The Pennsylvania State University
The Graduate School

MACHINE LEARNING METHODS IN PORTFOLIO REPLICATION

A Thesis in
Statistics
by
Cheng You

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2013

The thesis of Cheng You was reviewed and approved* by the following:

Jason Morton
Assistant Professor of Mathematics and Statistics
Thesis Advisor

David Hunter
Head of Department of Statistics, Professor of Statistics
Department Head

Debashis Ghosh
Professor of Statistics and Public Health Sciences

*Signatures are on file in the Graduate School.

Abstract

In this article, we address the problem of portfolio replication raised by researchers in finance. We develop a new machine learning algorithm L_1 Regularized Rolling Regression and also make inference on trading strategies based on the daily return and cumulative return space. We also incorporate hierarchical modeling and hidden Markov model to refine our results. Synthetic portfolio are constructed to simulate the performance of this parametric learning method. Real data analysis are shown to prove its capability of handling complex and low frequency data. This new method could also be generalized to unwind other problems of similar kind.

Table of Contents

List of Figures

List of Symbols

X^r	The $N \times p$ daily returns of assets
X^R	The $N \times p$ cumulative returns of assets
Y^r	The $N \times 1$ daily returns of a secret portfolio
Y^R	The $N \times 1$ daily returns of a secret portfolio
β^r	The $p \times 1$ current value weight of the portfolio
β^R	The $p \times 1$ initial value weight of the portfolio
$\beta^{c,r}$	The $p \times 1$ current value weight of the portfolio converted by the initial value weight β^R
β^w	The $p \times 1$ weighted value weight of the portfolio

Acknowledgments

First of all, I am deeply indebted to my thesis advisor, Dr. Jason Morton for his guidance and encouragement throughout the process of my research. Also, I would like to express my heartfelt gratitude to Dr. John Liechty for giving me invaluable insights into the regime switch modeling.
Most importantly, I would like to thank my parents for their unwavering love and support, without which I would not have made the completion of my thesis possible.
Last but certainly not least, I would like to thank my friends and colleagues for their love and help.

Chapter 1

Introduction

1.1 Problem Description

In this paper, we consider the problem of tracking the holdings of a particular portfolio. There is a secret portfolio which reports its returns daily. We do not know either which instruments the portfolio contains or what weight each instrument has. However, we do know the vast investments vehicles available in the markets, such as bonds and stocks. To make it more intuitive, we can imagine the problem like this. There is a black box which processes some information and reveals a certain piece of information at the end. The input of the black box is, for example, the returns of a large variety of investment instruments; the output of the black box is one series of returns generated by a hidden hand in the box. Now, we want to crack this black box and see what the hidden hand looks like. This problem is also called the inverse problem: extracting information about latent variables or functions and structural parameters from observed information.

1.2 Motivation

In order to solve this problem, people have developed many methods under different assumptions to make inferences of the latent variable or parameters. One very practical method is called portfolio replication. By replication, people can expect to obtain the same or almost the same return series with only a small subset of investment vehicles instead of a wide range of them. Hence, many are motivated to replicate the performance of a particular portfolio and currently using those replication models from academia. On one hand, we can expose to the same sources of risk and return while avoiding high fees and liquidity restrictions, typically imposed by the investment vehicles; on the other hand, we may wish to hedge the risk of a significant downturn in the fund or fund of funds that has been already invested. As we cannot know the portfolio's holdings, we are forced to estimate them statistically.

1.3 Two Questions

As we think deeper, there are two main questions that we face:

1. How can we locate and weight the instruments?

2. How can we detect the strategy changes?

We address these questions on the next chapter. The remainder of this article is organized as follows. Our methodology is elaborated in Chapter 2. Our new and exciting simulation results about weight tracking and strategy inference are described in Chapter 3. To demonstrate the efficiency, the application on a real data set is presented in Chapter 4. Finally, the conclusion is made in Chapter 5.

Chapter 2

Methodology

2.1 Two Answers

For the first question we raised in Chapter 1, most practical methods for portfolio replication available today make use of a rolling regression against fewer than ten factors that are selected manually in advance. Instead, we investigated the use of L_1 Regularized Rolling Regression to choose regressors from a large universe of investment vehicles and estimate those sparse coefficients simultaneously.

For the second problem, there are actually infinitely many strategies that are possibly adopted by different investors. In order to frame these strategies into one parametric space, we simplify them into two essential types of strategies that practitioners most often use.

2.1.1 The buy-and-hold strategy

Given the total initial portfolio value of V dollars and initial weight vector β, this strategy buys $V\beta_i$ of asset i at price S_i and simply holds the assets for a certain duration. A good replication method for the buy-and-hold portfolio should quickly converge to the true weights β in terms of the number of shares $\frac{V\beta_i}{S_i}$ initially bought of each asset and perform minimal or no further trades once the true portfolio is found. The weights in terms of the value of each asset that one holds will change over time.

2.1.2 The fixed-constantly-rebalanced strategy

Given the initial weight vector β, this strategy rebalances every period to hold asset i in proportion to β_i of the total value, selling and buying as necessary. A good replication method for the fixed-constantly-rebalanced strategy will try to find β and thereby match the rebalancing trades of the true strategy each period. These two strategies provide good toy models with which to develop our estimation procedures and extend well to strategies which are more realistic.

2.2 L_1 Regularized Rolling Regression

This section provides a detailed elaboration of the first question that we mentioned previously. In order to unwind the hidden process, we develop a statistical learning method called L_1 Regularized Rolling Regression. By saying L_1 regularized regression, we adopt LASSO method to find the sparse coefficient estimates. By saying rolling, we use a rolling window with respect to time to find the coefficient estimates in a certain period.

In other words, our algorithm adopts a rolling lasso estimator on the residuals to select which portfolio weights to consider updating. To start, at each rolling window of time periods, we regress the residuals of the portfolio returns on the vast selection of investment vehicles by LASSO estimator via LARS algorithm; Then, we roll the window by one time point, calculate the new residuals and do the same regression that we did in the first step by using the updated residuals; later, we add the remaining coefficient estimates by L_1 regularized regression on residuals back to the previous coefficient estimates.

Since the LASSO estimator is biased, we then use an unconstrained regression on the selected instruments to fix the final weight updates. For simplicity, we would like to have long only portfolios. Hence, all coefficients are constrained to be positive.

To illustrate a mathematical interpretation of our method, we have the following formulations.

Let us say we start our algorithm at time t. We denote Y_t as the portfolio return vector starting at time t and ending at time t plus the rolling window length, X_t

as the return matrix of a large group of investment instruments within the same period of Y_t and β_t as the true weight vector at time t.

$$Y_t = X_t\beta_t$$

We then perform LASSO regression of Y_t on X_t and obtain the LASSO estimator $\hat{\beta}_t^{Lasso}$.

$$\hat{Y}_t^{Lasso} = X_t\hat{\beta}_t^{Lasso}$$

To eliminate bias, we perform another regression by ordinary least squares based on the sparsity obtained in $\hat{\beta}_t^{Lasso}$.

$$\hat{Y}_t = X_t\hat{\beta}_t$$

Next, we move the rolling window by one time step, calculate the residuals and use them to obtain the coefficient estimate updates.

$$\hat{\epsilon}_{t+1} = \hat{Y}_{t+1} - X_{t+1}\hat{\beta}_t \tag{2.1}$$

$$\hat{\epsilon}_{t+1}^{Lasso} = X_{t+1}\hat{\beta}_{t+1}^{*,Lasso} \tag{2.2}$$

We perform another regression by ordinary least squares as before to get unbiased coefficient estimates.

$$\hat{\epsilon}_{t+1} = X_{t+1}\hat{\beta}_{t+1}^{*}$$

Now, we add the coefficient update back to the previous one and make it our new coefficient estimates.

$$\hat{\beta}_{t+1} = \hat{\beta}_t + \hat{\beta}_{t+1}^{*} \tag{2.3}$$

If we put all formulas together we can see that it gives us the desired coefficient estimates.

$$\begin{aligned}
\hat{\epsilon}_{t+1} &= X_{t+1}\hat{\beta}_{t+1}^{*} \\
\hat{Y}_{t+1} - X_{t+1}\hat{\beta}_t &= X_{t+1}\hat{\beta}_{t+1}^{*} \\
\hat{Y}_{t+1} &= X_{t+1}(\hat{\beta}_t + \hat{\beta}_{t+1}^{*})
\end{aligned} \tag{2.4}$$

Our detailed algorithm of L_1 Regularized Rolling Regression is as follows.
Input: Return series of investible assets (X_t) and benchmark asset ($b_t = X_{t,i}\hat{\beta}_{t,i}$). A rolling window w, relative sparsity parameter α and unconstrained regression inclusion threshold ϵ.
Output: Vector of portfolio weights β_t, with respect to either initial value (the buy-and-hold strategy) or current value (the fixed-constantly-rebalanced strategy), at each time point after the initial window w.
Initial value: Parameter vector $\beta_{t-1} = 0$
For each time t:
1. Compute residuals $e_{t-w,\cdots,t} = X_{t-w,\cdots,t}\beta_{t-1} - b_{t-w,\cdots,t}$, where $X_{t-w,\cdots,t}$ is the $w \times p$ matrix of returns on the p investible assets during w periods in the rolling window;
2. Let ξ be the L_1 Regularized coefficients of $e_{t-w,\cdots,t}$ regressed on $X_{t-w,\cdots,t}$ with sparsity parameter α;
3. Set $\gamma = \beta_{t-1} + \xi$;
4. Clip all elements of $\gamma < 0$ to 0;
5. Let $I = \{i \in \{1, \cdots, p\} | \gamma_i > \epsilon\}$, the set of assets appearing with nonzero coefficients in γ;
6. Let η be the coefficients of the unconstrained regression of $e_{t-w,\cdots,t}$ on $X_{(t-w,\cdots,t),I}$, the assets in I;
7. Set $\beta_t = \beta_{t-1} + \eta$.
Since people buy and sell assets at different time points of a day and there are always transaction fees occurred in the real world, we may add a certain amount of noises to the return series in order to make simulations more realistic later on. However, we assume that the aggregations are noise-free at the current stage.

2.3 Strategy Determination

This section addresses the second question that we asked in the previous Chapter. We intend to determine what strategy an investor adopts as the window rolls.
We assume that there are two fundamental strategies that an investor uses: the buy-and-hold strategy and the fixed-constantly-rebalanced strategy. For the sake of strategy inference, we attempt to understand the key properties of the two strategies.

For the buy-and-hold strategy, the percentages of initial value weight or shares of different assets remain unchanged. Thus, we consider the cumulative return,

$$R_t = \frac{S_t}{S_0} - 1, \qquad \text{where } S_t \text{ is the asset price at time t,}$$

as an analogy to price. The β^R estimates should be the initial weight vector of values of different assets in the portfolio.
To make it mathematically sound, we provide equations for validation in the following. We first define several terms that we are going to use.

$$\begin{aligned}
\beta_i^R &= \text{Inital weight of values allocated to asset } i \text{ at time 0, equivalently weight of shares} \\
X_{t,i}^R &= \frac{S_{t,i}}{S_{0,i}} - 1 = \text{Cumulative returns of asset } i \text{ at time } t \\
V_t &= \text{Total value of the portfolio at time } t \\
Y_t^R &= \text{Cumulative returns of the portfolio at time } t
\end{aligned}$$

Second, we decompose the cumulative return of the portfolio by the initial value weight and the individual cumulative returns of all investment vehicles.

$$\begin{aligned}
Y_t^R &= \frac{V_t}{V_0} - 1 \\
&= \frac{\sum_{i=1}^{p} \frac{\beta_i^R V_0}{S_{0,i}} S_{t,i}}{V_0} - 1 \\
&= \sum_{i=1}^{p} \beta_i^R (X_{t,i}^R + 1) - 1 \\
&= \sum_{i=1}^{p} \beta_i^R X_{t,i}^R
\end{aligned} \tag{2.5}$$

Hence, regressing Y^R on X_i^R can obtain the initial value weight, or equivalently to say, the weight of shares in the buy-and-hold strategy.
For the fixed-constantly-rebalanced strategy, the percentages of values of different assets remain unchanged. Hence, we use daily return,

$$r_t = \frac{S_t}{S_{t-1}} - 1, \qquad \text{where } S_t \text{ is the asset price at time t,}$$

to make inference. The β^r estimated is the weight with respect to the changing total value of the portfolio.

Likewise, we provide equations to validate the above claim. First, new terms are defined in the following.

$$\begin{aligned}
\beta_i^r &= \text{Weight of values allocated to asset } i \text{ at time t} \\
X_{t,i}^r &= \frac{S_{t,i}}{S_{0,i}} - 1 = \text{Daily returns of asset } i \text{ at time } t \\
V_t &= \text{Total value of the portfolio at time } t \\
Y_t^r &= \text{Daily returns of the portfolio at time } t
\end{aligned}$$

Second, we decompose the daily return of the portfolio by the current value weight and all the individual daily returns of investment vehicles.

$$\begin{aligned}
Y_t^r &= \frac{V_t}{V_{t-1}} - 1 \\
&= \frac{\sum_{i=1}^{p} \frac{\beta_i^r V_{t-1}}{S_{t-1,i}} S_{t,i}}{V_{t-1}} - 1 \\
&= \sum_{i=1}^{p} \beta_i^r (X_{t,i}^r + 1) - 1 \\
&= \sum_{i=1}^{p} \beta_i^r X_{t,i}^r \qquad (2.6)
\end{aligned}$$

Hence, regressing Y_r on X_i^r can obtain the current value weight in the fixed-constantly-rebalanced strategy.

When the rolling window is in the buy-and-hold region, the initial value weight β^R estimated by cumulative returns should be constant and hence the prediction errors will be small; meanwhile, the current value weight β^r estimated by daily returns should fluctuate and hence the prediction errors will be large.

When the rolling window is in the fixed-constantly-rebalanced region, the current value weight β^r estimated by daily returns should be constant and hence the prediction errors will be small; meanwhile, the initial value weight β^R estimated by cumulative returns should fluctuate and hence the prediction errors will be large.

2.4 Measures of Analysis

In order to determine which strategy the investor adopts, we consider several measures to analyze the estimation results.

2.4.1 Measure 1: Compare β^R and β^r

This is a simple and intuitive way. We observe whether the initial value weight β^R or the value weight β^r remains constant while the other varies. If β^R is relatively unchanged and β^r varies much, we claim that the investor is doing buy-and-hold; if it is the other way around, we claim that the investor is doing fixed-constantly-rebalanced.

2.4.2 Measure 2: Compare Errors

The span of X^R forms the cumulative return space; the span of X^r forms the daily return space. When the rolling window enters different regions, the two sets of errors generated by β^R and β^r in the two spaces should differ significantly. First, we can examine the error curves to determine region switching. Second, we propose to take the ratio of the absolute value of prediction errors in the daily return space versus prediction errors in the cumulative return space. In this case, it is foreseeable that the error ratio will be significantly large in the buy-and-hold region while it will be very small in the fixed-constantly-rebalanced region. However, the two sets of errors are in different spaces and may have a scaling problem. Therefore, we can transform β^R to $\beta^{c,r}$, compute errors in the daily return space and then take the ratio. We can validate the transformation via the following formula.

$$
\begin{aligned}
\gamma^R_{t,i} &= \text{Weight of shares of asset } i \text{ at time } t \\
\hat{\beta}^{c,r}_{t-1,i} &= \frac{V_{t,i}}{\sum_{i=1}^p V_{t,i}} \\
&= \frac{\frac{S_{t,i}\gamma_{t-1,i}}{\sum_{i=1}^p V_{t-1,i}}}{\sum_{i=1}^p \frac{S_{t,i}\gamma_{t-1,i}}{\sum_{i=1}^p V_{t-1,i}}} \\
&= \frac{\frac{S_{t,i}}{S_{t-1,i}} \frac{V_{t-1,i}}{\sum_{i=1}^p V_{t-1,i}}}{\sum_{i=1}^p \frac{S_{t,i}}{S_{t-1,i}} \frac{V_{t-1,i}}{\sum_{i=1}^p V_{t-1,i}}}
\end{aligned}
$$

$$= \frac{(X^r_{t,i}+1)\hat{\beta}^R_{t-1,i}}{\sum_{i=1}^{p}(X^r_{t,i}+1)\hat{\beta}^R_{t-1,i}} \tag{2.7}$$

2.4.3 Measure 3: Compare Regions

Since we have both estimators in the same space after transformation, we can also form a weighted LASSO estimator β^w to combine the two LASSO estimators. This can lead to a simplified model. For the sake of region determination, we can perform a hierarchical regression based on the two sets of daily returns of the portfolio. In different regions, only one set of daily returns of the portfolio can be very close to the true daily returns of the portfolio. Thus, the two regression curves should differ from each other significantly. We can also employ hidden Markov model to differentiate the regions, a common approach, especially in Bayesian analysis, to form regime switch models. This type of model enables us to make inference on the change of levels.

Chapter 3

Simulation

3.1 Setup and Parameters

To simulate and examine the performance of the method, we perform a synthetic portfolio replication. Our goal is to replicate a synthetic portfolio of five equally weighted assets drawn from S&P 500. The synthetic portfolio returns are regressed against all the S&P 500 assets from June 1, 2005 to November 31, 2008. In total, there are 881 daily returns and cumulative returns, respectively, for the 500 stocks except those that do not have full time history. We construct the synthetic portfolio to be fixed-constantly-rebalanced for most periods of time but to be buy-and-hold from Day 201 to 300. The larger the rolling window is, the larger the number of data points we have in each regression process. Hence, β's are smoother and false positives are fewer. However, a larger window also results in slower convergence and because time is valuable, slower convergence is detrimental to performance. Hence, it is necessary to select a well-balanced rolling window.

We make the following definition.

$$\begin{aligned}
X^r &= N \times p \text{ matrix of daily returns with row as time and column as each asset} \\
X^R &= N \times p \text{ matrix of cumulative returns with row as time and column as each asset} \\
Y^r &= N \times 1 \text{ vector of daily returns of the synthetic portfolio} \\
Y^R &= N \times 1 \text{ vector of cumulative returns of the synthetic portfolio}
\end{aligned}$$

In our case, N is 881 and p is 500 minus the number of stocks which have missing records during the whole time period.

In the fixed-constantly-rebalanced region, we set the current value weight $\beta_i^r = 0.2,\ i = 1, \cdots, 5$; in the buy-and-hold region, we set the initial value weight $\beta_i^R = 0.2,\ i = 1, \cdots, 5$.

Empirically, the rolling window is set at $w = 30$.

The relative sparsity parameter is set at $\alpha = 0.05$. From machine learning point of view, the relative sparsity parameter specifies the learning rate. If it is large, we obtain less sparsity and the learning process may be boosted or overly adjusted; if it is small, we obtain more sparsity and the learning process may be too slow. It is often debatable to use which learning rate but empirically we would like it to be fairly small, for instance, no more than 0.05, given that we have enough computational power.

The unconstrained regression inclusion threshold is set at $\epsilon = 10^{-6}$. Without loss of generality, we consider long-only portfolios in this simulation.

3.2 Compare Betas

We can plot β^r and β^R in the daily return graph and cumulative return graph, respectively.

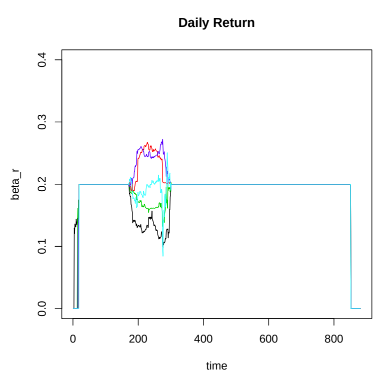

Figure 3.1. Rebalanced Weight β^r

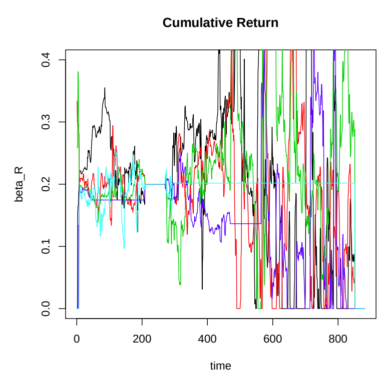

Figure 3.2. Initial Weight β^R

From the above graphs, we can observe that L_1 Regularized Rolling Regression reveals the initial value and current weights respectively when the rolling window is completely in each region, regardless of the adjustment time needed to achieve the stable weights. Based on the weight estimates, when the rolling window starts to enter the buy-and-hold region at time 171, β_r begins to fluctuate; when the rolling window enters the buy-and-hold region fully at time 200, β_R becomes stable as constant while β_r keeps fluctuating; when the rolling window starts to return to the fixed-constantly-rebalanced region at time 271, β_R begins to fluctuate; when the rolling window returns to the fixed-constantly-rebalanced region fully at time 301, β_r becomes constant again while β_R keeps fluctuating. We can use these phenomena directly to determine which strategy the investor uses and when he or she makes the change.

3.3 Compare Errors

For starters, we can examine the error graphs of β_r and β_R.

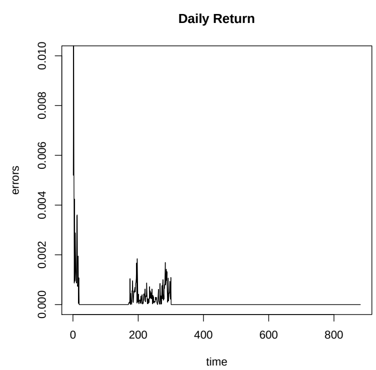

Figure 3.3. Error$_{\beta r}$

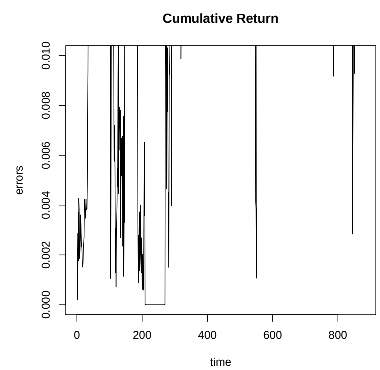

Figure 3.4. Error$_{\beta R}$

From the error graphs, we can also see that L_1 Regularized Rolling Regression determines two different regions. When the investor uses the fixed-constantly-rebalanced strategy, the errors with respect to β_r are small while the errors with respect to β_R are large.

For the scale of errors corresponding to β_r, when the rolling window starts to enter the buy-and-hold region at time 171 or leave at time 271, the magnitude of errors changes from 10^{-36} to 10^{-9}; when the rolling window completely returns to the rebalanced region at time 301, the magnitude of errors changes from 10^{-7} to 10^{-36}.

For errors corresponding to β_R, when the rolling window starts to return to the fixed-constantly-rebalanced region at time 271, the magnitude of errors changes from 10^{-29} to 10^{-7} and then immediately to 10^{-3}. Similarly, we can also use these phenomena to determine the region switch.

Let us examine the error graphs of β_r and $\beta^{c,r}$.

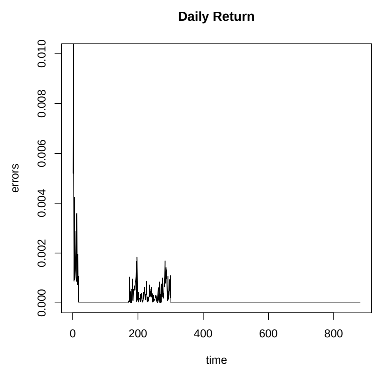

Figure 3.5. Error$_{\beta^r}$

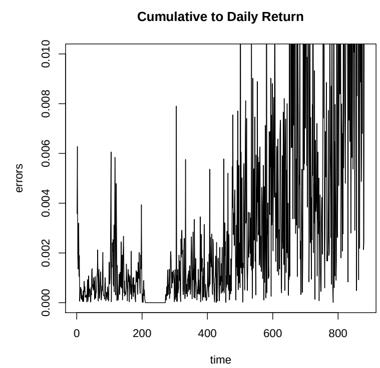

Figure 3.6. Error$_{\beta^{c,r}}$

The error and converted error graphs above gives us another perspective about the regime switch phenomena.

Second, we would like to see the error ratios with respect to β^r and $\beta^{c,r}$. We use converted errors of $\beta^{c,r}$ to compute the ratios. Notice that the converted errors produced by $\beta^{c,r}$ is off 1 index.

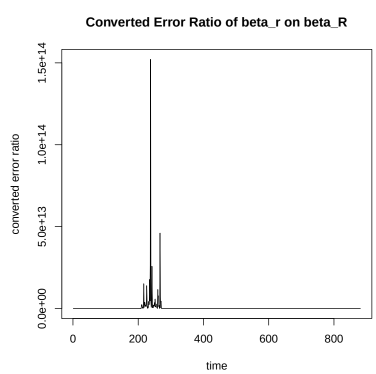

Figure 3.7. Error$_{\beta^r}$ On Error$_{\beta^{c,r}}$

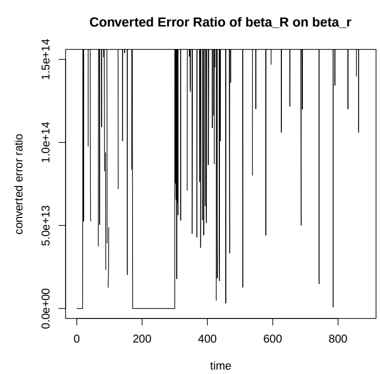

Figure 3.8. Error$_{\beta^{c,r}}$ On Error$_{\beta^r}$

To make the error ratios clearer, we can take the logarithm of the ratios.

From the error ratio graphs, we can observe that there are two different regions

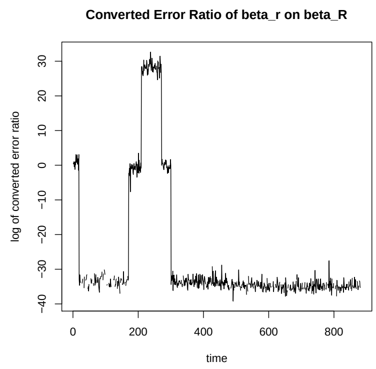

Figure 3.9. Log Error$_{\beta^r}$ On Error$_{\beta^{c,r}}$

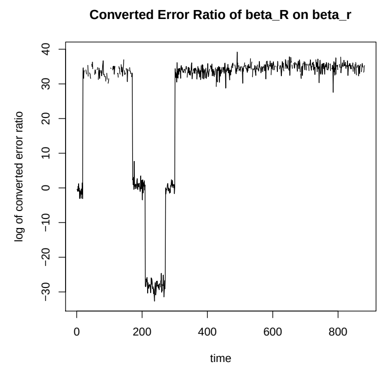

Figure 3.10. Log Error$_{\beta^{c,r}}$ On Error$_{\beta^r}$

similar to error graphs only in a finer fashion.

For converted errors ratio of daily return versus cumulative return, when the rolling window starts to enter the buy-and-hold region at time 171, the magnitude changes from 10^{-32} to 10^{-5}. In addition, when the rolling window starts to return to the fixed-constantly-rebalanced region at time 271, the magnitude changes from 10^{21} to 10^{-1}.

According to the fine difference of error ratios magnitude, we can make almost accurate inference about when region switch happens. In general, Figure 4.7 and 3.9 tells us when to stop constantly-rebalanced strategy and start using buy-and-hold; Figure 3.8 and 3.10 tells us when to stop buy-and-hold strategy and start using constantly-rebalanced.

3.4 Compare Regions

First of all, we would like to construct a combined LASSO estimator. To combine β^r and $\beta^{c,r}$, we propose a weighted LASSO estimator with a rolling window.

$$\beta_t^w = \frac{w_t^r}{w_t^r + w_t^{c,r}}\beta^r + \frac{w_t^{c,r}}{w_t^r + w_t^{c,r}}\beta_t^{c,r} \tag{3.1}$$

where

$$\beta_{t-1,i}^{c,r} = \frac{(X_{t,i}^r + 1)\hat{\beta}_{t-1,i}^R}{\sum_{i=1}^p (X_{t,i}^r + 1)\hat{\beta}_{t-1,i}^R}$$

$$w_t^r = \prod_{i=t}^{t+w} exp(-\text{Standardized Error}_i^r) \quad (3.2)$$

$$w_t^{c,r} = \prod_{i=t}^{t+w} exp(-\text{Standardized Converted Error}_i^{c,r}). \quad (3.3)$$

The graphs of weights are showed below. We can check and see that they are reasonable. Hence,

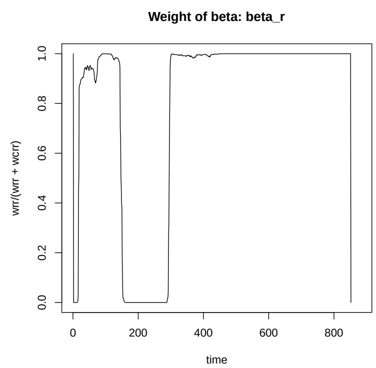

Figure 3.11. Weight of β^r

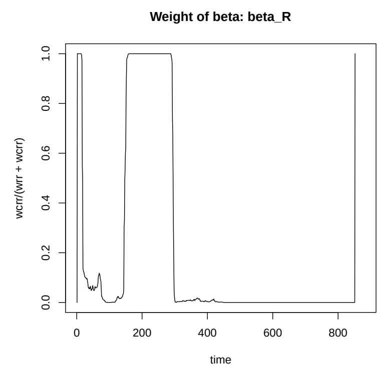

Figure 3.12. Weight of $\beta^{c,r}$

We can check and see that the weights are reasonable. Hence, the weighted LASSO estimator can be formed.

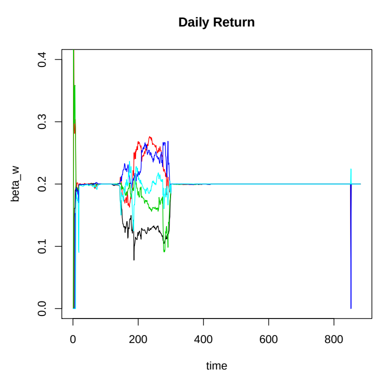

Figure 3.13. Combined Estimator β^w

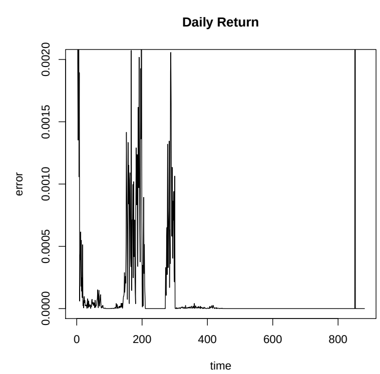

Figure 3.14. Error_{β^w}

The weighted LASSO estimator gives us another way of evaluating region switch. We can use either the different weights of β^r and $\beta^{c,r}$ or the errors of β^w to determine when the change happens and which strategy the investor uses.

Second, instead of estimating errors and combining estimators, we can simply apply the idea of hierarchical modeling. First, we obtain estimates of daily returns based on both estimators β^r and $\beta^{c,r}$; then, we regress the true daily returns on the two sets of estimates with a rolling window. The following are the relevant mathematical formulas.

$$Y^r_{t-w} = ew^r_{t-w}\hat{Y}^r_{t-w\cdots t} + ew^{c,r}_{t-w}\hat{Y}^{c,r}_{t-w,\cdots,t} \tag{3.4}$$

where

$$\begin{aligned}\hat{Y}^r_t &= X^r_t\hat{\beta}^r_t \\ \hat{Y}^{c,r}_t &= X_{r,t}\hat{\beta}^{c,r}_{t-1}\end{aligned}$$

ew^r_{t-w} and $ew^{c,r}_{t-w}$ are the coefficients of weights of the constantly-rebalanced strategy and buy-and-hold strategy, respectively. ew^r_{t-w} is in red and $ew^{c,r}_{t-w}$ is in black in the following graph.

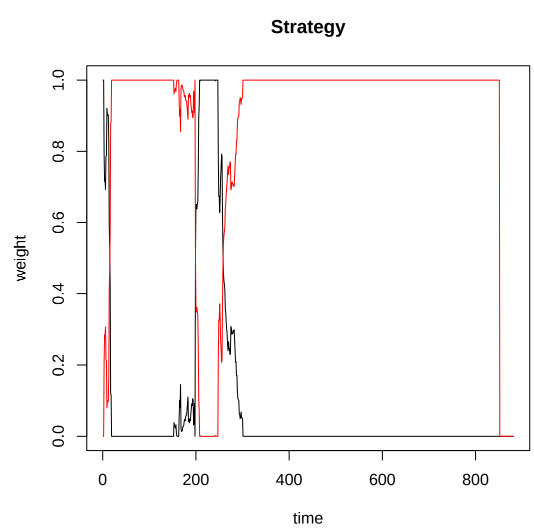

Figure 3.15. Hierarchical Regression Weight

From the hierarchical model, we can review the regions easily by looking at the coefficients estimation. Whichever coefficient is closer to 1 determines the strategy that is being used.

Last, we would also like to use hidden Markov model to determine the region since it is a standard approach to such problems. There are two steps in applying hidden Markov model. We first use EM algorithm to maximize log likelihood and then use Viterbi algorithm to determine regions. Take the fixed-constantly-rebalanced errors for example.

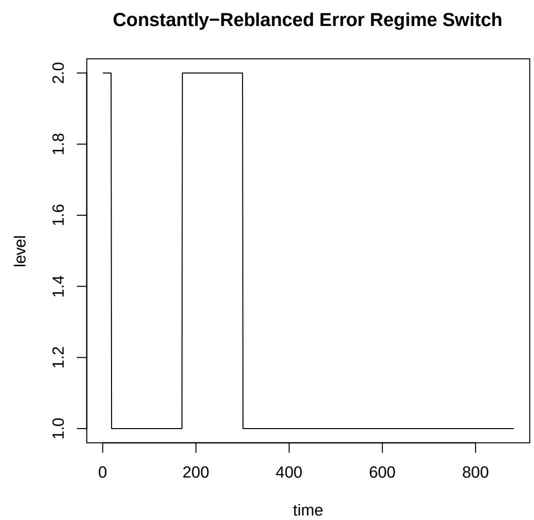

Figure 3.16. Viterbi Determined Region

We notice that hidden Markov model generates a very clean result of region switch, which works in our favor. It could be recommended to incorporate this kind of model to form a two-step modeling approach in order to determine region switch. It may also be considered to perform the two steps iteratively to obtain an equilibrium of the region switch model. More work needs to be done.

Chapter 4

Application

4.1 Data Description

In this chapter, we analyze the Venture Capital Research Index data from Thomson Reuters as an example of our application. The Thomson Reuters Venture Capital (Research) Index (In short, VC Index) tracks the monthly return of the venture capital universe and determine different industry sector weightings, which are then applied to investments in specific sector ETFs. The Thomson Reuters Standard Private Equity Data Feed (hereafter referred to as PE) defines the universe of companies and round events. Additional data are provided by Thomson One.

In our terminology, we define VC Index as our portfolio and the universe of PE as our universe of assets. Hence, the daily returns and cumulative returns can be defined as in Chapter 3. In this case, N is 213 and p is 417. The companies which have missing records during the whole time period are excluded from the universe. Without loss of economical sense and generality, we impose some constraints and assumptions. Since the investor wants diversification, we impose a cap of 0.2 for each asset. Since the investor concerns about the cost of short-selling, we still consider the long-only portfolio for replication.

4.2 Results

Because we only have monthly returns which is of low frequency, and few data points are available, we do not expect the well-behaved sparsity and stableness as in our simulation. However, this should not affect the usefulness of our result since hedge funds or ETFs rebalance their portfolio at least monthly. After tuning the parameters carefully, we have the following results.
First of all, we would like to look at the errors of fixed-constantly-rebalanced strategy and the converted errors of buy-and-hold strategy. To obtain better visualization, the smoothed errors are defined.

$$\text{Errors} = \frac{1}{w}(\sum_{i=t-w}^{t}(Y_i^r - X_i^r\hat{\beta}_{t-w}^r)^2)^{\frac{1}{2}}$$

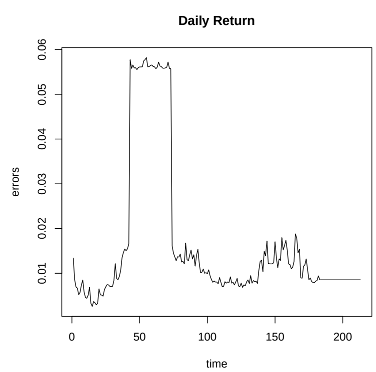

Figure 4.1. Error$_{\beta^r}$

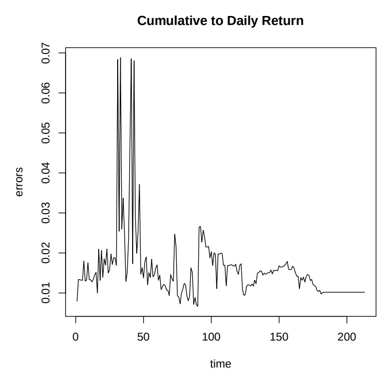

Figure 4.2. Error$_{\beta^{c,r}}$

From the graph, we would consider that from time 40 to 75, the investor should do a buy-and-hold strategy while using fixed-constantly-rebalanced strategy for the other period of time. Notice that both buy-and-hold strategy and fixed-constantly-rebalanced strategy give us small errors in some region. Thus, they are equivalent in some sense, regardless of off one index errors. This is because the initial value weights are changing every time when the window rolls. The initial value weights obtained in cumulative return space renew its initial time every time. It is in fact a rebalanced strategy as well.
Second, we want to plot the error ratios to see the switch.

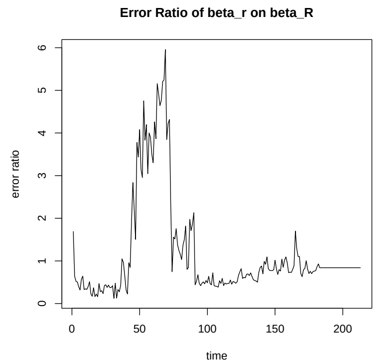

Figure 4.3. Error_{β^r} On $\text{Error}_{\beta^{c,r}}$

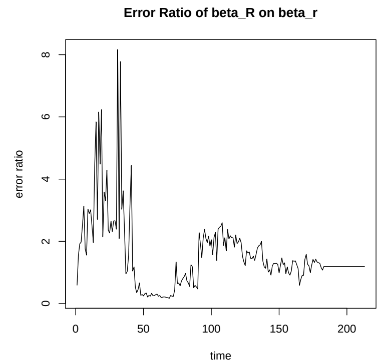

Figure 4.4. $\text{Error}_{\beta^{c,r}}$ On Error_{β^r}

We obtain similar results as in the previous graphs but only clearer for the switch area. Time 40 to 75 should be considered as the buy-and-hold region.

Third, we can construct the weighted LASSO estimator and determine the region switch by the two sets of weights of the newly combined estimator.

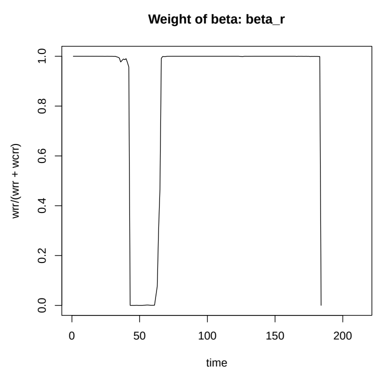

Figure 4.5. Weight of β^r

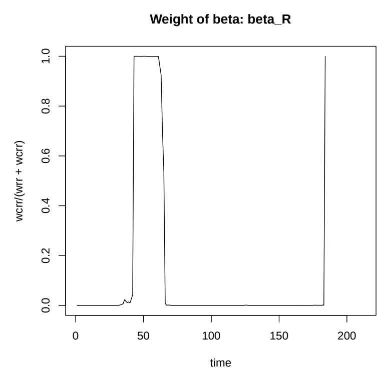

Figure 4.6. Weight of $\beta^{c,r}$

The weights with respect to each LASSO estimator show us which strategy to use in different time periods explicitly. For simplicity, we may employ the weighted LASSO estimator as our final stock picker.

Forth, we can use hidden Markov model approach as demonstrated in our simulation.

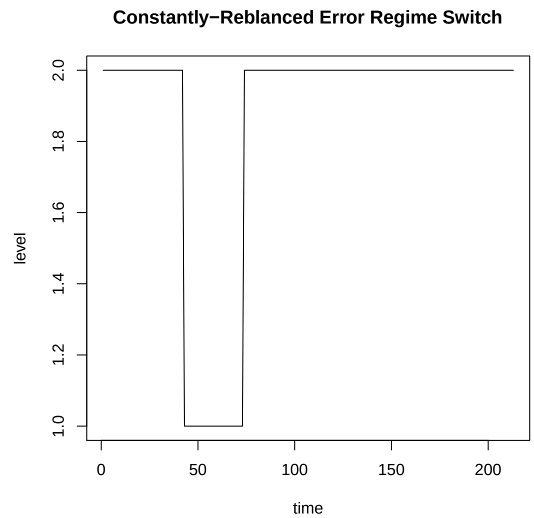

Figure 4.7. Viterbi Determined Region

The Viterbi algorithm provides us a good result of the region switch. We can use this result as our final strategy picker.

Chapter 5

Conclusion

5.1 Summary

In this thesis, we propose a new machine learning method of portfolio replication, L_1 Regularized Rolling Regression and make inference on the parametric trading strategies via two different return spaces. We demonstrate its efficiency through several different measures. In the simulation, L_1 Regularized Rolling Regression shows its fast computation and accurate properties undoubtedly and the strategy inference proves its good performance. In the real data analysis, although we do not obtain great sparsity and stableness, we can still make a lot of sense out of this new machine learning method. By incorporating other methods such as Hidden Markov Chain which are applied often in machine learning, we can have even better results of region switch. Overall, it is a good attempt to meet the frontier of machine learning and finance research.

5.2 Future Work

To begin with, instead of considering the time series regression methods by frequentist inference, we can also tackle the problem by Bayesian inference. After digging up the literatures, we can find that linear dynamic systems may be a good candidate. However, although it can estimate the distribution of each coefficient dynamically, it cannot obtain sparsity with the current techniques. Hence, we consider incorporating Bayesian Lasso into Dynamic Linear Models. The difficulty of this problem is that the Kalman Filtered mean is not easy to generalize since Bayesian Lasso requires a Laplace prior for coefficients. We are still trying to figure out.

Up to this point, we have only considered the parametric methods for portfolio replication. How about the nonparametric ones? We search again and have some interesting findings in image processing. Image data essentially resemble financial data. They are both high dimensional and massive. They both want to extract features, in our case, trading strategies. Hence, we are thinking that those non-parametric methods may be transplanted on our problem. There are two popular candidates in the image processing community: restricted Boltzmann machine and recurrent neural network. We are still looking into the details and hopefully some work will be done soon.

Bibliography

[1] Efron, B., I. Johnstone, T. Hastie, and R. Tibshirani (2004) "Least Angle Regression," *Annals of Statistics*.

[2] Tibshirani, R. (1996) "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*.

[3] Kothkari, A., A. Lai, and J. Morton (2008) "Portfolio Replication with Sparse Regression," *manuscript*.

[4] Ng, A. (2004) "L1 vs L2 regularization and rotational invariance," *In Proceedings of the Twenty-first International Conference on Machine Learning*.

[5] Lo, A. and J. Hasanhodzic (2007) "Can hedge-fund returns be replicated?: the linear case," *Journal of Investment Management*.

[6] Fung, W. and D. Hsieh (1997) "Empirical characteristics of dynamic trading strategies: the case of hedge funds," *Review of Financial Studies*.

[7] Khandani, A. and A. Lo (2007) "What Happened To The Quants In August 2007?" *Journal of Investment Management*.

[8] Brodie, J., I. Daubechies, D. Giannone, and I. Loris (2009) "Sparse and Stable Markowitz Portfolios," *National Academy of Sciences*.

[9] Khandani, A. and A. Lo (2011) "What Happened To The Quants In August 2007?: Evidence from Factors and Transactions Data," *Journal of Financial Markets*.

[10] Tropp, J. A. (2006) "Just relax: Convex programming methods for identifying sparse signals," *IEEE Transactions On Information Theory*.

[11] Tropp, J. A. and S. J. Wright (2010) "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE: Applications of Sparse Representation and Compressive Sensing*.