days in the full dataset. We repeat this experiment 100 times for different  $A$  and  $B$ .

Table 3.2: MAP scores for different methods.

SECTORS	FAC	CR	PCR	EVCR
CONSUMER DISC.	<b>0.73</b> $\pm 0.093$	0.34 $\pm 0.08$	0.45 $\pm 0.119$	0.32 $\pm 0.07$
ENERGY	<b>0.75</b> $\pm 0.049$	0.64 $\pm 0.069$	0.72 $\pm 0.064$	0.62 $\pm 0.066$
FINANCIALS	<b>0.64</b> $\pm 0.05$	0.45 $\pm 0.072$	0.58 $\pm 0.059$	0.49 $\pm 0.07$
HEALTH CARE	<b>0.65</b> $\pm 0.176$	0.4 $\pm 0.127$	0.54 $\pm 0.217$	0.33 $\pm 0.142$
INDUSTRIALS	<b>0.64</b> $\pm 0.123$	0.37 $\pm 0.104$	0.46 $\pm 0.095$	0.37 $\pm 0.092$
INFORMATION TECH.	<b>0.58</b> $\pm 0.07$	0.39 $\pm 0.058$	0.49 $\pm 0.065$	0.4 $\pm 0.066$
MATERIALS	<b>0.84</b> $\pm 0.095$	0.68 $\pm 0.114$	0.77 $\pm 0.116$	0.66 $\pm 0.127$

Table 3.2 compares the MAP scores (higher is better) for our factor model (FAC) with the scores for the benchmark methods: correlation (CR), partial correlation (PCR), and correlation of extreme values (EVCR). The results are averaged over 100 runs for each sector. The best result for each sector is in bold face. Results for FAC are statistically different ( $p$ -value  $< 0.001$ ) from the results of every other method under a paired t-test.

Our factor model consistently outperforms the other methods. EVCR often underperforms PCR, and at times, CR. We conjecture that the inflexibility of the parametric modeling of the tails and not considering the relationship between non-extreme values contribute to this failure. The poor performance of EVCR is striking because t-copula is widely used in financial risk assessment, especially in the pricing of collat-

Table 3.3: MAP scores for different sizes of unknown set  $B$ .

SECTORS	10%	20%	40%
CONSUMER DISC.	0.77	0.73	0.53
ENERGY	0.86	0.75	0.65
FINANCIALS	0.76	0.64	0.49
HEALTH CARE	0.82	0.65	0.52
INDUSTRIALS	0.77	0.64	0.47
INFORMATION TECH.	0.71	0.58	0.43
MATERIALS	0.92	0.84	0.76

eralized debt obligations (CDO) [65, 59].

Table 3.3 compares the MAP scores for different sizes of known and unknown sets. We change the size of  $B$  as a fraction of the total number of companies available, using 10%, 20%, and 40%. Set  $A$  contains the rest. Even when we use the information from only 60% of the companies and to rank the equities in the other 40%, our method remains effective. Notice that FAC for a known set of 60% of the data outperforms other methods for known set 80% of the data.

### Case study: Bank of America

Though univariate time series of daily equity returns lack significant autocorrelation and stationarity, multivariate time series of returns exhibits consistent correlation among themselves and with the market that persist over time (Figure 3-1). This characteristic is the fundamental behind portfolio diversification [16, 46, 48]. As an example, we look at an S&P500 constituent, Bank of America (BAC).

Between 2001 and 2011, BAC had 29 events, i.e., daily losses of at least 10%. Figure 3-2 shows the change in the parameters learned using the factor model for BAC. Notice that the market dependence (beta) rises after the 2008/9 crisis peaking at about 160% of that in 2001, while the inherent return (alpha) falls to  $-220\%$  of that in 2001. Further, the “herding effect,” as given by the spread in the weights, widens during the crisis of 2008/9. While BAC’s average connectedness changes only by a small amount, the spread gets significantly larger. This indicates that BAC becomes heavily connected to a smaller set of other equities in the financial sector.

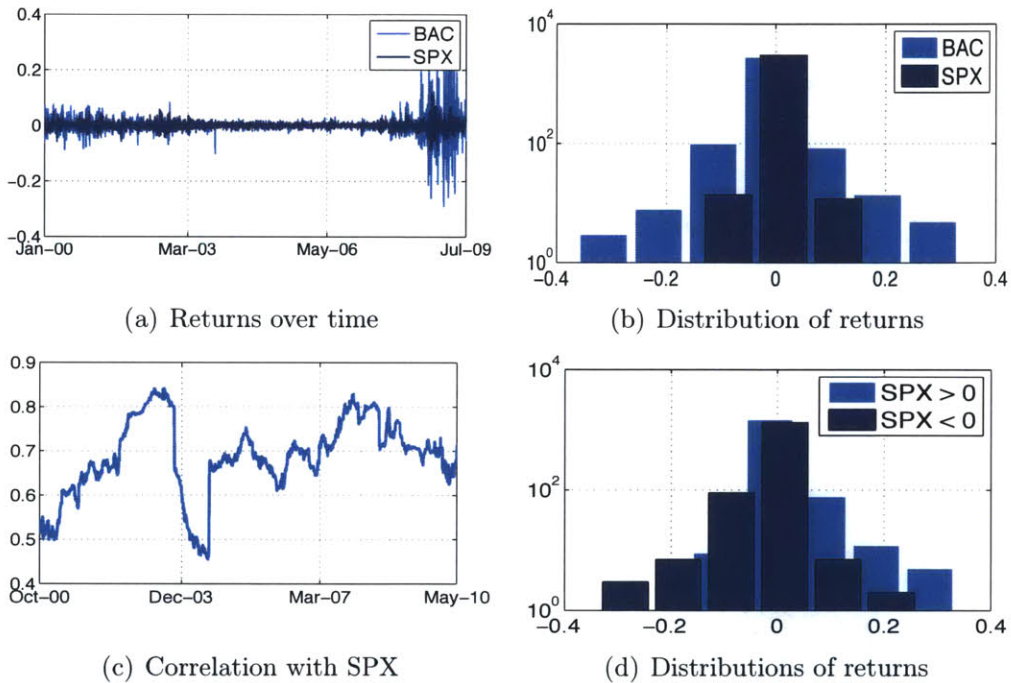
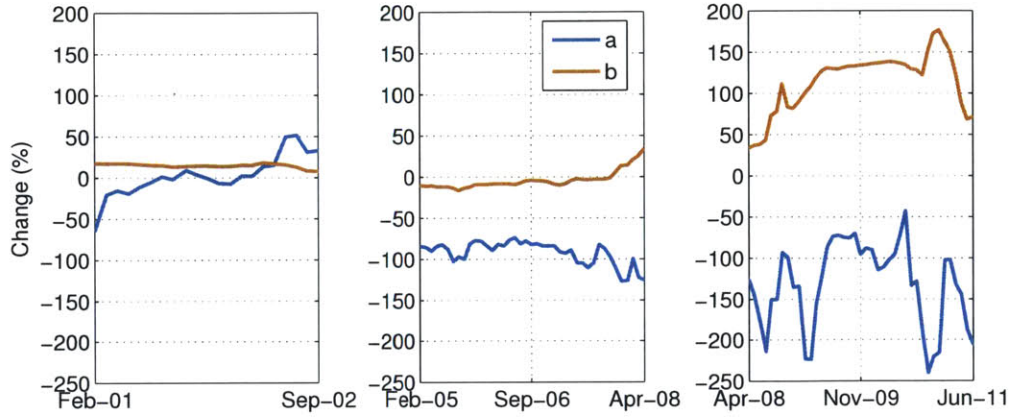
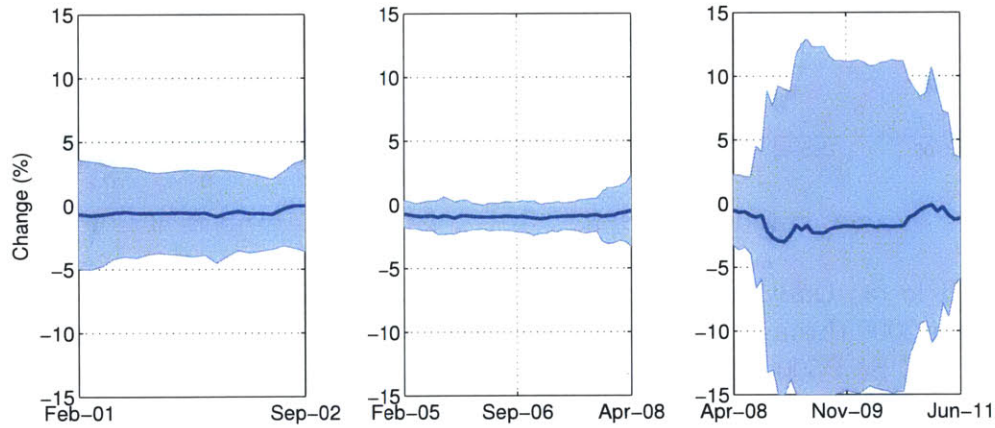


Figure 3-1: (a) Daily returns of Bank of America (BAC) and S&P500 Index (SPX) from year 2000 through 2011. (b) Distribution of the daily returns of Bank of America (BAC) and S&P500 Index (SPX) from the same period. (c) Correlations between BAC and SPX from year 2000 through 2012 computed on a sliding window of 200 days. (d) Conditional distributions of daily returns of BAC when SPX is positive (and negative) from the same period. Conditional distributions differ significantly from each other on tails.



(a) Market dependence ( $b_k$ )



(b) Co-movements with neighbors ( $w_{j,k}$ )

Figure 3-2: Connectedness estimated by the factor model for BAC: (a) Active return  $a$ , and market sensitivity  $b$  for BAC (Equation 3.1). (b) Mean and one standard deviation range of the correlation weights  $w_{j,k}$  between BAC and neighbors.

### 3.4.3 Portfolio Construction

The major application of our method is the reduction of large losses in equity portfolios. Since there is a tradeoff between risk and expected return, portfolio design usually starts by the investor choosing a desired level of expected return (or a risk tolerance). For a given desired expected return  $r_e$ , in the absence of any side information, the minimum variance portfolio (MVP) is the optimal portfolio [62]. For the MVP, portfolio weights  $\omega$  are derived by solving the optimization problem:

$$\begin{aligned} \min_{\omega} \quad & \frac{1}{2} \omega^T C \omega & (3.7) \\ \text{subject to} \quad & \sum_{j=1}^m \bar{r}_j \omega_j \geq r_e \\ & \sum_{j=1}^{m+1} \omega_j = 1; 0 \leq \omega_j \leq 1, j = 1, \dots, m \end{aligned}$$

Here,  $C$  is the covariance matrix of returns, and  $\bar{r}_j$  is the expected return of equity  $j$ . Typically, the covariance and the expected return are calculated from historical data. Here, we assume fixed capital (no leverage), no short positions, monthly (every 20 days) readjusted portfolio weights, and we ignore the costs of transactions.

When a portfolio is heavily diversified, the expected return is smaller. In our formulation the desired expected return  $r_e$  governs the amount of diversification. The range of achievable values for  $r_e$  is the minimum and maximum of the expected returns of the equities. Maximum expected return is achieved by owning the equity with the highest historical return. Minimum risk relative to the market is achieved by owning everything in that market.

Our method of minimizing the co-occurrences of large losses is beneficial when the portfolio is not significantly diversified, e.g., to increase the expected return. It has been shown that 90% of the maximum benefit of diversification is achieved with portfolios containing roughly 5% of the market constituents [79]. This led us to set  $r_e$  to the 95<sup>th</sup> percentile of the expected returns of the equities. This setting causes the optimization to choose about 3 – 5 equities per sector (5% to 10%).

We demonstrate our method’s utility by building portfolios with our connectedness matrix  $G$  (Section 3.3.4), and compare their performance to portfolios built using methods drawn from the financial literature. For the factor model, we use an approach similar to the traditional MVP except we replace the correlation estimates in  $C$  with connectedness matrix  $G$  and  $\bar{r}_j$  with active return  $a_j$ . We learn both  $G$  and  $a_j$  using our factor model. We use the same optimization as MVP, i.e., Equation 4.8. Here, we consider long-only portfolios, and therefore we would like to minimize the co-occurrences of large negative returns. We use  $f(x) = e^{-x/\gamma}$  in Equation 3.2. In Section 3.4.3, we discuss the choice of cost function in detail.

1.  $FAC_{0.1}$  is the factor model with cost function  $f(x) = e^{-x/0.1}$  applied. This model heavily focuses on minimizing the co-occurrences of large losses. This risk avoidance results in smaller overall return compared to  $FAC_{10}$ .
2.  $FAC_{10}$  is the factor model with cost function  $f(x) = e^{-x/10}$  applied. It captures the connections by focusing mainly on negative returns. It produces significantly larger overall returns, at the cost of larger worst case daily losses.

The cost functions for both  $FAC_{0.1}$  and  $FAC_{10}$  are normalized by their maximums, i.e.,  $f(x) \leftarrow \frac{f(x)}{\max_x f(x)}$ .

## Benchmark Methods

We compare the performance of our method in building sector wide portfolios with the following alternatives.

- COV: This is our baseline MVP portfolio that is built using the estimated covariance matrix  $C$ . This is the conventional approach.
- PCR: Partial correlation is substituted for correlation in estimating  $C$ .
- EVCR: Extreme value correlation is substituted for correlation in estimating  $C$ .

Later, when we evaluate the performance on market wide portfolios, we also consider the following benchmark portfolios to characterize the performance of our approach.

- **EW:** Equi-weighted portfolio, where the portfolio is rebalanced to equal weights. This portfolio incurs large turnovers. Therefore, in practice it is hard to implement. While this portfolio has been shown to possess good theoretical performance, i.e., extremely low volatility with high returns, large turnover will negate most of the returns via transaction costs.
- **MV:** Portfolio where the optimization minimizes conditional value at risk (CVaR) at 5% level [81]. CVaR, also known as expected shortfall, measures the likelihood (at a specific confidence level) that a specific loss will exceed the value at risk. For instance, CVaR at 5% level gives the expected return on the portfolio in the worst 5% of the cases [81].
- **SPX:** S&P500 index.

## Results

Table 3.4 summarizes the return characteristics for the three sectors with the most events. We re-weight the portfolio monthly, and estimate the returns daily. We use the following criteria to evaluate the performance of the portfolios.

- **Largest loss in a day:** Given by  $\min(r_t)$ .
- **The expected shortfall (also known as CVaR) at 5% level** gives the expected return on the portfolio in the worst 5% of the cases [81].
- **Max drawdown:** Maximum drawdown is the largest peak-to-subsequent-trough decline in cumulative return. It is given by  $M = \min_{1 \leq i \leq j \leq T} \prod_{t=i}^j (r_t + 1) - 1$ .
- **Annualized return:** This is the overall return from year 2001 to 2012, i.e.,  $R_T$  on December 31, 2012. Cumulative return  $R_T$  from day 1 to day  $T$  is given by  $R_T = \prod_{t=1}^T (r_t + 1) - 1$ . We quote annualized total return (i.e.,  $(R_T + 1)^{1/11} - 1$ ) in our results.
- **Information ratio:** Information ratio measures the excess return for additional risks taken [85]. It is given by  $I = E(r - r_\Lambda) / \sqrt{\text{var}(r - r_\Lambda)}$ , where  $r$  is the daily

return of the portfolio and  $r_A$  is the reference return (return on the S&P500 index). We quote annualized information ratio (i.e.,  $I\sqrt{252}$ ) in our results. A positive information ratio implies that excess return is greater than the additional risk taken.

- Sharpe ratio: The Sharpe ratio measures the risk-adjusted return [85]. It is given by  $S = E(r - r_f) / \sqrt{\text{var}(r)}$  where  $r_f$  is the risk free return (assumed to be 1% annually). We quote annualized Sharpe ratio (i.e.,  $S\sqrt{252}$ ) in our results.

Table 3.4 shows that by learning the connectedness between equities, our portfolios cannot only beat the market (positive information ratio), but also beat the optimal (minimum-variance) portfolios. We note that  $FAC_{0.1}$  reduces large losses better than  $FAC_{10}$ , and  $FAC_{10}$  achieves better overall returns than  $FAC_{0.1}$  as intended.

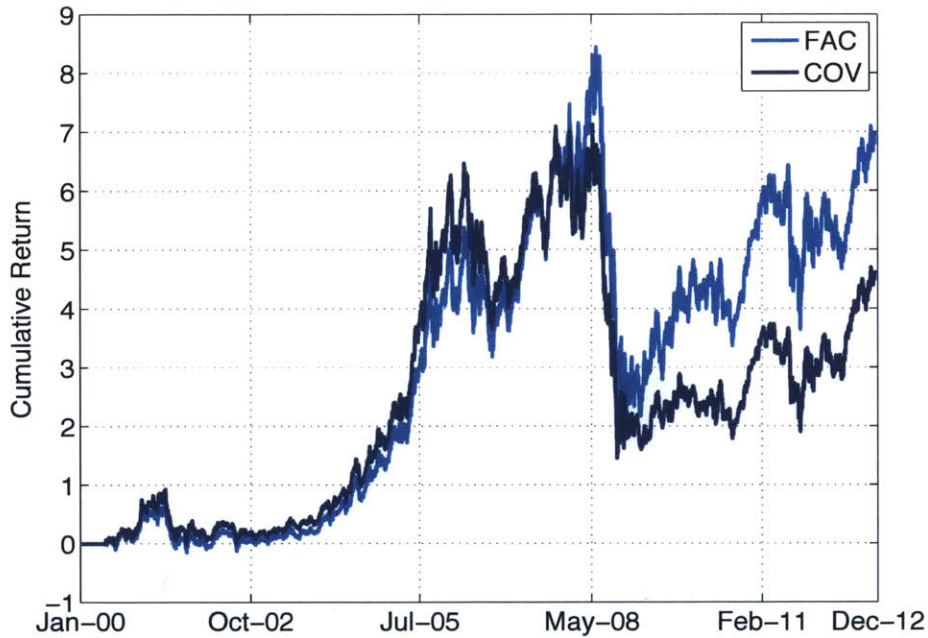
Table 3.4: Characteristics of portfolio returns

MEASURES	ENERGY			HEALTH CARE			MATERIALS		
	COV	$FAC_{0.1}$	$FAC_{10}$	COV	$FAC_{0.1}$	$FAC_{10}$	COV	$FAC_{0.1}$	$FAC_{10}$
LARGEST LOSS IN A DAY	-0.23	<b>-0.16</b>	-0.22	-0.11	<b>-0.08</b>	-0.11	-0.19	<b>-0.12</b>	-0.19
EXPECTED SHORTFALL (5%)	-0.05	<b>-0.04</b>	-0.06	-0.04	<b>-0.03</b>	-0.04	-0.06	<b>-0.04</b>	-0.06
MAX DRAWDOWN	-0.75	<b>-0.62</b>	-0.72	-0.55	<b>-0.46</b>	-0.53	-0.81	<b>-0.66</b>	-0.82
ANNUALIZED RETURN	0.16	0.16	<b>0.2</b>	0.21	0.13	<b>0.25</b>	0.22	0.12	<b>0.22</b>
INFORMATION RATIO	0.57	0.65	<b>0.67</b>	0.81	0.57	<b>0.96</b>	0.73	0.62	<b>0.75</b>
SHARPE RATIO	0.52	0.56	<b>0.59</b>	0.77	0.62	<b>0.86</b>	0.61	0.48	<b>0.62</b>

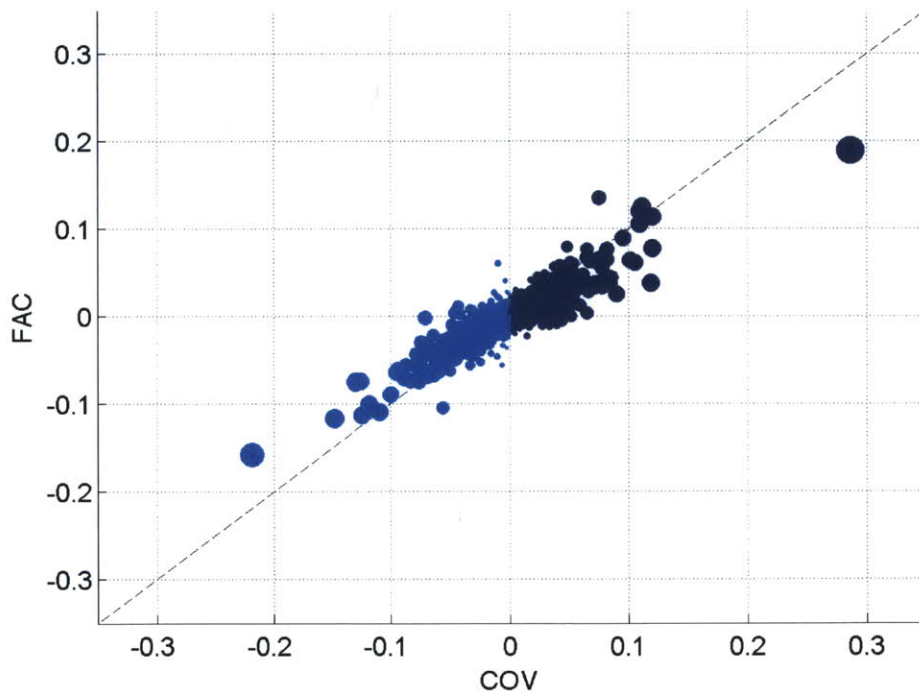
Figure 3-3 shows the impact of our method on returns in the energy sector. Until the 2008 crisis, because the energy sector remained calm, our FAC model performed comparably to COV. Note that 2001 crisis, unlike 2008 crisis, was limited to few sectors not including energy. After the collapse in May 2008, our model began learning new connectivities related to large negative returns and was able to reduce large losses (Figure 3-3(a)) late that year and going forward. It took about a year to learn the new model, but it persisted long enough to be useful. Figure 3-3(b) demonstrates the effectiveness of our method in making the large negative returns smaller without significantly affecting positive and small negative returns. The largest daily loss dropped 31%, i.e., from 22.6% to 15.6%.

Figure 3-4 shows the equity weights learned using the connectedness matrix for the energy sector. Until August 2008, our factor model based portfolio consistently fo-





(a) Cumulative Returns



(b) Daily Returns

Figure 3-3: Cumulative returns (a) of  $FAC_{10}$ , and daily returns (b) of  $FAC_{0.1}$  against that of COV for the energy sector. In figure (b), positive returns with COV are highlighted with the darker color. Size of the points correspond to the absolute values of the returns with COV. Returns above the line correspond to an improvement with  $FAC_{0.1}$ . A clockwise shift that is more prominent in the negative side (lighter region) is noticeable.

cused on two equities in the energy sector: Occidental Petroleum Corporation (OXY) and Sunoco (SUN). In the aftermath of the 2008 crisis, our model increases the diversification. Following that, the portfolio changed the focus to equities like Exxon Mobil Corporation (XOM) so that it wouldn't miss the opportunities that became available during the recovery.

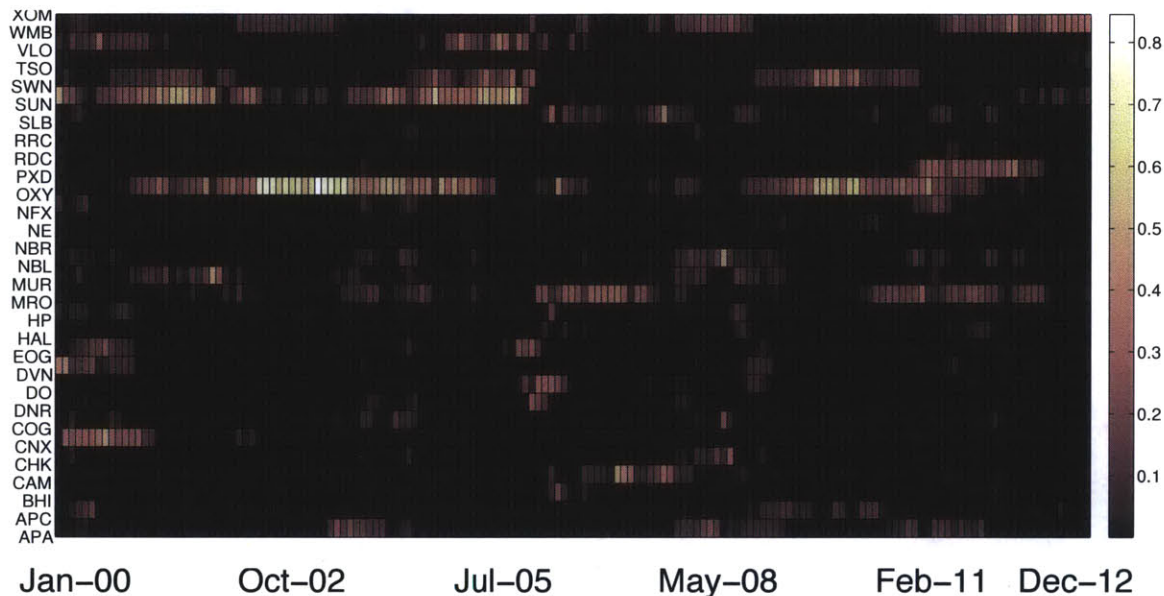


Figure 3-4: Heat map of the weights of the equities for the portfolios built for the energy sector.

Next, we build a market wide portfolio by combining the portfolios built for each sector weighted equally. Since, this design provide more diversification (via distributing across sectors) we set the desired return  $r_e$  to the 98<sup>th</sup> percentile of the expected returns of the equities. We compare our method with COV, PCR and EVCR. We also compare these “MVP-like” portfolios with other benchmark portfolios. In Table 3.5 we demonstrate the effectiveness of our model in constructing a market wide portfolio.  $FAC_{0.1}$  achieves an annualized information ratio of 1.4. It is significant that, the portfolio built with our method outperforms the “hard-to-beat” (and hard to implement) equal-weighted portfolio [76] in terms of annualized return and Sharpe ratio.

Table 3.5: Market wide portfolios.

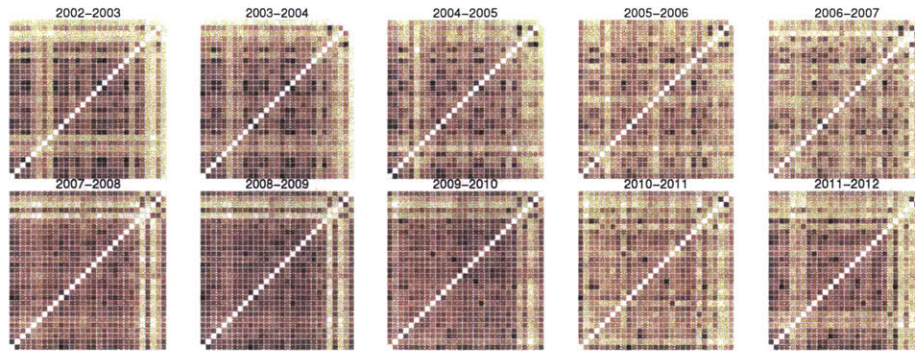
MEASURES	$FAC_{0.1}$	COV	PCR	EVCR	EW	MV	SPX
WORST DAY	<b>-0.09</b>	-0.15	-0.14	-0.15	-0.11	-0.11	-0.09
EXPECTED SHORTFALL(5%)	<b>-0.03</b>	-0.04	-0.04	-0.04	-0.03	-0.04	-0.03
MAX DRAWDOWN	<b>-0.54</b>	-0.67	-0.68	-0.6	-0.61	-0.65	-1.07
ANNUALIZED RETURN	<b>0.18</b>	0.17	0.17	0.16	0.14	0.15	-0.01
INFORMATION RATIO	1.06	0.91	0.88	0.85	<b>1.68</b>	1.5	N/A
SHARPE RATIO	<b>0.71</b>	0.63	0.61	0.6	0.6	0.61	0.1

## Temporal Evolution

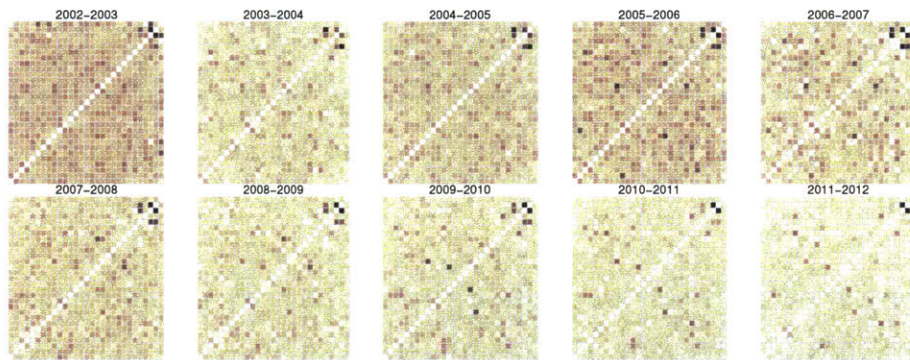
In this section, we look at the temporal evolution of the connectedness matrix. Figure 3-5 shows the heat map of correlation matrices and connectedness matrices from year 2002 to 2012.

In correlation matrices, we observe that the density of the connections drops from 2002-2003 to 2006-2007, and then increases dramatically over next two years. In connectedness matrices, we observe that the density of the connections steadily increases until 2006, drops slightly for 2006-2007, and then increases again for the next two years before coming down in 2012. We can attribute this pattern to the market crises of 2001 and 2008.

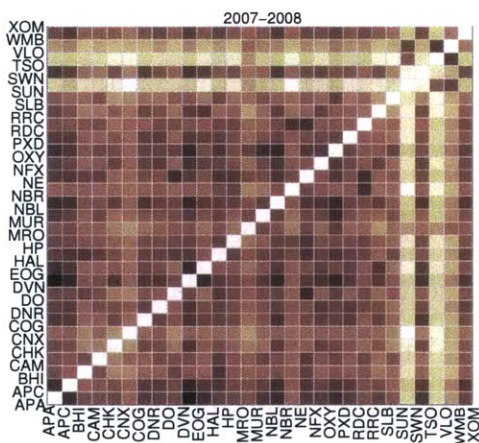
Unlike the 2008 crisis, the 2001 crisis didn't extend outside certain sectors, particularly to the energy sector. Increased density observed in the correlation matrix for the energy sector is probably due to the market dependence (equities become more dependent on the market during the times of crises) rather than direct relationships among equities. Because the connectedness matrix captures only direct relationships it avoids this issue. Unlike correlation matrices, connectedness matrices show an increasing trend in the density of connections from 2004 - 2008, indicating that the clusters of stocks in energy sector becoming connected, i.e., there is increased likelihood of experiencing co-occurrences of large losses. It is this information that help us minimize the co-occurrences of large losses as shown in Figure 3-3.



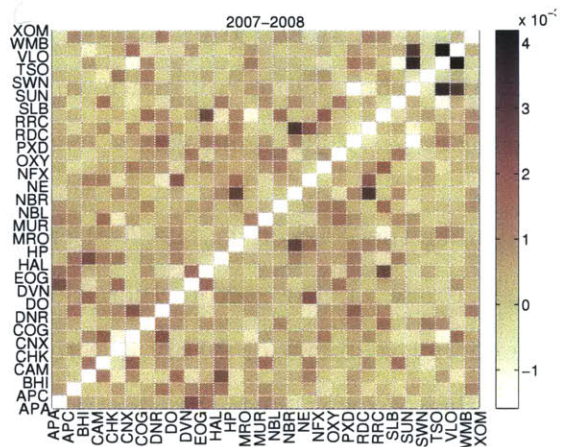
(a) Temporal Evolution of correlation matrix



(b) Temporal Evolution of connectedness matrix



(c) Correlation matrix for 2011-2012



(d) Connectedness matrix for 2011-2012

Figure 3-5: Correlation matrix (a) and connectedness matrix (b) computed on a 500 day window spanning two years for Energy sector. Enlarged view for 2007-2008 with the names of the stocks labeled in (c) - (d).

## Cost Function

When learning the weights for the factor model using least squares minimization, we use a cost function  $f(x)$  to capture the differential impact of certain ranges of returns (Equation 3.2). The flexibility in choosing the cost functions allows us to optimize different aspects of the problem.

For top-k ranking evaluation (Section 3.4.2), we define events at or below  $-10\%$  daily return. There, in order to achieve higher accuracy, we use  $f(x) = e^{-(x-r_\theta)^2}$ ;  $r_\theta = -0.1$ , because the maximum ambiguity is at the boundary.

For the portfolio construction task (Section 3.4.3), we consider long-only portfolios. There, in order to minimize the co-occurrences of large negative returns, we use  $f(x) = e^{-x/\gamma}$ . Table 3.6 summarizes the characteristics of the corresponding portfolio returns for the energy sector.

Table 3.6: Portfolio returns

MEASURES	COV	FAC	
		$\gamma = 0.1$	$\gamma = 10$
LARGEST LOSS IN A DAY	-0.23	<b>-0.16</b>	-0.22
EXPECTED SHORTFALL (5%)	-0.05	<b>-0.04</b>	-0.06
MAX DRAWDOWN	-0.75	<b>-0.62</b>	-0.72
ANNUALIZED RETURN	0.16	0.16	<b>0.2</b>
INFORMATION RATIO	0.57	0.65	<b>0.67</b>
SHARPE RATIO	0.52	0.56	<b>0.59</b>

We observe that with decreasing  $\gamma$ , the cost function becomes steeper, results in higher penalty for the squared differences (i.e., tighter fit) around large losses, and the model focuses more on minimizing the co-occurrences of large losses. This risk avoidance results in smaller overall return compared to the ones with larger  $\gamma$ .

As we stated earlier, an important aspect of our method is that it finds a balance between learning infrequent (but important) connections that are based on large losses and learning connections based on other daily returns. Thus, the portfolio built with our method achieves high risk adjusted returns and demonstrates that risk management can be a source of alpha. In contrast, the methods that solely focus on capturing rare events [25] and co-occurrences of rare events [77] (as represented by

EVCR in our experiments) miss out of common day occurrences.

## Reducing Survivorship Bias and Generalizability

We built our model and evaluated the portfolio performances using the data for equities that were in the S&P 500 continuously from 2001 to 2012. In this section we address two concerns: the survivorship bias that is inherent in the dataset we used and the generalizability of our model. Because of the survivorship bias the results are overstated for our method and benchmark models. For example, companies that left the S&P 500 between 2001 and 2012 are not part of our dataset, and these companies are likely to be more volatile than the ones included in our dataset.

In this section, we consider the full set of S&P 500 equities, i.e., when constructing a portfolio, we consider all the equities that exist in the S&P 500 list continuously throughout each training time period. The number of equities considered ranges from 415 to 477 in this set. This selection is free of survivorship bias. Next, in order to study the generalizability of our model parameters, we also consider the full set of S&P Smallcap 600 equities. The number of equities considered ranges from 404 to 512 in this set. Further, we also test our method on the data from both datasets for 2013.

We use the data from CRSP. We perform the portfolio construction described in Section 3.4.3. We use the same cost function and hyper parameters from the previous experiment.

Table 3.7 shows the impact of the survivorship bias on S&P 500 dataset. We observe that with survivorship bias removed annualized return drops from 18% to 13%. Note, however, that the relative improvement achieved by our method is greater on the data with survivorship bias addressed. For instance, the increase in the information ratio is dramatic.

Table 3.8 compares the return characteristics on the S&P 500 dataset and the S&P 600 Smallcap dataset. On both datasets,  $FAC_{0.1}$  achieves the best annualized returns while minimizing the risk. On the S&P Smallcap 600 dataset, COV and PCR achieve smaller annualized returns compared to that on the S&P 500 dataset,

Table 3.7: Impact of survivorship bias on Market wide portfolios: 2001-2012.

MEASURES	WITH SURVIVORSHIP BIAS			NO SURVIVORSHIP BIAS			SPX
	$FAC_{0.1}$	COV	PCR	$FAC_{0.1}$	COV	PCR	
WORST DAY	-0.09	-0.15	-0.14	-0.08	-0.09	-0.09	-0.09
EXPECTED SHORTFALL(5%)	-0.03	-0.04	-0.04	-0.02	-0.03	-0.03	-0.03
MAX DRAWDOWN	-0.54	-0.67	-0.68	-0.54	-0.71	-0.75	-1.07
ANNUALIZED RETURN	0.18	0.17	0.17	0.13	0.09	0.09	-0.01
INFORMATION RATIO	1.06	0.91	0.88	1.25	0.73	0.75	N/A
SHARPE RATIO	0.71	0.63	0.61	0.67	0.45	0.44	0.1

despite the fact that the S&P 600 Smallcap index (IJR) outperforms the S&P 500 index (SPX). This is probably due to historical estimates being better indicators of future returns for the S&P 500 equities compared to the S&P 600 Smallcap equities. However, we don't have a good explanation for why COV and PCR perform poorly compared to the index on S&P 600 Smallcap dataset. Our method achieves 14% annualized return on the S&P Smallcap 600 compared to 13% on the S&P 500. We attribute the improvement in annualized returns and risk measures on both datasets for our method to the explicit modeling of excess returns (alpha) and connections based on co-occurrences of large returns in our model.

Table 3.8: Market wide portfolios: 2001-2012.

MEASURES	S&P 500			SPX	S&P 600 SMALLCAP			IJR <sup>1</sup>
	$FAC_{0.1}$	COV	PCR		$FAC_{0.1}$	COV	PCR	
WORST DAY	-0.08	-0.09	-0.09	-0.09	-0.1	-0.1	-0.1	-0.11
EXPECTED SHORTFALL(5%)	-0.02	-0.03	-0.03	-0.03	-0.03	-0.04	-0.04	-0.04
MAX DRAWDOWN	-0.54	-0.71	-0.75	-1.07	-0.55	-0.86	-0.87	-0.73
ANNUALIZED RETURN	0.13	0.09	0.09	-0.01	0.14	0.03	0.03	0.1
INFORMATION RATIO	1.25	0.73	0.75	N/A	0.84	0.23	0.24	0.81
SHARPE RATIO	0.67	0.45	0.44	0.1	0.59	0.22	0.22	0.46

Next, we perform the portfolio construction for year 2013 on both datasets. Table 3.9 summarizes the return characteristics. All three methods achieve annualized returns as high as 30%. Note that 2013 was a particularly good year for the stock market, the S&P 500 index returned 29%. Yet, our method outperforms the alternatives with 49% annualized return on the S&P 500 equities and 34% annualized return on the S&P Smallcap 600 equities.

<sup>1</sup>iShares S&P 600 Smallcap index fund.

Table 3.9: Market wide portfolios: 2013.

MEASURES	S&P 500			SPX	S&P 600 SMALLCAP			IJR <sup>1</sup>
	$FAC_{0.1}$	COV	PCR		$FAC_{0.1}$	COV	PCR	
WORST DAY	-0.02	-0.03	-0.03	-0.03	-0.03	-0.4	-0.4	-0.03
EXPECTED SHORTFALL(5%)	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
MAX DRAWDOWN	-0.08	-0.13	-0.14	-0.1	-0.12	-0.11	-0.12	-0.09
ANNUALIZED RETURN	0.49	0.35	0.36	0.29	0.34	0.3	0.3	0.03
INFORMATION RATIO	2.66	0.75	0.8	N/A	0.37	0.15	0.12	1.39
SHARPE RATIO	3.33	2.28	2.29	2.37	2.01	1.65	1.61	2.54

### 3.5 Application Considerations

Minimum-variance portfolios built with our method achieve higher risk adjusted return than that built with traditional measures such as covariance. Like traditional minimum-variance portfolios that are built with covariance matrix, our method rebalances the portfolio monthly, and therefore incurs transaction costs. COV leads to slightly more transactions than our method, and would therefore incur higher transaction costs. Unlike equi-weighted portfolio, daily changes in portfolio weights are small for minimum-variance portfolios, and the results hold when we update the portfolio weights daily.

Relationships among equities can be viewed as occurring on three time scales: long-term (e.g., sector grouping), mid-term (e.g., based on fundamentals and technicals), and short-term (e.g., based on real-time news and announcements). In this work, we capture only the medium-term relationships. This may result in a slower response when market conditions change rapidly as in 2008 (Figure 3-3). Incorporating news and sentimental analysis into our model could address this (see Chapter 4).

We show that by limiting the large losses, risk management can be a source of excess returns. A portfolio manager might rely on non-linear payoff profiles to maximize the Sharpe ratio using risk management techniques. For example, he could “sharpen the Sharpe ratio” by truncating the right tail and the fat left tail of the return distribution by buying deep out of the money put and call options [39]. In contrast to this approach, our method provides not only high Sharpe ratio but also significantly high total return and information ratio.



A major application of our method is in finding the most suitable equity that can be added to an existing equity portfolio. Many institutional investors have policies in place that require them to maintain long only portfolios. Because of the size of their holdings and for other reasons (e.g., tax consideration) they might not be able to change their existing positions. In such cases, if they are to pick a new equity they could use our method to find the one that is least likely to increase the tail risk of their new portfolio. This is the problem definition that we began with in the first place (see Section 3.4.2).

During the 2008 crisis, all equities became heavily correlated as the market crashed, and market risk governed returns (Figure 3-2). Without short positions (or derivatives that simulate short positions), this kind of risk cannot be diversified away. Because we focus on minimizing the co-occurrences of large losses, our problem formulation does not permit negative weights in the portfolio optimization (Equation 4.8).

In such cases, we can hedge against such risk factors (market wide or sector wide movements) using future contracts. For example, if the daily return of our portfolio is  $r_t$  on day  $t$ , and risk factors are  $\Lambda_t^i; i = 1 \dots k$ , then we can construct a linear risk model by,

$$r_t = \alpha + \beta_1 \Lambda_t^1 + \dots + \beta_k \Lambda_t^k. \quad (3.8)$$

When such risk factors have a clear definition (e.g., market risk), are statistically related (high  $R^2$  and  $\beta$  for the linear risk model), and have tradable derivatives we can hedge away the risk [56]. For the portfolio built for energy sector ( $FAC_{10}$ ), if we consider the S&P500 index for market risk, we have a statistically significant connection (correlation = 0.6,  $\beta = 1.114$  and  $R^2 = 0.98$ ). Further, there exist many liquid futures and forward contracts on S&P500 index to capture the full economic effects.

The characterization in Equation 3.8 provides the following expected return and

variance for a portfolio:

$$E[r_t] = \alpha + \beta_1 E[\Lambda_t^1] + \dots + \beta_k E[\Lambda_t^k] \quad (3.9)$$

$$Var[r_t] = \beta_1^2 Var[\Lambda_t^1] + \dots + \beta_k^2 Var[\Lambda_t^k] \quad (3.10)$$

While hedging away the exposure to a risk factor will reduce the portfolio risk (by reducing  $Var[r_t]$ ), it will also result in reduced expected returns (by reducing  $E[r_t]$ ). Litterman (2006) argues that such risk exposures are “exotic betas” and not worth the premium [56]. As an alternative, investors could use a deep out of the money (DOTM) options to limit their risk to only significant changes in the risk factors.

### 3.6 Summary

We presented a method for learning connections between financial time series in terms of large losses. We use a factor model, and we modeled daily returns using three factors: active return, market sensitivity, and connectedness of returns. We learned these factors using a recursive regression. We employ a model independent approach to prioritize the connections on large losses. Thus, our method allows the user to balance the need to minimize the co-occurrences of large losses against the need to minimize overall connections. We solved the regression problem using an unconstrained least squares optimization that ensures that the resulting matrix is positive semi-definite so that it can be used in portfolio construction.

We evaluated our method in two ways. First, we evaluated its accuracy in producing a list of equities ordered by their likelihoods of having large losses, given information about the behavior of other equities. We then presented and demonstrated the potential real world utility of a method that constructs portfolios using the learned relationships. The performance of portfolios constructed using our methods were compared to the performance of portfolios constructed using conventional approaches, including traditional correlation-matrix based methods. Portfolios constructed using our method not only “beat the market,” but also beat the so-called

“optimal portfolios.” More importantly, portfolios constructed using our method significantly reduce the tail risk without sacrificing overall returns.



## Chapter 4

# LEARNING THE RELATIONSHIP BETWEEN EQUITY RETURNS AND NEWS RELEASES

In this chapter, we consider the task of learning the influence of news on the returns of equities using sentiment data from Thomson Reuters, called Thomson Reuters NewsScope dataset. On the firm specific news releases collected from more than 60 reporting sources, Thomson Reuters uses a proprietary natural language processing based algorithm to score the news feeds, and constructs a database of sentiment scores (positive or negative). Thus, Thomson Reuters NewsScope database assigns “hard” number to a “soft” concept of news sentiments.

A problem with using news releases to predict future return is that often, the information in the news has already been incorporated into the pricing of the equity. We describe a machine learning-based method to model a relationship between the equity return that is unexplained by the market return (excess return) and the amount of sentiment in the news releases that hasn’t been already reflected in the price of equities (excess sentiment).

We formulate the problem as recursive regression, and we use an unconstrained optimization to learn the model. We show that while using news directly yields an annualized return of 22% over a 10-year period, our proposed way of handling the

past boosts the annualized return to 34% over the same period.

## 4.1 Introduction

In finance, a widely used assumption is that financial markets are “informationally efficient.” This is known as the efficient market hypothesis (EMH). Under this hypothesis, the prices of assets reflect all past information, and in the strong form of the theory, also hidden (insider) information [31].

Empirical analyses have showed inconsistencies with EMH theory. There has been a growing interest in identifying various features (particularly sentiments of news releases) that could influence equity prices.

Using sentiment data from Thomson Reuters, we analyze the influence of news on the returns of equities. The data contain historical records of news feeds and corresponding sentiment scores as computed by a Thomson Reuter’s propriety algorithm.

A problem with using news releases to predict future return is that often the information in the news has already been incorporated into the pricing of the equity. This makes it look as if the equity prices “anticipate” future news flows. In order to address this, we build a model of the sentiment of the news releases based on the past returns and past news flows, and then use this model to discount the expected sentiment from future news releases. We then model the relationship between the equity return that is unexplained by the market return (excess return) and the sentiment that hasn’t been incorporated into the price of the equity (excess sentiment). Here, we hypothesize that the excess sentiment would be the dominant factor contributing to large excess returns. Learning the relationship between the excess return and the amount of sentiment that hasn’t been already reflected in the price of the equity, and exploiting this knowledge in portfolio construction is the focus of this chapter.

We formulate the problem as a recursive regression, and we use an unconstrained optimization to learn the model. Our experiments provide strong evidence that by exploiting these learned relationships we can build portfolios that outperform portfolios constructed using techniques drawn from the literature. As in the previous chapter,

the comparison is done using annualized return, minimum daily return, maximum drawdown, information ratio, and the Sharpe ratio [85].

## 4.2 Related Work

Our work lies at the intersection of statistical machine learning and behavioral finance. We review the literature related to our work on both behavioral finance and data mining applied to finance.

### 4.2.1 Efficient Market Hypothesis and Events

Eugene Fama, who won the Nobel prize in economics for his work on the efficient market hypothesis, shows that although anomalies such as over reaction and post-event reversals are common in short term, they are chance results, i.e., their opposites are equally likely [32]. His article "The Adjustment of Equity Prices to New Information" was the first study that sought to analyze how equity prices respond to an event [30].

### 4.2.2 Behavioral Finance and Data Mining

Duran (2007) analyzes the precursors and aftershocks of significant price changes, and finds the presence of over reactions in equity return time series in absence of any significant change in the fundamentals [28]. Stambaugh (2012) explores a broad set of anomalies in cross-sectional equity returns and the role played by investor sentiment. They show that a long-short strategy is more profitable following high levels of sentiment [91]. They show that the short leg of each strategy is more profitable following high sentiment, and that long legs of the strategies do not benefit from the sentiment.

Leonid Kogan's work in asset pricing models includes development of data-driven factor models to explain asset pricing anomalies [51, 52].

### 4.2.3 Sentiment Mining Applied to Finance

In the last few years, advances in computational power have made mining large numbers of text documents possible. NewsCATS, an automated text categorization system uses financial news to predict equity returns [67, 68]. Schumaker's works on financial news mining shows that machine learning methods can be used to predict future equity returns [84].

Researchers have found that the sentiment of tweets is correlated with large equity returns, and use message volume to predict next-day trading volume [90]. Researchers have also predicted the daily up and down changes in the closing value of the DJIA, by analyzing the sentiment of the text content of daily Twitter feeds [12]. Mao (2012) looks at the equities in the S&P 500 and the tweets mentioning them, and shows that the daily number of tweets is correlated with equity market statistics such as closing price and trade volume [61]. Furthermore, they show that twitter data can be used to predict equity market trends. Ruiz (2012) studies the correlation between micro-blogging activity and equity market events, defined as changes in the price and traded volume of equities [82].

Chan (2003) examines monthly returns following public news, and finds strong drift after bad news. He also observes that investors seem to react slowly to the information and that reversal after extreme price movements are unaccompanied by public news, supporting theories of investor over and under reaction [18]. Ryan (2004) focuses on the influence of firm-specific news releases on both price reaction and trading volume activity. They find that no less than 65% of significant price changes and trading volume movements in our sample can be readily explained by public domain information [83].

Investors' slow reactions to news and seemingly irrational behaviors in reacting to the news (under and over reaction) serve as the motivation to our work. While the literature often focuses on finding the correlation between news and price movements, we explore the opportunities for excess returns in the amount of sentiment that hasn't been already reflected in the price of the equity. To our best of knowledge there is no



previous work on this.

A downside of these works is that they learn the relationship between news and daily returns directly, and often, the information in the news would have already been incorporated into the pricing of the equity. Further, most of these works attempt to model the trends of the time series, and rarely support their claims with an out of sample portfolio construction. In our work, we try to quantify the excess sentiment in the news releases, and demonstrate its effectiveness by building portfolios using the proposed method.

## 4.3 Method

In this section, we first formally define our task and describe the evaluation methods. Thereafter, we describe our approach to modeling the excess return from news data. We then present a method to model the variance of excess returns that is attributable to news, and describe its use in building a portfolio.

### 4.3.1 Problem Formulation

Not all investors have the ability to trade on news releases immediately. Furthermore, most of the earnings announcements occur after the closing of the market. Therefore, we study the relationship between the intraday return (between the opening and closing of the market) and the news releases before the opening of the market. We note that lately increasing volume of off-market trading happens after the closing of the market. Thus the information in the news releases is absorbed into the pricing of the equities sooner than ever.

We formulate the learning problem as for each day  $T$  predicting which equities in a set  $A$  are most likely to experience a large return (called an event) on day  $T + 1$ . For each day, we define returns outside one standard deviation as events. We use the returns and sentiment data until day  $T$  to learn our model.

We evaluate our method in two ways. First, we evaluate its accuracy in producing a list of equities ordered by their likelihoods of having the corresponding events

(positive and negative), given the sentiment scores. We compare this list against the true occurrences of events using mean average precision (MAP) scores. Second, we present and demonstrate the potential utility of a method that constructs portfolios using the learned relationships. The performance of portfolios constructed using our methods are compared to the performance of portfolios constructed without news information and with portfolios constructed with methods using news information directly without accounting for the amount of sentiment that has already been reflected in the prices. For comparison we also provide benchmark portfolios: a equal-weighted portfolio (EW), and the S&P 500 index (SPX).

### 4.3.2 Modeling Excess Return

We use a factor model to describe the return of each equity in terms of the equity’s active return, market sensitivity, and the sentiment scores of the news releases for each of the equities. We then train a regression model on the historical data using regularized least squares and estimate the parameters using gradient descent. This method focuses on the excess return and the amount of sentiment that hasn’t been already reflected in the price of the equity.

We use the Thomson Reuters news analytics database that contains the news feeds and sentiments computed by Thomson Reuters’ language recognition algorithm. The database contains a series of news feeds for each equity with the time of origin. We use the following fields from the feed: equity identifier, date, positive sentiment score, negative sentiment score, and relevance score that measures how relevant the news feed is to the referred equity.

The dataset contain various types of news feeds: alerts, first-take articles, follow-ups, appends, and overwrites. We only look at the feeds that are alerts or first-take articles. For each equity, we only consider the feeds with relevance  $\geq 75\%$ . For equity  $j$  on day  $T$ , the positive sentiment score  $s_{T,j}^p$  is defined as the number of feeds with positive sentiment  $\geq 75\%$  from 16:00 EST (closing time of the US markets) on day  $T - 1$  to 09:30 EST on day  $T$ . Similarly, we compute the negative sentiment score  $s_{T,j}^n$ .

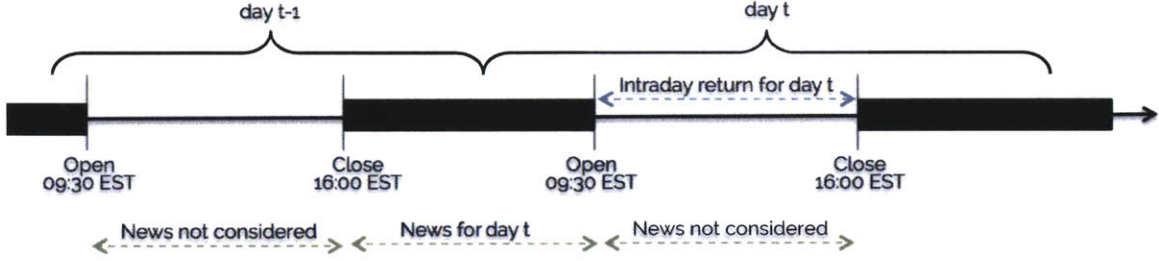


Figure 4-1: Timeline of return calculation. Intraday return of day  $t$  is matched with the news sentiments between 4:00pm EST on day  $t - 1$  to 09:30am EST on day  $t$ .

If the opening and closing prices of equity  $j$  on day  $T$  are  $p_{T,j}^o$  and  $p_{T,j}^c$ , the intraday return for equity  $j$  on day  $T$  is given by  $r_{T,j} = (p_{T,j}^c - p_{T,j}^o)/p_{T,j}^o$ . On day  $T + 1$ , we are given historical returns for  $m$  equities in a  $T \times m$  matrix  $\mathbf{R} = \{r_{t,j}\}; 1 \leq t \leq T, 1 \leq j \leq m$ . Figure 4-1 illustrates the timeline of return calculation. We are also given sentiment data  $\mathbf{S}^p = \{s_{t,j}^p\}, \mathbf{S}^n = \{s_{t,j}^n\}; 1 \leq t \leq T, 1 \leq j \leq m$ . We use index  $t$  for days, and  $j, k$  for equities.

First, we apply a GARCH model [13] to remove any serial dependencies. Therefore, we safely assume that the modeling errors are uncorrelated.

Factor model representations of returns are common in finance and econometrics [57, 49]. We model the return of equity  $k$  on day  $t$  by

$$\hat{r}_{t,k} = a_k + b_k r_{t,\Lambda} \quad (4.1)$$

By adding sentiment scores to this we obtain,

$$\hat{r}_{t,k} = \underbrace{a_k + b_k r_{t,\Lambda}}_{d_{t,k}} + w_k^p s_{t,k}^p + w_k^n s_{t,k}^n \quad (4.2)$$

In this model, we explicitly learn the factors for equity  $k$ : active return (alpha)  $a_k$ , sensitivity to the market  $b_k$ , and sensitivity to positive and negative sentiment scores of the news releases for the equity  $w_k^p$  and  $w_k^n$ . The S&P 500 index return ( $r_{t,\Lambda}$ ) is the average return of all equities on a given day.

We are interested in the influence of news flows on future returns, especially on

large returns. In the data, we observe that often equity prices tend to “anticipate” future news flows, and this pattern is more prominent for negative news. Furthermore, news articles frequently refer to the large returns of the immediate past. In order to address this, we model the sentiment scores based on the past returns (last 5 days, i.e,  $t - 5 : t - 1$ ) and past news flows, and then use the excess sentiment in predicting the future returns.

We model the expected positive sentiment by

$$\hat{s}_{t,k}^p = u_{1,k}^p + u_{2,k}^p \bar{s}_{t5,k}^p + u_{3,k}^p \bar{s}_{t20,k}^p + \sum_{\rho=-5:-1} u_{\rho,k}^p (r_{t+\rho,k}^f - d_{t+\rho,k}^f)$$

Here,  $\bar{s}_{t5,k}^p$  and  $\bar{s}_{t20,k}^p$  are mean positive sentiment scores in the past 5 and 20 days for equity  $k$ .  $r_{t+\rho,k}^f = (p_{T+1,j}^o - p_{T,j}^o)/p_{T,j}^o$  is the daily return between the opening prices, and  $d_{t+\rho,k}^f$  is the corresponding explained return. In this model, we learn the weights  $u_{1,k}^p$ ,  $u_{2,k}^p$ ,  $u_{3,k}^p$ , and  $u_{\rho,k}^p$ . We model the expected negative sentiment score  $\hat{s}_{t,k}^n$  similarly.

We update Equation 4.2 with the excess sentiment scores,

$$\hat{r}_{t,k} = a_k + b_k r_{t,\Lambda} + w_k^n (s_{t,k}^n - \hat{s}_{t,k}^n) + w_k^p (s_{t,k}^p - \hat{s}_{t,k}^p) \quad (4.3)$$

Estimating the weights directly on the observed data is prone to overfitting. Therefore, we learn the parameters by solving the following regularized least squares problem:

$$\min_{a^*, b^*, w^*, u^*} \sum_{\substack{t=1:T \\ k=1:m}} (r_{t,k} - \hat{r}_{t,k})^2 + \lambda (a_k^2 + b_k^2 + |w|^2 + |u|^2) \quad (4.4)$$

The regularization parameter  $\lambda$  is determined by cross validation. We use gradient descent to minimize the regularized square errors. For each  $1 \leq t \leq T$ ,  $1 \leq k \leq m$ ,  $x \in \{p, n\}$ , and  $-5 \leq \rho \leq -1$ , we update the parameters by:

- $a_k \leftarrow a_k + \eta(e_{t,k} - \lambda_1 \cdot a_k)$
- $b_k \leftarrow b_k + \eta(e_{t,k} \cdot r_{t,\Lambda} - \lambda_1 \cdot b_k)$
- if  $s_{t,k}^x > 0$

$$\begin{aligned}
& - w_k^x \leftarrow w_k^x + \eta(e_{t,k}(s_{t,k}^x - \hat{s}_{t,k}^x) - \lambda_2 \cdot w_k^x) \\
& - u_{1,k}^x \leftarrow u_{1,k}^x - \eta(e_{t,k} \cdot w_k^x + \lambda_3 \cdot u_{1,k}^x) \\
& - u_{2,k}^x \leftarrow u_{2,k}^x - \eta(e_{t,k} \cdot \bar{s}_{t_5,k}^x \cdot w_k^x + \lambda_3 \cdot u_{2,k}^x) \\
& - u_{3,k}^x \leftarrow u_{3,k}^x - \eta(e_{t,k} \cdot \bar{s}_{t_{20},k}^x \cdot w_k^x + \lambda_3 \cdot u_{3,k}^x) \\
& - u_{\rho,k}^x \leftarrow u_{\rho,k}^x - \eta(e_{t,k} \cdot (r_{t+\rho,k}^f - d_{t+\rho,k}^f) \cdot w_k^x + \lambda_3 \cdot u_{\rho,k}^x)
\end{aligned}$$

Here,  $\eta$  is the learning rate, and  $e_{t,k} \stackrel{def}{=} (r_{t,k} - \hat{r}_{t,k})$ . We use the last 400 days in the historical data to train our model. We iterate until convergence for the initial estimate of the parameters. The model is updated daily to make predictions for the next day. Since each day, the new training set differs from the previous day's by only two days, convergence is achieved within a few iterations.

Finally, using this learned relationship, on day  $T$  we predict the excess return  $\tilde{r}_{T+1,k}$  for day  $T+1$  by,

$$\tilde{r}_{T+1,k} = a_k + w_k^n (s_{T+1,k}^n - \hat{s}_{T+1,k}^n) + w_k^p (s_{T+1,k}^p - \hat{s}_{T+1,k}^p) \quad (4.5)$$

### 4.3.3 Variance Decomposition

Variance measures the variability (volatility) from an average or mean, and volatility is a measure of risk. In finance, the variance statistic is used to determine the risk an investor might take on when purchasing a specific equity [40].

We can express the return on day  $t$  by  $r_t = \sum_k X_{k,t} f_{k,t} + u_t$ , where  $f_{k,t}$  are factor returns (e.g., market return, sentiment scores) on day  $t$ ,  $X_{k,t}$  are exposures to the factors, and  $u_t$  is the idiosyncratic return that cannot be explained by the factors. Then, if we assume that the idiosyncratic returns are not correlated with factor returns, the covariance matrix  $\mathbf{V}$  is given by

$$\mathbf{V} = \mathbf{X}\mathbf{F}\mathbf{X}^T + \mathbf{\Delta} \quad (4.6)$$

Here,  $\mathbf{F}$  is the covariance matrix of factor returns and  $\mathbf{\Delta}$  is a matrix of residual variances. If the portfolio holding (weights for each equity) is  $\omega$ , the variance of the

portfolio (total risk)  $\sigma_p^2$  is given by [40],

$$\sigma_p^2 = \underbrace{\omega \mathbf{X} \mathbf{F} \mathbf{X}^T \omega^T}_{\sigma_f^2} + \underbrace{\omega \Delta \omega^T}_{\sigma_r^2} \quad (4.7)$$

Here, total risk  $\sigma_p^2$  is expressed in terms of factor risk  $\sigma_f^2$  and residual risk  $\sigma_r^2$ , where factor risk incorporates both market risk and news risk. For equity  $k$ , when we substitute excess return  $\tilde{r}_{k,t}$  in place of return  $r_{k,t}$ , factor risk will be made of news risk alone.

When the expected excess return  $\mathbf{r} = \{\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_m\}$  and expected covariance  $\mathbf{C}$  are given, for desired return  $r_e$ , the optimal portfolio weights are derived by the variance (risk) minimization [62]:

$$\begin{aligned} \min_{\omega} \quad & \frac{1}{2} \omega^T \mathbf{C} \omega \\ \text{subject to} \quad & \sum_{k=1}^{10} \tilde{r}_k \omega_k \geq r_e \\ & \sum_{k=1}^{10} \omega_k = 1; 0 \leq \omega_k \leq 1, k = 1, \dots, 10 \end{aligned} \quad (4.8)$$

Here,  $\omega$  are the portfolio weights.

We substitute covariance  $\mathbf{C}$  with factor covariance  $\mathbf{X} \mathbf{F} \mathbf{X}^T$  from Equation 4.7, thus obtaining an optimization minimizing the news risk,  $\sigma_f^2$ .

## 4.4 Experiments

To fully characterize our method, we evaluate our method in two ways. First, we evaluate its accuracy in producing a list of equities ordered by their likelihoods of having corresponding events (when predicting the winners the positive events), given the sentiment scores. Second, we present and demonstrate the potential real world utility of a method that constructs portfolios using the learned relationships. The performance of portfolios constructed using our methods are compared to the performance of portfolios constructed without news information, and to portfolios constructed

with methods using news information directly without accounting for the amount of sentiment that has already been reflected in the prices.

#### 4.4.1 Data

We use the equity price data used in Chapter 3. The set of companies are from nine sectors: consumer discretionary, consumer staples, energy, financials, health care, industrials, information technology, materials, and utilities. We exclude the telecommunications sector because it has less than 5 equities that were in it from 2003 to 2012.

We obtain the sentiment scores from Thomson Reuters news analytics database<sup>1</sup> that contains the news feeds as well as sentiments computed by a proprietary Thomson Reuters' language recognition algorithm. The database contains a series of news feeds for each equity with the time of origin. We use the following fields from the feed: equity identifier, date, positive sentiment score, negative sentiment score, and a relevance score that measures how relevant the news feed is to the referred equity. The dataset contain various types of news feeds: alerts, first-take articles, follow-ups, appends, and overwrites. We only look at the feeds that are alerts or first-take articles.

#### 4.4.2 Methods

The methods *News* and *News+* correspond to Equation 4.2 and Equation 4.3 respectively. The baseline method *Alpha* orders the equities based on their active returns  $a_k$  in Equation 4.1. All three methods are trained using the regularized regression formulation of Equation 4.4.

---

<sup>1</sup>[http://thomsonreuters.com/products/financial-risk/01\\_255/News\\_Analytics\\_-\\_Product\\_Brochure-\\_Oct\\_2010\\_1\\_.pdf](http://thomsonreuters.com/products/financial-risk/01_255/News_Analytics_-_Product_Brochure-_Oct_2010_1_.pdf)

### 4.4.3 Top-k Ranking

Given the data for days 1 to  $T$ , we predict which equities will have positive and negative events (returns outside 1 standard deviation in each direction) on day  $T + 1$ . We produce an ordered list of equities, ranked by their likelihoods of having events on day  $T + 1$  based on their predicted returns  $\hat{r}_{T+1}$ .

#### Evaluation

As in Chapter 3, we use mean average precision (MAP) to evaluate the correctness of the ordered list. We produce a top-5 list, and evaluate with  $MAP(5)$ .

#### Experimental Results

We ran our experiments from 2003 to 2012, so that at the start of the experiment we have at least a year of historical data to train on. For each day, we trained the model using the data from previous 400 days (excluding that day), and predict the expected return for that day. Then, we produced an ordered list of equities, ranked by their likelihoods of having events on that day, and evaluated it using average precision (AP). We repeated this experiment for each day to obtain mean average precision (MAP). Thus, we evaluated our method for one-step ahead predictions on out-of-sample data while preserving the temporal causality. Such rolling-origin evaluations are commonly used to evaluate time series predictions [35, 23].

Table 4.1 compares the  $MAP(5)$  scores (higher is better) for our method with the scores of a baseline method. The best result for each sector is in bold face. To evaluate the statistical significance of the improvement of one method over the other we perform a paired sign test<sup>2</sup>. Except for the values highlighted (\*), the p-value of the tests for each of the methods against the baseline method is significant (i.e.,  $\leq 0.001$ ). The p-value of the test between *News+* and *News* is significant for all values.

---

<sup>2</sup>Sign test is a non-parametric test which makes very few assumptions about the nature of the distributions under test. It is used to test the hypothesis that the difference median is zero between the continuous distributions of two random variables, in the situation where paired samples are drawn from these random variables.



Table 4.1: MAP scores across sectors for events based on returns.

SECTORS	POSITIVE EVENT			NEGATIVE EVENT		
	ALPHA	NEWS	NEWS+	ALPHA	NEWS	NEWS+
CONSUMER DISCRETIONARY	0.1	0.11	<b>0.2</b>	0.12	0.13	<b>0.22</b>
CONSUMER STAPLES	0.12	0.12*	<b>0.2</b>	0.14	0.14*	<b>0.22</b>
ENERGY	0.18	0.18*	<b>0.27</b>	0.2	0.21	<b>0.29</b>
FINANCIALS	0.12	0.12*	<b>0.25</b>	0.14	0.14	<b>0.26</b>
HEALTH CARE	0.1	0.12	<b>0.21</b>	0.13	0.13	<b>0.21</b>
INDUSTRIALS	0.11	0.12	<b>0.22</b>	0.13	0.13*	<b>0.23</b>
INFORMATION TECHNOLOGY	0.12	0.14	<b>0.22</b>	0.13	0.14	<b>0.24</b>
MATERIALS	0.17	0.19	<b>0.27</b>	0.18	0.19	<b>0.28</b>
UTILITIES	0.18	0.19	<b>0.26</b>	0.21	0.21	<b>0.29</b>
MARKET	0.08	0.09	<b>0.09</b>	0.09	0.1	<b>0.1</b>

While it is unsurprising that the sentiment scores computed from the news feeds allow us to predict the equities having positive and negative events (winners and losers) more accurately, the results show that the improvement achieved by the way we discount the past in predicting the future is significant. In each sector, *News+* outperforms *News* by more than 80% in MAP score.

In Figure 4-2, we compare  $MAP(k)$  scores for positive events at different cutoffs  $k$ . As the cutoff increases, the MAP scores of these methods converge with that of the baseline method. *News+* and *News* use sentiment data to capture large returns better. They perform less well on finding the patterns on smaller returns. We attribute this to the fact that in general many factors influence the changes in equity prices and for smaller returns noise dominates the model we learn.

In Figure 4-3, we compare  $MAP(k)$  scores for positive events at different cutoffs  $k$  and different event levels. For positive and negative events at higher z-score levels (i.e., returns outside 3 standard deviation) this cutoff is hardly relevant. MAP score measures both the ability to identify the likely candidate, and the rank of the candidate in the ordered list. Because these events are rare, in this case it is the ability to identify the likely candidates, and not the ordering within the selected group that matters.

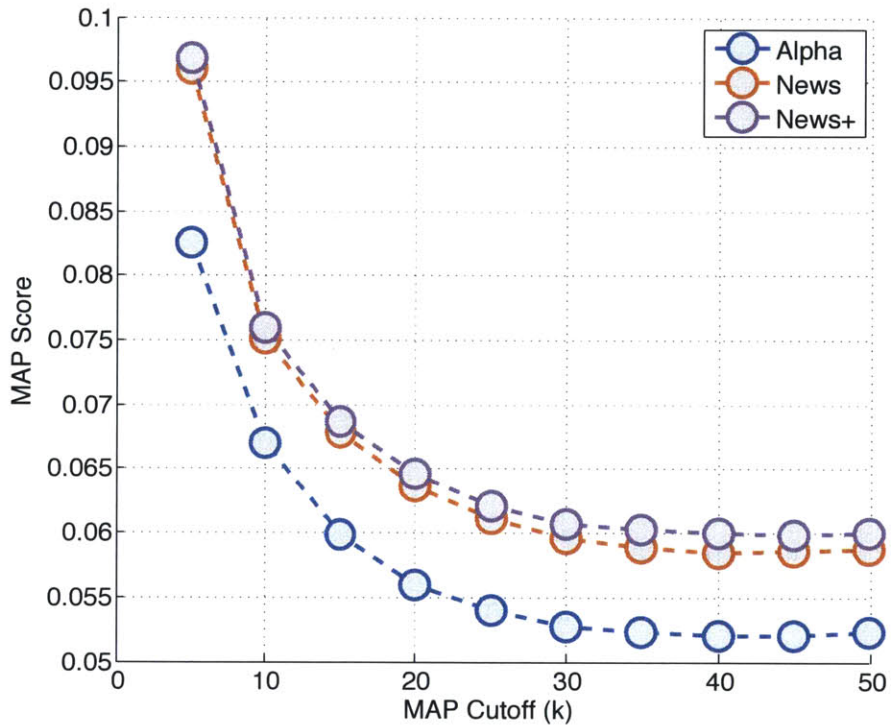


Figure 4-2: Market wide MAP scores for positive events at different cutoffs (k) for MAP(k).

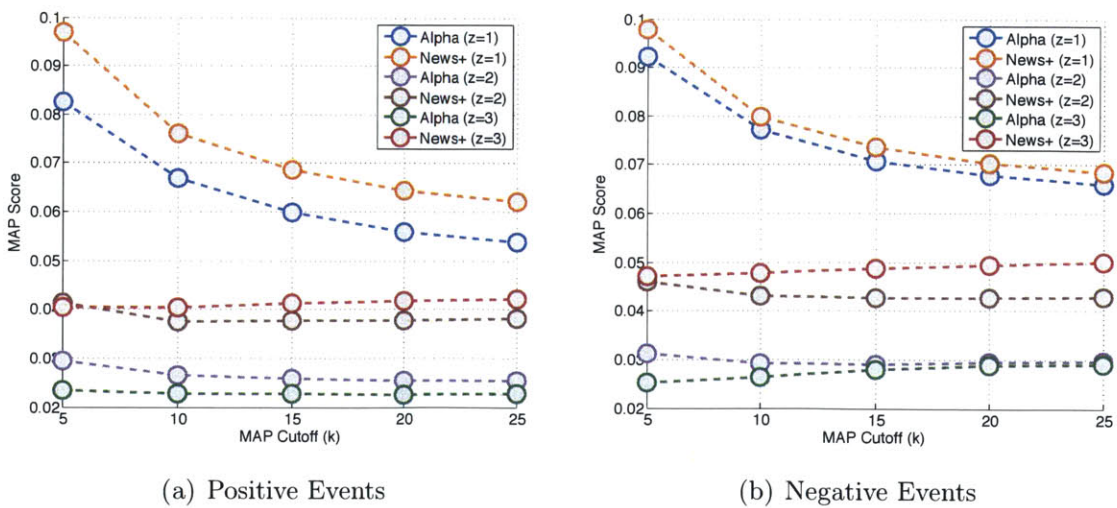


Figure 4-3: Market wide MAP scores for positive events(a) and negative events (b) at different cutoffs (k) and events levels (z).

## Explained Variance

For a regression model  $Y \sim f(X)$ , the coefficient of determination  $R^2$  describes the proportion of variance in  $Y$  explained by  $X$ . The maximum possible improvement in  $R^2$  that is contributed by the independent variable  $j$  is given by,

$$\widehat{R}^2 = \frac{R^2 - R_{-j}^2}{1 - R_{-j}^2} \tag{4.9}$$

Here  $R_{-j}^2$  is the coefficient of determination of the regression omitting  $X_j$ .

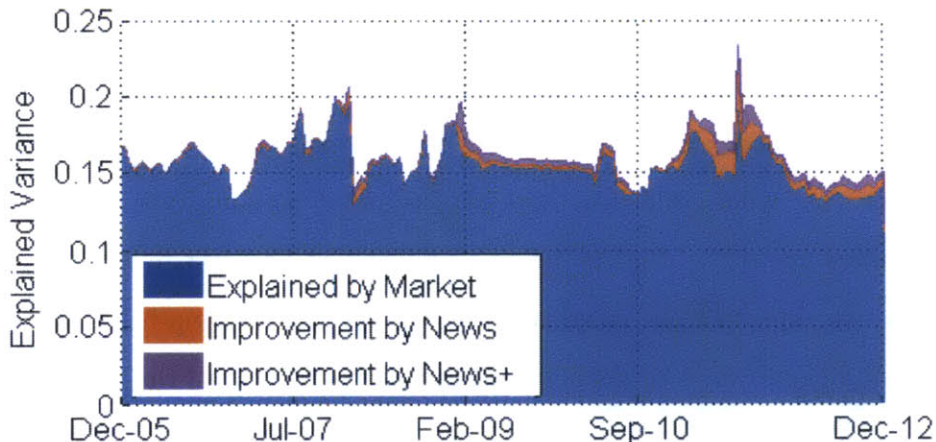


Figure 4-4: Improvement in  $R^2$  for BAC

As an example, we look at the models learned for an S&P 500 constituent: Bank of America (BAC). Figure 4-4 shows  $R^2$ , the variance explained by the market return (Equation 4.1),  $\widehat{R}^2$ , the improvement in coefficient of determination achieved by including sentiment scores (Equation 4.2), and additional improvement achieved by discounting the sentiment that has already been reflected in the price (Equation 4.3). We note that the variance explained by sentiment scores (*News*) varies over time, and peaks during high news activity periods for BAC such as the end of 2008 and 2011. On the other hand, variance explained by excess sentiment (*News+*) becomes prominent after large shifts in explained variance and following the peaks of variance explained by sentiment scores.

#### 4.4.4 Portfolio Construction

The major application of our method is in exploring the opportunities for excess returns in the amount of sentiment that hasn't been already reflected in the price of the equity. Classic portfolio theory argues that maximum expected return is achieved by owning the equity with the highest historical return. Minimum risk relative to a market is achieved by owning everything in that market [93]. As in Chapter 3, we build portfolios containing the top-10 equities across the market.

First, we build a long only portfolio (LONG) containing the top-5 equities ranked by their predicted returns. Equities are weighted equally within each portfolio. Here, we assume fixed capital (no leverage), no short positions, portfolio weights are readjusted daily, and we ignore the costs of transactions. For comparison we also provide benchmark portfolios: an equal-weighted portfolio (EW) constructed using the same intraday returns, where the portfolio is rebalanced to equal weights daily, and the S&P 500 index (SPX).

Second, we find optimal weightings for these equities (top-5 based on predicted returns) using Equation 4.8. We set  $r_e$  to the 50<sup>th</sup> percentile of the expected returns of the equities. We use the factor covariance (Equation 4.7) estimated from the historical data (last 400 days) in place of covariance in Equation 4.8. Thus, we obtain portfolio weights, and build a weighted portfolio (LONG-Weighted) with top-5 equities based on their predicted returns.

We run our experiments from 2003 to 2012. For each day, we train the model using the data from previous 400 days (excluding that day), and predict the expected return for that day. Then we update the portfolio weights for that day. We re-weight the portfolio daily, and estimate the returns daily. Thus, as before, we evaluate our method for one-step ahead predictions on out-of-sample data while preserving the temporal causality.

As in Chapter 3, we use the following criteria to evaluate the performance of the portfolios.

- Largest loss in a day: Given by  $\min(r_t)$ .

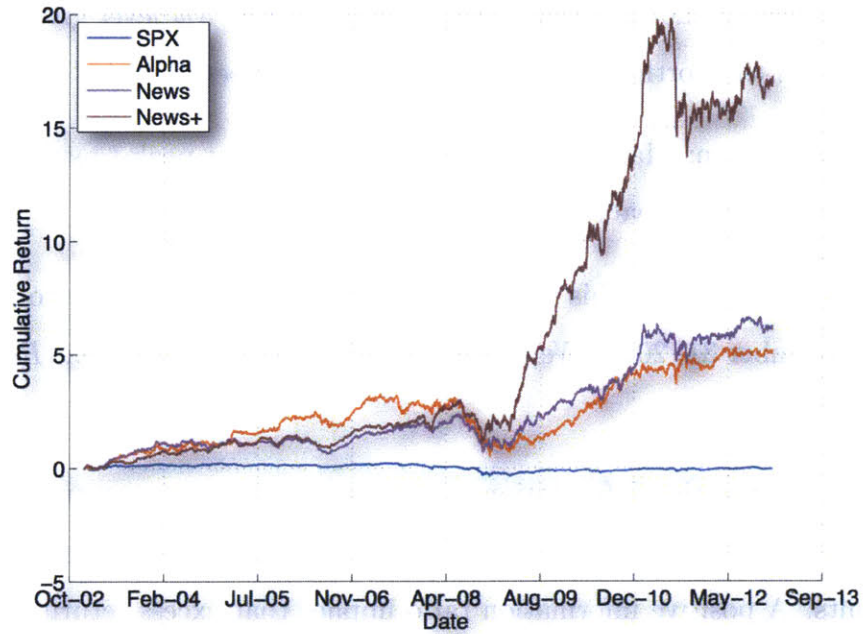
- The expected shortfall (also known as CVaR) at 5% level gives the expected return on the portfolio in the worst 5% of the cases [81].
- Max drawdown: Maximum drawdown is the largest peak-to-subsequent-trough decline in cumulative return.
- Annualized return: This is the overall return from year 2001 to 2012, i.e.,  $R_T$  on December 31, 2012. We quote annualized total return (i.e.,  $(R_T + 1)^{1/11} - 1$ ) in our results.
- Information ratio: Information ratio measures the excess return for additional risks taken [85]. We quote annualized information ratio (i.e.,  $I\sqrt{252}$ ) in our results. A positive information ratio implies that excess return is greater than the additional risk taken.
- Sharpe ratio: The Sharpe ratio measures the risk-adjusted return [85]. We quote annualized Sharpe ratio (i.e.,  $S\sqrt{252}$ ) in our results.

Table 4.2 shows that by exploiting the information in the sentiment scores computed from news feeds, we can beat the market (positive Information Ratio) and other benchmark portfolios. The weighted portfolio with *News+* results in 38.13% improvement over equal-weighted for *News+* in terms of annualized returns, compared to 16.26% improvement with *News*. This demonstrates the former's ability in modeling the news risk better than the latter.

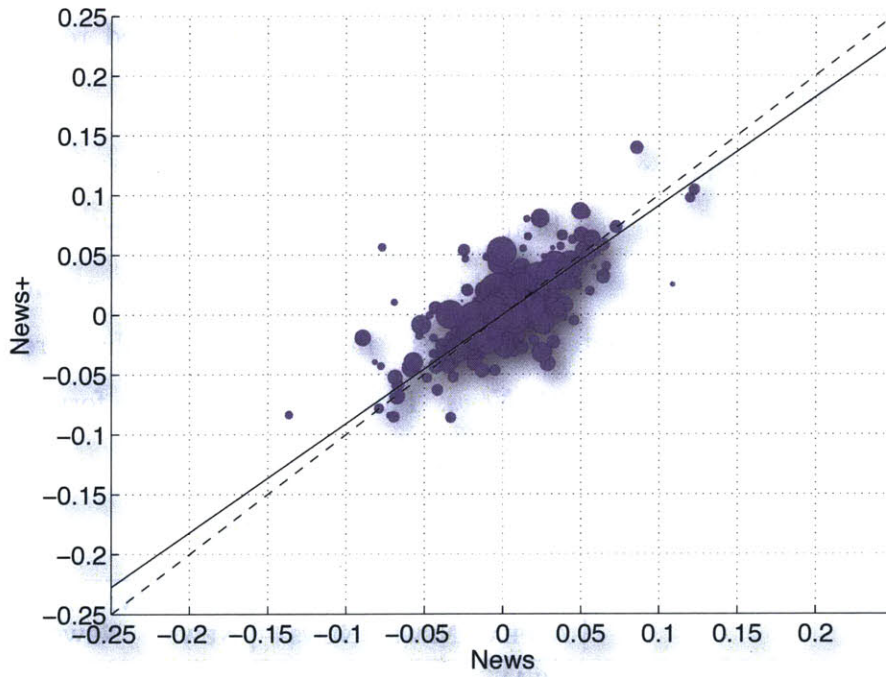
Table 4.2: Portfolio characteristics.

MEASURES	LONG			LONG-WEIGHTED		EW	SPX
	ALPHA	NEWS	NEWS+	NEWS	NEWS+		
WORST DAY	-0.11	-0.13	-0.12	-0.14	-0.09	-0.09	-0.09
EXPECTED SHORTFALL (AT 5%)	-0.04	-0.04	-0.04	-0.04	-0.03	-0.03	-0.03
MAX DRAWDOWN	-0.65	-0.63	-0.61	-0.58	-0.53	-0.63	-0.79
ANNUALIZED SHARPE RATIO	0.77	0.75	0.9	0.9	1.27	0.41	0.34
ANNUALIZED IR	1.08	1.02	1.24	1.12	1.59	1.19	N/A
ANNUALIZED RETURN (%)	20.61	18.88	24.39	21.95	33.69	6.07	5

Figure 4-5 shows the cumulative return achieved by our methods against the S&P 500 index (SPX). First, we note the significantly high returns generated by



(a) Cumulative returns



(b) Daily returns

Figure 4-5: Cumulative returns (a), and daily returns (b) of *News* and *News+* models for the *LONG-Weighted* portfolio. In figure (b), size of the points correspond to total number of news releases for that day. Returns above the dotted line correspond to an improvement with *News+* over *News*. A clockwise shift (solid line) indicates that the large returns for *News+* are moderated on both positive and negative sides. Also, there is a small (barely noticeable) shift indicating higher average returns.

our methods after 2006 (particularly after 2009), perhaps related to the increase in the coverage and utility of the sentiment scores computed by Thomson Reuters. Furthermore, starting mid 2009 *News+* outpaces *News*, probably because with the increasing number of news releases it becomes more important to account for the sentiment that has already been incorporated into the equity prices. Since 2011 *News* and *News+* underperform, perhaps indicating the vanishing opportunities to exploit the sentiments in firm specific news.

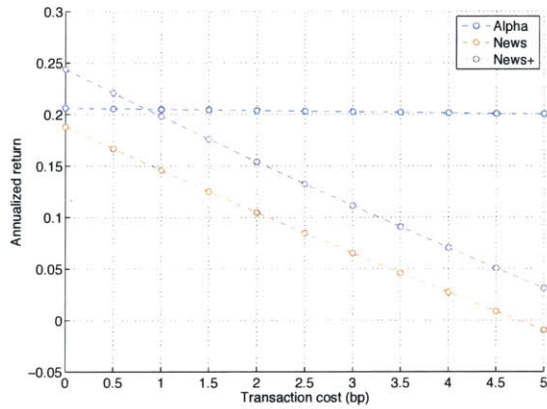
In Figure 4-5(b), we can observe the places where this differential impact is significant. Note that, the days with large numbers of news releases often produce small returns as indicated by the accumulation of large circles around the center. These likely correspond to the earning announcements, which are often expected and already priced into the equities by the market. It is on those days where a small number of perhaps unexpected news releases occur that our methods achieve large returns.

## 4.5 Application Considerations

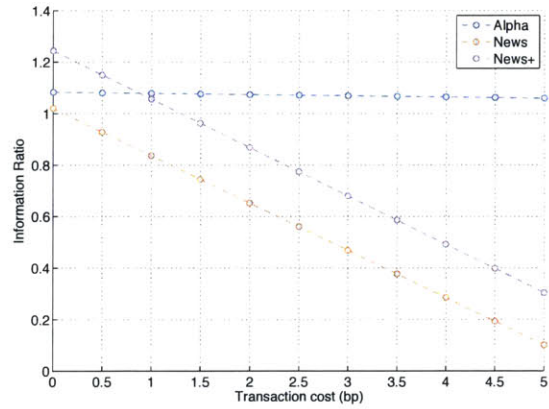
Our method rebalances the portfolio daily, and therefore incurs transaction costs. These costs are ignored here, therefore absolute returns are overstated for both methods. Figure 4-6 shows the annualized returns and information ratios for long portfolios with these methods when transaction cost is taken into account. While *News* is comparable to *Alpha* when zero-transaction cost is assumed, it immediately becomes less appealing with increasing transaction costs. *News+* becomes comparable to *Alpha* at 1 basis point (0.01%). Therefore, retail investors could hardly use our method alone.

In our methods, we pick only top-5 equities based on their predicted returns. While such a concentrated portfolio yields higher returns, it also poses higher risk. Figure 4-7 shows the information ratio achieved for different holding sizes (i.e., top- $k$  positions for different  $k$ ). We notice that at  $k = 25$  we achieve the highest information ratio across the range of transaction costs that we considered.

The average turnover of LONG-Weighted portfolio is 0.02, 0.91, and 0.90 for *Alpha*, *News*, and *News+* respectively. The settlement of the trades can take as long as three

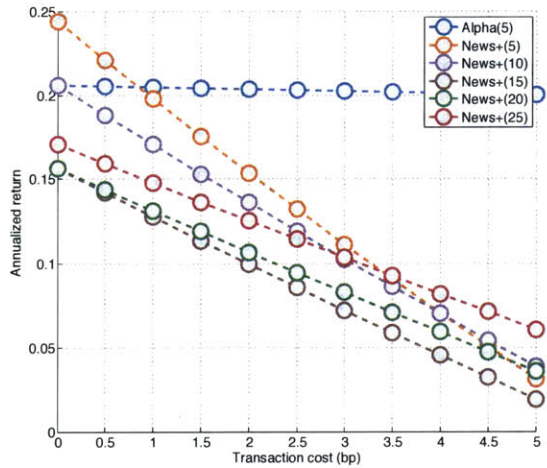


(a) Annualized returns

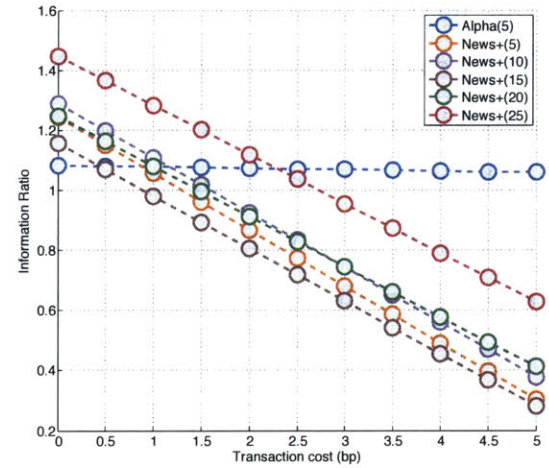


(b) Information Ratios

Figure 4-6: The annualized returns (a), and information ratios (b) for LONG portfolios built with our methods *Alpha*, *News*, and *News+* under different transaction costs measured in basis points (1/100th of a percent).



(a) Annualized returns



(b) Information Ratios

Figure 4-7: The annualized returns (a), and information ratios (b) for different holding sizes of LONG portfolios with *News+* compared against the long only portfolio with *Alpha* of size 5.



days, and a retail investor would need a margin account to support turnovers as high as 0.8. Hence, the investor would also incur corresponding borrowing (margin) fees.

In financial time series analysis, patterns can be learned in four time scales: long-term (fundamental based), mid-term (technical and macro-based), short-term (event based), and ultra-short-term (based on inefficiencies in statistical arbitraging). In this work, we capture short-term patterns. When patterns on other time scales dominate (for instance the crash in 2008, which is a mid-term macro-based pattern), our method might fail to distinguish the causes. We intend to address this in future work by building a more comprehensive model using news-based macroeconomic sentiment measures.

## 4.6 Interday returns

In our work we considered intraday returns (daily returns between the opening and closing of the market) so that we can capture the casual information in the news effectively. Here, we look at the influence of our method in predicting the interday return (daily returns between the successively closings of the market).

We define it more concretely as follows. If the closing prices of equity  $j$  on day  $T$  and day  $T + 1$  are  $p_{T,j}^c$  and  $p_{T+1,j}^c$ , the interday return for equity  $j$  on day  $T$  is given by  $r_{T,j} = (p_{T+1,j}^c - p_{T,j}^c)/p_{T,j}^c$ . On day  $T + 1$ , we are given historical returns for  $m$  equities in a  $T \times m$  matrix  $\mathbf{R} = \{r_{t,j}\}; 1 \leq t \leq T, 1 \leq j \leq m$ . Figure 4-8 illustrates the timeline of return calculation. For equity  $j$  on day  $T$ , the positive sentiment score  $s_{T,j}^p$  is defined as the number of feeds with positive sentiment  $\geq 75\%$  from 16:00 EST (closing time of the US markets) on day  $T$  to 16:00 EST on day  $T + 1$ . Similarly, we compute the negative sentiment score  $s_{T,j}^n$ .

Figure 4-9 shows the annualized return and information ratio for the LONG portfolio. Here, the portfolio is built using methods *Alpha*, *News*, and *News+* for interday returns at different lags. Negative lag (lead) is related to the news sentiments corresponding to past returns. Instantaneous reaction to news is assumed at lag 0, and the corresponding returns (although highly attractive) are thus improbable in practice.

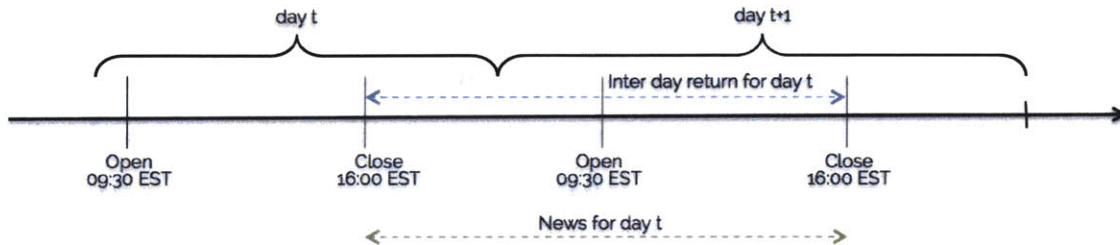


Figure 4-8: Timeline of return calculation. Interday return of day  $t$  is matched with the news sentiments between 4:00pm EST on day  $t$  to 4:00pm EST on day  $t + 1$ .

Positive lag gives the causal predictive power of news sentiments on interday returns. For next day predictions (lag 1), *News+* achieves 43% annualized return and 1.6 information ratio. We note that the predictive power of our news decays rapidly, and by the second day it is comparable to that of *Alpha*. Interestingly, (hypothetical) information ratio at  $-1$  lag is as high as 1.8, indicating prescient price reaction against the forthcoming news release.

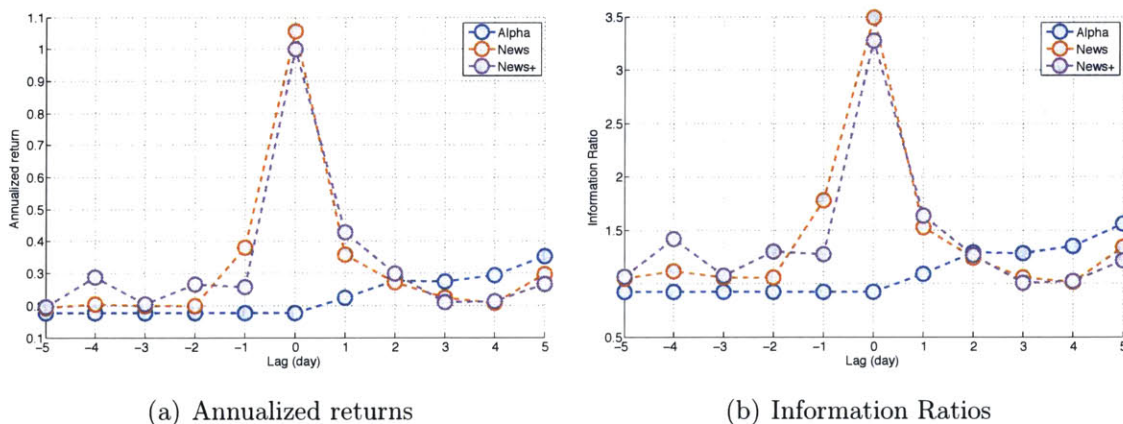


Figure 4-9: The annualized returns (a), and information ratios (b) for LONG portfolios built with methods *Alpha*, *News*, and *News+* for interday returns under different lags.

## 4.7 Summary

We presented a method for learning connections between financial time series and news releases. A problem with using news releases to predict future return is that

often the information in the news has already been incorporated into the pricing of the equity. In order to address this, we build a model of the sentiment of the news releases based on the past returns and past news flows, and then use this model to discount the expected sentiment from future news releases. Our method is a machine learning-based model that captures the relationship between the equity return that is unexplained by the market return (excess return) and the amount of sentiment that hasn't been already reflected in the price of the equity. We modeled returns using three factors: active return, market sensitivity, and connectedness of returns. We learned these factors using a recursive regression. We solved the regression problem using an unconstrained least squares optimization.

We evaluated our method in two ways. First, we evaluated its accuracy in producing a list of equities ordered by their likelihoods of having large gains and large losses, given the sentiments computed from the news releases for each of the equities. We then presented and demonstrated the potential real world utility of a method that constructs portfolios using this model. We show that while using news directly yields an annualized return of 18.88%, our proposed way of handling the past boosts the annualized return to 24.39% over a 10-year period. Furthermore, we also demonstrate that our method more accurately models the variance attributable to news (news risk), resulting in annualized return as high as 33.69% compared to 21.95% for the same portfolio construction that uses news directly.



# Chapter 5

## LEVERAGING NETWORK STRUCTURE IN FINANCIAL TIME SERIES

In Chapters 3 and 4, we presented techniques for modeling pairwise connections between equities, and demonstrated their utility in building a diversified portfolio. Building on this work, in this chapter we consider the task of leveraging network structure to model connections that span multiple equities, such as clusters and hierarchies in a network of financial time series. Specifically, we discuss how the topological properties of the network<sup>1</sup> can be used to obtain a segmentation of equities that has advantages over the traditional sector-based grouping.

### 5.1 Introduction

It is known that risk is not uniformly spread across financial markets [78]. The root cause of this pattern is that multiple equities are simultaneously affected by latent external factors (e.g., macro-economic factors such as oil price), resulting in the prices of a set of equities moving together. In order to construct a well-diversified portfolio to reduce investment risk, we group equities that are affected similarly by external factors, and diversify across groups. Sector grouping provides a static segmentation that is often used for this purpose.

---

<sup>1</sup>Throughout the chapter we use network and graph interchangeably.

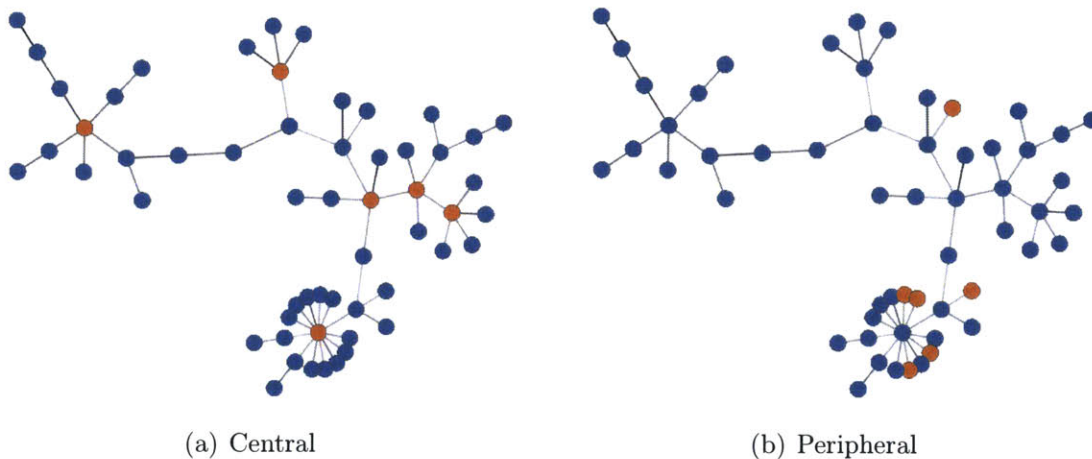


Figure 5-1: The financial filtered network built with the equities from the financial sector for years 2006-2007. 6 of the most central and 6 of the most peripheral equities as identified by their node degrees (ties are broken randomly) are highlighted in orange color.

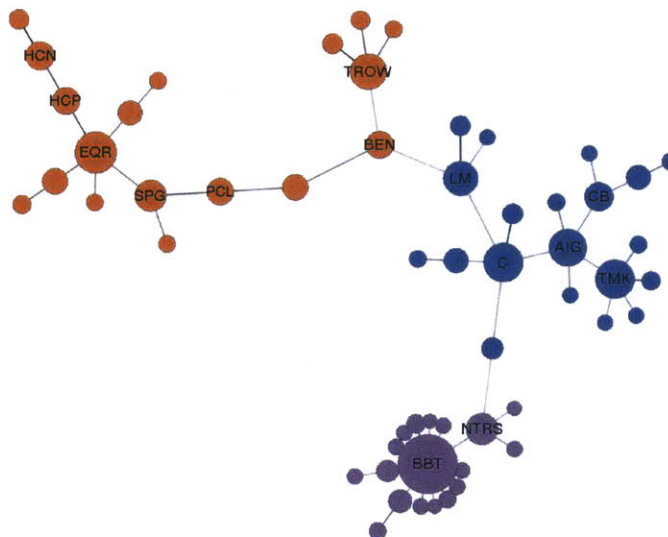


Figure 5-2: The financial filtered network built with the equities from the financial sector for years 2006-2007. Three clusters as identified by spectral clustering on this network are indicated by the node colors. Size of the node corresponds to the degree of the node.

Recently, researchers have proposed using financial filtered networks to model the heterogeneous spreading of the risk across equities [78, 75]. A financial filtered network is built by retaining the highest correlated links of a multivariate time series, where the network represents the pairwise connection (e.g., correlation) between the time series. The topology of such a network encodes the high order dependency structure (transitive connections) of the equities. The underlying idea is that equities differently positioned within this financial network exhibit different patterns of behavior, and therefore selecting equities from different regions of the network leads to a well-diversified portfolio. Compared to sector grouping, this approach is dynamic (can be updated frequently), and data driven. Furthermore, this approach can be optimized for different objective functions, for example, the equities could be clustered based on the co-occurrences of large losses.

Our goal is to use such a network structure to quantitatively distinguish sets of equities, where sets are, loosely speaking, orthogonal to each other based on the returns of the equities in a set. We focus on two major ideas. First, we explore the use of centrality and peripherality measures based on the network topology to identify a set of equities that are the most (and least) connected with the remaining equities (Figure 5-1). Second, we use graph clustering to find regions that are sparsely connected with each other (Figure 5-2).

A major idea of this work is to use the topology of a financial filtered network to encode the high order dependency structure (transitive connections) of the financial equities. Although correlation is generally used in the literature to represent the strength of the link between two nodes in a financial filtered network, it may not be the correct linkage measure. A correlation coefficient says nothing about whether other equities contribute to the observed relationship between the two equities. This is because, correlation is estimated between two time series independently of the rest of the time series in the multivariate time series. In contrast, partial correlation measures the degree of association between two time series while discounting the influence of other time series in the multivariate time series. Thus, it removes the transitive connections from the pairwise association. Therefore, in a financial filtered

network built with partial correlation, transitive connections will be encoded by the network topology (e.g.,  $A \rightarrow B \rightarrow C$ ). In contrast, in a correlation network, they will be encoded via a direct link (e.g.,  $A \rightarrow C$ ). The factor model ( $FAC_{10}$ ) presented in Section 3.4.3 is similar to partial correlation (see Section 3.3.3), but in addition, it specifically removes the market component and strongly weights the connections based on large returns. Throughout this chapter, we compare the effectiveness of these three different linkages in encoding higher order connections in financial time series.

The remainder of the chapter is organized as follows. In Section 5.2 we discuss the related work. Section 5.3 presents a list of linkage measures, and describes how to construct a financial filtered network using these measures. In Section 5.4, we discuss a list of topological measures, and describe a method to find two distinct (central and peripheral) regions in the network using these topological measures. In Section 5.5, we present a method to segment the network into dissimilar regions using a graph clustering algorithm. Then, we compare the use of these clusters for diversifying a portfolio of equities as opposed to diversifying a portfolio by holding equities from different sectors. In Section 5.6, we present graph sparsification as an alternative to constructing a filtered network. In Section 5.7, we discuss the application considerations. In Section 5.8, we provide a summary.

## 5.2 Related Work

In recent years, there have been numerous works on network-based approaches applied to several fields from biology [37], social [89, 2], and financial systems [15, 78, 75]. Several researchers have used network structures to model stock markets in which the nodes represent different equities and the links relate to correlations of returns [96, 14, 15, 78, 75].

In our work, we extend the ideas proposed in Pozzi, et al (2013), and Peralta and Zareei (2014). Pozzi, et al (2013) asserts that investments in equities that occupy peripheral, i.e., poorly connected, regions in financial filtered networks are most



successful in diversifying the risk [78]. Peralta and Zareei (2014) shows that highly central equities are more systemic, i.e., market and time dependent, due to their embeddedness into the market structure. This in turn increases the portfolio's variance. They prove that the centrality of each individual stock in that network and the weight it receives in an optimal allocation of resources (in Markowitz's mean variance optimization) are related [75]. They show that optimal portfolios assign wealth toward the periphery of the market structure. However, they also point out that the equities with higher centrality status are more stable (i.e. retain their centrality status for consecutive years) than the peripheral equities.

In our work, we show that highly central equities, even after the market dependency is excluded, are more systemic, i.e, the market movements significantly influence their returns. They offer high returns at the expense of high variance and increased risk (measured by the worst day performance, expected shortfall, and maximum drawdown). We show that well-chosen peripheral equities can offer similar risk adjusted returns but at a reduced risk. We also show that graph clustering helps us identify uncorrelated equities within central region. Our work has the following areas of applications: identifying an alternative segmentation to sector grouping, leveraging network structure to build diversified portfolios, and discovering stable dependencies for risk management.

## 5.3 Financial Filtered Networks

First, we describe a method to construct financial filtered networks. Then, we discuss the empirical observations made on the resulting networks.

### 5.3.1 Methods

The first step in constructing a financial filtered network is to compute the edge weights of the graph. Then, we apply network filtering to sparsify the graph such that the underlying structure can be exposed and studied.

## Linkage Measures

We compare the following three linkage measures.

- Correlation(CR): Standard implementation.
- Partial correlation(PCR): This captures the connection between two return time series, while controlling for the remaining time series.
- $FAC_{10}$ : This is the factor model (Equation 3.4a) with cost function  $f(x) = e^{-x/10}$  applied. We use the method presented in Section 3.4.3. This method is similar to partial correlation (see Section 3.3.3), but in addition, it specifically removes the market component and more strongly weights the connections based on large returns.

When filtering the network with a minimal spanning tree, it is common to convert the linkage measures to distances. For example, if the correlation coefficient between equities  $i$  and  $j$  is  $c_{ij}$ , we convert it to the corresponding Euclidean distance  $d_{ij} = \sqrt{2(1 - c_{ij})}$ . The resulting distance matrix  $D$  is used hereafter.

## Network Filtering

There are different methods in the literature for filtering a network to identify its underlying structure: minimum spanning trees (MST) and planar maximally filtered graphs (PMFG) are the most common. MST and PMFG are based on constructing a constrained graph that retains the largest connections (where connection is defined in terms of a linkage method) between connected nodes. The MST is a subgraph of the PMFG. Both of them are planar, i.e., they can be drawn on a surface without links crossing. For a network of  $N$  nodes, MST is a tree with  $N - 1$  links and PMFG has  $3(N - 2)$  links. For a connected planar graph, MST has the minimum number of links and PMFG has the maximum number of links [95, 78]. There are also other methods that combine linkage estimation with network filtering: partial correlation threshold network (PCTN), and partial correlation planar graph (PCPG) [95, 48]. In

our work, we only consider MST because of its simplicity, and clarity in visualizing the connections.

### 5.3.2 Empirical Observations

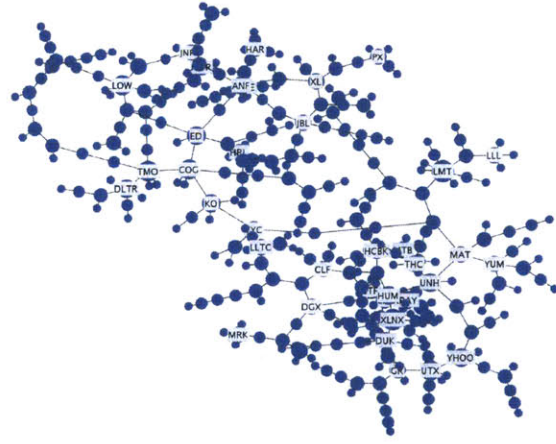
Figure 5-3 presents the minimal spanning trees generated by different linkages for the years 2001-2002. All three trees have the same number of nodes and the same number of links. We note that dense bubbles in the correlation tree are replaced with deeper branches in the partial correlation tree. We hypothesize that since correlation estimates already incorporate the transitive effects, a financial network built using correlation will be less effective in encoding the transitive effects in the topology. The  $FAC_{10}$  tree is even deeper than the partial correlation tree. This is probably because we explicitly discount the market component in  $FAC_{10}$ . In addition, because  $FAC_{10}$  focuses more on the connection between large returns, technology equities that faced large drops in 2001-2002 become the central nodes (e.g., FFIV, AKAM, YHOO, AES) in the corresponding graph.

Financial networks built with each linkage method tell different versions of the 2008 crisis (Figure 5-4).

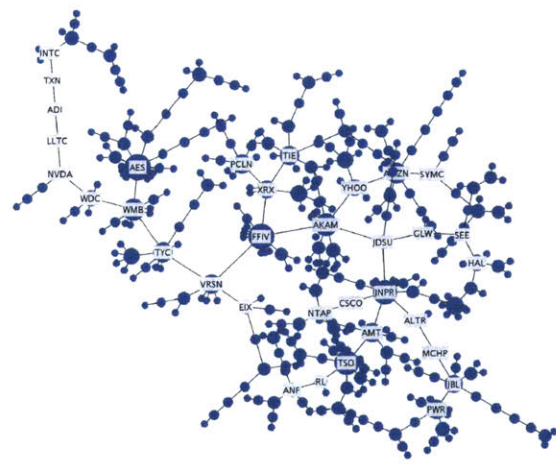
- According to the correlation network, during the bubble, companies that were directly related to mortgage and retail/commercial banking were highly correlated. For instance, the following equities became the central nodes: BBT (BB&T Corporation, a holding company of commercial bank subsidiaries), BEN (Franklin Resources), WFC (Wells Fargo), BAC (Bank of America), C (Citi Group), HIG (Hartford Financial Services group), USB (U.S. Bancorp), AXP (American Express), and GE (General Electric, a diversified manufacturing company as well as a financial services company). Market movement was based on these entities. In the aftermath of the crisis (2007-2008), the focus shifted to non-financial blue chip equities (large, well-established, and financially sound companies).
- Companies from the financial sector are largely missing in the  $FAC_{10}$  network



(a) Correlation



(b) Partial Correlation

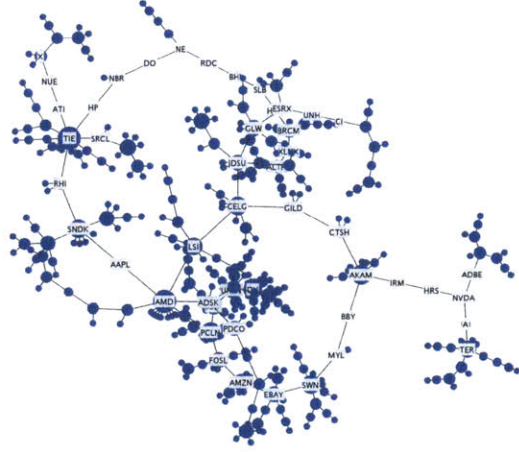


(c)  $FAC_{10}$

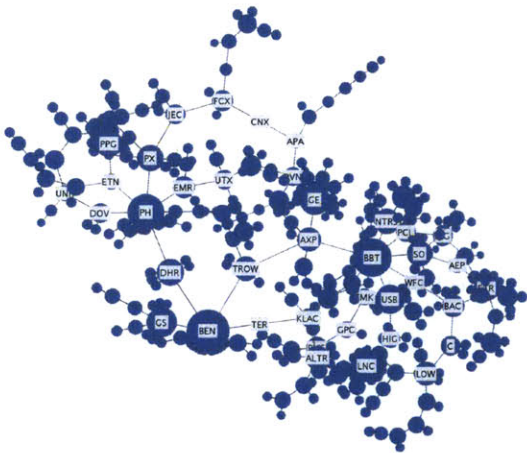
Figure 5-3: Minimal spanning tree network generated by different linkages for 2001-2002. Some nodes along the backbone of the network are labeled. Size of the node corresponds to the degree of the node. Note that the correlation tree contains the most dense regions where nodes cluster together, and  $FAC_{10}$  tree contains the deepest branches.



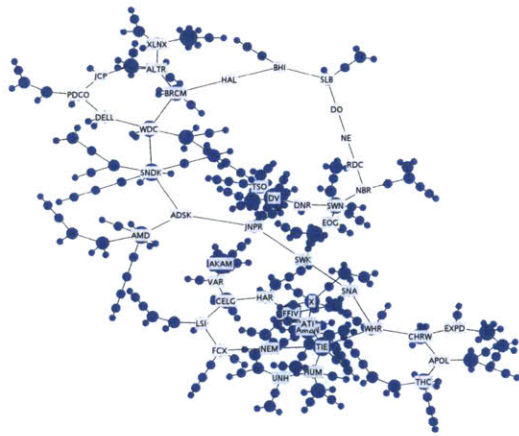
(a) Correlation (2005-2006)



(b)  $FAC_{10}$  (2005-2006)



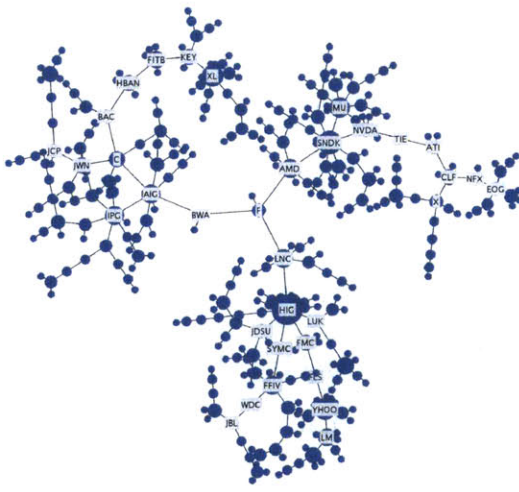
(c) Correlation (2006-2007)



(d)  $FAC_{10}$  (2006-2007)



(e) Correlation (2007-2008)



(f)  $FAC_{10}$  (2007-2008)

Figure 5-4: The network topologies of correlation and  $FAC_{10}$  during the 2008 financial crisis. Highly central nodes are labeled. Size of the node corresponds to the degree of the node.

for the years 2005-2007. This is because they were the market drivers (i.e., market dependent), and their connections weren't based on losses. In 2007-2008, several companies faced significant drops in their prices (drops several times greater than that of the S&P 500 index), most notably HIG (Hartford Financial Services Group), F (Ford), and AIG (American International Group). They take the center stage in the  $FAC_{10}$  network in 2007-2008.

Notable absentees in this list are some of the companies that are not part of our data set (because they weren't in S&P 500 from 2001 to 2012), e.g., Lehman Brothers and General Motors.

## 5.4 Separating Peripheral and Central Regions

Our goal is to use the network structure to quantitatively distinguish equities that have, loosely speaking, orthogonal returns. To this aim, we explore the use of centrality and peripherality measures based on the network topology to identify a set of equities that are the most (and least) connected with the remaining equities. Then, we analyze the characteristics of the portfolios built using these two distinct sets of equities.

### 5.4.1 Methods

We use the following topological measures to quantify the centrality and peripherality of the nodes.

- **Degree (D):** Degree of a node is the number of edges connected to that node. This helps us identify the hubs.
- **Eccentricity (E):** Eccentricity of a node is the maximum length of the shortest paths that connect the node to any other node. Here, the length of a path is the summation of link weights along the path.
- **Closeness (C):** Closeness of a node is the average length of the shortest paths that connect the node to any other node.

- **Eigenvector centrality (EC):** Eigenvector centrality measures the centrality of a node as a weighted sum of the centrality of its neighboring nodes [69]. It is based on the principle that a highly central node becomes central by being connected with many other nodes, especially with nodes that are themselves highly central ones. Google’s PageRank algorithm is a variant of Eigenvector centrality that ranks websites [70].

We consider a network  $G = \{V, A\}$ , where  $V$  is the set of nodes and  $A$  is the adjacency matrix of the graph whose diagonal entries are set to zero to prevent self-loops. Eigenvector centrality  $\Omega_i$  of node  $i$  is given by,

$$\Omega_i = \lambda^{-1} \sum_j A_{ij} \Omega_j \tag{5.1}$$

When written in matrix form, the eigenvector  $\Omega_k$  of  $A$  corresponding to the largest eigenvalue  $\lambda_k$  gives the centrality score.

$$A\Omega = \lambda\Omega \tag{5.2}$$

While these topological measures are related, each of them captures different properties of the network than others. We can notice this in Table 5.1, where we show the cross correlation of these measures across all the equities by the start of 2012 (computed from years 2010 - 2011).

Table 5.1: Correlation matrix of topological measures for year 2012

	ECCENTRICITY	CLOSENESS	DEGREE	EIGENVECTOR CENTRALITY
ECCENTRICITY	1.00	0.91	-0.21	0.4
CLOSENESS	0.91	1.00	-0.24	0.46
DEGREE	-0.21	-0.24	1.00	-0.62
EIGENVECTOR CENTRALITY	0.4	0.46	-0.62	1.00

In order to make use of the additional information offered by each of the topological measures, Matteo, et al (2010) and Pozzi, et al (2013) suggest using a combination of them. We use the following combination called periphery score  $P_i$ , where we rank

the equities by the average z-scores of the topological measures.

$$P_i = zscore(E_i) + zscore(C_i) - zscore(D_i) + zscore(EC_i) \quad (5.3)$$

### 5.4.2 Empirical Observations

Figure 5-5 presents the whisker in/out graph to give an at-a-glance summary of each stock's relative presence in the most peripheral set of stocks and the most central set of stocks for years 2003 - 2012. There is a significant difference in the selection across different measures. Since periphery score is a combination of other topological measures, its selection is more stable (i.e. selecting the same set of stocks for consecutive years). In the rest of the chapter, we consider only the periphery score.

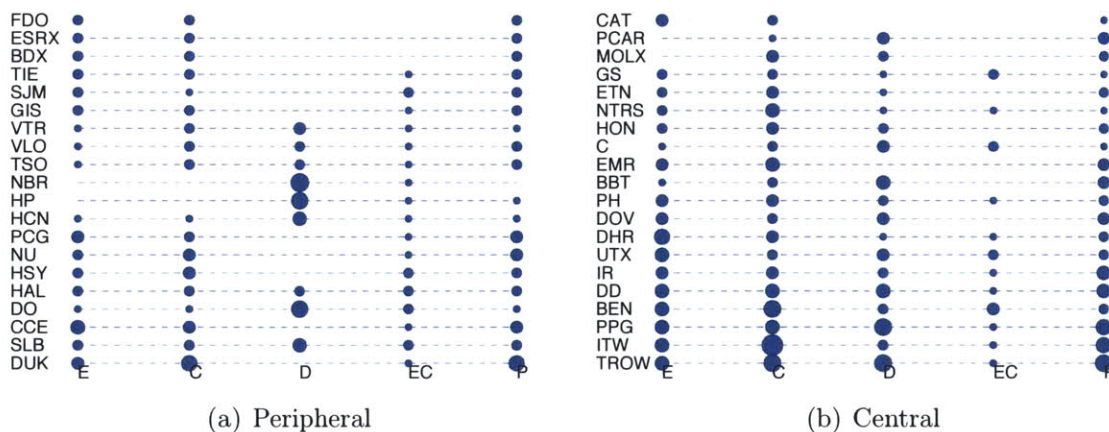


Figure 5-5: The 10 most central and the 10 most peripheral equities as identified for each year from 2003 to 2012 by different topological measures: E (Eccentricity), C (Closeness), D (Degree), EC (Eigenvector centrality), and P (Periphery score). The size of the circle denotes how many years the equity was in the top-10 list. Note that central set is more stable than the peripheral set.

Figure 5-6 presents the whisker in/out graph to give an at-a-glance summary of each stock's relative presence in the most peripheral set of stocks and the most central set of stocks as identified by periphery score with different linkage methods for years 2003 - 2012. We observe that equities in the most central set are more stable, i.e., they retain their increased central status for consecutive years, whereas peripheral equities change their positions more frequently.



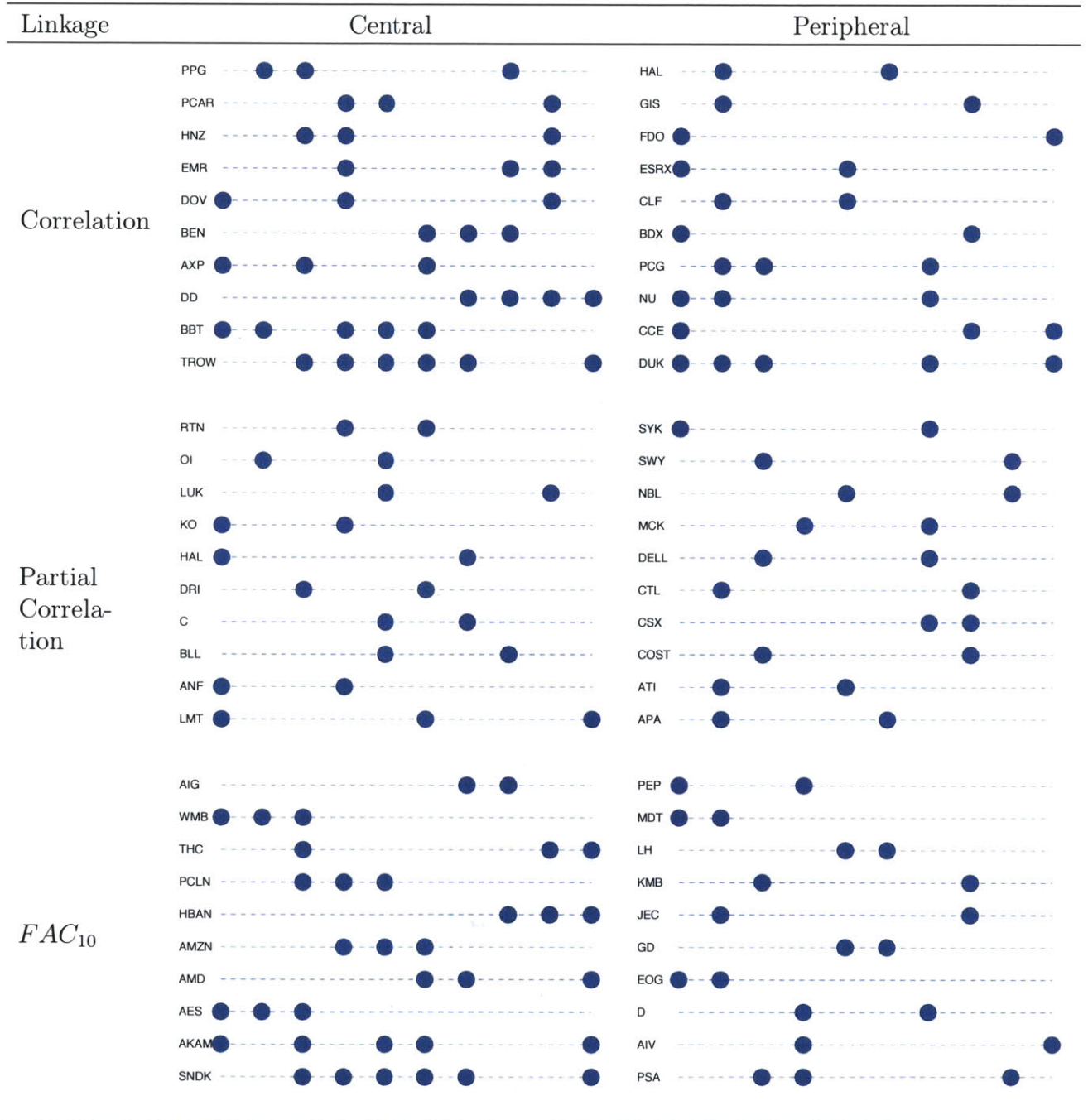
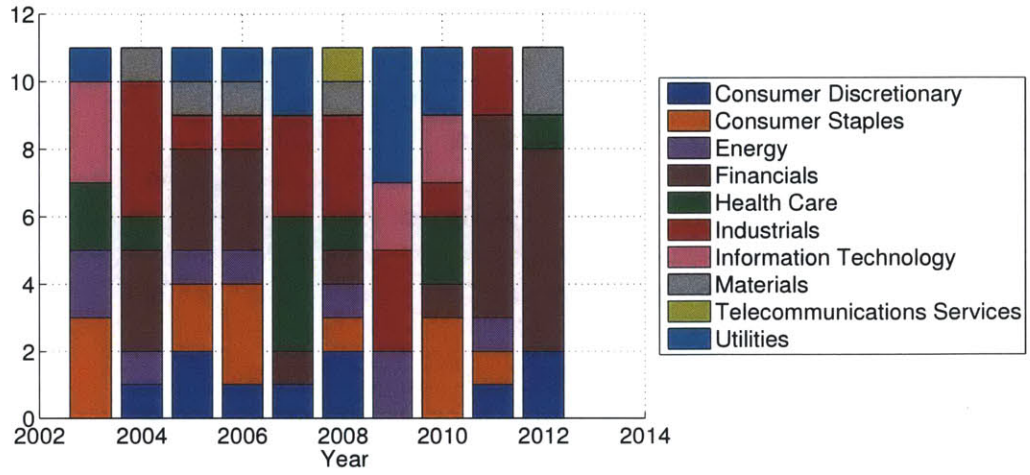
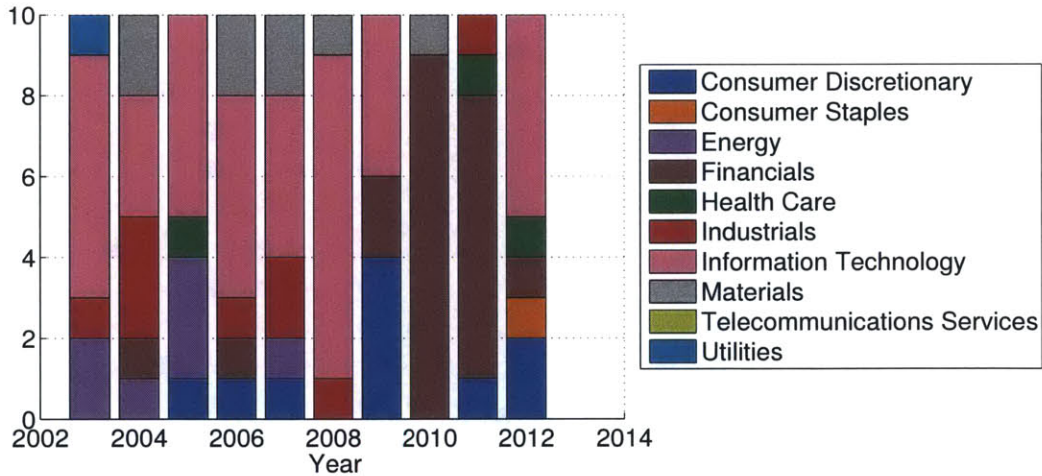


Figure 5-6: Composition of the 10 most central equities and the 10 most peripheral equities as estimated by periphery score using different methods for linkage for years 2003-2012. For each year, a filled circle marks the presence of the equity in the top-10 set.



(a) Top-10 (Peripheral)



(b) Bottom-10 (Central)

Figure 5-7: Sector composition of the 10 most central equities and the 10 most peripheral equities as estimated by periphery score using  $FAC_{10}$  for linkage. The year corresponds to the period of two years until that point that was used to generate the network.

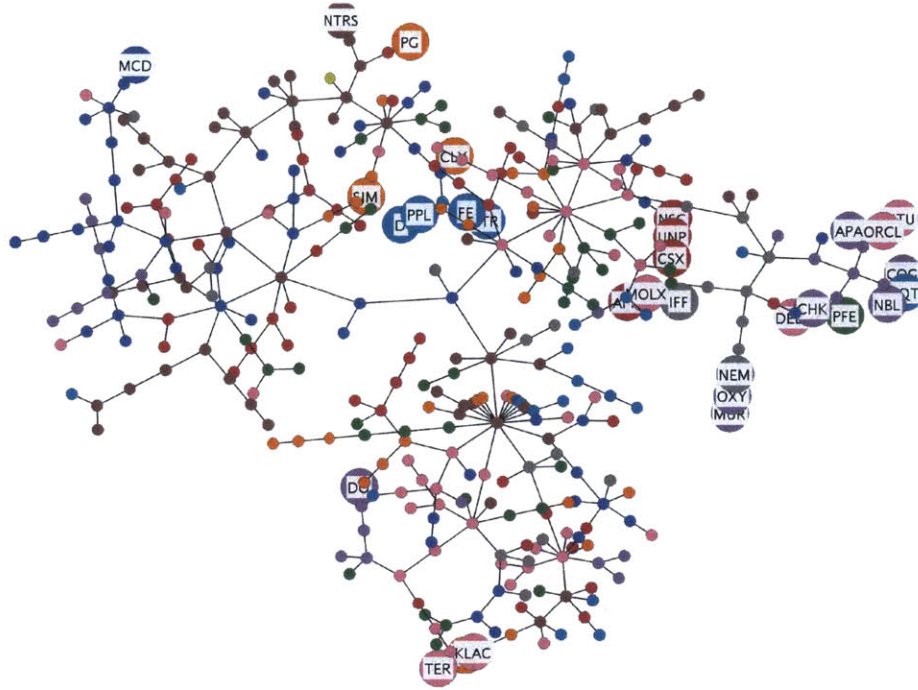
Figure 5-7 shows the sector composition of these sets as estimated by the periphery score using  $FAC_{10}$ . The information technology sector dominates the central region except for the aftermath of the 2008 crisis when the financial sector becomes dominant. In 2009-2010, the financial sector had a few big movers such as AIG (American International Group), C (Citi Group), and BAC (Bank of America).

Figure 5-8 shows the network and the set of equities identified as most central and most peripheral with  $FAC_{10}$  as the linkage metric in years 2007-2008. While the set of peripheral equities shows significant diversity in its sector composition, the set of central equities is dominated by the financial sector (in brown color), the information technology sector (in pink color), and the manufacturing sector (in blue color), such as C, AIG, LNC, F, etc. The only stock from the financial sector in the peripheral set is NTRS (Northern Trust), which had a quick rebound in its excess daily returns compared to the rest of the sector.

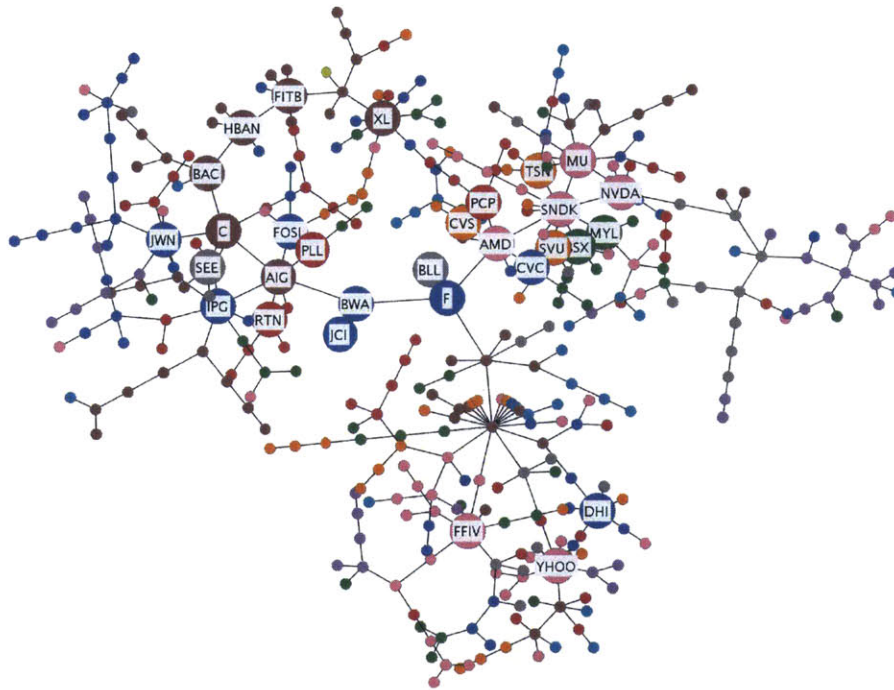
Figure 5-9 shows some of the characteristics of the equities in the 10 most central and the 10 most peripheral sets as estimated by periphery score using correlation for linkage. Central stocks seem to have high beta, and to be more market dependent. However, none of the stock characteristics offers a consistent separation of these two sets.

### 5.4.3 Portfolio Returns

In Figure 5-10, we compare the return characteristics of the 10 most central and the 10 most peripheral equities based on linkages computed by various methods together with periphery score as the topological measure for 2003 (network constructed from 2001-2002). The network topology constructed with correlation as the linkage measure results in a poor separation based on the periphery score. When  $FAC_{10}$  is used for linkage measure, we achieve visibly better separation between these two sets of equities based on their mean and standard deviation of daily returns compared to the ones with correlation and partial correlation. While top-10 equities, and their means and standard deviations of returns change yearly,  $FAC_{10}$  offers better separation compared to correlation and partial correlation in all years.



(a) Top-30 (Peripheral)



(b) Bottom-30 (Central)

Figure 5-8: Minimal spanning tree network generated with  $FAC_{10}$  as linkage metric for 2007-2008. Colors correspond to the sector the stock belongs to as identified in Figure 5-7. The 30 most central and the 30 most peripheral equities as identified by the periphery score are highlighted.

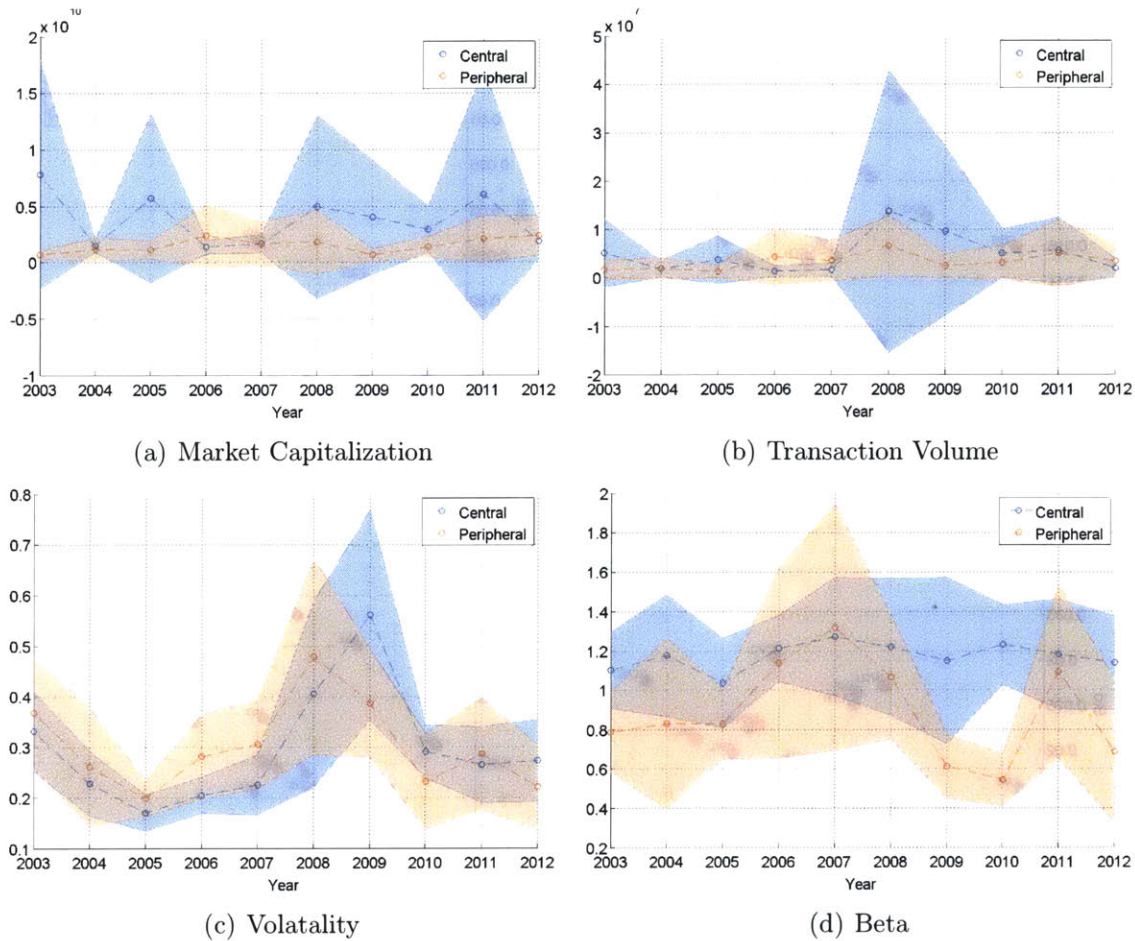
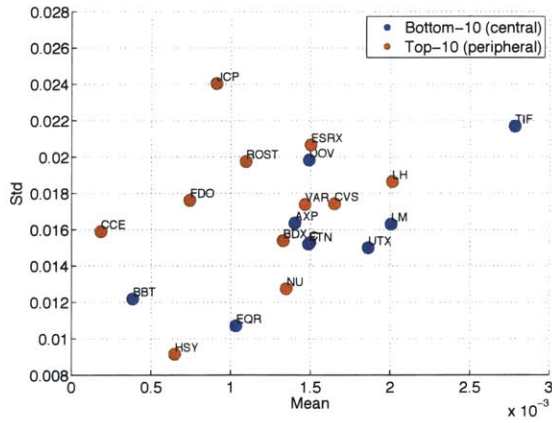


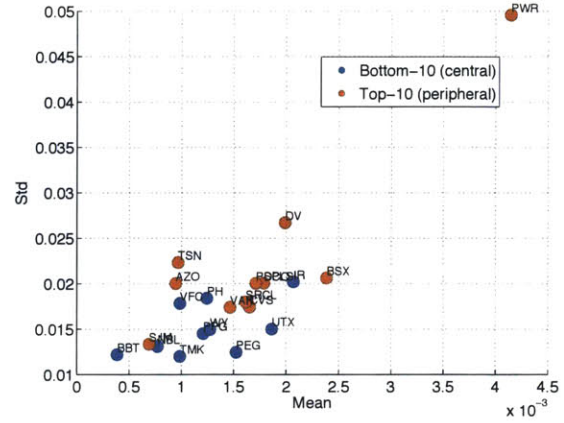
Figure 5-9: The characteristics of the equities in the 10 most central and the 10 most peripheral sets as estimated using periphery score using correlation for linkage. Figures show the average and one standard deviation of market capitalization (a), transaction volume (b), annualized volatility (c), and beta (market dependency) (d).

Inspired by this, we build two long only portfolios, one made of the 10 most central and another with the 10 most peripheral equities based on our linkage methods together with the periphery score as the topological measure. As in the previous experiment, at the start of each year, we estimate the network parameters using the previous two years of historical daily returns. We identify the 10 equities for each group, and equally distribute the capital across them and hold until the end of the year.

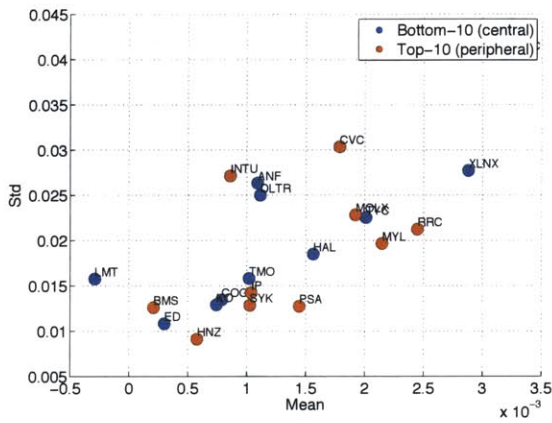
Table 5.2 presents the return characteristics. Pozzi, et al (2013) claims that peripheral equities offer great diversification resulting in higher returns and higher risk adjusted returns (Sharpe ratio) than central equities. They use correlation for linkage



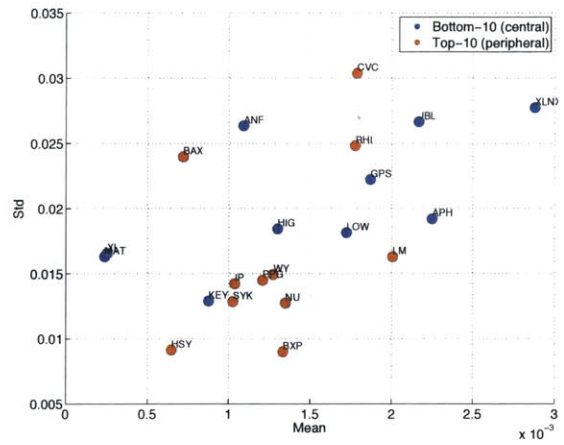
(a) Correlation



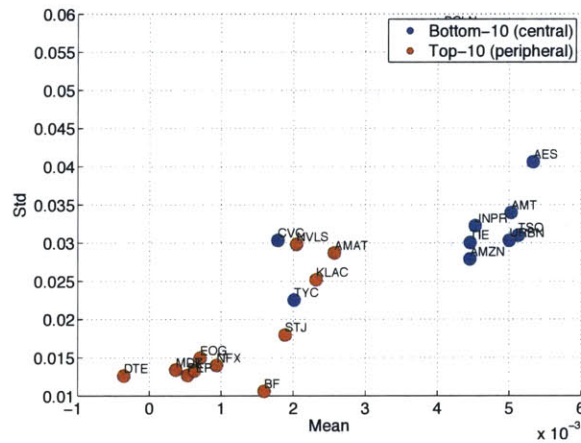
(b) Correlation (market component removed)



(c) Partial correlation



(d) Partial correlation (market component removed)



(e)  $FAC_{10}$

Figure 5-10: Mean and standard deviation of daily returns of the 10 most central equities and the 10 most peripheral equities as estimated by periphery score for year 2003 (network constructed from 2001-2002). For completeness, we provide the top-10 stocks obtained with correlation and partial correlation when market component is removed from the daily returns before building the network.

Table 5.2: Central/peripheral portfolios

MEASURES	CR		PCR		FAC <sub>10</sub>	
	CENTRAL	PERIPHERAL	CENTRAL	PERIPHERAL	CENTRAL	PERIPHERAL
WORST DAY	-0.13	-0.16	-0.12	-0.1	-0.12	-0.1
EXPECTED SHORTFALL (5%)	-0.04	-0.04	-0.04	-0.04	-0.05	-0.03
MAX DRAWDOWN	-0.75	-0.74	-0.63	-0.71	-0.79	-0.65
ANNUALIZED RETURN	0.14	0.1	0.18	0.12	0.26	0.14
INFORMATION RATIO	0.95	0.4	1.25	0.8	1.08	1
SHARPE RATIO	0.61	0.5	0.76	0.62	0.85	0.71

in their work. However, we observe that the portfolio made of central equities achieves higher return and risk adjusted return (Sharpe ratio) across all linkage methods. In contrast, the portfolio of central equities performs poorly in some risk measures - notably in the worst day performance. This observation is further supported by Figure 5-10, where we can see that the daily returns of the central equities have a higher mean and standard deviation compared to the daily returns of peripheral equities.

## 5.5 Graph Clustering

In the previous section, we looked at the return characteristics of two portfolios built using two distinct set of equities (the most central, and the most peripheral). Next, we present a segmentation method based on graph clustering to improve portfolio diversification as opposed to the traditional sector-based groupings.

### 5.5.1 Methods

A typical approach for finding different regions of a network is graph clustering. Spectral clustering is a well known method based on the minimum k-cut [37]. For an undirected graph  $G = \{V, E\}$ , where  $V$  is the set of nodes and  $E$  is the set of Edges, if  $D$  is the edge cost matrix such that  $D_{ij} = c(E_{ij})$  for a cost function  $c : E \rightarrow [0, \infty)$ , we partition  $V$  into a collection of  $k$  clusters  $P = \{V_1, V_2, \dots, V_k\}$ . The problem of finding  $k$  clusters can be stated by the following objective function,

$$\min c(P) = \sum_{x \neq y} \sum_{i \in V_x, j \in V_y} D_{ij}, \quad (5.4)$$

such that no cluster has more than  $\ell$  nodes. The smallest value of  $\ell$  for which this is solvable is  $\ell = \lceil |V|/k \rceil$ . There exists an efficient approximate solution to this optimization based on spectral factorization [41, 55].

**Input:** A multivariate time series  $X \in \mathbb{R}^{n \times p}$ .

**Step 1:** Let  $A$  be the adjacency matrix for the chosen linkage method, and compute  $D$  such that  $D_{ij} = \sqrt{(2(1 - A_{ij}))}$ .

**Step 2:** Normalize  $D$  to make it doubly stochastic so that when the degree of a node is high, the costs associated with its edges are low,

$$D \leftarrow D / (\max_j \sum_i D_{ij})$$

$$D \leftarrow D + (1 - \sum_i D_{ij}) I$$

where  $I$  is an identity matrix of the same size as  $D$ .

**Step 3:** Compute the  $k$  eigenvectors  $U = \{e_1, \dots, e_k\} \in \mathbb{R}^{p \times k}$  of  $D$  corresponding to the  $k$  largest eigenvalues.

**Step 4:** Cluster  $U$  into  $k$  clusters using a clustering algorithm such as k-means.

**Output:** Partitions  $P = \{V_1, \dots, V_k\}$ .

This algorithm does not guarantee that the cluster sizes will not exceed  $\ell$ . Although, it can be achieved by forcing the k-means to produce a balanced set of clusters [99]. In our experiments, we consistently obtained a balanced set of clusters with traditional k-means algorithm (Figure 5-11).

### Connectivity Measure

The fundamental objective of clustering is to group similar equities together in the same cluster, and increase the dissimilarities across the clusters. To accomplish this, we use the connectivity measure proposed in Billio, et al (2012), defined as the fraction



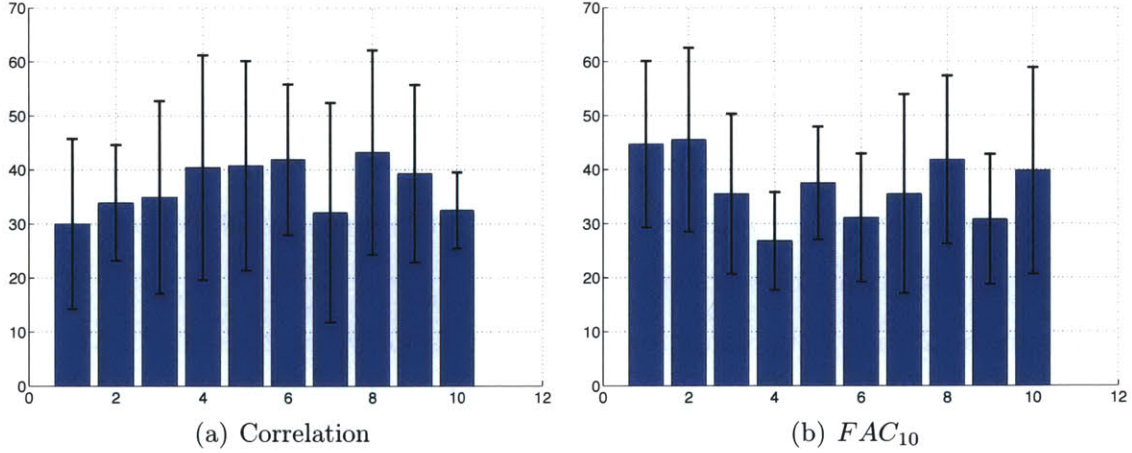


Figure 5-11: Histogram of cluster sizes for clusters identified by spectral clustering with different linkage measures: correlation and  $FAC_{10}$ . The bar chart gives the mean number of cluster members for each cluster across years, and error bars correspond to one standard deviation. We match the cluster labels for consecutive years by greedily optimizing the Jaccard similarity coefficient of the cluster memberships. The Jaccard coefficient measures similarity between two sets ( $A$  and  $B$ ), and is defined as the size of the intersection divided by the size of the union of the sets  $J(A, B) = |A \cap B| / |A \cup B|$  [54].

of the total volatility captured by the first few eigenvalues [10]. When equity returns move together, the first few eigenvalues of the correlation matrix will capture a large portion of the total volatility of the set. For the aggregate return of the set  $R^s = \sum_i^M R^i$ , where  $R^i$  is the return of an individual equity, the variance of the system  $\sigma_s^2$  is given by

$$\sigma_s^2 = \sum_i^M \sum_j^M \sum_k^M \sigma_i \sigma_j L_{ik} L_{jk} \lambda_k. \quad (5.5)$$

Here,  $L_{ik}$  is the factor loading of equity  $i$  corresponding to the  $k$ -th eigenvalue  $\lambda_k$ , and  $\sigma_i$  is the standard deviation of returns  $R^i$ . Then, the fraction of the variance explained by the first  $m$  eigenvalues  $h_m$  is given by,

$$h_m = \frac{\sum_k^m \lambda_k}{\sum_k^M \lambda_k} \quad (5.6)$$

## 5.5.2 Empirical Observations

As in the previous experiment, at the beginning of each year, we build the financial filtered network using the previous two years of historical daily returns. Then, we filter the graph, and segment it into 10 clusters (since there are 10 sectors). Here, we use spectral clustering for segmentation.

Figure 5-12 shows the composition of clusters for years 2003-2012. We match the cluster labels for consecutive years by greedily optimizing the Jaccard similarity coefficient of the cluster memberships. The Jaccard coefficient measures similarity between two sets ( $A$  and  $B$ ), and is defined as the size of the intersection of the sets divided by the size of the union of the sets  $J(A, B) = |A \cap B| / |A \cup B|$  [54]. We see that with correlation each cluster has a dominant sector attached to it. But with  $FAC_{10}$  and  $FAC_{10}^s$ , the membership is more diverse across sectors, implying the connections they capture are more orthogonal to sector grouping than the connections captured by correlation.

Table 5.3: Cluster connectivity

MEASURES	INTRA-CLUSTER (HIGHER IS BETTER)		INTER-CLUSTER (LOWER IS BETTER)	
	$h_1$	$h_2$	$h_1$	$h_2$
SECTOR GROUPING	0.67	0.78	0.48	0.58
CR + MST + GRAPH CLUSTERING	0.67	0.79	0.54	0.63
PCR + MST+ GRAPH CLUSTERING	0.53	0.7	0.49	0.59
$FAC_{10}$ + MST+ GRAPH CLUSTERING	0.53	0.7	0.45	0.56

In Table 5.3, we compare the average intra-cluster connectivity and average inter-cluster connectivity (connectivity among the most central equity of each cluster) for the years 2003-2012. Because the number of members per sector varies from 5 to 60, and because there is a significant overlap across different clusters produced by our methods, we select the 5 most central equities per cluster in this evaluation. We observe that graph clustering improves cluster separation (lower inter-cluster connectivity), but the resulting clusters are less cohesive (lower intra-cluster connectivity) than traditional sector grouping.

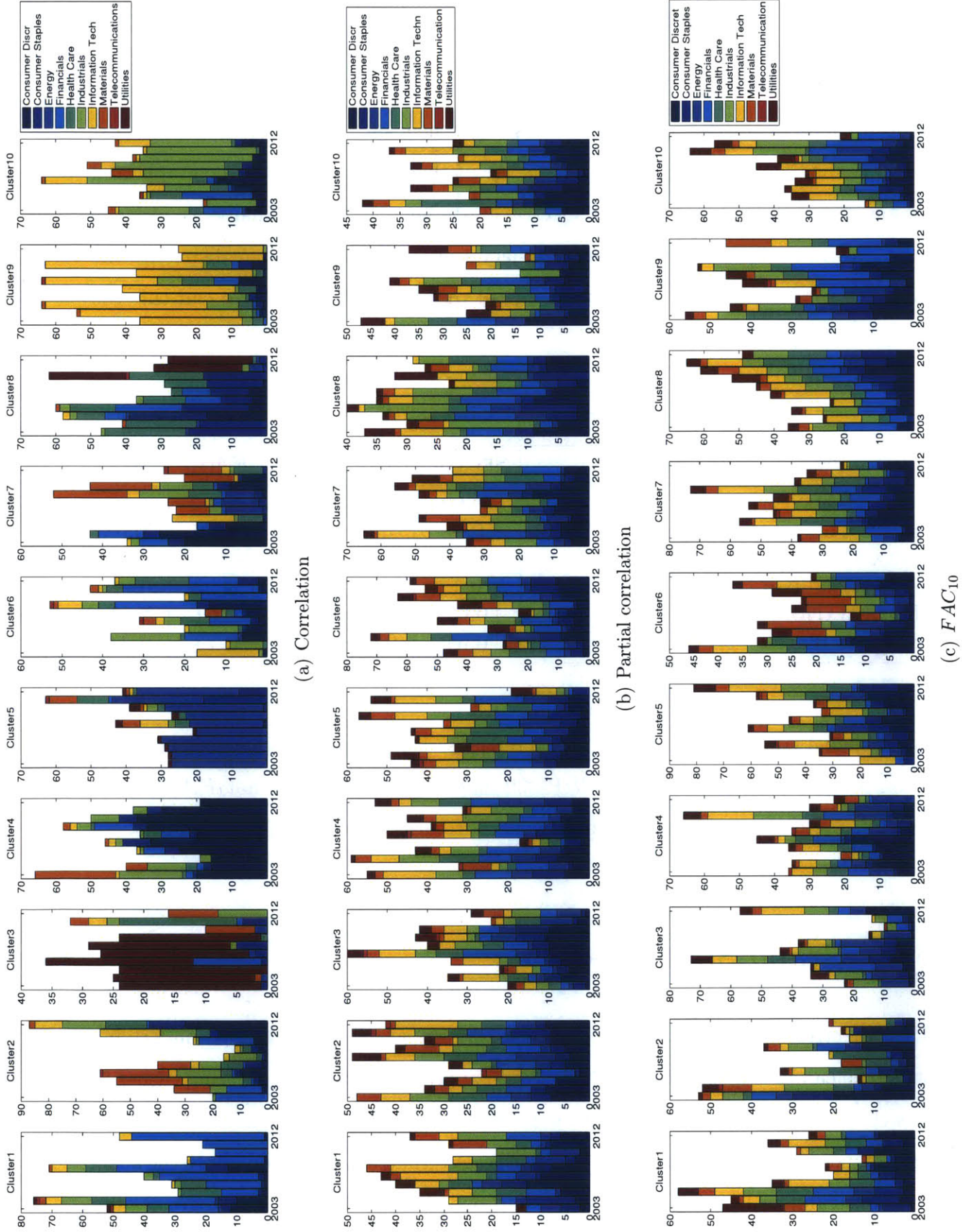


Figure 5-12: Sector composition of the clusters for years 2003-2012 for three different linkage methods. Cluster labels across consecutive years are matched by optimizing Jaccard coefficient of the cluster memberships.

### 5.5.3 Portfolio Returns

Next, we discuss how these clusters could be used to build an index portfolio similar to sector indices. First, as in the previous experiment, at the beginning of each year we build the financial network using a linkage method. Then, we filter the graph, and cluster it into 10 clusters using the spectral clustering algorithm. Finally, we pick the most central equity as given by the periphery score  $p_i$  for each cluster, distribute the capital across these 10 equities, and hold till the end of the year. We call this the cluster index portfolio.

Table 5.4: Cluster index portfolios

MEASURES	SECTOR	CR	PCR	$FAC_{10}$	SPX
WORST DAY	-0.09	-0.12	-0.1	-0.12	-0.09
EXPECTED SHORTFALL (5%)	-0.03	-0.04	-0.04	-0.05	-0.03
MAX DRAWDOWN	-0.71	-0.68	-0.61	-0.77	-0.79
ANNUALIZED RETURN	0.1	0.14	0.14	0.22	0.05
INFORMATION RATIO	0.67	1.01	1	0.91	N/A
SHARPE RATIO	0.54	0.63	0.68	0.73	0.34

Table 5.4 compares the return characteristics. For completeness, we also provide a comparison with the returns of S&P 500 index (SPX). We note that  $FAC_{10}$  gives the best performance in terms of both over all returns and risk adjusted returns.

## 5.6 Graph Sparsification

The objective of network filtering is to sparsify the graph such that the underlying structure can be exposed and studied. When applying the network filtering earlier in this chapter, we retained the highest connected links while constraining some overall property of the network, such as the planarity. Alternatively, we could apply the sparsity constraint in learning the linkages, and our method of learning linkages readily supports this.

### 5.6.1 Method

There have been methods proposed in the literature to sparsify inverse covariance method estimation [74, 72]. Because inverse covariance defines the dependencies of the graph of a multivariate time series, a sparse inverse covariance matrix leads to fewer dependencies in the graph, and thus to graph sparsification.

Following this approach, we rewrite Equation 3.3 by replacing the  $\ell_2$  norm by the  $\ell_1$  norm. When the  $\ell_1$  norm is used for regularization, it leads to solutions with fewer nonzero parameter values, effectively reducing the number of variables upon which the given solution is dependent. We make use of the *Adaptive Lasso* estimator proposed by Zou (2006) for the case of dependent data [101]. We choose the penalty on each of the weights as the reciprocal of the absolute value of a pre-estimator, which is obtained by applying Equation 3.3 with the  $\ell_2$  norm. We call the corresponding linkage method  $FAC_{10}^s$ . Since the resulting graph is already sparse, we don't apply graph filtering.

### 5.6.2 Portfolio Returns

In Table 5.5, we provide the results of the previous experiment with  $FAC_{10}^s$ . We also provide comparisons with the returns corresponding to a network built using sparse partial correlation estimates ( $PCR^s$ ). For  $PCR^s$ , we use the method proposed in Pavlenko, et al (2012) [72]. As in the case of  $FAC_{10}^s$ , the resulting network is already sparse, and therefore we don't apply network filtering to this method.  $FAC_{10}^s$  gives better results compared to  $FAC_{10}$ . Further, instead of relying on an arbitrary constraint to filter the graph, with  $FAC_{10}^s$  we let the sparsity constraint of the optimization guide the process.

## 5.7 Application Considerations

The index portfolios built using financial networks achieve higher risk adjusted returns than the market index. For example, the partial correlation network based cluster

Table 5.5: Cluster index portfolios with graph sparsification

MEASURES	$PCR$	$PCR^s$	$FAC_{10}$	$FAC_{10}^s$
WORST DAY	-0.1	-0.09	-0.12	-0.09
EXPECTED SHORTFALL (5%)	-0.04	-0.04	-0.05	-0.04
MAX DRAWDOWN	-0.61	-0.8	-0.77	-0.54
ANNUALIZED RETURN	0.14	0.16	0.22	0.22
INFORMATION RATIO	1	0.88	0.91	1.56
SHARPE RATIO	0.68	0.7	0.73	0.91

index portfolio and  $FAC_{10}$  based cluster index portfolio achieve 3 times and 4 times the annualized return of S&P 500 index respectively. Because these portfolios are held for a year, they incur very small transaction costs.

Relationships among equities can be viewed as occurring on three time scales: long-term, mid-term and short-term. In this chapter, we capture only the long-term relationships. This may result in a slower response when underlying conditions change. In contrast, more frequent updates will lead to higher transaction cost. By incorporating these considerations into the model, for a given transaction cost an optimal update frequency can be found.

We show that by constructing a well-diversified portfolio, risk management can be a source of excess returns. As with the portfolios presented in Chapter 3, the portfolios described in this chapter are long only portfolios. Therefore, they are exposed to market risk. Without short positions (or derivatives that simulate short positions), this kind of risk cannot be diversified away.

## 5.8 Summary

We explored the task of leveraging network structure to study the dependencies in financial time series. We compared the use of different linkage metrics to encode the higher order connections in financial time series via the topological structure of a network. We discussed how the topological properties of a financial filtered network could be used to obtain an improved segmentation of equities, thus improving the portfolio selection process.

For example, we showed that highly central equities, even after the market dependency is excluded, are more systemic. They offer high returns at the expense of high variance and increased risk (measured by worst day performance, expected shortfall, and maximum drawdown), and are significantly influenced by the market movements. We also showed that peripheral equities, when they are uncorrelated among themselves, can offer similar risk adjusted returns but at a reduced risk.

Then, we presented a segmentation method based on graph clustering as opposed to the traditional sector based groupings. We showed that a portfolio built using this method provides better risk adjusted returns compared to sector based grouping and the S&P 500 index.





# Chapter 6

## CONCLUSION AND FUTURE WORK

In this thesis, we showed how cross-sectional patterns on financial time series could be learned using techniques from data mining and machine learning. We also showed how these patterns could be leveraged to build a well-diversified portfolio, and that these portfolios performed well on out of sample data in terms of minimizing risk and achieving high returns.

We discussed three types of complimentary methods for leveraging data to learn cross-sectional patterns in financial time series. These methods shared the common goal of constructing a selective but well-diversified portfolio that performs well out of sample.

- We presented a method for learning connections between financial time series in terms of large losses. We used a factor model, and we modeled daily returns using three factors: active return, market sensitivity, and connectedness of returns. We learned these factors using a recursive regression. We employed a model independent approach to prioritize the connections on large losses. Thus, our method allows the user choose an objective function that balances the need to minimize the co-occurrences of large losses against the need to minimize overall connections. We showed that portfolios constructed using this method not only “beat the market,” but also beat the so-called “optimal portfolios.” More importantly, portfolios constructed using our method significantly reduced tail

risk without sacrificing overall returns.

- We presented a method for learning connections between financial time series and news releases. A problem with using news releases to predict future return is that often the information in the news has already been incorporated into the pricing of the equity. We modeled the sentiment of the news releases based on the past returns and past news flows, and then used this model to discount the expected sentiment from future news releases. Thus, we modeled the excess returns in terms of excess sentiment. We showed that while using news directly generated an annualized return of 22% over a 10-year period, our proposed way of handling the past increased the annualized return to 34% over the same period.
- We discussed how leveraging network structure to model the dependencies in financial time series could be useful in obtaining an improved segmentation of equities. We showed that segmenting equities in this way can be used to improve the portfolio selection process. For example, we showed that highly central equities offered high returns at the expense of high variance and increased risk (measured by worst day performance, expected shortfall, and maximum draw-down). We also showed that peripheral equities, when they were uncorrelated among themselves, could offer similar risk adjusted returns but at a reduced risk compared to central equities. Then, we showed that a segmentation method based on graph clustering could be used as an alternative to the traditional sector based groupings.

Across all three tasks discussed in this thesis, the appropriate modeling of the connections between equities has been crucial to good performance.

Specifically, we showed the importance of regularized learning, frequent model updates, and flexible models that utilize techniques from multiple areas.

## 6.1 Future Directions

The idea of modeling the “correlation” structure by learning the connections could be used in many other areas in finance. Specifically, addressing the issue of *tail risk* when learning the connections could be useful in modeling counter party risk in brokerages, inter-bank lending for retail banks, and unwinding risk for hedge funds.

For some applications, we update the portfolio daily, which results in high turnover. This in turn results in increased transaction costs and margin requirements. Though, a short-term portfolio, by definition, needs to be updated frequently, we can update the portfolio weights more intelligently by factoring in these additional costs into the model.

We used Thomson Reuters NewsScope dataset together with daily returns data. There are several other datasets that we could use to learn the connections between equities. Investor sentiment can be gleaned from Twitter feeds. Compared to the Thomson Reuters NewsScope data, Twitter feeds are more frequent, real time, and available at “zero cost.” However, the advantage offered by Twitter feeds of being “the wisdom of crowds,” may very well be its downside when it becomes the “the madness of mobs.’

Generally, we considered long-only portfolios. When all equities are heavily correlated, as during the 2008 crisis, market risk governs most of the returns. Without short positions (or derivatives that simulate short positions), this kind of risk cannot be diversified away. We can incorporate these strategies into our model. In Section 3.5, we discussed how we could hedge against a known risk factor (e.g., market risk). Such strategies will have additional restrictions embedded. For instance, there are fees associated with stock borrowing to facilitate short selling, and during certain times (especially when short-selling is most needed) regulative restrictions can be imposed on short-selling. Incorporating these concerns into the model would make the model more versatile.

In financial time series analysis, patterns can be learned in four time scales: long-term (fundamental based), mid-term (technical and macro-based), short-term (event

based), and ultra-short-term (based on inefficiencies in statistical arbitraging). In this thesis, we captured the short-term (Chapter 4), mid-term (Chapter 3), and long-term (Chapter 5). An intriguing line of future work is to use these strategies and others such as equal weighted portfolio in an ensemble. Then, we can learn to weight each of them based on investor preferences and macroeconomic situations.

Like correlation, the connection we learn is not an observable variable, but an average estimated over a time period. With the availability of intraday data of stock quotes at millisecond levels, researchers have proposed methods to treat volatility of daily returns (average variance of daily returns) as an observable [6, 3, 22]. We can extend this to the problem of modeling the “correlation” structure on the daily returns, which will open an exciting avenue of research. Then, the latent connections captured by the methods proposed in this thesis could be treated as observables.

In the world of finance, where the actors are eager to exploit any inefficiency in the system aggressively, potential benefit of a model will quickly decay with time. This feedback necessitates constant updates to the model, and furthermore, requires fresh ideas to be profitable. If such feedbacks can be factored into the model, the useful lifespan of the model could be lengthened.

Finally, the act of predicting the future and acting upon this information changes the world, i.e., the circumstances that initially led to the specific prediction. This is particularly the case for large institutional investors who would trade in large amounts. Some of the effects of this issue is factored into the transaction cost, but addressing this issue specifically would be an interesting problem.

# BIBLIOGRAPHY

- [1] Vikas Agarwal and Narayan Y Naik. Risks and portfolio decisions involving hedge funds. *Review of Financial studies*, 17(1):63–98, 2004.
- [2] Yuan An, Jeannette Janssen, and Evangelos E Milios. Characterizing and mining the citation graph of the computer science literature. *Knowledge and Information Systems*, 6(6):664–678, 2004.
- [3] Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Paul Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, 2003.
- [4] Andrew Arnold, Yan Liu, and Naoki Abe. Temporal causal modeling with graphical granger methods. In *Proceedings of the 13th ACM SIGKDD, KDD '07*, pages 66–75. ACM, 2007.
- [5] K H Bae. A New Approach to Measuring Financial Contagion. *Review of Financial Studies*, 16(3):717–763, July 2003.
- [6] Ole E Barndorff-Nielsen. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280, 2002.
- [7] Christopher B Barry. Portfolio analysis under uncertain means, variances, and covariances. *The Journal of Finance*, 29(2):515–522, 1974.
- [8] Suleyman Basak and Alexander Shapiro. Value-at-risk-based risk management: optimal policies and asset prices. *Review of Financial studies*, 14(2):371–405, 2001.
- [9] R.M. Bell and Y. Koren. Scalable collaborative filtering with jointly derived neighborhood interpolation weights. *ICDM 2007. Seventh IEEE International Conference on*, pages 43–52, 2007.
- [10] Monica Billio, Mila Getmansky, Andrew W Lo, and Lorian Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559, 2012.
- [11] M O Boldi and A C Davison. A mixture model for multivariate extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):217–229, 2007.

- [12] Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.
- [13] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.
- [14] Giovanni Bonanno, Guido Caldarelli, Fabrizio Lillo, and Rosario N Mantegna. Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, 68(4):046130, 2003.
- [15] Giovanni Bonanno, Guido Caldarelli, Fabrizio Lillo, Salvatore Miccichè, Nicolas Vandewalle, and Rosario Nunzio Mantegna. Networks of equities in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2):363–371, 2004.
- [16] A. Borodin, R. El-Yaniv, and V. Gogan. Can we learn to beat the best stock. *J. Artif. Intell. Res. (JAIR)*, 21:579–594, 2004.
- [17] E. Castillo and A.S. Hadi. Fitting the generalized pareto distribution to data. *Journal of the American Statistical Association*, 92(440):1609–1620, 1997.
- [18] Wesley S Chan. Stock price reaction to news and no-news: drift and reversal after headlines. *Journal of Financial Economics*, 70(2):223–260, 2003.
- [19] Don M Chance and Robert Brooks. *Introduction to Derivatives and Risk Management*. Cengage Learning, 2009.
- [20] P. Chen and H. Chihying. Learning causal relations in multivariate time series data. *Economics: The Open-Access, Open-Assessment E-Journal*, 1:11, 2007.
- [21] U. Cherubini, E. Luciano, and W. Vecchiato. *Copula methods in finance*. Wiley, 2004.
- [22] Roxana Chiriac and Valeri Voev. Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947, 2011.
- [23] Todd E Clark and Kenneth D West. Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of Econometrics*, 135(1):155–186, 2006.
- [24] S.G. Coles and J.A. Tawn. Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 377–392, 1991.
- [25] Stuart Coles, Joanna Bawa, Lesley Trenner, and Pat Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.
- [26] Laurens De Haan and Ana Ferreira. *Extreme value theory: an introduction*. Springer, 2007.

- [27] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the  $1/n$  portfolio strategy? *Review of Financial Studies*, 22(5):1915–1953, 2009.
- [28] Ahmet Duran and Gunduz Caginalp. Overreaction diamonds: Precursors and aftershocks for significant price changes. *Quantitative Finance*, 7(3):321–342, 2007.
- [29] Frank J Fabozzi, T Daniel Coggin, Bruce Collins, and John Ritchie. *Investment management*, volume 2. Prentice Hall, 1995.
- [30] Eugene Fama, Lawrence Fisher, Michael Jensen, and Richard Roll. The adjustment of stock prices to new information. *International Economic Review*, 10, 1969.
- [31] Eugene F Fama. The behavior of stock-market prices. *The Journal of Business*, 38(1):34–105, 1965.
- [32] Eugene F Fama. Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3):283–306, 1998.
- [33] Bruce J Feibel. *Investment performance measurement*, volume 116. John Wiley & Sons, 2003.
- [34] Soheil Feizi, Daniel Marbach, Muriel Médard, and Manolis Kellis. Network deconvolution as a general method to distinguish direct dependencies in networks. *Nature biotechnology*, 2013.
- [35] Robert Fildes, Michèle Hibon, Spyros Makridakis, and Nigel Meade. Generalising about univariate forecasting methods: further empirical evidence. *International Journal of Forecasting*, 14(3):339–358, 1998.
- [36] K J Forbes and R Rigobon. No contagion, only interdependence: measuring stock market comovements. *The Journal of Finance*, 57(5):2223–2261, 2002.
- [37] André Fujita, Patricia Severino, Kaname Kojima, João R Sato, Alexandre G Patriota, and Satoru Miyano. Functional clustering of time series gene expression data by granger causality. *BMC systems biology*, 6(1):137, 2012.
- [38] Lorenzo Garlappi, Raman Uppal, and Tan Wang. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies*, 20(1):41–81, 2007.
- [39] William Goetzmann, Jonathan Ingersoll, Matthew I Spiegel, and Ivo Welch. Sharpening sharpe ratios. Technical report, National bureau of economic research, 2002.
- [40] Richard C Grinold and Ronald N Kahn. *Active portfolio management*, volume 2. McGraw-Hill New York, 2000.

- [41] Joao P Hespanha. An efficient matlab algorithm for graph partitioning. *Dept. Elect. Comput. Eng., Univ. California, Santa Barbara, CA, USA, Tech. Rep*, 2004.
- [42] Douglas W Hubbard. *The failure of risk management: why it's broken and how to fix it*. John Wiley and Sons, 2009.
- [43] Ravi Jagannathan and Tongshu Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1684, 2003.
- [44] Dennis W Jansen, Kees G Koedijk, and Casper G De Vries. Portfolio selection with limited downside risk. *Journal of Empirical Finance*, 7(3):247–269, 2000.
- [45] Michael C Jensen, Fischer Black, and Myron S Scholes. The capital asset pricing model: Some empirical tests. 1972.
- [46] A Kalai and S Vempala. Efficient algorithms for universal portfolios. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pages 486–491, 2000.
- [47] M.G. Kendall and A. Stuart. *The Advanced Theory of Statistics. Vol. 2: Inference and: Relationship*. Griffin, 1973.
- [48] Dror Y Kenett, Michele Tumminello, Asaf Madi, Gitit Gur-Gershgoren, Rosario N Mantegna, and Eshel Ben-Jacob. Dominating Clasp of the Financial Sector Revealed by Partial Correlation Analysis of the Stock Market. *PLoS ONE*, 5(12):e15032, December 2010.
- [49] A. Khandani and A. Lo. What happened to the quants in august 2007? *Available at SSRN 1015987*, 2007.
- [50] Roger W Klein and Vijay S Bawa. The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, 3(3):215–231, 1976.
- [51] Leonid Kogan and Dimitris Papanikolaou. Growth opportunities, technology shocks, and asset prices. *The Journal of Finance*, 2013.
- [52] Leonid Kogan and Mary H Tian. Firm Characteristics and Empirical Factor Models: A Data-Mining Experiment. *SSRN Electronic Journal*, 2012.
- [53] CNV Krishnan, Ralitsa Petkova, and Peter Ritchken. Correlation risk. *Journal of Empirical Finance*, 16(3):353–367, 2009.
- [54] Michael Levandowsky and David Winter. Distance between sets. *Nature*, 234(5323):34–35, 1971.
- [55] Chansook Lim, Stephan Bohacek, João P Hespanha, and Katia Obraczka. Hierarchical max-flow routing. In *GLOBECOM*, page 6, 2005.



- [56] Andrew W Lo. *Hedge Funds: An Analytic Perspective (Revised and Expanded Edition)*. Princeton University Press, 2010.
- [57] F. Longin and B. Solnik. Correlation structure of international equity markets during extremely volatile periods. *Available at SSRN 147848*, 1999.
- [58] Francois Longin and Bruno Solnik. Extreme correlation of international equity markets. *The Journal of Finance*, 56(2):649–676, 2001.
- [59] D. MacKenzie. End-of-the-world trade. *London Review of Books*, 30(9):24–26, 2008.
- [60] Tanja Magoc. *New algorithms for optimal portfolio selection*. Proquest, Umi Dissertatio, 2011.
- [61] Yuexin Mao, Wei Wei, Bing Wang, and Benyuan Liu. Correlating S&P 500 stocks with Twitter data. In *HotSocial '12: Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research*, 2012.
- [62] H. Markowitz. Portfolio selection: efficient diversification of investments. *New Haven, CT: Cowles Foundation*, 94, 1959.
- [63] Alexander J McNeil. Extreme value theory for risk managers. *Departement Mathematik ETH Zentrum*, 1999.
- [64] Alexander J McNeil and Rüdiger Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3):271–300, 2000.
- [65] D. Meneguzzo and W. Vecchiato. Copula sensitivity in collateralized debt obligations and basket default swaps. *Journal of Futures Markets*, 24(1):37–70, 2004.
- [66] Robert C Merton. An analytic derivation of the efficient portfolio frontier. *Journal of financial and quantitative analysis*, 7(04):1851–1872, 1972.
- [67] M.-A Mittermayer and G.F Knolmayer. NewsCATS: A News Categorization and Trading System. In *Data Mining, 2006. ICDM '06. Sixth International Conference on Data Mining*, 2006.
- [68] Marc-André Mittermayer and Gerhard Knolmayer. *Text mining systems for market response to news: A survey*. Institut für Wirtschaftsinformatik der Universität Bern, 2006.
- [69] Mark EJ Newman. The mathematics of networks. *The new palgrave encyclopedia of economics*, 2:1–12, 2008.
- [70] Lawrence Page. Method for node ranking in a linked database, September 4 2001. US Patent 6,285,999.

- [71] L'uboš Pástor. Portfolio selection and asset pricing models. *The Journal of Finance*, 55(1):179–223, 2000.
- [72] Tatjana Pavlenko, Anders Björkström, and Annika Tillander. Covariance structure approximation via glasso in high-dimensional supervised classification. *Journal of Applied Statistics*, 39(8):1643–1666, 2012.
- [73] D. Pekasiewicz. *Methods of multivariate statistical analysis and their applications*. Wydawnictwo Uniwersytetu Łódzkiego, 2007.
- [74] Jie Peng, Pei Wang, Nengfeng Zhou, and Ji Zhu. Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, 104(486), 2009.
- [75] Gustavo Peralta and Abalfazl Zareei. A network approach to portfolio selection. *Available at SSRN 2430309*, 2014.
- [76] Y. Plyakha, R. Uppal, and G. Vilkov. Why does an equal-weighted portfolio outperform value-and price-weighted portfolios? *Available at SSRN 1787045*, 2012.
- [77] Ser-Huang Poon, Michael Rockinger, and Jonathan Tawn. Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *Review of financial studies*, 17(2):581–610, 2004.
- [78] F Pozzi, T Di Matteo, and T Aste. Spread of risk across financial markets: better to invest in the peripheries. *Scientific reports*, 3, 2013.
- [79] F.K. Reilly and K.C. Brown. *Investment Analysis and Portfolio Management*. South-Western Pub, 2011.
- [80] A.J. Richards. Comovements in national stock market returns: Evidence of predictability, but not cointegration. *Journal of Monetary Economics*, 36(3):631–654, 1995.
- [81] R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.
- [82] Eduardo J Ruiz, Vagelis Hristidis, Carlos Castillo, Aristides Gionis, and Alejandro Jaimes. Correlating financial time series with micro-blogging activity. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 513–522. ACM, 2012.
- [83] Paul Ryan and Richard J Taffler. Are economically significant stock returns and trading volumes driven by firm-specific news releases? *Journal of Business Finance & Accounting*, 31(1-2):49–82, 2004.
- [84] Robert P Schumaker and Hsinchun Chen. Textual analysis of stock market prediction using breaking financial news. *ACM Transactions on Information Systems*, 2009.

- [85] W.F. Sharpe. The Sharpe Ratio, 1994.
- [86] William F Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk\*. *The journal of finance*, 19(3):425–442, 1964.
- [87] William T Shaw. Monte carlo portfolio optimization for general investor risk-return objectives and arbitrary return distributions: a solution for long-only portfolios. *arXiv preprint arXiv:1008.3718*, 2010.
- [88] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000.
- [89] Malcolm K Sparrow. The application of network analysis to criminal intelligence: An assessment of the prospects. *Social networks*, 13(3):251–274, 1991.
- [90] Timm O Sprenger, Andranik Tumasjan, Philipp G Sandner, and Isabell M Welpe. Tweets and trades: The information content of stock microblogs. *European Financial Management*, 2013.
- [91] Robert F Stambaugh, Jianfeng Yu, and Yu Yuan. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2):288–302, 2012.
- [92] C. Stărică. Multivariate extremes for models with constant conditional correlations. *Journal of Empirical Finance*, 6(5):515–553, 1999.
- [93] Robert A Strong. *Portfolio construction, management, and protection*. Thomson/South-Western, 2006.
- [94] R M Stulz. Rethinking risk management. *Journal of applied corporate finance*, 9(3):8–25, 2005.
- [95] Michel Tumminello, Tomaso Aste, Tiziana Di Matteo, and RN Mantegna. A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30):10421–10426, 2005.
- [96] N Vandewalle, F Brisbois, and X Tordoir. Non-random topology of stock markets. *Quantitative Finance*, 1(3):372–372, 2001.
- [97] Zhenyu Wang. A shrinkage approach to model uncertainty and asset allocation. *Review of Financial Studies*, 18(2):673–705, 2005.
- [98] H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, pages 817–838, 1980.
- [99] Shi Zhong and Joydeep Ghosh. Scalable, balanced model-based clustering. In *SDM*, pages 71–82. SIAM, 2003.

- [100] M. Zhu. Recall, precision and average precision. *Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, 2004.*
- [101] Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429, 2006.