Pattern-Based Method for Reducing Drawdowns in Stock Index Investment

LIU, Lili

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Master of Philosophy
in
Computer Science and Engineering

The Chinese University of Hong Kong
May 2015

Thesis Assessment Committee

Professor Lai Wan Chan (Chair)
Professor Benjamin Wan Sang Wah (Thesis Supervisor)
Professor Kwong Sak Leung (Committee Member)
Professor Qing Li (External Examiner)

Abstract

Pattern-Based Method for Reducing Drawdowns in Stock Index Investment

by

LIU, Lili

Master of Philosophy

Department of Computer Science and Engineering

The Chinese University of Hong Kong

Drawdown is the percentage loss of wealth from the previous peak. It is an important downside-risk metric to measure the performance of portfolio investment. Constructing a portfolio with small drawdowns and high returns is the consistent goal of all investors. However, great drawdowns are not rare in financial history, and they happened on stock indices. To the best of our knowledge, there are no effective and practical drawdown-reduction approaches.

In this thesis, we propose a novel pattern-based drawdown-reduction model to reduce future drawdowns of stock index investment. It firstly extracts two kinds of relevant features, i.e., *ddFeature* and *priceFeature*, from historical stock index price series, then trains a prediction model based on these pattern-and-drawdown pairs in history, and finally makes investment decisions according to a binary investment function. The effects of parameters involved with feature extraction and investment decision are well studied, and an adaptive investment method is also proposed. Experimental results on three different datasets, S&P 500, NASDAQ Composite and HSI, show that the proposed method achieves both small drawdowns as well as high cumulative return.

摘要

基於模式的降低股票指數跌幅的投資方法

劉麗麗

香港中文大學

計算機科學與工程學系

哲學碩士

跌幅是指當前投資資產相對於之前資產最高值的損失百分比。它是衡量投資風險的重要指標。因而，構建一個跌幅小且收益高的投資組合是所有投資人共同的願望。然而，在金融市場中，大的跌幅時有發生，甚至在股指上也會出現。研究發現，在當前的投資管理中，還沒有行之有效的降低跌幅的方法。

本論文中，我們提出了一種基於模式的降低股指跌幅的模型。該模型首先從股指的歷史日收盤價序列中提取了兩種相關特征：跌幅特征（$ddFeature$）和價格特征（ $priceFeature$），然後利用提取的特征-跌幅對來訓練預測模型，最後，利用二值決策函數根據預測的跌幅做出最終的投資決策。特征提取過程和投資決策過程中用到的參數也進行了詳細的研究。基於此，我們提出了一種自適應參數的投資模型。實驗結果表明，該方法在三種不同的數據庫（標普500，納斯達克綜合指數以及恆生指數）上取得了較高的收益和非常低的跌幅。

Acknowledgements

I would like to express my sincere appreciation to my adviser, Prof. Benjamin W. Wah, for his support and guidance. I would like to especially acknowledge his methodology of research. He always emphasizes focusing on the most important problems and provides guidance to solve those problems based on their nature, which benefits me a lot.

I would like to thank my committee members, Prof. Kwong Sak Leung, Prof. Lai Wan Chan and Prof. Qing Li for their valuable suggestions.

I also would like to thank my friends in CUHK, especially Jingxi Xu, Zhitang Chen, Mengyi Zhang and Weicong Liu.

Lastly, I am deeply appreciate of the love of my parents, my husband and my brother.

Contents

Chapter 1

Introduction

In financial industry, drawdown (DD) is usually defined as the percentage loss of current wealth from the maximum wealth till now [YZ12]. Specifically, given a wealth path $W = \{W_t | 0 \leq t \leq T\}$, the drawdown at time t ($t \in [0, T]$) is defined as $DD_t(W) = (1 - \frac{W_t}{M_t}) \times 100\%$, where W_t is the wealth value at time t and $M_t = max\{W_\tau\}_{\tau=0}^{t}$ is the maximum wealth from time 0 to time t . It can be seen from the formula that the drawdown value is in the interval of 0% to 100%.

Figure 1.1 shows an example of computing drawdown values. The top figure shows a simulated wealth path from W_0 to W_{10}. The x axis is the time t, and the y axis is the wealth value. The bottom figure shows the corresponding drawdown values.

The maximum drawdown ($MaxDD$) of a given wealth path $W = \{W_t | 0 \leq t \leq T\}$ is defined as the largest peak to trough percentage loss $MaxDD(W) = max\{DD_\tau(W)\}_{\tau=0}^{T}$. It reflects the maximum percentage loss of the wealth path. The maximum drawdown value in Figure 1.1 is 26.7%, which is marked by red dot in the bottom figure. It is computed by the peak at time 4 and the trough at time 7.

Drawdown is one of the most important measure for assessing downside-risk associated with an investment. And it is widely used in investment management. The drawdown of a portfolio determines not only the investors' wealth accumulation, but also the fund managers' career development.

Before investing money to a financial product, all investors have their expec-

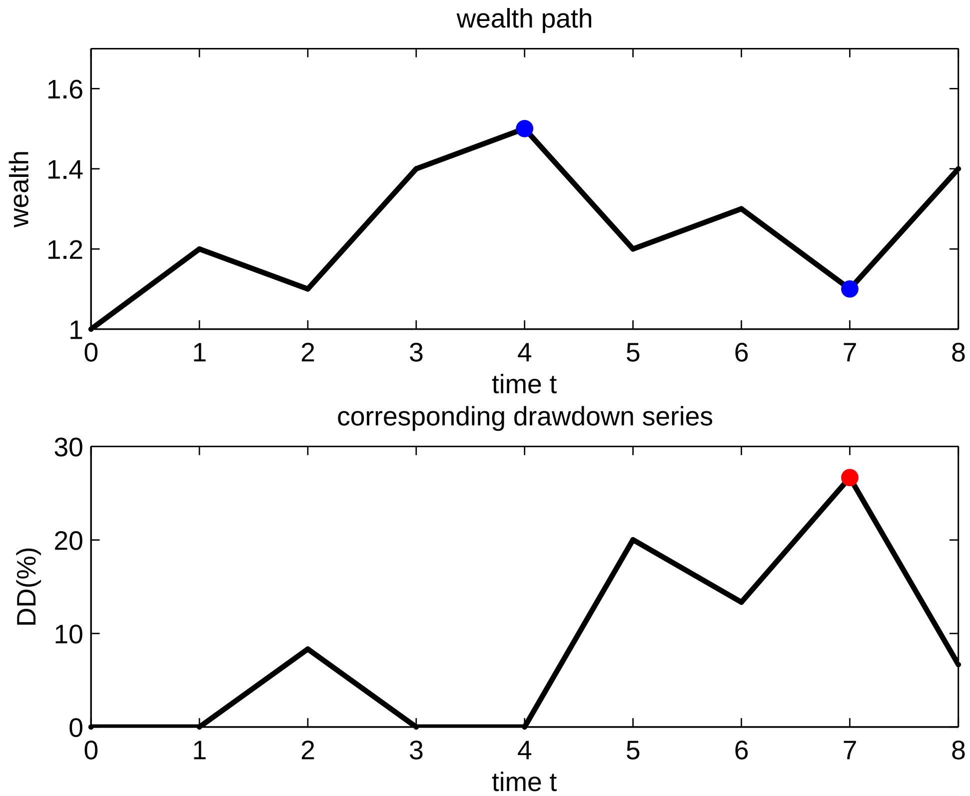


Figure 1.1: The simulated wealth path and corresponding drawdown (%) series.

tation and risk tolerance of the financial product. At least, they expect to earn more money than fixed interest rate by putting money in the bank. Meanwhile, they know the inevitable risk accompany with the product and set the risk tolerance according to their financial condition and personal preference. Drawdown is so straightforward that it is easy to be understood, especially for investors who may have no financial background knowledge. If the behavior of the product deviates from the expectation and challenges the bottom line of drawdown tolerance, investors may decide to withdraw their money.

Besides, drawdown is also an important measure for the performance of fund managers. The responsibility of fund managers is to trade the money of clients and earn money in the form of management fees from the clients' accounts. In another word, the clients' accounts are the only source of income of fund managers. Losing these accounts will directly lead to a collapse in their salaries, and even worse, it may lead to the end of their careers. Therefore, the primary concern of fund managers is to keep the existing clients and attract more new ones. To satisfy the expectation of clients, they have to strictly control the drawdown level. Consequently, drawdown is a key factor to measure the performance of fund managers, and determines their salaries and bonuses. According to Chekhlov et al. [CUZ05], a 50% drawdown is unlikely to be tolerated for any average account, and an account with 20% drawdown may be shut down.

For fund managers, they naturally expect to win all the time. The reality is that no matter what strategy is used, loss is inevitable. This is due to the uncertainty nature of investment. What's worse, great drawdowns were not rare in financial history. For example, Figure 1.2 shows the cumulative return and the corresponding drawdown of passive investing [1] on daily S&P 500 stock index from 1990/01/02 to 2014/03/14. It can be seen that there are two great drawdowns: the first one started at March 24, 2000 and ended at October 09, 2002 accompanying with 49% drawdown; the second one started at October 09, 2007 and lasted till March 09, 2009, and the corresponding drawdown was even large, i.e., about 57%. The two periods are illustrated by two pairs of red lines in the top figure and

[1] Passive investing is a widely used investing strategy. It mimics the specified index, and expects to obtain the same profit as the return of the index itself. Therefore, it actually implements the "buy-and-hold" strategy, that is, one invests his/her wealth on some assets at initial time and holds the portfolio all the time.

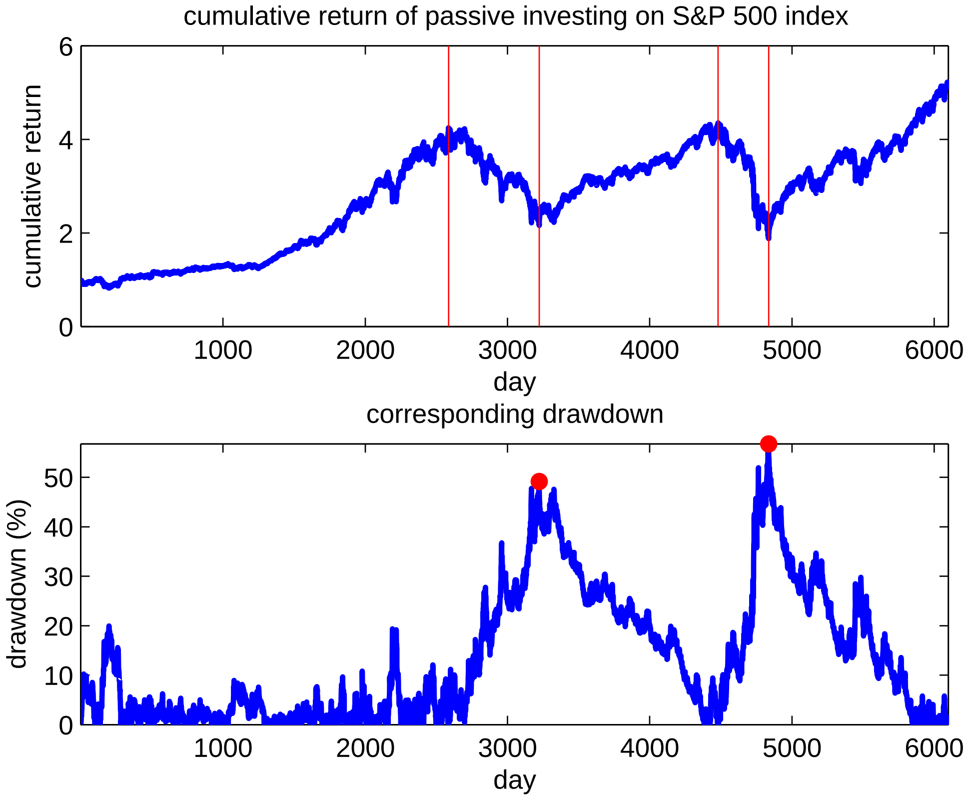


Figure 1.2: Cumulative return and corresponding drawdown series of passive investing on daily S&P 500 index from 01/02/1990 to 03/14/2014.

the corresponding drawdown are marked by red dots in the bottom figure. Based on random walk model of stock prices, Zhou and Zhu [ZZ09] point out that the probability of a 50% drawdown happened in stock market within a century is about 90%. These great drops severely breakthrough the tolerance of investors, which may lead to devastating effects to investors and fund managers, and even lead to the bankruptcy of investment companies.

Except for the long-term drawdown behavior, the moving yearly, i.e., 250-day, maximum drawdown values is shown in Figure 1.3. It shows even considering the drawdown in one year interval, the maximum drawdown is very large for some periods, which is intolerable by investors.

Facing the severe reality, reducing drawdown is the primary task for all fund managers. Actually, small drawdowns and high cumulative return are the two

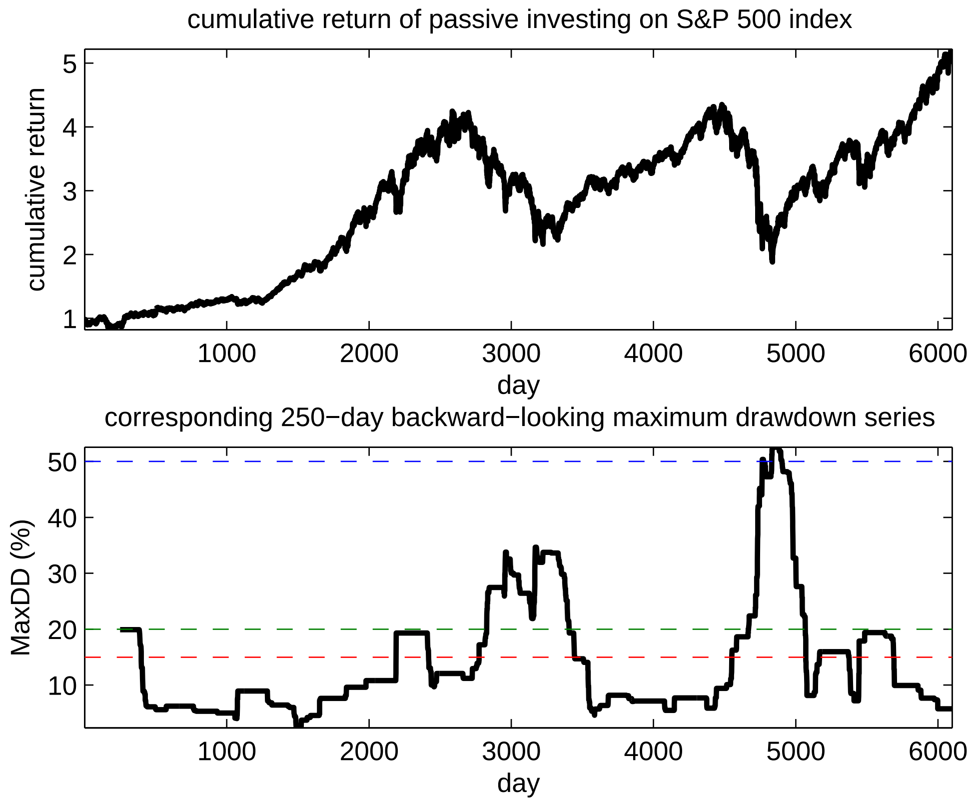


Figure 1.3: Cumulative return and corresponding moving yearly (i.e., 250-day) backward-looking maximum drawdown series of passive investing on daily S&P 500 stock index from 01/02/1990 to 03/14/2014.

ultimate goals of all fund managers. An investment with small drawdowns and low returns is not attractive as the profit is too small; while an investment with large drawdowns and high returns is unrealistic because investors will withdraw all their money when a large drawdown occurs.

1.1 Related Work

Based on the importance of reducing drawdown, researchers make efforts on solving portfolio optimization problem with constraint on drawdown. Grossman and Zhou [GZ93] pioneer the mathematical formulation of portfolio optimization under drawdown constraint. They model the problem as to maximize the long-term growth rate of the expected utility of final wealth based on the constraint that the loss of current wealth is no more than a fixed percentage of the maximum value achieved till current time. The model assume continuously rebalancing between a risk-free asset and a risky asset. And the price process of risky asset follows a geometric Brownian motion with known, fixed drift and volatility parameters, while the risk-free one has a constant return. Based on the assumption, Grossman and Zhou [GZ93] give exact analytical solution by dynamic programming.

Two years later, Cvitanic and Karatzas [CK95] extend the continuous optimal drawdown control problem to multiple risky assets. However, Klass and Nowicki [KN05] point out that a discrete implementation of [GZ93] can result in the loss of optimality. This is caused by the delay effect of discrete data. As a generalization, Cherny et al. [CO13] provide a new approach to solve the Grossman and Zhou [GZ93] model by finding an equivalent unconstraint problem.

All of the above studies focus on finding analytical solution to a well-formulated optimization problem based on strong assumptions about the assets, and the solution contains the drift and volatility parameters. It is well-known that for real stock price path, these parameters cannot be known in advance. Although these parameters can be estimated from past data, future data may have different drift and volatility parameters. A lot of studies corroborate that stocks have significant different behaviors at bull and bear markets.

It can be seen in Figure 1.4 that the daily price process of S&P 500 index has different behaviors among different regimes, where the regimes are marked by red

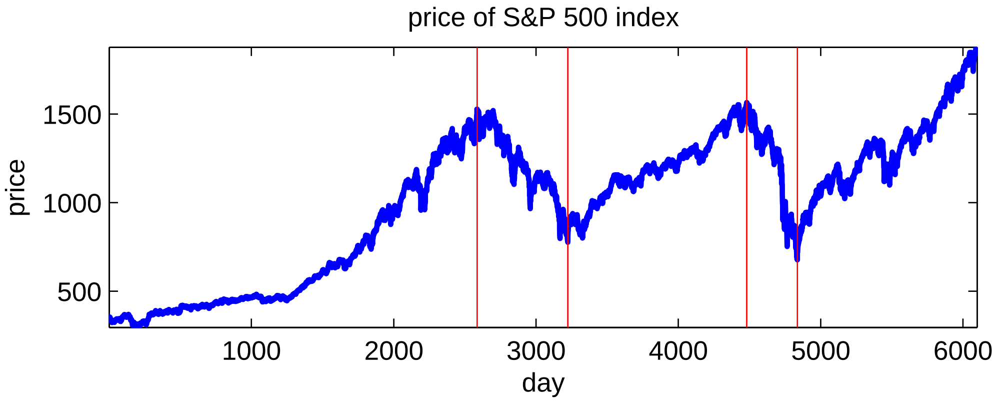


Figure 1.4: Price process of S&P 500 index.

lines. And the regime changing point cannot be known in advance. Considering both the estimation error and the time-varying property of the parameters, the analytical solution cannot be used to real stock data.

Different from Grossman and Zhou model, Chekhlov et al. [CUZ03, CUZ05] propose Conditional Drawdown (CDD), which is a concept similar as Conditional Value-at-risk (CVaR), and solve the optimization problem of maximizing the average return with CDD constraint. In order to solve the problem, some modifications are made. Firstly, uncompounded cumulative return is used, that is, the cumulative return is calculated by addition of return without considering the compounded rate of return. Secondly, the drawdown, which is defined as absolute drawdown (AD), is regarded as the distance between previous maximum and current wealth, i.e., $AD_t = max\{W_\tau\}_{\tau=0}^{t} - W_t$, instead of percentage loss. Based on the simplifications, the optimization problem can be solved by linear programming. Although the simplifications make the problem solvable, the uncompounded return and absolute drawdown are far from reality.

Figure 1.5 shows the cumulative return and corresponding drawdown, absolute drawdown of compounded and uncompounded return of passive investing on S&P 500 index. It can be seen that the cumulative returns, drawdown and absolute drawdown of compounded and uncompounded approaches vary a lot. In real stock investment, it is well known that the returns are compounded. The uncompounded return is not an appropriate approximation of it.

Besides, absolute drawdown is not a reasonable measure for risk. For example,

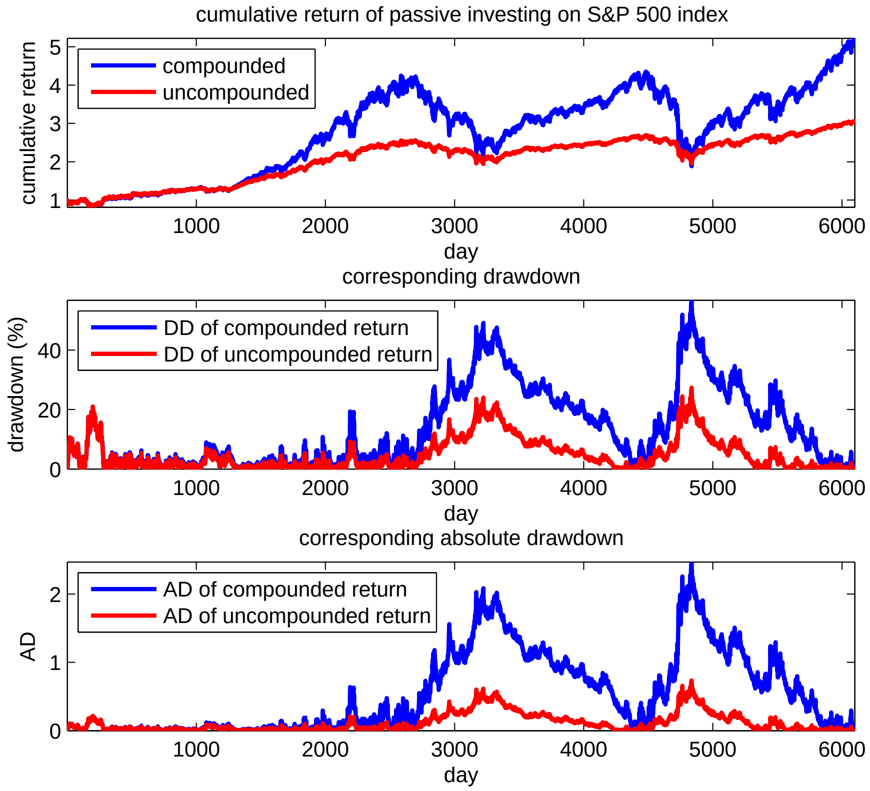


Figure 1.5: Illustration of cumulative return and corresponding drawdown, absolute drawdown (AD) of compounded and uncompounded return.

Table 1.1 shows the drawdown and absolute drawdown of two portfolios. Portfolio A drops from 1 to 0.5, with a 50% drawdown, which means large risk; while portfolio B drops from 100 to 99.5, with drawdown value of 0.5%, which is a small risk. However, the absolute value for the two portfolios are the same; i.e., 0.5. It is absolutely not reasonable in real investment.

Inspired by CDD, in 2014, Goldberg and Mahmoud [GM14] propose a new kind of drawdown measure Conditional Expected Drawdown (CED), which is actually the tail mean of maximum drawdown distribution. Although the proposed CED with good mathematical properties, it still considers absolute drawdown instead of drawdown. It has the same problem as is illustrated in Table 1.1.

To summarize, most previous researchers focus on mathematical formulation

Table 1.1: Example of drawdown and absolute drawdown of two portfolios.

Portfolio	Previous Maximal Wealth	Current Wealth	DD	AD
A	1	0.5	0.5	0.5
B	100	99.5	0.005	0.5

and theoretical analysis of portfolio optimization problem with drawdown constraint. In order to solve the proposed optimization problem, they even sacrifice the practicability by applying assumptions about the distribution of future data which actually cannot be predicted from historical data. As a result, they only provide the mathematical solutions to the proposed optimization problems without giving any drawdown control results on real stock data.

Although investment on single asset can be regarded as an special case of portfolio, due to the strong assumptions in the previous works, no effective approaches have been proposed to reduce drawdown for a single asset in real financial industry. In industry, the commonly used method is to place stop-loss order. A stop-loss order is set when or after buying a stock, and is triggered when the price drops to the specified price. The stop-loss price is usually calculated according to percentage loss, e.g., 20% of price drawdown with inception of buying time. However, the setting of stop loss price is not easy. If stop loss order is used appropriate, more loss can be avoided; or it may cause to sell just before the stock rebounds. Besides, it may cause disaster when stock price drops deeply and recover in short time (known as flash crash).

For example[2], during 2010 flash crash, Proctor & Gamble (PG), a very stable stock with low volatility, went from \$60 to \$40 and back in less than 30 minutes. If one manager had set a stop loss price at \$50, the order is triggered when the price of PG dropped to \$50, and the stock would be sold at the next available price. And the PG price rebounded immediately, which led the manager lose a lot of money and unable to buy the stock back at a cheap price.

To illustrate the performance of investment with stop-loss criteria, we design three investment methods based on different stop-loss criteria; i.e., $DD = 10\%$, $DD = 20\%$ and $DD = 30\%$. That is, we initially hold the asset and sell it when the drawdown reaches 10%, 20% and 30% respectively. After selling the asset, we

[2]http://arborinvestmentplanner.com/stop-loss-orders-investment-tool-value-investors-avoid/

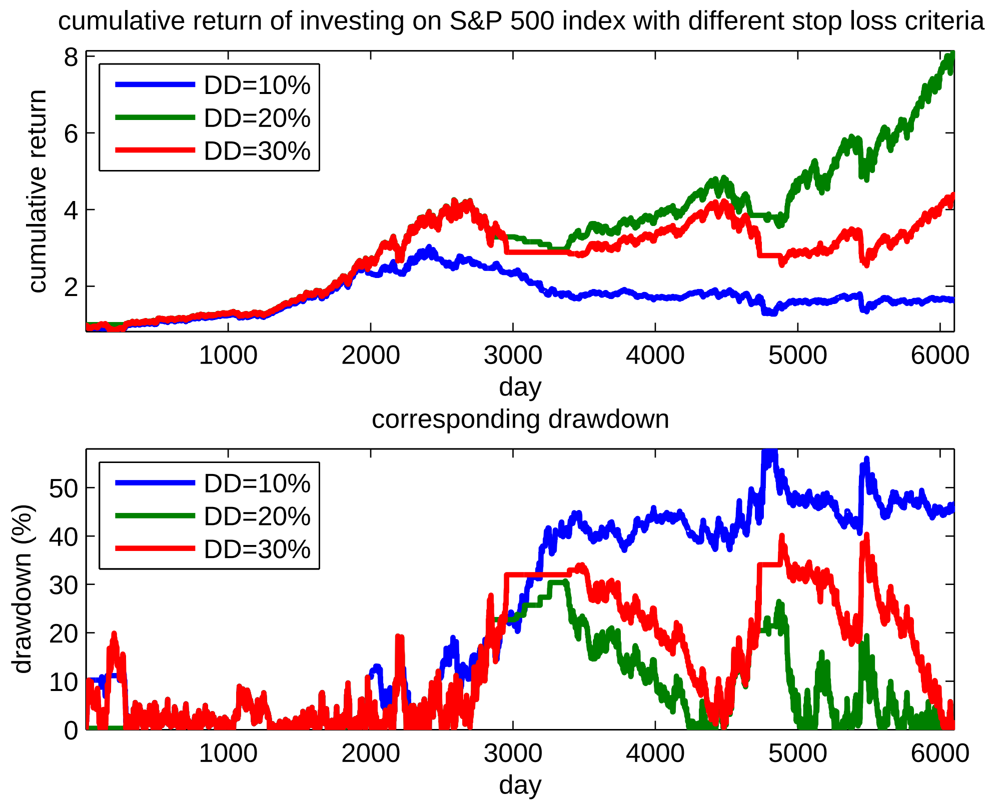


Figure 1.6: Cumulative return of investing on daily S&P 500 index based on three stop loss criteria.

reinvest when the price increases 10%, 20% and 30% respectively.

Figure 1.6 shows the cumulative return of investing on daily S&P 500 index based on the three strategies. As daily adjusted close price is used, there is no flash crash phenomenon in the data. It can be seen that when the stop-loss criterion—DD is too small, the delay is small; however, the cumulative return is terrible; while the delay is large and the cumulative return is bad when DD is too big. Only when the DD is set appropriately, the cumulative return and the delay effect are better, e.g., $DD = 20\%$. However, the corresponding drawdown is still very large, i.e., higher than 30%, during some periods.

Although the commonly used stop-loss criteria can reduce drawdown to some extent for the stock index, the drawdown is still not satisfying during some periods. Facing the problem that the passive investing method and stop-loss investment method on stock index having great drawdowns, in this thesis, we proposed a practical stock index investment approach based on buying and not buying strategies, which leads to both small drawdowns and high cumulative return.

1.2 Analysis

In order to achieve both small drawdowns and high cumulative return on stock index investment, intuitively, two tasks should be solved firstly. One is to predict the return of future window, and the other one is to predict the drawdown of future window.

In order to reduce future drawdown, one naive question is whether we can we find any past patterns that have correlation with future drawdown on the price path. If correlations can be found, the investment strategies can be made accordingly. However, before that, the first problem is to make clear how to assess the future drawdown. Why this problem is important?

To answer this question, an example is given. Figure 1.7 shows the path and the corresponding drawdown of two simulated price paths. How to compare the two paths? Which one is better in terms of drawdown? Should average drawdown or maximum drawdown be used to assess them? Actually, according to the aforementioned concerns, maximum drawdown is the primary one that people care about. It is because when the maximum drawdown exceeds some predefined maximum limit, investors may withdraw their money and, consequently, managers have no money to continue investing. Therefore, Path B is better as it has smaller maximum drawdown.

A 100-day backward-looking maximum drawdown of S&P 500 index is shown in Figure 1.8. It can be seen that successive sliding windows may have the same maximum drawdown. In addition, the maximum drawdown sequence has clustering behavior. There are obvious periods with high maximum drawdown and periods with low maximum drawdown.

Suppose the predicted maximum drawdown is $\widehat{MaxDD}$ and the predicted return is $\widehat{return}$, the best investment decision is illustrated in Table 1.2. That is, only when the predicted $\widehat{MaxDD}$ is small and $\widehat{return}$ is positive, money will be invested in the next window. Otherwise, investing money will lead to high

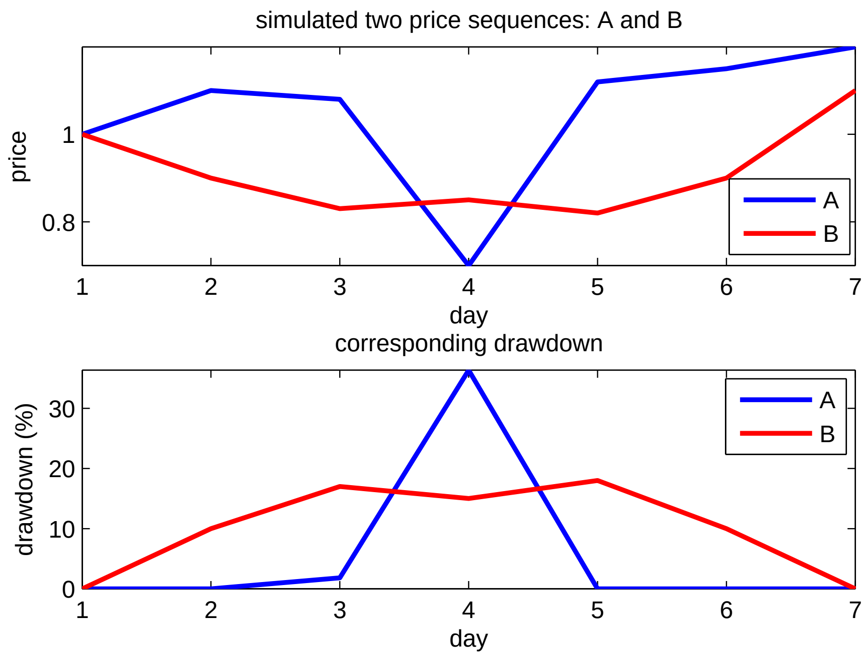


Figure 1.7: Example of assessing drawdown sequence.

drawdown or negative return, which is not desirable by investors.

However, we have the following observations from the data, which will help us to refine this problem. Figure 1.9 shows the moving 50-day return and corresponding 50-day maximum drawdown series of passive investing on S&P 500 index. It shows the positive returns and negative returns are totally mixed together, while most 50-day maximum drawdown values are small, which is less volatile than the return series. Therefore, it is easier to predict the maximum drawdowns than predict the value/sign of returns.

Table 1.2: The best investment decision based on the predicted $\widehat{MaxDD}$ and $\widehat{return}$ pairs.

($\widehat{MaxDD}$,$\widehat{return}$)	Invest
(small, +)	Yes
(small, -)	No
(large, +)	No
(large, -)	No

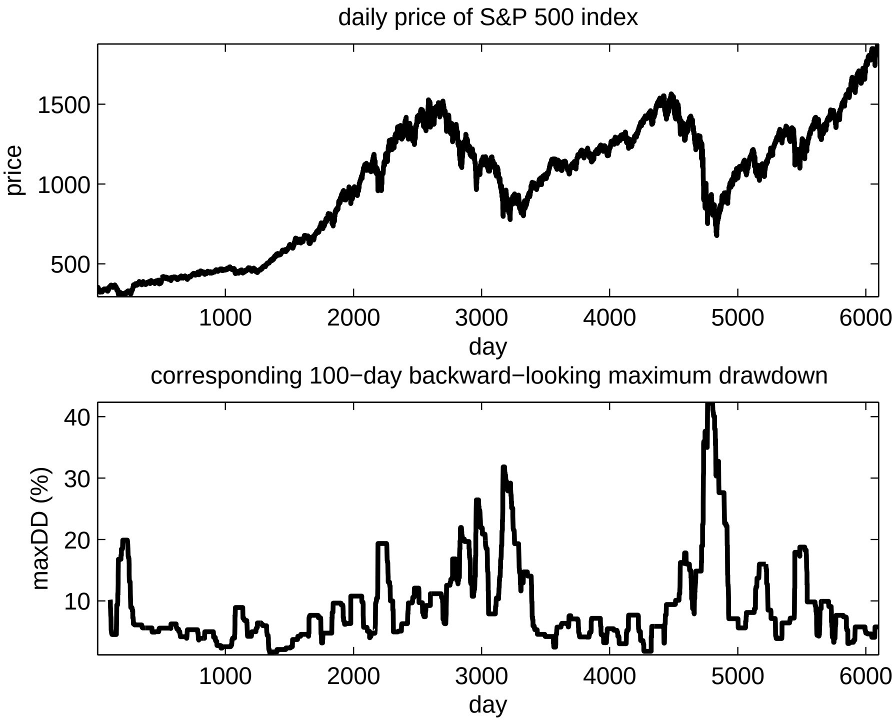


Figure 1.8: Daily price and corresponding 100-day backward-looking maximum drawdown.

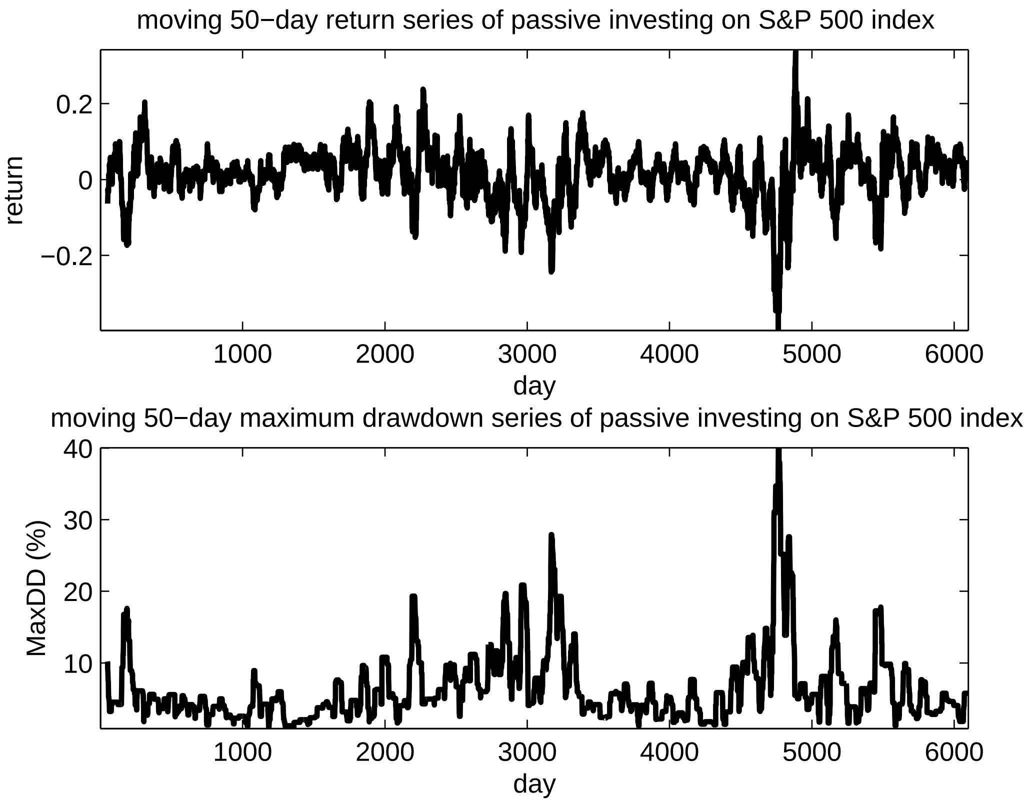


Figure 1.9: Moving 50-day return and corresponding 50-day maximum drawdown series of passive investing on S&P 500 index.

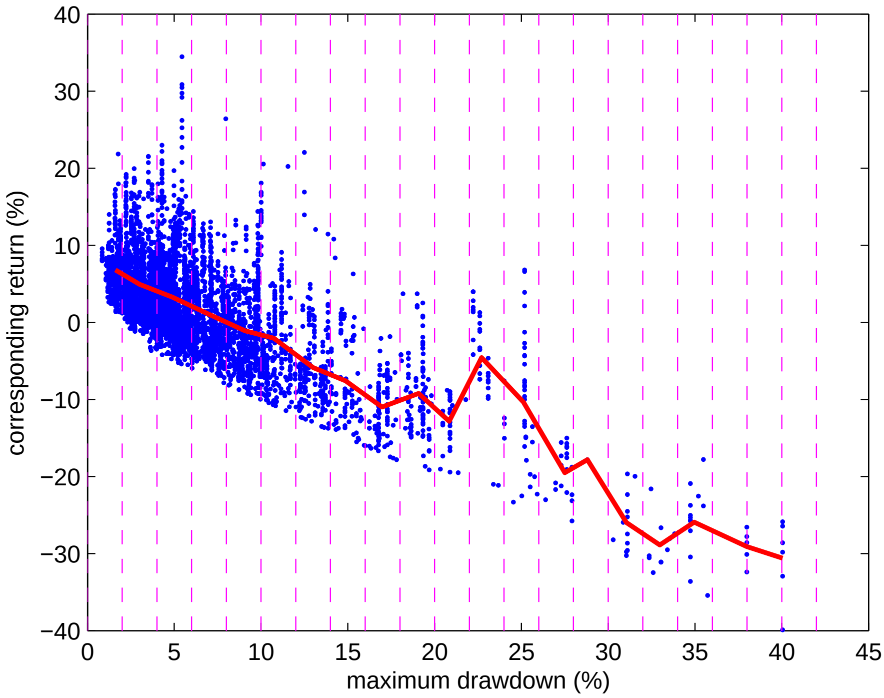


Figure 1.10: Maximum drawdown and corresponding return pairs of all the 50-day sliding windows of passive investing on S&P 500 index.

In addition, the corresponding $(MaxDD, return)$ pairs of Figure 1.9 is presented in Figure 1.10. The piecewise mean of the maximum drawdown and corresponding return pairs is presented by red curve. The length of each piece is two units of maximum drawdown, which is marked by dashed lines. It can be seen that, generally, the maximum drawdown and corresponding return are negatively correlated. Specifically, when maximum drawdown is small, the corresponding return is high; and when maximum drawdown is big, the corresponding return is low. If the $MaxDD$ is known, the mean value of return can be inferred. Therefore, in this thesis, the problem is simplified to one task, that is, to predict the maximum drawdown of future window.

1.3 Observations

Before presenting the proposed investment method, we firstly show some observations of the data. It is crucial because observation is the first step to find potential properties that actually determine the investment method. In order to reduce future drawdowns, basically, the main task is to investigate the correlation between past data and the corresponding future maximum drawdown. And then the most appropriate approach can be designed accordingly.

In this section, we depict several observations that inspire us to find past patterns that have correlation with future maximum drawdown. The observations will be statistically verified in Chapter 2. Daily adjusted close price of S&P 500 index from 1990/01/02 to 2014/03/14 is applied as an example. Similar properties can be found in other stock indices, e.g., NASDAQ and Hang Seng index. It should be noted that, in order to focus on drawdown control, we only consider stock index as a portfolio. Actually the proposed investing approach can be used on other stock portfolios. As portfolio selection is another complicated topic, we do not consider portfolio selection in this thesis.

In the following parts, we will focus on finding past patterns that have correlation with future maximum drawdown.

1.3.1 Past Maximum Drawdown

When trying to find something in past which has correlation with future maximum drawdown, the counterpart in the nearest past is the most straightforward one.

Actually, we can find evidence from Figure 1.8. The maximum drawdown series has clustering behavior. During some period, the maximum drawdown in adjacent windows is a good indicator of the future maximum drawdown. For example, Figure 1.11 shows the price and the corresponding drawdown of two adjacent 100-day windows. The maximum drawdown in the first window is 5.6%, while the maximum drawdown in the second window is also 5.6%.

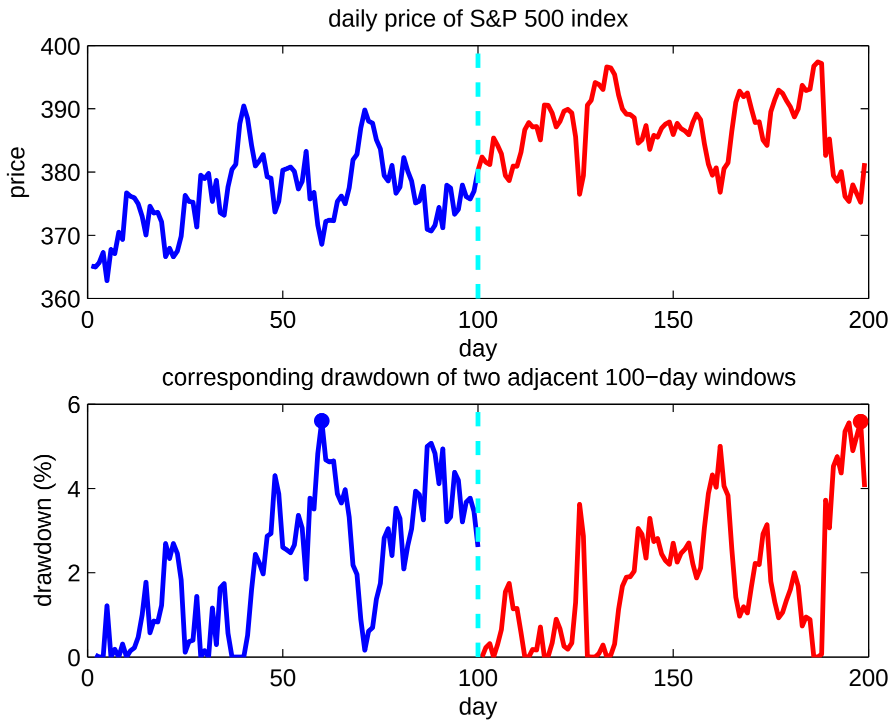


Figure 1.11: Price and corresponding drawdown of two adjacent 100-day windows.

1.3.2 Past Drawdown Series

For some period, the nearest maximum drawdown in the past is a good indicator of future maximum drawdown. However, there are still some periods that the past maximum drawdown cannot represent the future maximum drawdown. One example can be seen in Figure 1.12. The future maximum drawdown is 7.1%, while the past maximum drawdown is 22.6%. For this case, the past maximum drawdown is not a good indicator of future maximum drawdown.

Actually, the drawdown series in past window is known. In last section, only the maximum drawdown in past is considered. However, we can observe that the second separated maximum drawdown (5.3%), marked by red circle, is more approaching with the future maximum drawdown. Here, the second separated maximum drawdown is defined as the maximum drawdown that computed without the maximum drawdown duration. And the maximum drawdown duration is the time it takes from the beginning of the drop to the new high. For example, in Figure 1.12 the maximum drawdown duration in past window is marked by magenta line. Without the maximum drawdown duration, the maximum drawdown is 5.3%, which is named as the second separated maximum drawdown.

Similarly, the third separated maximum drawdown can also be studied. That is, correlation may exist between other statistics of past drawdown series and future maximum drawdown. The statistically analysis of the correlation will be presented in Chapter 2.

Till now, one question may be raised: why do not use the raw past drawdown series directly? The past drawdown series contains abundant information, however, it contains a lot of noise simultaneously. Therefore, we cannot use it directly.

1.3.3 Past Price Pattern

Price pattern analysis plays an important role in technical analysis. And it is widely used to forecast the direction of price. It inspires us to infer that the past price pattern provides useful information to predict future maximum drawdown.

Because the raw price sequence includes a lot of data points and noise, a simplification method is necessary to preprocess the raw price sequence. Piecewise

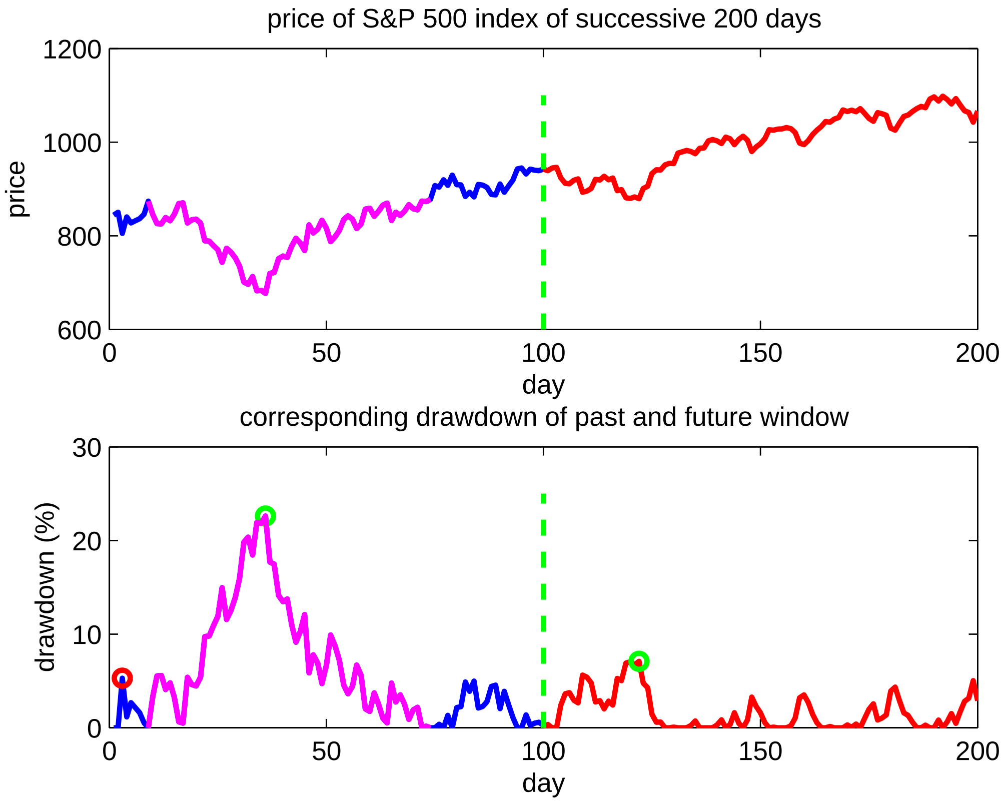


Figure 1.12: Example to show the drawdown sequence of past and future window.

linear approximation is the most commonly used representation for simplification. Figure 1.13 shows examples of 3-line approximation of the past price sequence. The piecewise linear approximation algorithm will be described in Chapter 2. It can be seen that the future maximum drawdown in Figure 1.13(a) is high, while it is small in Figure 1.13(b), and the past price patterns of the two figures vary a lot. Therefore, we infer that the past price pattern provides useful information to discriminate large and small future drawdowns. Again, this inference will be verified in Chapter 2.

1.4 Research Goals

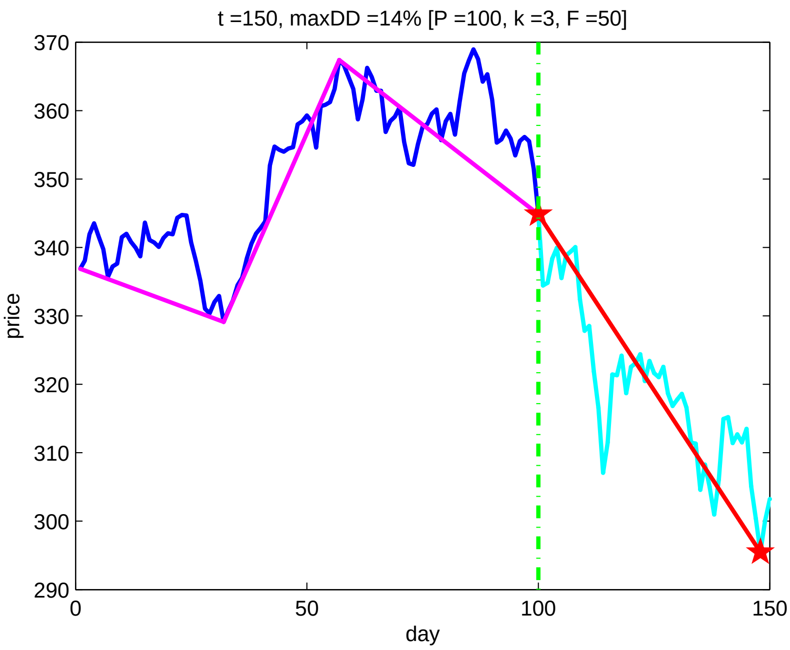
As proposed in Section 1, our overall goal is to design a practical investment approach which leads to both small drawdowns and high cumulative return. To achieve the overall goal, we need to predict the future maximum drawdown firstly and then make investment strategy accordingly. To predict future maximum drawdown, we assume the data series is stable, and the future data is an i.i.d. sampling of the training data, which is also the commonly used assumptions of all the time series prediction problems. In the thesis, we use this assumption to generate the decision points.

Specifically, we consider the following subgoals:

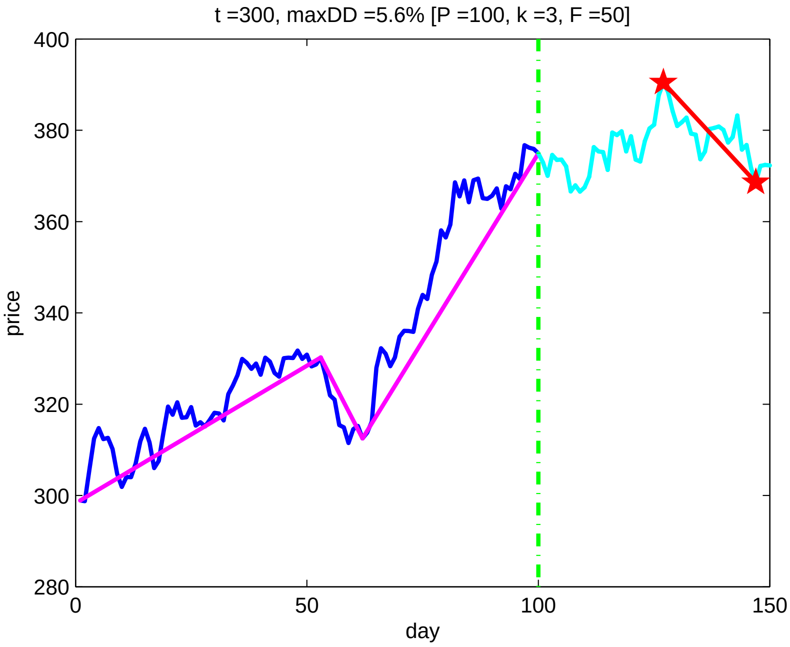
- find past patterns that have correlation with future maximum drawdown.
- use powerful machine learning method to study the potential correlation between past pattern and corresponding future maximum drawdown.
- based on the found correlations and investors' preference, make appropriate investment decisions.

1.5 Thesis Outline

This thesis is organized as follows: Chapter 2 describes the proposed pattern-based drawdown reduction method. Chapter 3 presents the experimental results on real



(a)



(b)

Figure 1.13: Examples of past price pattern and corresponding future maximum drawdown, where past window length (P) is 100 and the future window length (F) is 50.

stock index data. Three different stock indices are considered, i.e., S&P 500, NASDAQ, and Hang Seng index. Besides, the influence of changing parameter values are also studied in detail. And, comparisons are made to show the effectiveness of the proposed method. Chapter 4 summarizes the proposed method and discusses about the future work.

Chapter 2

Pattern-Based Drawdown Reduction Method

This thesis proposes a pattern-based investment model to reduce future drawdowns. To avoid noise and keep useful information of past adjacent window, we extract two kinds of features, named as drawdown feature (*ddFeature*) and price feature (*priceFeature*). Besides, the combined feature *comFeature* which is actually the joint feature of (*ddFeature, priceFeature*) is also considered. Support vector machine (SVM), which is corroborated to be a promising method for the prediction of financial time series, is applied to study the potential correlation between extracted feature and corresponding future maximum drawdown. The trained SVM using data in learning window will be applied to predict future maximum drawdowns. And suitable strategy will be taken accordingly to avoid future high drawdowns.

In this section, we describe the feature extraction process and present statistically demonstration of the correlation between the extracted feature and future maximum drawdown. We then depict the machine learning algorithm to solve this problem. Based on the prediction result, the investment decision strategy is presented accordingly. In addition, the implementation of the investment method is also described.

Table 2.1: The pseudocode for *ddFeature* extraction algorithm.

***ddFeature* extraction algorithm**
1. Compute the corresponding drawdown series;
2. $i = 0$, $ddFeature = NULL$;
3. Find the maximum drawdown $MaxDD$ and the corresponding maximum drawdown duration;
4. $i++$, $ddFeature = MaxDD$;
5. Eliminate the maximum drawdown duration from the drawdown series;
6. If no data remains, go to Step 10;
7. Find the maximum drawdown $MaxDD$ from the remaining drawdown series and the corresponding maximum drawdown duration;
8. $i++$, $ddFeature = \{ddFeature, MaxDD\}$;
9. If $i == n$, the *ddFeature* extraction algorithm is finished; else go to Step 5.
10. $n-i$ 0s will be added to *ddFeature* to keep the length of the extracted feature is n.

2.1 Feature Extraction

In order to use historical data to make good inference about future maximum drawdown, feature extraction from historical data is the first step. Based on the observations in Section 2 of Chapter 1, we extract the following features from the nearest past window.

2.1.1 *ddFeature*

Drawdown feature is the first n separated maximum drawdowns extracted from the past window. Consequently, there are n features extracted, which are denoted by *ddFeature*. Given the raw price series, the algorithm to extract *ddFeature* is shown in Table 2.1.

Figure 2.1 shows an example of the drawdown feature *ddFeature* with $n = 3$.

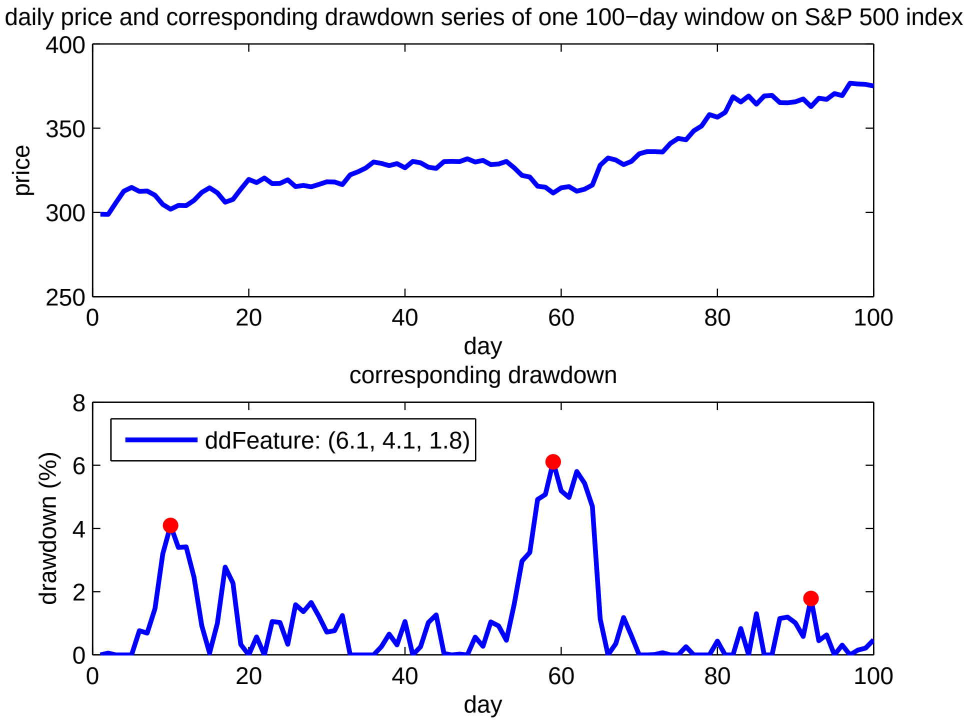


Figure 2.1: Example of extracting *ddFeature* from one 100-day window of S&P 500 index. The three red dots are the extracted first three separated maximum drawdowns.

The extracted first three separated maximum drawdowns are marked by red dots.

The dynamic backward-looking correlation coefficient of 1000 sliding 100-day past and future maximum drawdown pairs is illustrated in Figure 2.2. It can be seen that the correlation is varying with time. And the correlation is significant for most of the period, which can be seen from the p-value series where 78% of the p-value series is less than 0.05. Therefore, the past maximum drawdown is an effective indicator of future maximum drawdown for most of the periods.

To show the effectiveness of all the extracted features, we present the dynamic correlation between each feature and future maximum drawdown. Suppose $n = 3$, the 3 dynamic backward-looking correlation coefficient series of 1000 sliding windows is presented in Figure 2.3. All the correlation changes with time. It is glad to see that during some period that one feature has weak correlation with future maximum drawdown, some other features have high correlation, e.g., the period

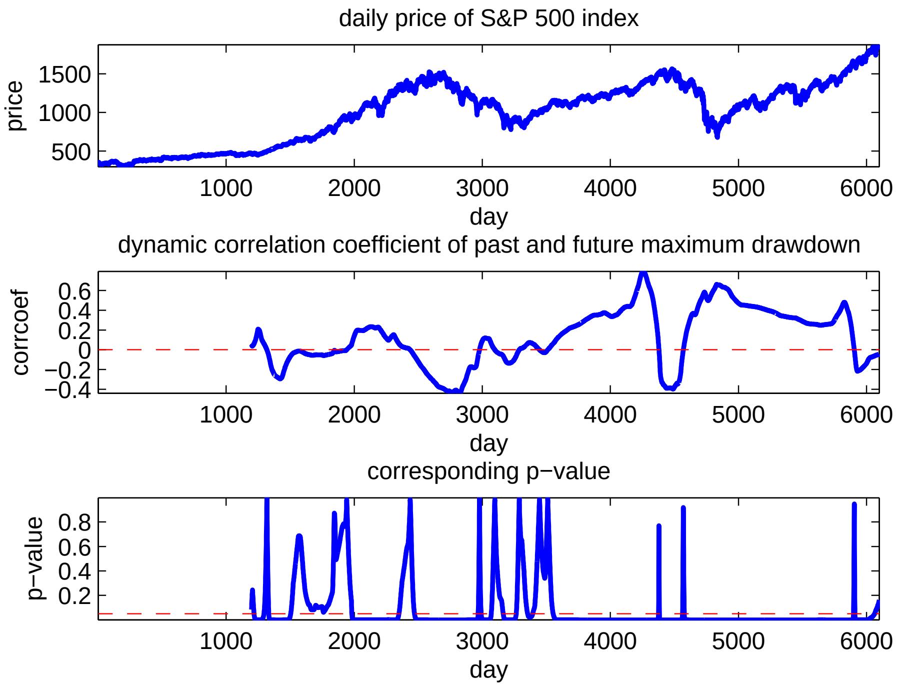


Figure 2.2: Dynamic backward-looking correlation coefficient of 1000 sliding past and future maximum drawdown pairs, where past window length $P = 100$ and future window length $F = 100$.

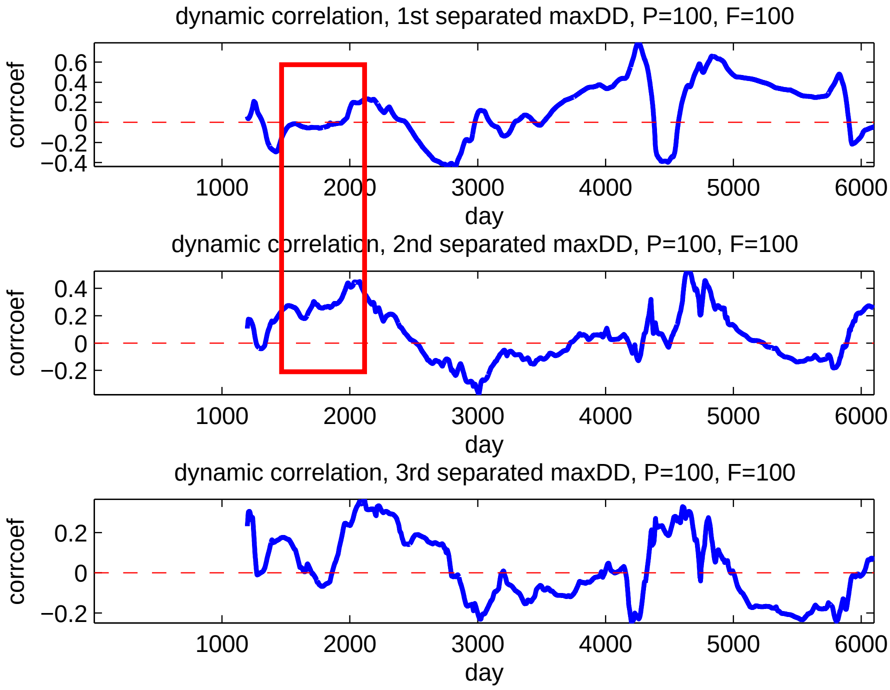


Figure 2.3: 3 dynamic backward-looking correlation coefficient series of 1000 sliding windows on S&P 500 index dataset with the same past and future window length $P = F = 100$.

marked by red rectangle, which provides possibility that with carefully selection of the features, the selected candidates may exhibit high correlation through the whole period. It is verified by Figure 2.4. Figure 2.4 plots the dynamic correlation coefficient series which is composed by the maximum of the 3 correlation coefficients in Figure 2.3 for each time slot. It appears that more than 99% of the p-values are less than 0.05, which shows that by selecting the best feature, much higher correlation can be found through the whole period. In addition, it should be noted that the correlation coefficient in Figure 2.4 is actually the low bound because only one best feature is considered. If the combination of different features is considered, the correlation will be not worse than Figure 2.4.

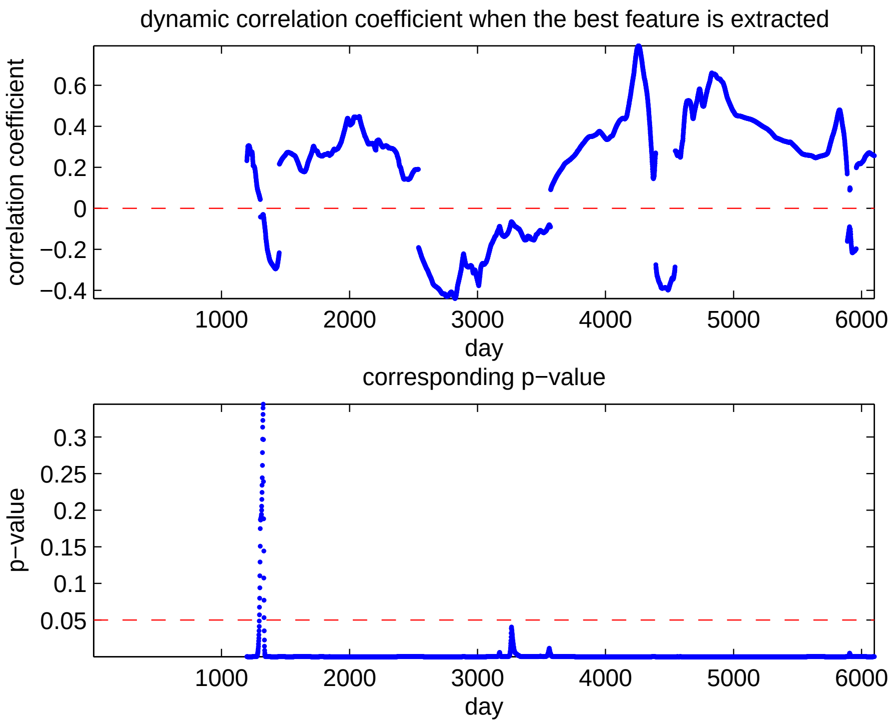


Figure 2.4: Dynamic correlation coefficient series and corresponding p-value series when the best feature in Figure 2.3 is applied for each time slot.

Table 2.2: The pseudocode for *priceFeature* extraction algorithm.

***priceFeature* extraction algorithm**

1. Put the data at the $(2i-1)$th and $2i$th time points into one segment, thus the original price series is divided into $n/2$ segments.
2. Calculate the cost of merging two adjacent segments, namely merging cost. Here we define the merging cost as the sum of the squared distance between the price data within this segment and the corresponding line.
3. Merge the two adjacent segments having the smallest merging cost together.
4. If the current segment number is greater than k, go to Step 2, otherwise go to Step 5.
5. Connect the neighbor segments in a head-to-head manner, that is, the head of the former line is connected with the head of the latter line, forming a piecewise k-line segments.

2.1.2 *priceFeature*

Features extracted from piecewise linear approximation of the raw price series in past adjacent window is called price feature, which is denoted as *priceFeature*. The bottom-up segmentation algorithm [KCHP01] is one popular time series segmentation algorithms. In this thesis, we choose the bottom-up approach to conduct the time series segmentation. To keep the extracted features with same dimension, we set the same number of segments. Suppose the price series contains n data points and the expected segments is k, the steps is shown in Table 2.2.

The segments k plays an important role. Actually, the approximation error is increasing with k. If k is large, in extreme case, $k = n$, the approximation error is 0. However, this representation has many noises. To eliminate noise in the original price series, we need to use few lines to get the sketch of the original series and the corresponding approximation error will increase. Here, approximation error is not the main concern. We expect the approximation has small noise. The piecewise linear approximation effectively decreases the noise in the original price series.

When the approximation is obtained, we use two values to represent each line, i.e., the length of the line and the return $r_L = (\frac{P_T}{P_S} - 1)$, where S and T are the

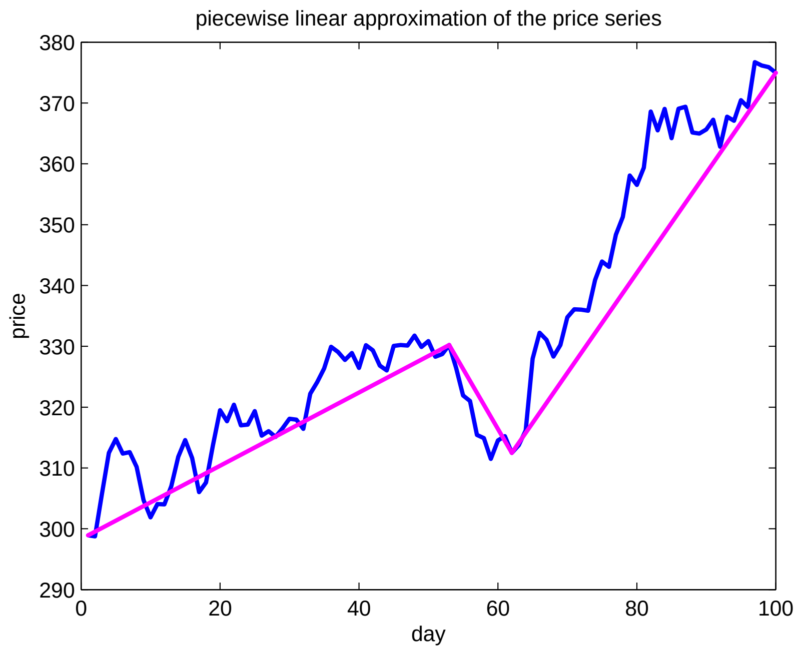


Figure 2.5: Piecewise linear approximation of the price series with three segments.

starting and terminal point of the line. Therefore, $2k$ features can be extracted from the linear segments. Here, we denote the $2k$ features as *priceFeature*. Figure 2.5 shows 3-line approximation of the price series. Correspondingly, 6 values are computed to represent the three lines, i.e., $(52, 0.105, 9, -0.054, 38, 0.20)$.

To show the correlation between the *priceFeature* and future maximum drawdown, Figure 2.6 presents the dynamic correlation coefficient of each dimension of *priceFeature* and future maximum drawdown, where the *priceFeature* is extracted using 3 segments. The correlation is not as high as *ddFeature*, however, the correlation is still significant for most of the period, which can be seen from the corresponding p-value series in Figure 2.7.

2.1.3 *comFeature*

ddFeature is a feature extracted from drawdown series, while *priceFeature* is extracted from the raw price data. They contain different information.

To better utilize both of the two kinds of features, the combined feature *com-*

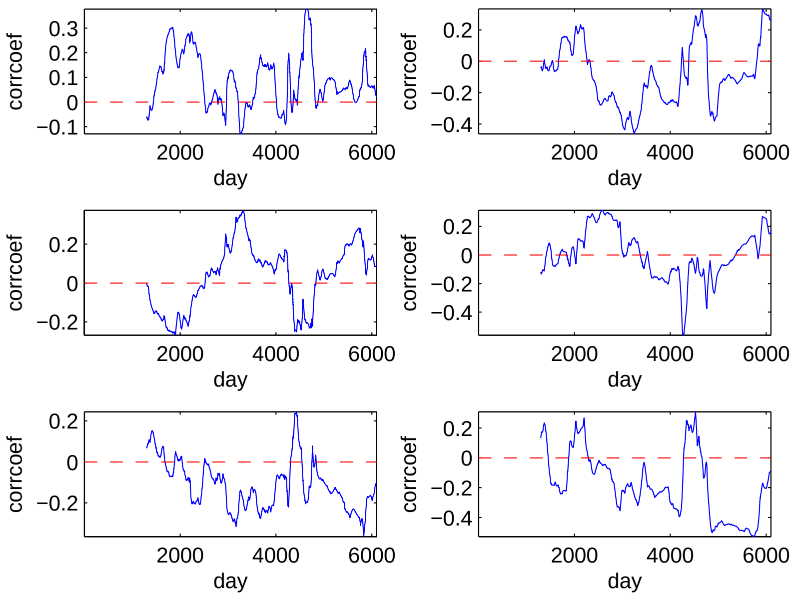


Figure 2.6: The dynamic backward-looking correlation coefficient series of each dimension of *priceFeature* and future maximum drawdown with 3 segments. Each subfigure shows the dynamic correlation coefficient series of one dimension of *priceFeature* and future maximum drawdown. The correlation coefficient is computed based on 1000 sliding windows.

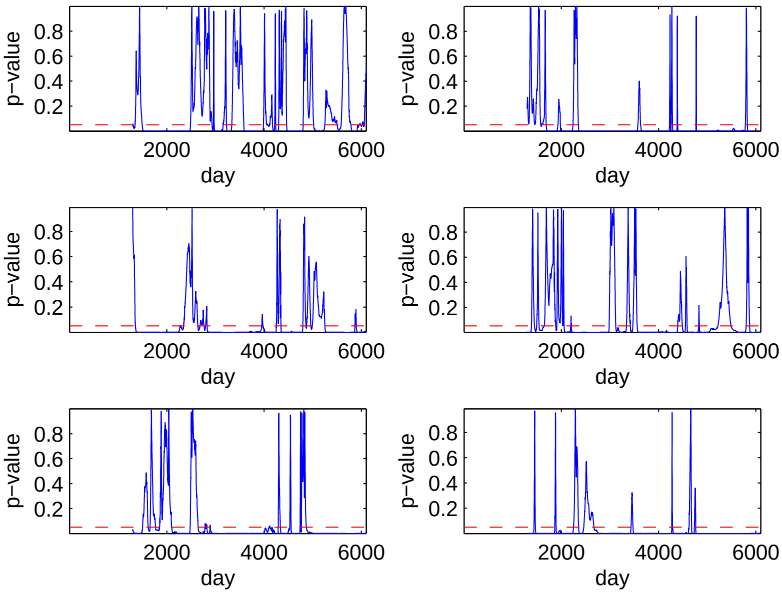


Figure 2.7: The corresponding p-value series of the correlation coefficient in Figure 2.6. Each subfigure shows the p-value of using one dimension of *priceFeature*.

Feature, which is the joint feature of *ddFeature* and *priceFeature* (*ddFeature, priceFeature*), is also considered in this thesis.

2.2 Learning Method

Once the features in adjacent window are obtained, measures should be taken to discriminate different future maximum drawdowns according to their extracted adjacent features. Therefore, to best mine the nonlinear correlation between the extracted features and corresponding future maximum drawdowns, one strong machine learning algorithm is needed.

Support vector machine (SVM), a novel and powerful learning algorithm, was proposed by Vapnik and his colleagues in 1995 [Vap95, VV98]. Although it is a specific neural network algorithm, it has some unique merits that the traditional neural networks do not have. Specifically, neural networks implement the empirical risk minimization principle by minimizing the empirical error which is prone to have over-fitting problem. In contrast, SVM is trained based on the structural risk minimization principle which seeks to minimize an upper bound of the generalization error which is demonstrated to be very resistant to over-fitting problem. In addition, training SVM is a uniquely solvable linearly constrained quadratic programming problem, so that the solution is globally optimal; however, training neural networks requires nonlinear optimization which has the danger of getting stuck at local minima. Due to the excellent properties, SVM have been widely used to solve many machine learning problems, such as pattern recognition and time series prediction.

Recently, a lot of applications of SVM to financial forecasting problems have been presented[TC01, CT01, Kim03, HNW05]. And Sapankevych and Sankar [SS09] summarizes the application of SVM to time series prediction, including financial data. Based on real financial data, it is demonstrated that SVM outperforms traditional neural networks and many other algorithms.

Based on the successful application of SVM in financial forecasting, we use SVM as the learning algorithm. Specifically, support vector regression (SVR), which aims to solve the general regression problem, is applied. It depends on our goal of forecasting the future maximum drawdown based on extracted features. The maximum drawdown is continuous, therefore, the regression model is more suitable and flexible than support vector classification (SVC), which is another type of SVM.

In the following, the basic theory of ϵ-SVR will be presented. And the LIBSVM implementation [CL11] is applied in this thesis.

SVR uses linear model to implement nonlinear regression through some nonlinear mapping of the original feature vector into a high-dimensional feature space.

Consider a set of training data $\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_l, y_l)\}$, where $\boldsymbol{x}_i \subseteq R^n$ is a feature vector and $y_i \subseteq R^1$ is the target output. the standard form of SVR [VV98] is

$$\begin{aligned}
\min_{\boldsymbol{w},b,\boldsymbol{\xi},\xi*} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{l}\xi_i + C\sum_{i=1}^{l}\xi_i^* \\
s.t. \quad & \boldsymbol{w}^T\phi(\boldsymbol{x}_i) + b - y_i \leq \epsilon + \xi_i \\
& y_i - \boldsymbol{w}^T\phi(\boldsymbol{x}_i) - b \leq \epsilon + \xi_i^* \\
& \xi_i, \xi_i^* \geq 0,\ i = 1, \ldots, l
\end{aligned} \tag{2.1}$$

where $\phi(\boldsymbol{x}_i)$ maps $\boldsymbol{x}_i$ into a higher-dimensional space and $C > 0$, $\epsilon > 0$. This problem can be solved by the dual problem, and the solution can be represented as follows:

$$y = \sum_{i=}^{l}(-\alpha_i + \alpha_i^*)K(\boldsymbol{x}_i, \boldsymbol{x}) + b \tag{2.2}$$

where $K(*)$ is the kernel function, α_i, α_i^* and b are parameters. Use Eq. 2.2, the y value can be predicted.

In this thesis, $\boldsymbol{x}$ is the extracted feature vector, i.e., *ddFeature*, *priceFeature* and *comFeature*, and y is the expected maximum drawdown.

2.3 Investment Strategy

When the future maximum drawdown is forecasted, appropriate investment decision should be taken accordingly. Here, we consider 'no-short' strategy. That is, if the predicted future maximum drawdown is large, no money will be invested for the next F days; while if the predicted future maximum drawdown is small, we will buy and hold the index for the next F days. That is, the same decision is executed for the F-day window. If all the money are invested in one day, it means, the next investment decision can be made till F days later. It is not flexible. To avoid this situation, we use an aggregation investment approach. Specifically, before investing, we firstly split the original wealth $W_0 = 1$ to F parts. For each of the first F days, use only one part of the wealth, i.e., $1/F$. And for all the other following days, the cumulative return of one part can be computed just before the day, and this part of cumulative return becomes available and will be used for the next F-days window.

Then, we can use the following binary investment decision function $f(\widehat{MaxDD})$ as follows:

$$f(\widehat{MaxDD}) = \begin{cases} 1 & \widehat{MaxDD} \leq \delta \\ 0 & \widehat{MaxDD} > \delta \end{cases} \tag{2.3}$$

where $\widehat{MaxDD}$ is the predicted future maximum drawdown, and δ is the threshold to decide whether to invest or not. Here, 1 means investing all available capital for the next period; while 0 means no money will be invested for the next period.

To achieve small drawdowns, the value of δ should be as small as possible. In the extreme case, by setting $\delta = 0$, no money will be invested for the whole period and, consequently, the maximum drawdown of the investment is zero. However, the profit is also zero. Such a δ value cannot guarantee both high cumulative return and small drawdowns. In contrast, if δ is too large, e.g., $\delta = 100\%$, the generated cumulative is actually the same as the original passive investing result.

Actually, the maximum drawdown of future window affects the maximum drawdown of the whole investment lifetime; while the return during the future window determines the cumulative return.

As mentioned in Section 1.2, the maximum drawdown and corresponding return are negatively correlated in general, which is shown in Figure 2.8. In Figure 2.8, the piecewise mean of the maximum drawdown and corresponding return pairs is presented by the red curve. The length of each piece is two units of maximum drawdown, which is marked by dashed lines.

To obtain as high cumulative return as possible, the chosen δ value should select as many investing windows with positive returns as possible and select as few investing windows with negative returns as possible. Basically, the chosen δ value should lie in the pieces that the mean returns are positive, i.e., the value of δ should not be too large. Otherwise, if δ value is too large, the cumulative return will decrease.

Considering both of the two goals, small drawdowns and high cumulative return, the suitable δ should be in the range of 0 to 7 for Figure 2.8. To achieve as high cumulative return as possible, the fixed maximum value of suitable δ is applied to conduct the experiment results with *ddFeature*, *priceFeature* and *com-*

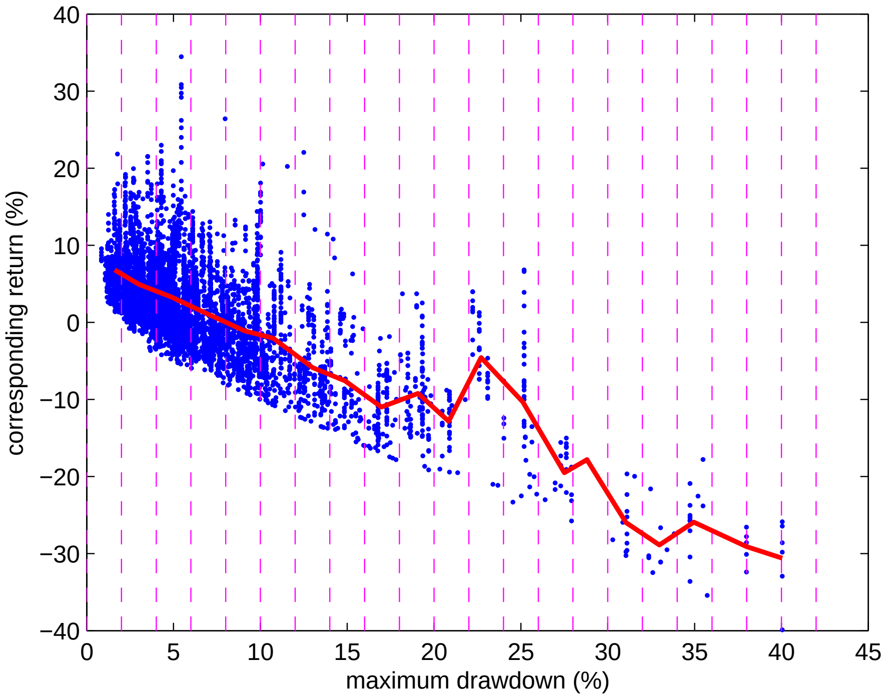


Figure 2.8: Maximum drawdown and corresponding return pairs $(MaxDD, return)$ of all the 50-day sliding windows of passive investing on S&P 500 index dataset.

Feature in Section 3.

In addition, the investment strategy with adaptive δ value is also studied, where the adaptive δ value is computed by finding the maximum suitable δ value in nearest past training window. As the original price series varying with time, time-varying δ is more reasonable.

2.4 Implementation

Based on the extracted features, learning method, and investment strategy, we present our investment procedure in this section. To maintain the latest training data, a sliding window method is applied. The procedure will repeat till the end of the dataset. To facilitate description, the illustration of one decision procedure is shown in Figure 2.9. The whole procedure of pattern-based investment method

Table 2.3: The pseudocode for the whole pattern-based investment procedure.

Pattern-based investment algorithm

1. $t = N + P + F - 1$, i.e., the first day to make an decision.
2. For t, find the corresponding learning window. The learning window contains the nearest N past and future window pairs with respect to different reference time, which are shown by the blue and yellow windows in part "Learning Window" of Figure 2.9.
3. Compute the N feature and maximum drawdown pairs, denote (*feature,MaxDD*). Here, the feature can be *ddFeature*, *priceFeature* and *comFeature*. The N pairs compose the training set.
4. Train the SVM use the training set.
5. Apply the trained SVM to forecast the maximum drawdown in future window, i.e., the yellow window in part "Prediction" in Figure 2.9, using the current feature extracted from the most nearest adjacent window, i.e., the blue window in part "Prediction" in Figure 2.9.
6. Compare the predicted maximum drawdown in future window and the preset investment threshold δ, and make investment decision accordingly for the next F days.
7. If reach the last decision window of the dataset, the investment process is finished; else $t = t + 1$, go to Step 2.

is presented in Table 2.3.

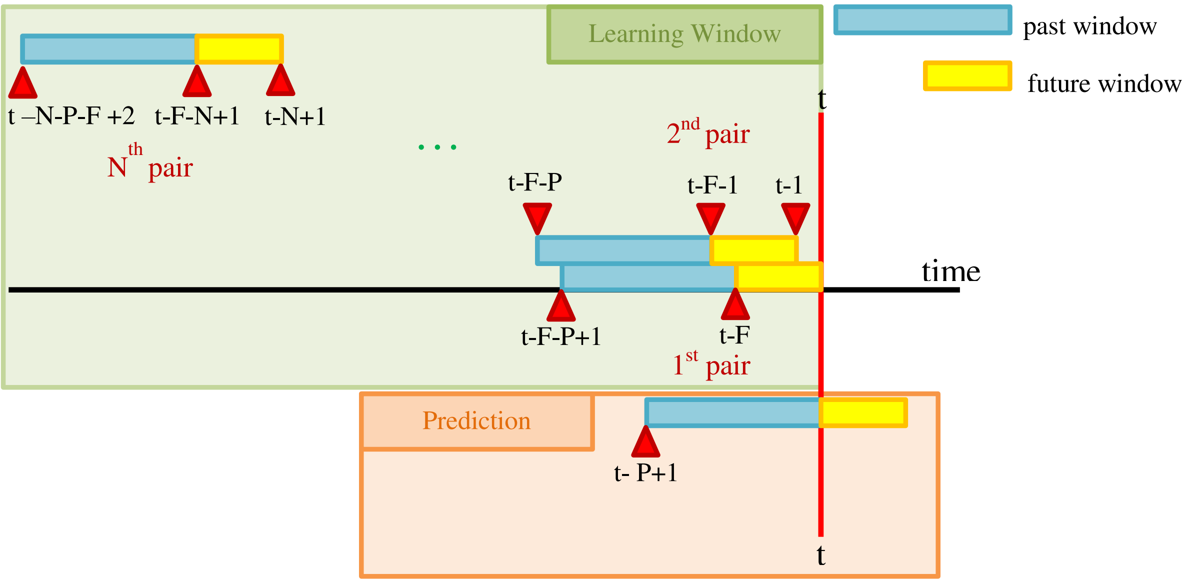


Figure 2.9: Illustration of the investment procedure. Suppose the window length for past window is P and the window length for future window is F.

Chapter 3

Experimental Results

In this section, we describe our experimental results on how to use the extracted patterns to reduce potential future drawdowns. The effects of changing the parameters are also presented. To better catch the dynamic behavior, adaptive investment is used to catch the most recent information. And we also compare the proposed method with risk-free investment and moving average investment. In addition, we discuss the effectiveness and limitations of the proposed pattern-based drawdown reduction approach.

3.1 Experiments

The experiments are designed to examine the performance of extracted features and investment strategy with different parameters. A lot of parameters needs specified for feature extraction and investment function establishment. For example, the past window size P, the future window size F, and the number of training samples N for feature extraction. In addition, for *ddFeature* and *priceFeature*, they have personalized parameters, i.e., n and k. For investment function, the threshold δ should also be considered. The effects of parameter changes will be studied and analyzed. Based on our analysis, adaptive parameter investment method is also presented. At last, we make comparisons and discussions.

In this thesis, we consider real historical daily prices in stock market. In general, we employ three stock indices datasets, which are summarized in Table

Table 3.1: Summary of the three stock indices datasets in our experiments.

Dataset	Region	Time Frame	# Trading Days
S&P 500	US	Jan. 2nd 1990-Mar. 14 2014	6099
NASDAQ Composite	US	Jan. 2nd 1990-Mar. 14 2014	6099
HSI	HK	Jan. 2nd 1990-Mar. 14 2014	6027

3.1. The first one is the S&P 500 index from January 2, 1990 to March 14, 2014. Short for the Standard & Poor's 500, S&P 500 is a stock market index based on the market capitalizations of 500 large companies listed on the NYSE or NASDAQ. It is one of the most commonly followed stock indices and is regarded as one of the best representation of the U.S. stock market. Due to the dot-com bubble in 2000-2001 and the subprime mortgage crisis in 2008-2009, the drawdowns of S&P 500 index during these two periods can be as large as 50%.

To avoid bias from data snooping, we use two additional indices, NASDAQ Composite index and HSI from January 2, 1990 to March 14, 2014. The NASDAQ Composite is a stock market index including more than 3,000 components of the NASDAQ stock market. Since both U.S. and non-U.S. companies are considered, it is not exclusively a U.S. index. Meanwhile, it is well-recognized as an indicator of the technology companies and growth companies. HSI, short for Hang Seng Index, is a stock market index in Hong Kong. The 50 constituent stocks represent about 60% of capitalisation of the Hong Kong Stock Exchange. Since HSI is a market index in Asia, it has much different drawdown behavior. However, both of them have great drawdowns (Figure 3.1). The historical adjusted close prices of the three datasets are downloaded from Yahoo! Finance.

The performance is measured by cumulative return ($CumRet$), maximum drawdown ($MaxDD$) and average drawdown ($AvDD$).

Cumulative return is one of the standard criteria to evaluate the performance of an investing strategy. It is actually the wealth achieved till the end of the whole trading period by the strategy. In this thesis, we simply set the initial wealth $W_0 = 1$ and consequently W_n is the cumulative return at the end of the nth trading day. For the same risk level, the higher the value of the cumulative wealth, the more preferable performance is.

To measure risk, maximum drawdown is an excellent way, which is commonly

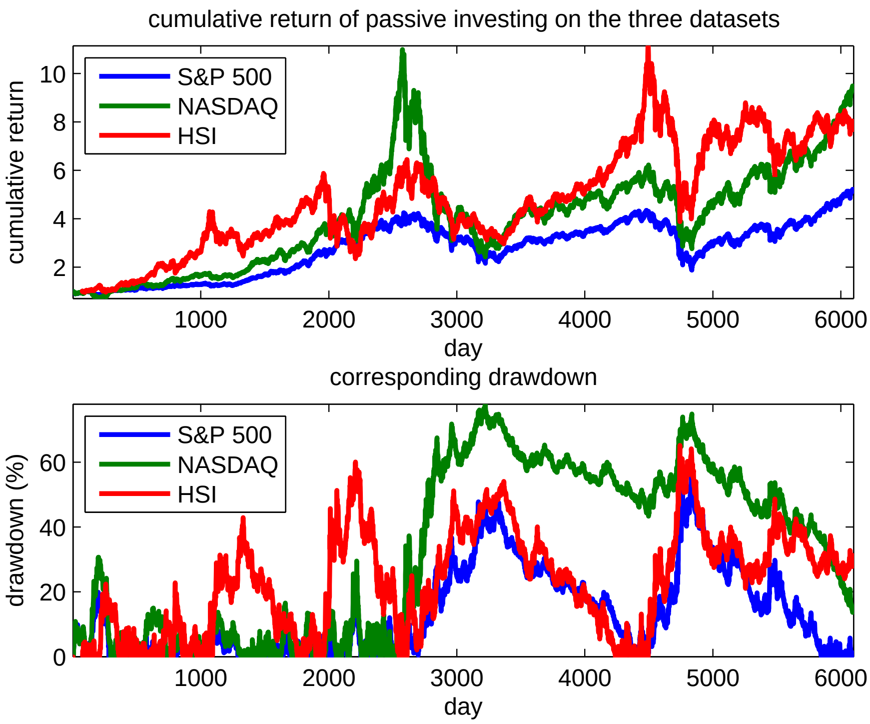


Figure 3.1: Cumulative return and corresponding drawdown series of passive investing on the S&P 500, NASDAQ Composite and HSI datasets.

Table 3.2: The performance of passive investing on S&P 500, NASDAQ Composite and HSI datasets, respectively.

Dataset	$CumRet$	$MaxDD(\%)$	$AvDD(\%)$
S&P 500	5.1	56.8	13.1
NASDAQ Composite	9.2	77.9	32.6
HSI	7.6	65.2	23

used in investment community. Specifically, maximum drawdown measures the downside risk, which is exactly the risk investors cares about. Given two wealth series with similar cumulative wealth, the smaller the value of the maximum drawdown, the more preferable the one is. In addition, average drawdown, which is defined as the mean of the drawdown series, is recorded. Maximum drawdown describes the maximum downside risk, while average drawdown presents the downside risk in average level. Given similar level of cumulative return and maximum drawdown, the wealth series with small average drawdown is more preferable.

As cumulative return and drawdown are two different dimensions considered by investors, it is not reasonable to combine them to generate one unique metric. Investors may have different preferences on the return and drawdown. If there are two strategies: one generates relatively high cumulative return and high maximum drawdown, and the other generates relatively low cumulative return and low maximum drawdown, we cannot tell which one is better because they are not comparable. Therefore, in this thesis, we present the results of cumulative return and drawdown values respectively.

Besides, the moving yearly maximum drawdown is also presented. This is because yearly maximum drawdown is an important indicator to measure the performance of fund managers.

The cumulative return, maximum drawdown and average drawdown values of passive investing on the three datasets are shown in Table 3.2. And the moving yearly maximum drawdown is shown in Figure 3.2. Even for one year time duration, the maximum drawdowns of all the three datasets are higher than 50%. It means more than half of the wealth will disappear of passive investing, which is disaster for investors. Hence, measures should be taken to prevent these great drawdowns.

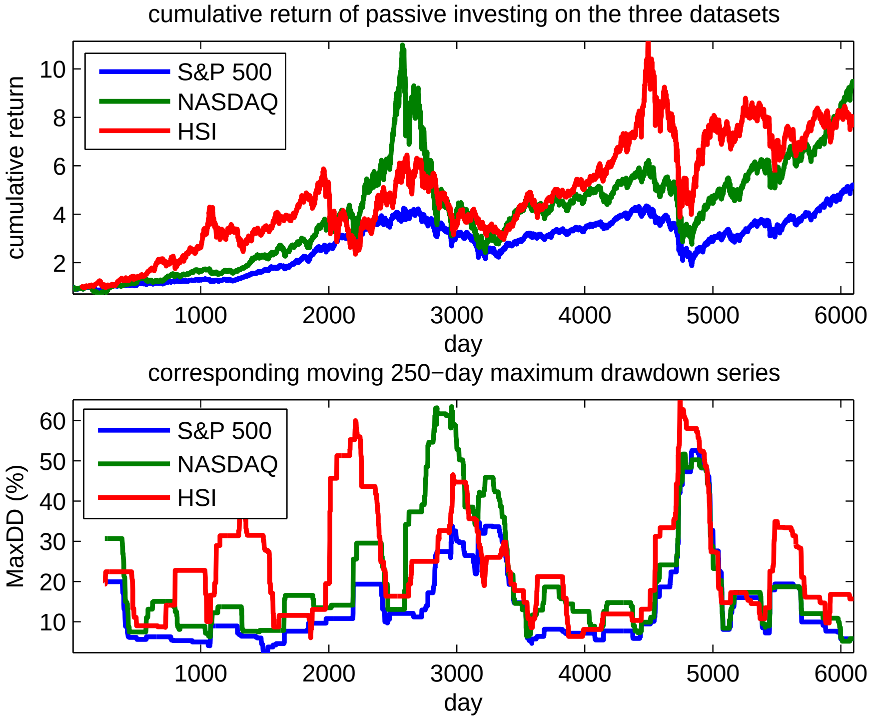


Figure 3.2: Cumulative return and corresponding moving yearly maximum drawdown series of passive investing on the S&P 500, NASDAQ Composite and HSI datasets.

In this thesis, all computations of the experiment are performed in MATLAB. And, as the LIBSVM [CL11] is used to implement ϵ-SVR, we set the parameters '-s 3 -t 2 -d 3 -p 0.1 -e 0.001 -h 1'. '-s 3' means ϵ-SVR is used; '-t 2' means the radial basis function (RBF) is used as the kernel function. And '-d 3' means the degree in kernel function is set 3; '-p 0.1' means the ϵ in loss function of ϵ-SVR is set 0.1; and '-h 1' means to use the shrinking heuristics. It is actually the best one of parameter combinations.

3.2 Experimental Results

In this section, the effect of each parameter for *ddFeature*, *priceFeature* and the combined feature *comFeature* is presented. In addition, the effect of the investment strategy threshold δ will also be presented. In order to keep the results of different trails comparable, all the trails that varying one parameter and fixed the other parameters have the same investment periods. Besides, we assume the initial investment capital is high and the transaction cost is negligible. Therefore, in this thesis, the effect of transaction cost is not considered. That is, we assume the transaction cost is zero.

Usually, mean square error is used to measure the performance of prediction. However, small forecast error doesn't mean profits. To intuitively show the prediction performance of future maximum drawdown, we directly show the predicted maximum drawdown series. Besides, we present the values of cumulative return, maximum drawdown and average drawdown with fixed threshold value δ to study the effects of parameters. The δ is an empirical value which is given based on the property of each dataset. Specifically, the setting of δ value needs to satisfy two criteria: firstly, it should keep the generated maximum drawdown of the whole investment period less than an expected threshold (as pointed out by Chekhlov et al. [CUZ05], a 20% drawdown is intolerant by most investors, we use 20% as the threshold for maximum drawdown); secondly, it should keep the corresponding piecewise future returns positive, as is mentioned in Section 2.3, Chapter 2. Applying the two criteria, the suitable δ value for S&P 500 dataset is from 0 to 7, and the suitable δ value for NASDAQ Composite and HSI are both in the range of 0 to 9. To achieve as high cumulative return as possible, the maximum of suitable

Table 3.3: Parameters considered to extract *ddFeature*

Parameter	Description
N	number of training samples
P	past window size
F	future window size
n	the number of separated maximum drawdowns

δ is used.

The effects of other possible values of δ is also investigated by fixing all the other parameters.

3.2.1 Experimental Results with *ddFeature*

ddFeature is a n-dimension feature composed by the first n separated maximum drawdown of past window. The relevant parameters are listed in Table 3.3.

To investigate the effect of parameters, the performance with varying N, P, F and n for the three datasets are presented in Table 3.4-3.7. In Table 3.4, the number of training samples N is varying while all the other parameters are fixed, i.e., $P = 100$, $F = 50$, $n = 3$ and personalized $\delta = 7$, 9 and 9 for S&P 500, NASDAQ Composite and HSI datasets, respectively[1]. Compared with passive investing, all of the trails greatly reduce the maximum drawdown and average drawdown values, especially for S&P 500 and NASDAQ Composite datasets. But not all of the trails can generate much higher cumulative returns than the passive investing. Although some of the trails achieve higher cumulative returns than passive investing strategy, the improvement is small.

Of the results, it appears the optimal choice of N for S&P 500 index dataset is 200, which achieves the highest cumulative return, minimal maximum drawdown and minimal average drawdown. However, there is no optimal N which generates consistently optimal cumulative return, maximum drawdown and average drawdown for the other two datasets. For NASDAQ Composite dataset, the trails at $N = 50$ and $N = 100$ achieve the same minimal maximum drawdown (16.1%)

[1] To guarantee all the trials with different parameter and the passive investing strategy having the same decision days, we set the first decision is made on day $max(N) + max(P) + max(F)$. The same method is used when using varying parameters, therefore, the statistical result of passive investing in different tables is a little different.

Table 3.4: Results using *ddFeature* with varying N and fixed $P = 100$, $F = 50$, $n = 3$ and personalized δ on the three datasets, respectively. PI is passive investing strategy.

Dataset	$\delta(\%)$	N	$CumRet$	$MaxDD(\%)$	$AvDD(\%)$
S&P 500	PI		4.9	56.8	13.7
	7	50	3.9	14.7	3.5
		100	5.1	17.4	2.8
		150	5.9	16.5	2.5
		200	**6.1**	**13.9**	**2.3**
		250	5.5	17.4	3.4
		300	5.4	16.8	3.4
NASDAQ Composite	PI		8.3	77.9	34.5
	9	50	5.1	**16.1**	5.3
		100	7.1	**16.1**	4.7
		150	**10.5**	22.3	**4.1**
		200	8.3	24.6	6.0
		250	7.6	25.3	6.4
		300	8.1	22.1	5.3
HSI	PI		5.4	65.2	24.4
	9	50	**4.8**	32.5	10.7
		100	3.7	**26.3**	**8.3**
		150	2.8	32.9	15.9
		200	2.6	37.1	16.5
		250	2.7	38.7	16.6
		300	3.1	45.5	15.6

while the trail at $N = 150$ achieves the optimal cumulative return (10.5) and average drawdown (4.1%). For HSI dataset, the trail at $N = 50$ achieves the highest cumulative return (4.8) while the trail at $N = 100$ achieves the optimal maximum drawdown (26.3%) and average drawdown (8.3%). However, the cumulative return of optimal trail for HSI dataset is not as high as passive investing. That is, for different datasets, the optimal number of training samples N is varying.

Table 3.5 shows the results with varying past window length P and fixed $N = 200$, $F = 50$, $n = 3$. Compared with passive investing, all the trails obtain better maximum drawdown and average drawdown. However, the improvement of cumulative return is little. It appears the optimal choice of P is $P = 150$ for HSI dataset. Although the trail achieves the optimal cumulative return, maximum

Table 3.5: Results using *ddFeature* with varying P and fixed $N = 200$, $F = 50$, $n = 3$ and personalized δ on the three datasets, respectively. PI is passive investing strategy.

Dataset	$\delta(\%)$	P	$CumRet$	$MaxDD(\%)$	$AvDD(\%)$
S&P 500	PI		4.9	56.8	13.7
	7	50	5.8	14.8	3.1
		100	6.1	**13.9**	**2.3**
		150	**6.2**	20.5	3.6
		200	5.1	15.8	3.0
NASDAQ Composite	PI		8.3	77.9	34.5
	9	50	7.4	21.6	**4.3**
		100	**8.3**	24.6	6.0
		150	6.0	22.4	7.0
		200	7.5	**20.7**	4.9
HSI	PI		5.4	65.2	24.4
	9	50	3.6	36.6	17.4
		100	2.6	37.1	16.5
		150	**3.7**	**22.5**	**9.8**
		200	3.2	31.9	14.0

drawdown and average drawdown simultaneously, the generated cumulative return on HSI is still very low. All of the trails on S&P 500 dataset generate much higher cumulative return, lower maximum drawdown and lower average drawdown than passive investing. For NASDAQ Composite dataset, the trail at $N = 100$ generates similar cumulative return with passive investing, but it achieves much lower maximum drawdown and average drawdown.

The results with varying features n is presented in Table 3.6. It appears changing the number of features n does not have a significant impact on the results. All the trails exhibit similar values with vary n values.

The results with varying F is presented in Table 3.7. As the suitable values of δ depends on the value of F (ref. Section 2.3, Chapter 2), when F is changing, the threshold δ should be adjusted accordingly. Compared with passive investing, all of the trails with varying F achieves better performance on the three metrics on S&P 500 dataset; while the trails at $F = 50$, 75 and 100 obtain better performance on NASDAQ Composite dataset, especially, the trail at $F = 100$ generates the optimal results. Besides, it appears the trails at $F = 75$ achieves high cumulative

Table 3.6: Results using *ddFeature* with varying n and fixed $N = 200$, $P = 100$, $F = 50$ and personalized δ on the three datasets, respectively. PI is passive investing strategy.

Dataset	$\delta(\%)$	n	$CumRet$	$MaxDD(\%)$	$AvDD(\%)$
S&P 500	PI		4.9	56.8	13.5
	7	2	5.6	17.3	**2.0**
		3	6.2	13.9	2.3
		4	**6.3**	**13.4**	2.2
		5	6.1	15.3	2.7
NASDAQ Composite	PI		8.8	77.9	33.9
	9	2	8.7	**21.8**	**5.4**
		3	**9.0**	24.7	5.9
		4	8.4	26.1	5.9
		5	7.8	25.2	6.7
HSI	PI		5.9	65.2	24.0
	9	2	2.7	**35.2**	17.8
		3	2.8	37.4	16.2
		4	**3.0**	38.1	**14.3**
		5	**3.0**	37.5	**14.3**

return, low maximum drawdown and low average drawdown for the three datasets simultaneously.

Comparing the results in Table 3.4 to 3.7, the changing of parameters of N, P and F have high impact than n. And the combination of parameters $N = 200$, $P = 100$, $n = 3$ and $F = 75$ generates high cumulative return, low maximum drawdown and low average drawdown consistently for the three datasets simultaneously. Besides, the combinations of $N = 200$, $P = 100$, $n = 3$ and $F = 75$ or $F = 100$ generate not only very high cumulative return and low average return, but also generate very low maximum drawdown values, i.e., the generated maximum drawdown values are less than 20%, for S&P 500 and NASDAQ Composite datasets. According to Chekhlov et al. [CUZ05], a 20% drawdown is intolerant by investors. The two trials successfully keep the drawdown less than 20%. In addition, the combinations of $N = 200$, $P = 100$, $n = 3$ and $F = 25, 75$ are the only two trails that generates much higher cumulative returns than passive investing on HSI dataset. It appears that it is much difficult to generate high cumulative return for HSI dataset by changing the value of parameters of $ddFeature$.

Table 3.7: Results using *ddFeature* with varying F and fixed $N = 200$, $P = 100$ and $n = 3$ on the three datasets, respectively. PI is passive investing strategy.

Dataset	$\delta(\%)$	F	$CumRet$	$MaxDD(\%)$	$AvDD(\%)$
S&P 500	PI		4.9	56.8	13.7
	5	25	5.1	27.5	5.9
	7	50	6.1	**13.9**	**2.3**
	11	75	**6.5**	15.0	3.9
	11	100	6.0	14.1	3.8
	13	125	5.5	18.7	5.1
	15	150	5.0	19.3	6.2
NASDAQ Composite	PI		8.3	77.9	34.5
	5	25	5.6	27.7	5.8
	9	50	8.3	24.6	6.0
	11	75	8.6	17.5	4.4
	13	100	**12.1**	**14.4**	**2.0**
	23	125	7.1	42.7	15.8
	23	150	5.2	44.8	17.8
HSI	PI		5.4	65.2	24.4
	7	25	5.7	41.8	16.8
	9	50	2.6	37.1	16.5
	17	75	**5.9**	32.9	12.7
	19	100	5.1	31.3	**10.1**
	19	125	3.0	**29.8**	13.8
	19	150	2.0	31.4	17.1

To study the performance of prediction, the real maximum drawdown and the predicted maximum drawdown series of three parameter combinations with good performance on the three datasets respectively are shown in Figure 3.3. For most days, the predicted maximum drawdown is similar with the real one. As we make investment decisions based on a binary function of the predicted maximum drawdown, when the real and prediced maximum drawdowns are both on the same side of the δ line in the figure, it does not affect the result. However, there are some days that the predicted maximum drawdown is far away from its real value. What's worse, they distributes on different sides of the δ line, which leads to wrong investment decision. Actually, delay effect is obvious in the predicted maximum drawdown series. It is because at these days, the data changes severely. As we always use the most recent historical data to train the model, it cannot generalize well to future data for these days.

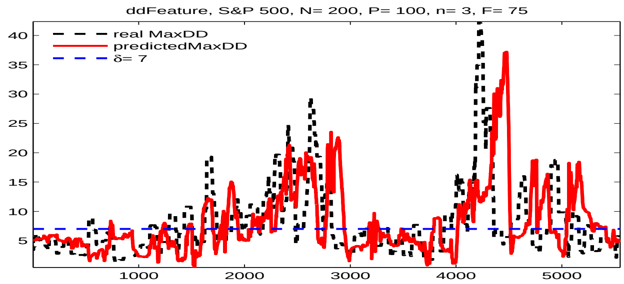
And the corresponding generalized cumulative return series and corresponding yearly maximum drawdown series of on the three datasets are presented in Figure 3.4. The generalized cumulative return series are smoother than the passive investing. However, the trails do not generate consistently higher returns at any period, which shows the inefficacy of the proposed model. As the model reduces drawdown values, it also reduces returns simultaneously.

3.2.2 Experimental Results with *priceFeature*

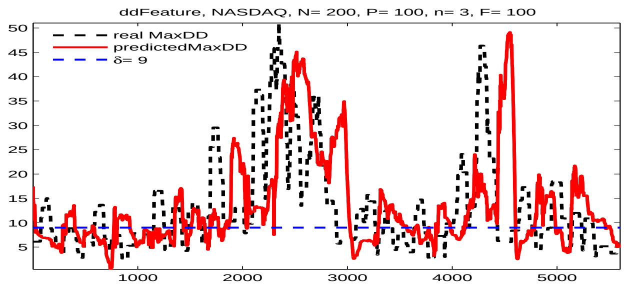
In the following part, the experiment results using pattern *priceFeature* on S&P 500, NASDAQ Composite and HSI datasets are presented. The *priceFeature* is extracted from the piecewise linear approximation of the original price series, which is detailed described in Section 2.1.2, Chapter 2. To investigate the sensitivity of generalization to *priceFeature* parameters, the following relevant parameters in Table 3.8 are studied.

The results with varying N, P, k and F on the three datasets are presented in Table 3.9 to 3.12. In each table, one parameter varies while all the other parameters are fixed.

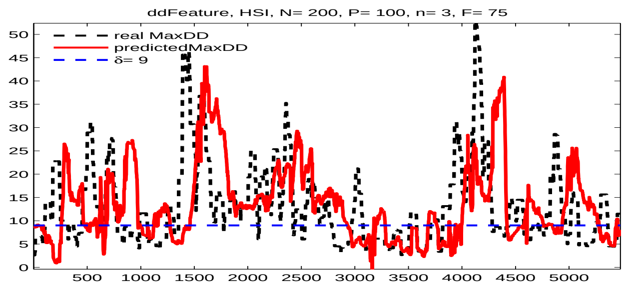
Table 3.9 shows the results with varying number of training samples N and fixed $P = 100$, $F = 50$, $k = 3$. Similar as results of *ddFeature*, all the trails signif-



(a)

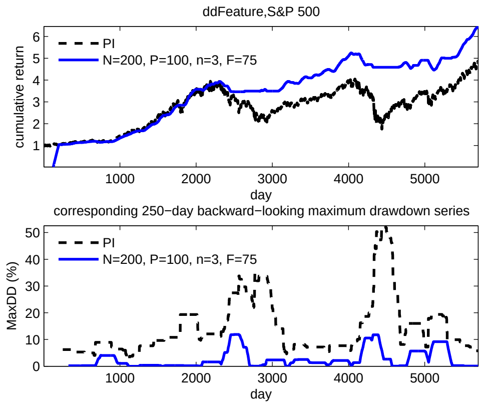


(b)

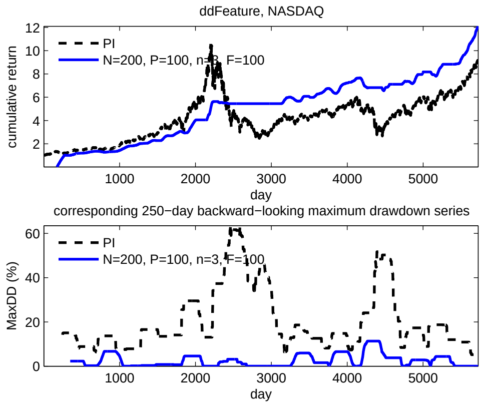


(c)

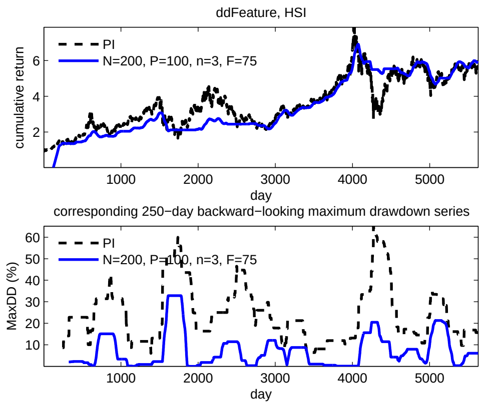
Figure 3.3: The real maximum drawdown and the predicted maximum drawdown series using *ddFeature* on S&P 500 dataset at $N = 200$, $P = 100$, $n = 3$ and $F = 75$ (a), NASDAQ Composite dataset at $N = 200$, $P = 100$, $n = 3$ and $F = 100$ (b) and HSI dataset at $N = 200$, $P = 100$, $n = 3$ and $F = 75$ (c). (The label for x-axis is "day", while the label for y-axis is "MaxDD".)



(a)



(b)



(c)

Figure 3.4: The generalized cumulative return and corresponding drawdown series using *ddFeature* on S&P 500 dataset at $N = 200$, $P = 100$, $n = 3$ and $F = 75$ (a), NASDAQ Composite dataset at $N = 200$, $P = 100$, $n = 3$ and $F = 100$ (b) and HSI dataset at $N = 200$, $P = 100$, $n = 3$ and $F = 75$ (c).

Table 3.8: Parameters considered to extract *priceFeature*

Parameter	Description
N	number of training samples
P	past window size
F	future window size
k	number of segments

icantly reduce the maximum drawdown and average drawdown values, especially for S&P 500 and NASDAQ Composite datasets. But there are only a few of trails that generate higher cumulative returns than passive investing. Of the results, it appears the trails at $N = 250$ and 300 achieves both high cumulative return and small maximum drawdown on S&P 500 and NASDAQ Composite datasets. And the maximum drawdowns are less than 20%. For HSI dataset, the trail at N = 200 achieves the highest cumulative return (4.1) while the trail at N = 150 achieves the optimal maximum drawdown (29.0%) and average drawdown (13.6%). However, the cumulative return of optimal trail for HSI dataset is not as high as passive investing.

The results with varying P and fixed $N = 200$, $F = 50$, $k = 3$ is presented in Table 3.10. It appears the trials at $P = 50$ and $P = 100$ generates the optimal results for the three datasets. However, the improvement of the trails on cumulative return value is very little.

The results with varying k and fixed $N = 200$, $P = 100$, $F = 50$ is presented in Table 3.11. It appears changing the number of features does not have a significant impact on the results. All the trails with varying k are similar.

The results with varying F is presented in Table 3.12. It appears that changing future window size F have significant impact on the results. The trails at $F = 75$ generates the highest cumulative returns consistently for the three datasets. And the trails at $F = 50$ generates the lowest maximum drawdown and average drawdown for NASDAQ Composite and HSI datasets. The optimal maximum drawdown is achieved at $N = 100$ for S&P 500 dataset.

Compare the results in Table 3.9 to 3.12, it should be noted that using *price-Feature*, the combination of parameters at $N = 200$, $P = 100$, $n = 3$ and $F = 75$ is the only trail that generates higher overall performance than passive investing.

Table 3.9: Results using *priceFeature* with varying N and fixed $P = 100$, $F = 50$, $k = 3$ and personalized δ on the three datasets, respectively. PI is passive investing strategy.

Dataset	δ(%)	N	$CumRet$	$MaxDD$(%)	$AvDD$(%)
S&P 500	PI		4.9	56.8	13.7
	7	50	3.8	16.7	3.9
		100	3.9	18.8	3.9
		150	5.4	17.7	**2.8**
		200	5.2	19.3	3.8
		250	5.8	16.6	3.5
		300	**6.1**	**12.1**	3.5
NASDAQ Composite	PI		8.3	77.9	34.5
	9	50	4.5	17.9	4.8
		100	3.6	21.5	8.9
		150	5.7	16.2	5.7
		200	7.2	**15.6**	4.4
		250	8.6	15.9	**4.2**
		300	**9.0**	17.5	4.4
HSI	PI		5.4	65.2	24.4
	9	50	3.5	42.0	15.1
		100	3.6	38.4	15.4
		150	3.6	**29.0**	**13.6**
		200	**4.1**	29.5	13.6
		250	3.7	34.3	14.4
		300	3.1	41.4	19.9

Table 3.10: Results using *priceFeature* with varying P and fixed $N = 200$, $F = 50$, $k = 3$ and personalized δ on the three datasets, respectively. PI is passive investing strategy.

Dataset	$\delta(\%)$	P	$CumRet$	$MaxDD(\%)$	$AvDD(\%)$
S&P 500	PI		4.9	56.8	13.7
	7	50	**5.6**	20.4	4.1
		100	5.2	**19.3**	**3.8**
		150	5.0	21.1	4.4
		200	3.8	32.3	9.0
NASDAQ Composite	PI		8.3	77.9	34.5
	9	50	**8.4**	15.9	**3.6**
		100	7.3	**15.6**	4.4
		150	6.4	18.2	6.7
		200	5.2	35.1	11.0
HSI	PI		5.4	65.2	24.4
	9	50	3.3	35.8	19.0
		100	**4.1**	**29.5**	**13.6**
		150	3.0	36.6	15.3
		200	2.7	35.4	18.3

Table 3.11: Results using *priceFeature* with varying k and fixed $N = 200$, $P = 100$, $F = 50$ and personalized δ on the three datasets, respectively. PI is passive investing strategy.

Dataset	$\delta(\%)$	k	$CumRet$	$MaxDD(\%)$	$AvDD(\%)$
S&P 500	PI		4.9	56.8	13.5
	7	2	5.0	18.4	3.8
		3	5.3	19.3	3.8
		4	**5.8**	**15.1**	**2.9**
		5	5.8	15.6	3.2
NASDAQ Composite	PI		8.8	77.9	33.9
	9	2	7.3	15.8	5.5
		3	7.9	15.6	**4.4**
		4	8.1	**15.5**	4.6
		5	**8.4**	15.7	4.8
HSI	PI		5.9	65.2	24.0
	9	2	2.8	39.2	18.5
		3	**4.4**	**29.3**	**13.4**
		4	4.2	34.9	16.2
		5	4.1	34.8	16.4

Table 3.12: Results using *priceFeature* with varying F and fixed $N = 200$, $P = 100$ and $n = 3$ on the three datasets, respectively. PI is passive investing strategy.

Dataset	$\delta(\%)$	F	$CumRet$	$MaxDD(\%)$	$AvDD(\%)$
S&P 500	PI		4.9	56.8	13.7
	5	25	4.3	25.2	6.5
	7	50	5.2	19.3	**3.8**
	11	75	**5.6**	28.5	6.5
	11	100	5.4	**16.5**	5.2
	13	125	4.5	25.9	8.2
	15	150	4.5	26.4	8.4
NASDAQ Composite	PI		8.3	77.9	34.5
	5	25	5.7	18.6	3.8
	9	50	7.3	**15.6**	**4.4**
	11	75	**9.9**	24.4	8.0
	13	100	9.4	23.1	7.6
	23	125	6.9	42.9	12.9
	23	150	5.0	43.7	16.3
HSI	PI		5.4	65.2	24.4
	7	25	4.8	44.1	19.6
	9	50	4.1	**29.5**	**13.6**
	17	75	**5.8**	36.2	14.9
	19	100	4.2	38.9	16.5
	19	125	2.8	37.7	17.7
	19	150	1.9	37.8	20.9

Besides, the real maximum drawdown and the predicted maximum drawdown series of three parameter combinations with good performance on the three datasets respectively are shown in Figure 3.5. Similar as using *ddFeature*, most of the predictions are near their true values. However, for the days that the real maximum drawdowns are extreme values or the real maximum drawdowns are following the extreme values, the prediction cannot perform well. It also shows delay effects on these days. Compare the result using *ddFeature* in Figure 3.3(c) and the result using *priceFeature* in Figure 3.5(c)(both of them have the same parameter setting), the predicted maximum drawdown series using *priceFeature* is much smoother. Actually, the same phenomenon can be found on the other datasets with different parameter settings. As the features of *ddFeature* are the adjacent maximum drawdowns, the prediction of *ddFeature* relies much on the adjacent maximum drawdown value. Therefore, when an extreme maximum drawdown occur in adjacent window, the predicted value is also extreme with high probability.

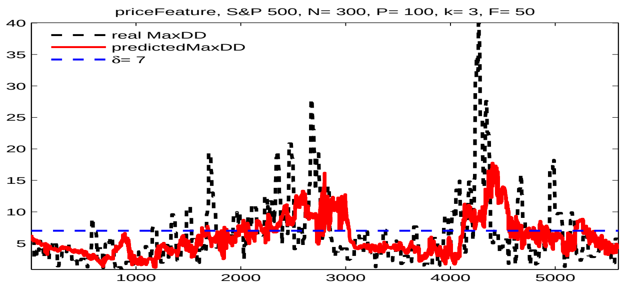
Figure 3.6 shows the corresponding generalized cumulative return and corresponding yearly maximum drawdown series on the three datasets respectively. The yearly maximum drawdown is greatly reduced on the three datasets. And the trails generates both high cumulative return and low yearly maximum drawdowns for S&P 500 and NASDAQ datasets. But the maximum drawdown on HSI dataset is still a little high for some periods.

Compare the results of using *ddFeature* and *priceFeature*, although they generate different predicted maximum drawdown series, they perform similar in terms of cumulative return and maximum drawdown on the three datasets. This is because both of them make wrong investment decisions for some common period due to the same delay effects. As the data varies a lot at some periods, the prediction model cannot work well.

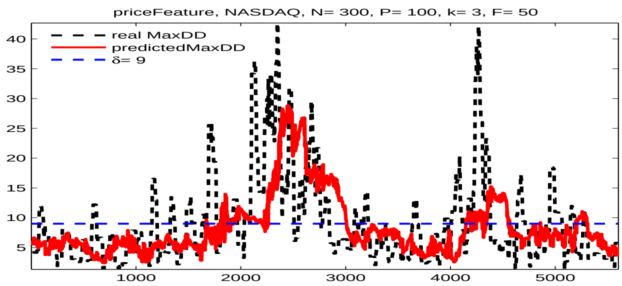
3.2.3 Experimental Results with *comFeature*

The *comFeature* is the joint feature of *ddFeature* and *priceFeature* $(ddFeature, priceFeature)$. It contains features of both *ddFeature* and *priceFeature*. Therefore, the relevant parameters includes all parameters of *ddFeature* and *priceFeature* (Table 3.13).

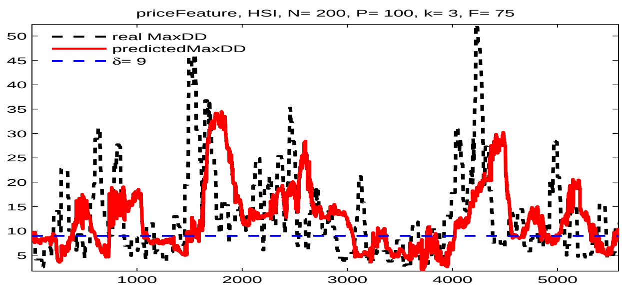
To compare the prediction performance of using the three features, we firstly



(a)

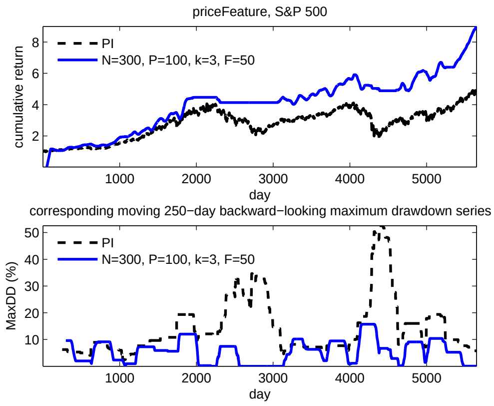


(b)

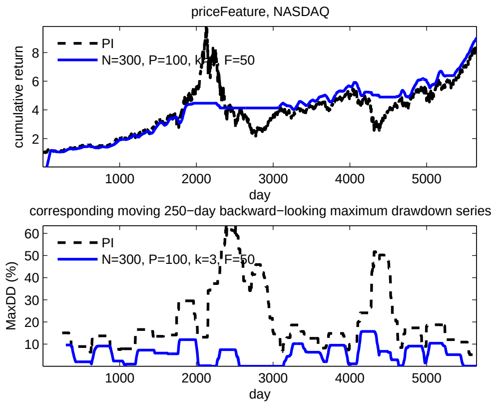


(c)

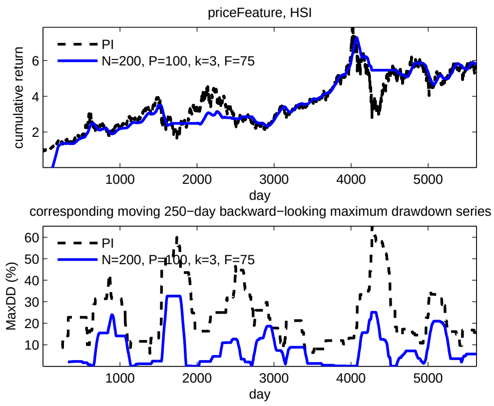
Figure 3.5: The real maximum drawdown and the predicted maximum drawdown series using *priceFeature* on S&P 500 dataset at $N = 300$, $P = 100$, $k = 3$ and $F = 50$ (a), NASDAQ Composite dataset at $N = 300$, $P = 100$, $k = 3$ and $F = 50$ (b), HSI dataset at $N = 200$, $P = 100$, $k = 3$ and $F = 75$ (c). (The label for x-axis is “day”, while the label for y-axis is “MaxDD”.)



(a)



(b)



(c)

Figure 3.6: The generalized cumulative return and corresponding yearly maximum drawdown series using *priceFeature* on S&P 500 dataset at $N = 300$, $P = 100$, $k = 3$ and $F = 50$ (a), NASDAQ Composite dataset at $N = 300$, $P = 100$, $k = 3$ and $F = 50$ (b), HSI dataset at $N = 200$, $P = 100$, $k = 3$ and $F = 75$ (c).

Table 3.13: Parameters considered to extract *comFeature*

N	number of training samples
P	past window size
n	the number of separated maximum drawdowns
k	number of segments of *priceFeature*
F	future window size

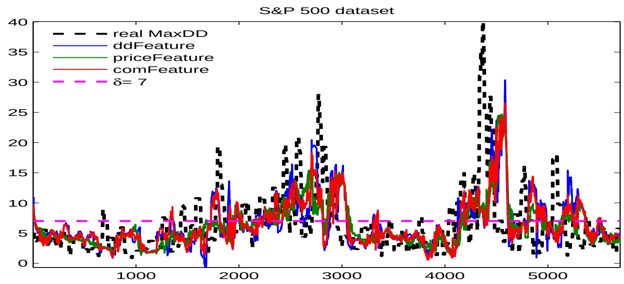
show the predicted maximum drawdown series in Figure 3.7 with fixed $N = 200$, $P = 100$, $F = 50$, $n = 3$ and $k = 3$ on the three datasets. The predicted maximum drawdown series of using *ddFeature*, *priceFeature* and *comFeature* are similar for most of the period. More specifically, the predicted maximum drawdown series of using *comFeature* lies in between of the predicted maximum drawdown series of using *ddFeature* and *priceFeature*.

Therefore, based on the similar predicted series, the cumulative return and drawdown series of using varying features are similar too, which is shown in Table 3.14 and Figure 3.8. It appears that the performance of using *comFeature* is just in between of that of separated *ddFeature* and *priceFeature*. Due the occurrence of extreme maximum drawdown values at some days, none of *ddFeature* and *priceFeature* can generalize well. Therefore, the combined feature cannot make improvement for these days. It should be noted that, although the results are based on fixed parameter at $N = 200$, $P = 100$, $F = 50$, $n = 3$ and $k = 3$, the same phenomenon can be observed with other parameter settings.

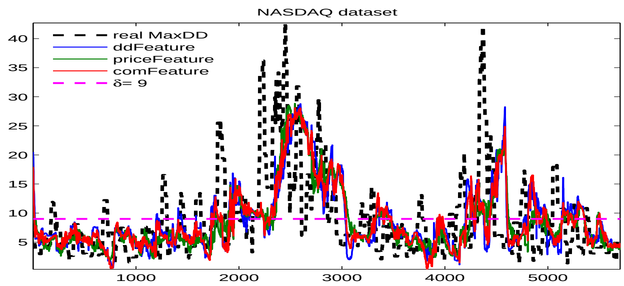
3.2.4 Experimental Results with Investment Decision Threshold

In last three parts, we investigate the impact of feature relevant parameters, suppose a fixed and suitable investment threshold δ is given. In this section, the effect of changing the value of δ with fixed all the other feature parameters will be studied.

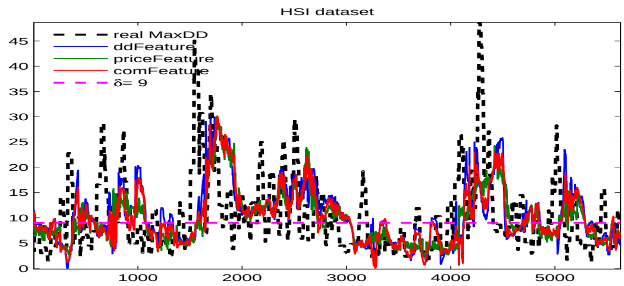
Given fixed $N = 200$, $P = 100$, $F = 50$, $n = 3$, $k = 3$, the corresponding generalized cumulative returns and maximum drawdowns with varying δ for using *ddFeature*, *priceFeature* and *comFeature* on the three datasets are shown in Figure



(a)

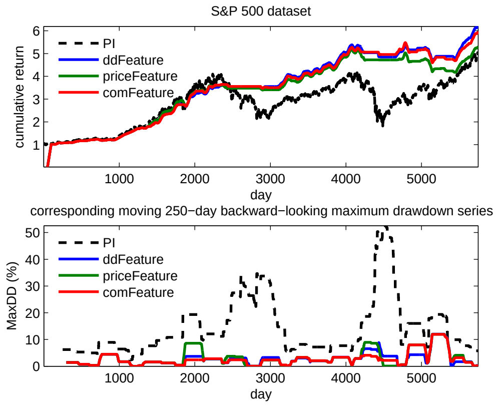


(b)

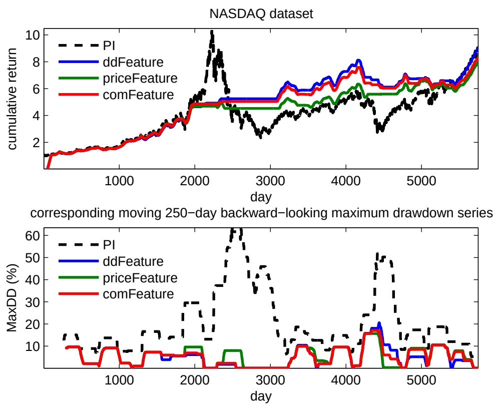


(c)

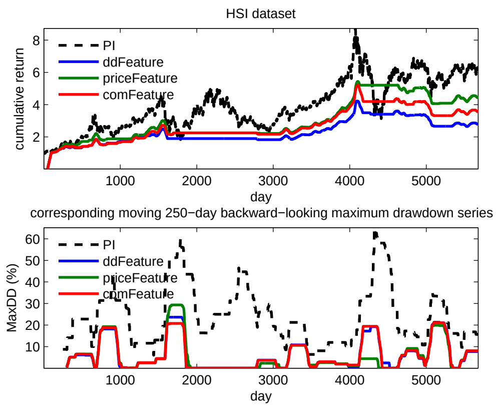
Figure 3.7: The real maximum drawdown and the predicted maximum drawdown series using *ddFeature*, *priceFeature* and *comFeature* with fixed $N = 200$, $P = 100$, $F = 50$, $n = 3$ and $k = 3$ on S&P 500 dataset (a), NASDAQ Composite dataset (b), and HSI dataset (c).(The label for x-axis is "day", while the label for y-axis is "MaxDD".)



(a)



(b)



(c)

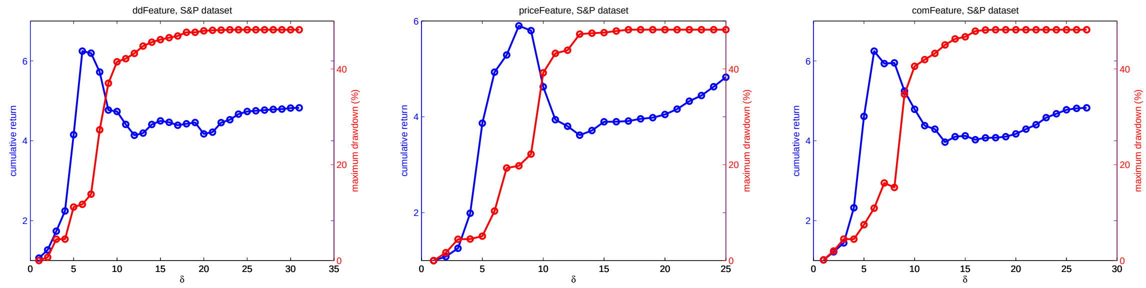
Figure 3.8: The generalized cumulative return and corresponding yearly maximum drawdown series of the *ddFeature*, *priceFeature* and *comFeature* with fixed $N = 200$, $P = 100$, $F = 50$, $n = 3$ and $k = 3$ on S&P 500 dataset (a), NASDAQ Composite dataset (b), and HSI dataset (c).

Table 3.14: The results with varying features and fixed $N = 200$, $P = 100$, $F = 50$, $n = 3$ and $k = 3$.

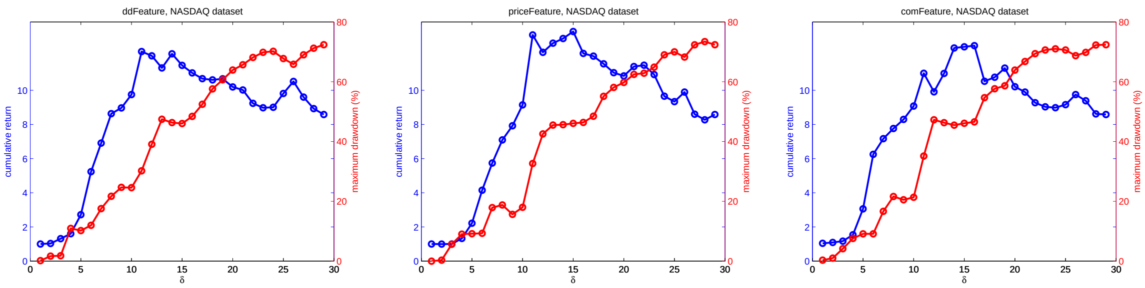
Dataset	$\delta(\%)$	Feature	$CumRet$	$MaxDD(\%)$	$AvDD(\%)$
S&P 500	7	PI	4.9	56.8	13.5
		ddFeature	**6.2**	**13.9**	**2.3**
		priceFeature	5.3	19.3	3.8
		comFeature	5.9	16.2	2.5
NASDAQ Composite	9	PI	8.8	77.9	33.9
		ddFeature	**9.0**	24.7	5.9
		priceFeature	7.9	**15.6**	**4.4**
		comFeature	8.3	20.5	5.3
HSI	9	PI	5.9	65.2	24.0
		ddFeature	2.8	37.4	16.2
		priceFeature	**4.4**	**29.3**	**13.4**
		comFeature	3.5	36.3	13.9

3.9. The δ value has a significant impact on cumulative return and maximum drawdown, especially when δ is small. It appears, generally, the maximum drawdown value is increasing with δ value for all features on all datasets. Therefore, to achieve small maximum drawdown value, δ should be as small as possible. Meanwhile, the cumulative returns increase with δ value when δ is small and decrease when δ becomes larger. That is, a very small δ generates small cumulative return, which is not a good choice; while a very big δ value generates both small cumulative return and large maximum drawdown, which is not a good choice either. When δ is in a suitable small range, e.g., 0 to the values that generates the peak cumulative return, both the cumulative return and maximum drawdown values increase with δ value. It means there is no δ value that generates both the largest cumulative return and lowest maximum drawdown simultaneously. Investors have to make tradeoff between the two goals, cumulative return and maximum drawdown.

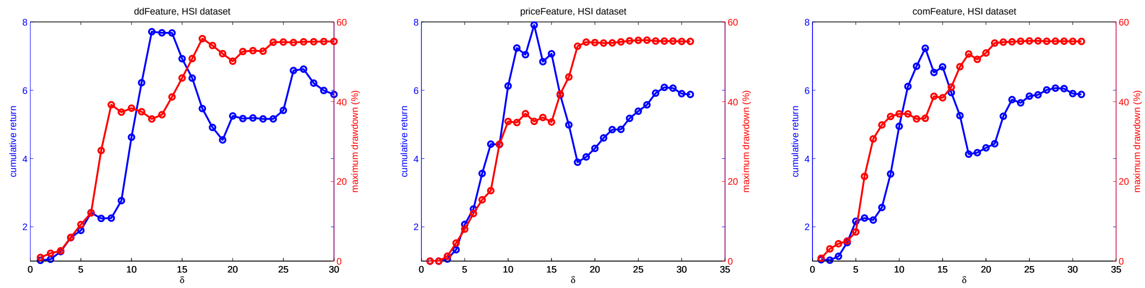
To clearly present the investing results along the whole period, Figure 3.10 shows the generalized cumulative return and corresponding yearly maximum drawdown series using *ddFeature* with varying δ values and fixed $N = 200$, $P = 100$, $F = 50$, $n = 3$ on S&P 500 dataset. The δ value actually determines the investing performance. Specifically, if δ is small, the generated cumulative return series is smooth, but with little profit; while if δ is too large, it is actually similar as the



(a)



(b)



(c)

Figure 3.9: The cumulative return and maximum drawdown values with varying δ and fixed $N = 200$, $P = 100$, $F = 50$, $n = 3$ and $k = 3$ on S&P 500 dataset (a), NASDAQ Composite dataset (b), and HSI dataset (c).

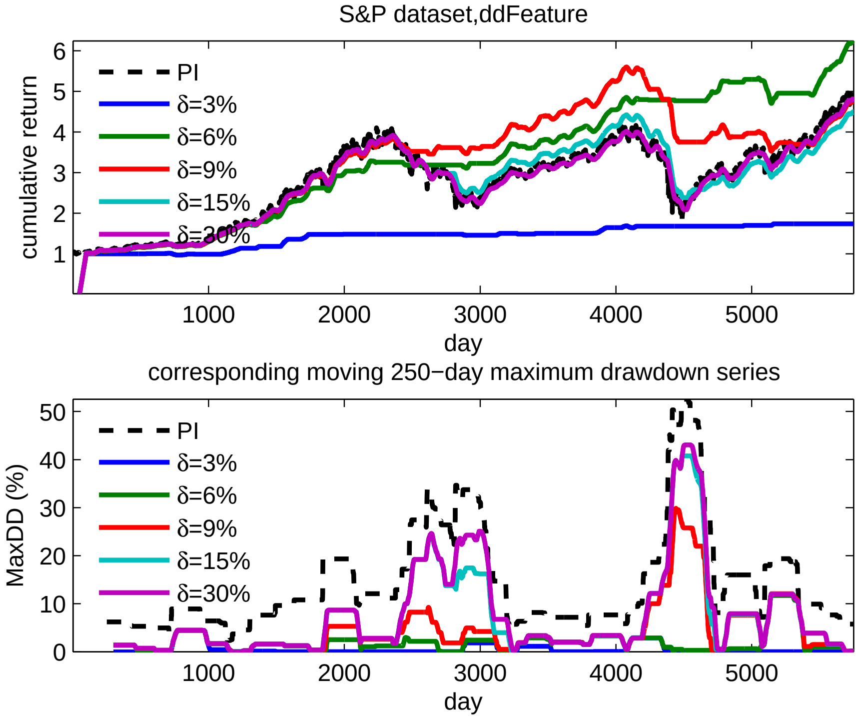


Figure 3.10: The generalized cumulative return and corresponding yearly maximum drawdown series of varying δ and fixed $N = 200$, $P = 100$, $F = 50$, $n = 3$ on S&P 500 dataset, using *ddFeature*.

result of passive investing. And the same property can be observed for all the other features on all the datasets.

3.2.5 Experimental Results with Adaptive Investment

In this section, the result of using time-varying investment decision threshold δ is presented. To obtain the adaptive δ value at each day, the suitable δ values in the corresponding N training samples is computed. Applying the same methodology in Section 2.3, Chapter 2, i.e., the maximum of the suitable δ value is used as the adaptive threshold for the next decision.

Firstly, the adaptive δ value series is shown in Figure 3.11. It should be noted the adaptive δ value do not change the predicted maximum drawdown series, it only affects the investment decision. Therefore, the predicted maximum drawdown series is actually the same as the ones in Figure 3.7. The generated adaptive δ series varies a lot with time. Basically, it is smaller than the fixed δ at the periods that the real maximum drawdowns are small; and it is bigger than the fixed δ at the periods that the real maximum drawdowns are large.

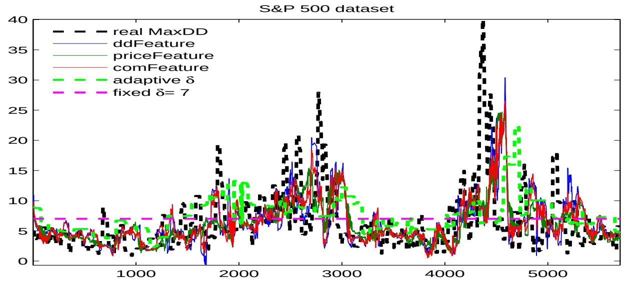
And the performance of using the adaptive δ series is shown in Table 3.15 and Figure 3.12. Compare Table 3.15 and Table 3.14, it appears that the trails generates higher cumulative return and higher maximum drawdown for all features on NASDAQ Composite and HSI datasets. As both the generated cumulative return and maximum drawdowns are higher than the results with fixed δ, it actually cannot be recognized as an improvement. And for S&P 500 dataset, the trails with adaptive δ generates higher cumulative return and smaller maximum drawdown on *priceFeature* and *comFeature*, but it does not outperform the trail with fixed δ on *ddFeature*. Compare Figure 3.12 and Figure 3.8, it appears that the trails with adaptive δ generates much higher returns before the peaks of price series, while it usually has higher drawdowns after the peaks, e.g., the days around 2200 on NASDAQ Composite. This is, at some periods (i.e., the period before the peaks of price series), the relatively large adaptive δ is beneficial to make the investment decisions; however, for some other periods (i.e., the period after the peaks), the large adaptive δ leads to large drawdowns. Essentially, it also have delay effect, which is also attributed to the data property. The adaptive δ computed based on the historical training samples cannot generalize well to future when the data varies a lot.

3.2.6 Comparison

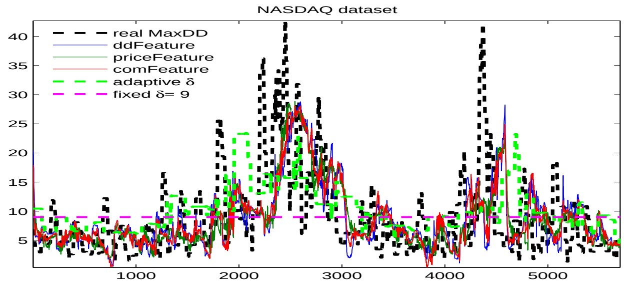
To show the advantages of the proposed method, in this section, we compare the proposed method with two kinds of popular investment methods. One is risk-free investment, the other is moving average investment. Both of them are widely used in financial industry.

Comparison with Risk-free Investment

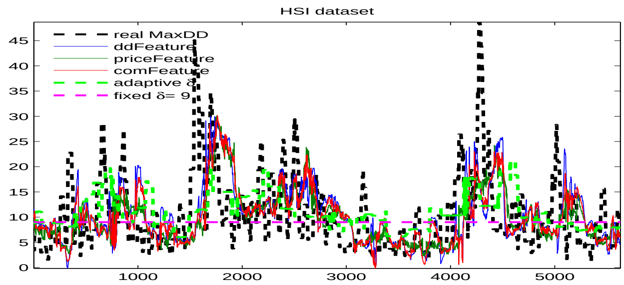
Firstly, we compare the proposed methods with the risk-free investment. Suppose the annual risk-free rate is 5%, that is, it generates 5 percent profit with 0 risk



(a)

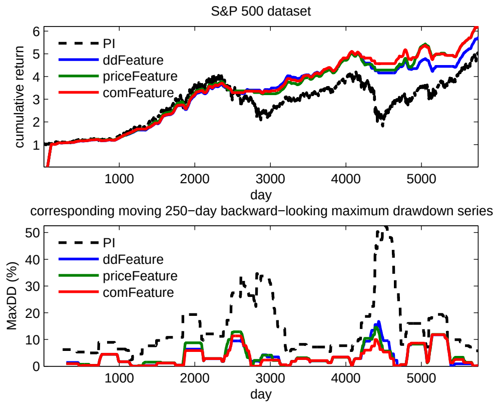


(b)

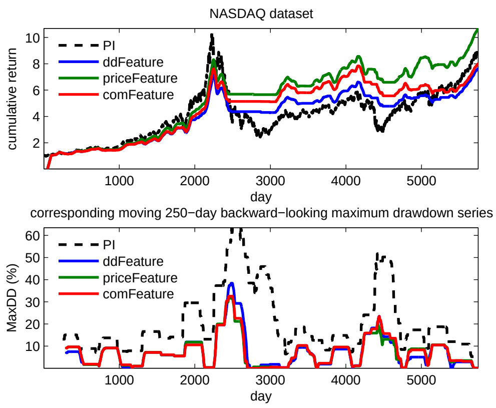


(c)

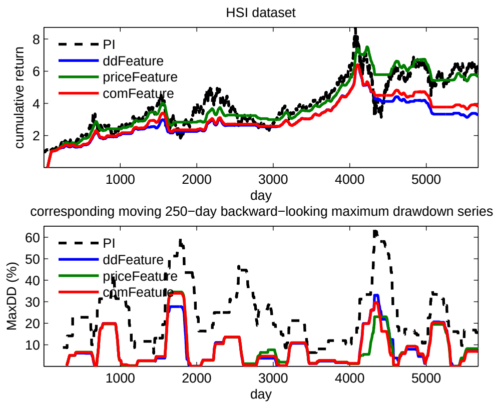
Figure 3.11: The real maximum drawdown and the predicted maximum drawdown series using the three features and adaptive δ value on S&P 500 dataset (a), NASDAQ Composite dataset (b), and HSI dataset (c).(The label for x-axis is "day", while the label for y-axis is "MaxDD".)



(a)



(b)



(c)

Figure 3.12: The generated cumulative return and corresponding moving yearly maximum drawdown series using adaptive δ value on S&P 500 dataset (a), NASDAQ Composite dataset (b), and HSI dataset (c).

Table 3.15: The results with adaptive δ values and fixed $N = 200$, $P = 100$, $F = 50$, $n = 3$ and $k = 3$ of using *ddFeature*, *priceFeature* and *comFeature* on the three datasets.

Dataset	Feature	$CumRet$	$MaxDD(\%)$	$AvDD(\%)$
S&P 500	PI	4.9	56.8	13.5
	ddFeature	5.7	17.9	4.2
	priceFeature	6.2	14.9	4.2
	comFeature	**6.2**	**11.8**	**3.2**
NASDAQ Composite	PI	8.8	77.9	33.9
	ddFeature	7.6	39.4	16.2
	priceFeature	**10.7**	**32.0**	**10.9**
	comFeature	8.0	32.9	13.7
HSI	PI	5.9	65.2	24.0
	ddFeature	3.2	50.9	16.4
	priceFeature	**5.7**	**34.6**	**14.1**
	comFeature	3.8	43.1	17.5

every year. As the testing days is 22.8 years, the cumulative return of risk-free rate investment is 3.04 with 0 drawdown.

It is clear that the cumulative returns of the proposed method on the three datasets are greater than the risk-free investment, while the drawdown levels are higher than the risk-free one. As return and drawdown are two different metrics, we cannot judge which one is better. However, due to the uncertainty of stock investment, the risk of stock investment is definitely greater than 0. Under acceptable drawdown level, the proposed method achieves much high cumulative return.

Comparison with Moving Average Investment

Moving average is an important technical indicator and the moving mean return is regarded as showing the momentum of the market [Che09]. The moving mean return is the moving average of the last k returns, which is defined as $\bar{r}_t^k = \frac{r_t + r_{t-1} + \cdots + r_{t-k+1}}{k}$, where r_t is the return given by $r_t = \frac{p_t - p_{t-1}}{p_{t-1}}$, where p_t is the price at time t. Generally, if the current $\bar{r}_t^k$ is greater than 0, the momentum is positive, and we make investment for the next day; else it indicates an negative momentum, and we do not invest money for the next day.

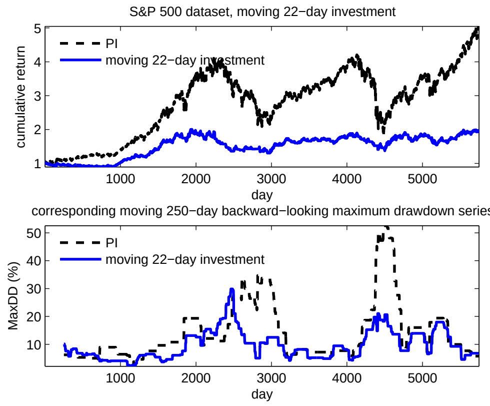
Figure 3.13 shows the results using the moving average method on the three datasets. In the experiment, we set $k = 22$, which is the one that generates the best result for the three datasets. Compared the results with the ones in Figure 3.4 and Figure 3.6, the maximum drawdown levels of using moving average method are higher than that of the proposed method, while the cumulative returns are similar for the NASDAQ composite and HSI datasets. However, the cumulative return of the moving average method is significantly worse than that of the proposed method on S&P 500 dataset. Therefore, the proposed method is outperform the conventional moving average method.

3.3 Discussion

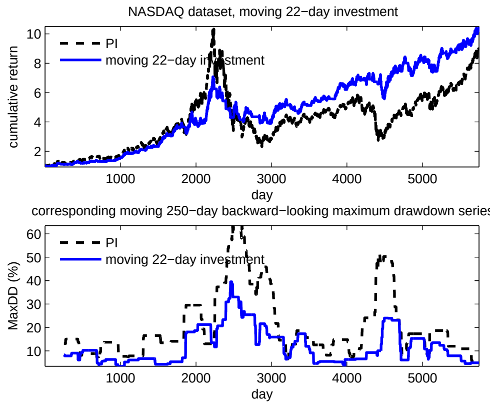
The experimental results on S&P 500, NASDAQ Composite and HSI datasets prove that the proposed pattern-based investment method can reduce the drawdowns of return series. However, there are several difficulties that limit the efficacy of the proposed model.

The efficacy of the proposed model depends on the parameter values. And different datasets have different performance given the same parameter combinations. Even for the same dataset, the data varies with time, which leads to the change of suitable parameter combinations. In this thesis, there are a lot of involved parameters, i.e., N, P, F, n, k and δ, and all the parameters are empirical values. If the parameters are not set appropriately, the model may not work any more. Therefore, the parameter setting is important and difficult. It depends on the property of dataset at different time and depends on different preference of return and risk. Although an adaptive δ value is attempted, the progress is not significant.

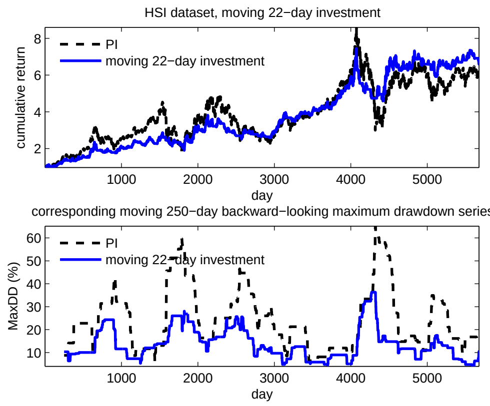
It has delay effect when peaks or troughs occurred, which shows the ineffectiveness of the proposed model. During such periods, the model learned based on past training data cannot catch the changes in future. To solve this problem, one solution is to reduce the length of training data, e.g., the number of training samples N, the past window size P and the future window size F, which is expected to focus on most recent data and eventually decrease delay effect. However, such a solution will lead the model more sensitive to noise, which may be not a good



(a)



(b)



(c)

Figure 3.13: The generated cumulative return and corresponding moving yearly maximum drawdown series using moving 22-day (i.e., monthly) investment method on S&P 500 dataset (a), NASDAQ Composite dataset (b), and HSI dataset (c).

choice for the overall period.

The model cannot generate consistently higher returns for any time, which means, for some subperiods, it maybe not as good as passive investing. The main reason of such a result is caused by the delay effect of training and predicting data. Therefore, the main task is to find methods to reduce delay effect.

Although prediction accuracy is not equivalent to return and drawdown values, improving prediction accuracy will be useful to find a better tradeoff between return and drawdown. In this thesis, there are only two kinds of features are studied, which are both based on the daily close price. To achieve better prediction result, more information, such as volume, interest rate, may be helpful.

Overall, in comparing the empirical results from the passive investing model and the proposed pattern-based drawdown reduction model, the latter is markedly more smoother. The generated cumulative return is higher and the corresponding drawdown series are much smaller. It is clear that the proposed *ddFeature* and *priceFeature* are effective features and the proposed investing strategy is useful to reduce potential drawdowns.

Chapter 4

Conclusion

Drawdown, which is the percentage loss of wealth from the previous peak, is an important downside risk metric in the investment area. To construct a well-performing investment strategy, potential drawdowns should be considered, and measures should be taken to avoid future drawdowns.

In this thesis, we have proposed a novel machine learning approach to address the drawdown reduction problem. Firstly, two kinds of features, *ddFeature* and *priceFeature* are extracted based on the correlation between them and future maximum drawdown. Then a powerful machine learning method SVM is applied to train the prediction model. Based on the prediction result, an aggregation investment method is applied to make investment decision for each day. And experimental results have shown the efficacy of the proposed model.

Our results are compared to those obtained using passive investing strategy, risk-free investment and moving average investment. And it is clear from experimental evidence that the proposed pattern-based drawdown reduction model smoother the return and achieves very small drawdown series.

One future direction is to study time-varying algorithms based on the time-varying property of the data. Although we assume the future maximum drawdown is an i.i.d. sampling from training data. The real data doesn't follow the assumption all the time. Therefore, the prediction fails during these periods. Studying time-varying algorithms will solve the problem.

The second direction is to study using more sophisticated investment decision functions and allow short selling in the investment strategies. Although the binary decision function achieves good performance, it is not flexible and it does not consider the confidence of the predicted data. Besides, by allowing short selling, the profit will be much high. Therefore, in the future, we will construct a investment function allowing short selling based on the predicted value and its confidence.

The third direction is to study using feedback consideration when training the SVM model. Feedback is a good way to correct the over-fitting problem introduced by learning model. Using feedback consideration may bring improvement to our original algorithm.

Another direction is to study the generalizability of the proposed model to individual stocks. In stock exchange market, there are thousands of stocks, while the number of stock indices is very limited. To make investment more flexible, investing on individual stocks is better and more appealing. Individual stocks are more volatile than stock index, which brings more opportunity and risk. If the proposed model works well on individual stocks, more profit will be generated. Besides, the generalizability of the proposed model to portfolio management is also a future direction. During the portfolio management process, drawdowns are unavoidable, and great drawdowns are not rare in the investment area. If the proposed model works on portfolio management, the generated returns will be smoother and with less risk.

The fifth one is to mine the potential features that are useful to make prediction on future maximum drawdown, such as volume, interest rate, and the price series of constituent stocks.

Overall, our experimental results have demonstrated that the proposed pattern-based drawdown reduction model is a promising investment strategy, which achieves both smoother and higher returns and smaller drawdowns.

Bibliography

[Che09] S.-S. Chen. Predicting the bear stock market: Macroeconomic variables as leading indicators. *Journal of Banking & Finance*, 33(2):211–223, 2009.

[CK95] Jaksa Cvitanic and Ioannis Karatzas. On portfolio optimization under "drawdown" constraints. *IMA Volumes in Mathematics and its Applications*, 65:35–35, 1995.

[CL11] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[CO13] Vladimir Cherny and Jan Obłój. Portfolio optimisation under nonlinear drawdown constraints in a semimartingale financial model. *Finance and Stochastics*, 17(4):771–800, 2013.

[CT01] Lijuan Cao and Francis EH Tay. Financial forecasting using support vector machines. *Neural Computing & Applications*, 10(2):184–192, 2001.

[CUZ03] Alexei Valerievich Chekhlov, Stanislav Uryasev, and Michael Zabarankin. Portfolio optimization with drawdown constraints. In Bernd Scherer, editor, *Asset and Liability Management Tools*, pages 263–278. 2003.

[CUZ05] Alexei Chekhlov, Stanislav Uryasev, and Michael Zabarankin. Drawdown measure in portfolio optimization. *International Journal of Theoretical and Applied Finance*, 8(01):13–58, 2005.

[GM14] Lisa R Goldberg and Ola Mahmoud. On a convex measure of drawdown risk. *Available at SSRN 2430918*, 2014.

[GZ93] Sanford J Grossman and Zhongquan Zhou. Optimal investment strategies for controlling drawdowns. *Mathematical Finance*, 3(3):241–276, 1993.

[HNW05] Wei Huang, Yoshiteru Nakamori, and Shou-Yang Wang. Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, 32(10):2513–2522, 2005.

[KCHP01] Eamonn Keogh, Selina Chu, David Hart, and Michael Pazzani. An online algorithm for segmenting time series. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 289–296. IEEE, 2001.

[Kim03] Kyoung-jae Kim. Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1):307–319, 2003.

[KN05] Michael J. Klass and Krzysztof Nowicki. The grossman and zhou investment strategy is not always optimal. *Statistics & Probability Letters*, 74(3):245 – 252, 2005.

[SS09] Nicholas I Sapankevych and Ravi Sankar. Time series prediction using support vector machines: a survey. *Computational Intelligence Magazine, IEEE*, 4(2):24–38, 2009.

[TC01] Francis EH Tay and Lijuan Cao. Application of support vector machines in financial time series forecasting. *Omega*, 29(4):309–317, 2001.

[Vap95] Vladimir Vapnik. *The nature of statistical learning theory.* springer, 1995.

[VV98] Vladimir Naumovich Vapnik and Vlamimir Vapnik. *Statistical learning theory*, volume 2. Wiley New York, 1998.

[YZ12] Zhaoji Yang and Liang Zhong. Optimal portfolio strategy to control maximum drawdown-the case of risk based dynamic asset allocation. *Available at SSRN 2053854*, 2012.

[ZZ09] Guofu Zhou and Yingzi Zhu. Why is the recent financial crisis a 'once-in-a-century' event? *Available at SSRN: 1462790*, 2009.