ON SOME ASPECTS OF PORTFOLIO MANAGEMENT

By

Mengrong Kang

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirement
for the degree of

Statistics-Master of Science

2013

ABSTRACT

ON SOME ASPECTS OF PORTFOLIO MANAGEMENT

By

Mengrong Kang

We study the on-line portfolio and the stochastic portfolio investment algorithms and test them with historical data sets. With regard to the stochastic portfolio we develop an optimal formula to manage the portfolio with daily trading in terms of the weights that are assigned to the different stocks in the portfolio. The implementation of the optimal stochastic portfolio depends on good estimation of the parameters that are in our case drifts and volatilities. We present some procedures to estimate the parameters dynamically. The problem of estimating drifts is inherently very hard as the noise (volatility) overwhelms the drifts. Volatilities are easier to estimate than the drifts and we can take advantage of the unique properties of the Brownian motion process to get pretty good estimates taking into account the decreasing effects of older financial data. Then we apply Karush–Kuhn–Tucker Theorem to get the weights of the optimal stochastic portfolio using the estimators. Finally we compare the results of the stochastic portfolio to that of the on-line portfolio using real stock data that now is widely available. In some cases the results achieved by the stochastic portfolio on real historical data are stunning.

ACKNOWLEDGEMENTS

Thanks to my thesis adviser Professor Shlomo Levental for teaching and advising me with this interesting thesis topic, and spending a lot of time to discuss and provide comments about the topic. Thanks to Professor Raoul LePage for helpful comments on the on-line portfolio. Thanks to Lening Kang for providing us with the stock market data used in our experiments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

There are many methods to manage portfolios. In this work we will concentrate on two methods: the on-line portfolio method and the stochastic portfolio method. We prefer those two methods because they are more mathematical / statistical in nature. Each of them has its own advantages and disadvantages and this will be one of the main topics covered in this work. We will focus mainly on the stochastic portfolio where the problems are both probabilistic and statistical and there is some interaction between the two. We will develop some interesting procedures of implementation the theory. We also tested the procedure on real historical data and found some very interesting results.

Here are some historical remarks of earlier works in portfolio management. The classical theory of portfolio management is based on methods that consider the mean and variance/correlation of stock prices. The mean-variance approach is the basis of the Sharpe Markowitz (Sharpe et al. [12-14] [18], 1959) theory of investments in the stock market which is used by business analysts to develop a single period equilibrium model, the Sharpe Lintner capital asset pricing model (CAPM) (Sharpe [18], 1964; Lintner [11], 1965; R. C. Merton [15], 1973). The first moment (mean) of the random annual return of a portfolio gives us information on the expected long term behavior under i.i.d assumptions of the prices relatives, when we keep the weight of stocks fixed in time. However, in stock markets one normally reinvests every day so that the total wealth achieved is a product of the individual wealth achieved on each day. Besides, Rosenberg and Ohlson [17], 1976, showed that the dynamic interaction between investors' behavior and the behavior of stocks led to internal inconsistencies in the continuous time CAPM. Also, it is now well-known that the future behavior of stock markets is not independent of the past. For the above reasons, distributional methods (Kelly [9], 1956; Bell and Cover[2], 1980; Cover [3], 1984; Cover and Gluss [4] , 1986; Algoet [1] , 1992), that use adaptive investment strategies for rebalanced portfolios, have been developed.

Here is the way the thesis is organized. In Section 2, we describe the method of on-line portfolio [8] which does not use any distribution assumptions. In Section 3, we use a distribution assumption, Brownian motion based, and we present the target that the portfolio manager has to optimize in terms of the weight of the portfolio. We also show how to achieve this target using Karush–Kuhn–Tucker (KKT) Theorem. In Section 4, we deal with estimating of the parameters in the stochastic model, which is crucial in the implementing optimal solution. Finally, there is a short summary in Section 5.

2. On-line Portfolio

This method was initiated by Cover's seminal paper [5] who presented the "universal portfolio". His approach, while superior in theory as we learn from the theorems that are proved in the paper, is in reality not easy to implement. Helmbold, Schapire, Singer and

Warmuth [8] managed to create a version of the universal portfolio, known as “on-line portfolio”, that is much simpler to execute, even though their theoretical results are not as good as Cover’s. Ironically they got better results than the universal portfolio when they experimented with some real market data. They described an on-line portfolio selection using multiplicative updates, and achieved almost the same wealth as the best constant rebalanced portfolio (i.e. the weights of the stocks in the portfolio are constants) which, in particular, is better than the best performing stock in the portfolio.

The following simple example (Helmbold et al. [7] , 1996) demonstrates the power of constant-rebalanced portfolio strategies. Assume that two investments are available. The first asset is a risk-free, no-growth investment stock whose value never changes. The second investment is a hypothetical highly volatile stock. On even days, the value of this stock doubles and on odd days its value is halved. The relative prices of the first stock can be described by the sequence $1, 1, 1, \dots$ and the those of the second by the sequence $\frac{1}{2}, 2, \frac{1}{2}, 2, \dots$ Neither investment alone can increase in value by more than a factor of 2, but a strategy combining the two investments can grow exponentially. One such strategy splits the investor’s total wealth evenly between the two investments, and maintains this even split at the end of each day. On odd days the relative wealth decreases by a factor of $\left(\frac{1}{2}\right) * 1 + \left(\frac{1}{2}\right) * \left(\frac{1}{2}\right) = \frac{3}{4}$. However, on even days the relative wealth grows by $\left(\frac{1}{2}\right) * 1 + \left(\frac{1}{2}\right) * 2 = \frac{3}{2}$. Thus, after two consecutive trading days the investor's wealth grows by a factor of $\left(\frac{3}{4}\right) * \left(\frac{3}{2}\right) = \frac{9}{8}$. It takes only twelve days to double the wealth, and over $2n$ trading days the wealth grows by a factor of $\left(\frac{9}{8}\right)^{n}$.

In this work we represent a portfolio of stocks as a vector of relative prices $x=(x_1,x_2,\ldots,x_N)$ where x_i is a relative price of the i-th stock, i.e. the ratio of the next day's opening price to its opening price on the current day, and N is the number of stocks. Let Z be the value of the portfolio, and let w_i be the weight (proportion) invested in the i-th stock. We assume that $w_i > 0$ and $\sum w_i = 1$. We can represent Z as a positive valued combination of assets that have the identity

$$Z = w_1 Z + w_2 Z + \cdots + w_N Z$$

Then the ratio of the portfolio value at the next day to the value at the current day can be defined as

$$z = w \cdot x = \sum w_i x_i$$

Since we are going to trade once a day, we will assume

$$x_i = \frac{dX_i}{X_i} + 1$$

where X_i is the price of i-th stock.

We are going now to describe the on-line portfolio selection. Let N denote the number of stocks in the portfolio, $\beta = \left(\frac{N^2 \log N}{8T}\right)^{\frac{1}{4}}$ where T is the number of trading days, and let $1 = (1, \ldots, 1)^t$ denote a vector of length N, then the on-line portfolio algorithm modifies the price relatives at the t-th day as

$$\widetilde{x^t} = \left(1 - \frac{\beta}{N}\right) x^t + \left(\frac{\beta}{N}\right) 1 \ ,$$

and then select portfolio weights by using the vector

$$\widetilde{w^{t+1}} = (1 - \beta) w^{t+1} + \left(\frac{\beta}{N}\right) 1 \ ,$$

where

$$w_i^{t+1} = \frac{w_i^t \exp\left(\frac{\eta \tilde{x}_i^t}{w^t \cdot \tilde{x}^t}\right)}{\sum_i w_i^t \exp\left(\frac{\eta \tilde{x}_i^t}{w^t \cdot \tilde{x}^t}\right)} \ ,$$

$$\eta = \sqrt{\frac{8(\beta^2)\log N}{(N^2)T}} \ .$$

The following theorem is proved in Helmbold et al paper.

Theorem. If w^1 is the uniform proportion vector, $T \geq 2N^2 \log N$, then we have

$$\sum_{t=1}^{T} \log\left(\tilde{w}^t \cdot x^t\right) \geq \sum_{t=1}^{T} \log\left(u \cdot x^t\right) - 2(2N^2 \log N)^{\frac{1}{4}} T^{\frac{3}{4}}$$

where u is the weights of the optimal constant-rebalanced portfolio.

In Helmbold et al paper, the on-line portfolio algorithm performs better than the universal portfolio when tested on historical data of several portfolios. This means that the log-return after T days using the updated weight vectors, $\sum_{t=1}^{T} \log\left(\tilde{w}^t \cdot x^t\right)$, is larger than the one achieved by Cover's algorithm. The result seems to be inspiring. But when studying the updated weights $\widetilde{w}^t$, we find that the weights always remain around the initial weights we choose on the first day. That is if we set $\eta = 0.05$, which Helmbold et al suggested, even when considering an extreme situation in which the portfolio has a very bad stock and a very good stock.

For example, we study the HP, Kodak and money market portfolio in 20 years (from 1992 to 2011) using on-line portfolio method with initial weight 1/3 for each asset. The annual returns of HP, Kodak and money market are 0.08698516, -0.1855094 and 0.03463049, while the annual return of the on-line portfolio is 0.02337952, far less than the annual return of HP stock and even the riskless rate. See Table 1a for the weight summary. Although the portfolio tends to invest more in the better stock, HP, the difference between the weights invested in different stocks is quite small. But comparing the prices of these two stocks during those 20 years (Figure 1a), it shows clearly that investing much more in HP than in Kodak during the last 10 years is a better option.

In another example, we tested the on-line portfolio on IBM and Ford in 20 years with initial weight 1/2 for each asset. The annual returns of IBM and Ford are 0.1181272 and 0.05999252, while the annual return of the on-line portfolio is 0.1135267, still less than the annual return of IBM stock. See Table 2b for the weight summary. Although the portfolio tends to invest more in the better stock, IBM, the difference between the weights invested in different stocks is tinny that can be ignored. But comparing the prices during those 20 years shown in Figure 1a, clearly investing much more in IBM than in Ford is a better option.

Table 1a

	Money Market	HP	Kodak
Min.	0.3209	0.3323	0.2944
1st Qu.	0.3258	0.3419	0.3196
Median	0.3288	0.3469	0.3253
Mean	0.3289	0.3476	0.3235
3rd Qu.	0.3313	0.3538	0.3316
Max.	0.3429	0.3630	0.3358

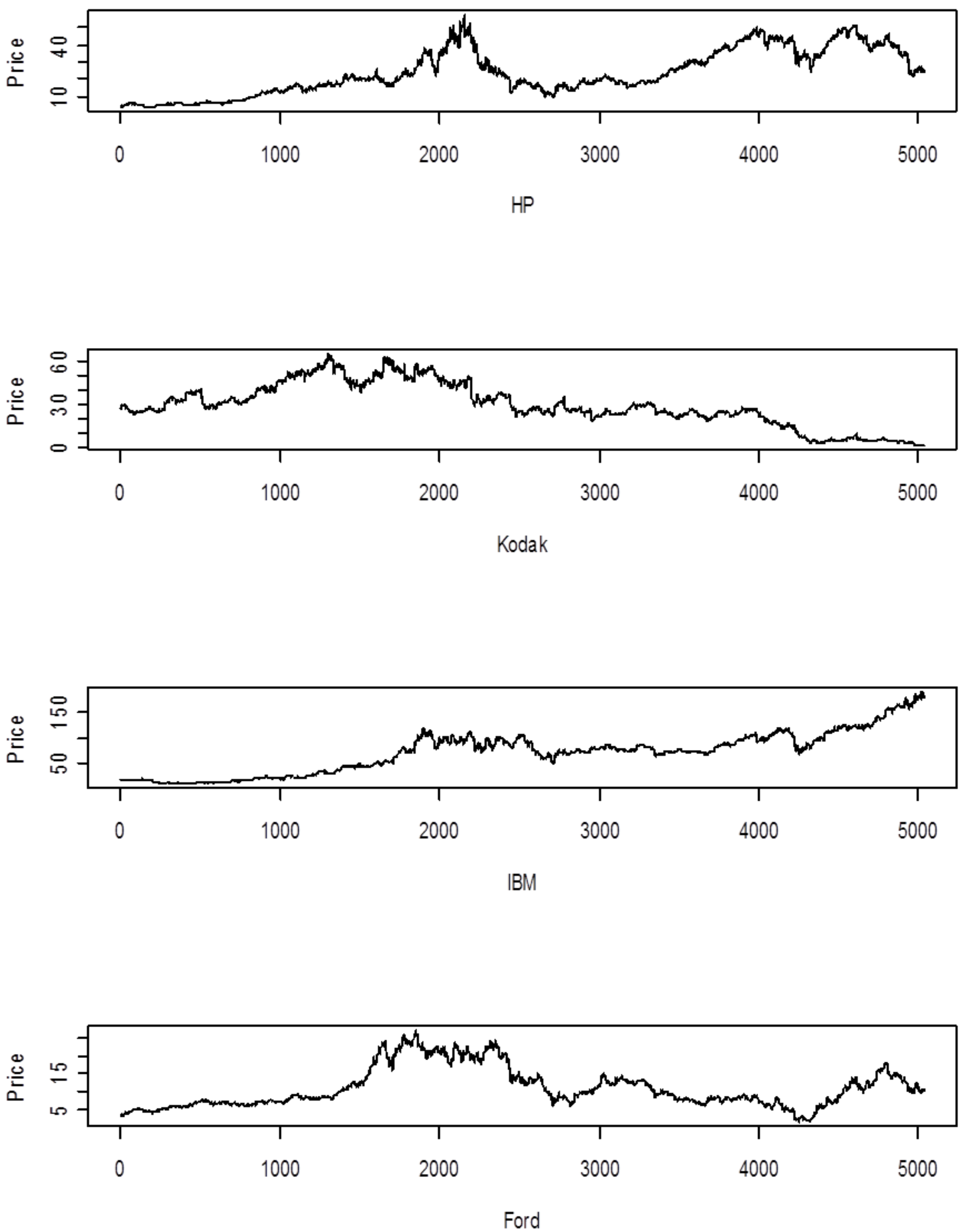
Weight summary of on-line portfolio: money market, HP and Kodak

Table 2b

	IBM	Ford
Min.	0.4846	0.4747
1st Qu.	0.4946	0.4933
Median	0.5004	0.4996
Mean	0.5007	0.4993
3rd Qu.	0.5067	0.5054
Max.	0.5253	0.5154

Weight summary of on-line portfolio: IBM and Ford

Figure 1a



Stock price of HP, Kodak, IBM and Ford

The main reason that the on-line portfolio did not perform satisfactorily in the example above is that the on-line portfolio selection using multiplicative updates follows Kivinen and Warmuth[10] with the idea that good performance can be achieved by choosing a vector w^{t+1} that is "close" to w^{t} . The limitation of this approach results sometimes in a situation, especially when the market has high level of volatility, in which the portfolio cannot change fast enough and fails to compete against the best single constant-rebalanced portfolio or even the money market for a long time.

3. A Stochastic Portfolio and its Optimization

From the Section 2, we learn the importance of taking into account the volatility of the market. As we saw the on-line portfolio did not perform well in an environment of high volatility. In order to solve this problem we will use a stochastic model. The first approach in this direction is the Shape Markowitz mean-variance method which deals with discrete-time. When it comes to continuous-time, which is more appropriate, Fernholz and Shay [6] used stochastic portfolio theory to emphasize the long term performance of portfolios. Typically, the objective is to maximize annual average return of the portfolio in the long run, i.e. maximize $\frac{\log(Z_T)}{T}$, where Z_T is the value of the portfolio after T years. We suppose that Z can be expressed as an Ito differential

$$dZ = w_1 Z dX_1/X_1 + w_2 Z dX_2/X_2 + \cdots + w_N Z dX_N/X_N$$

The stochastic portfolio theory assumes that the logarithm of each stock price follows a diffusion process with random drift and volatility. Also, the covariances between stock prices are random. The model of stock prices is as follows:

$$\frac{dX_t^i}{X_t^i} = \alpha_t^i dt + \sigma_t^i dB_t^i, \quad i = 1, \dots, N$$

$$dB_t^i dB_t^j = \rho_t^{ij} dt, \quad i \neq j, \left|\rho_t^{ij}\right| \leq 1$$

where B_t^i is a standard Brownian motion. We also have money market that pays interests rate denoted by r_t, which is considered as a riskless rate of return. Let X_t^0 be the money accumulated in the money market from an initial investment of $1 :

$$\frac{dX_t^0}{X_t^0} = r_t dt$$

When we are using the model, we will trade once a day and we will assume that during a day the drifts, volatilities and covariances are constants. In what follows, we sometimes drop the time notation with the understanding that everything happens during a day.

From the assumption, the stock prices can be easily formed as below:

Applying Ito's Lemma, we get for each stock price(omitting the i notation)

$$X(t) = X(0)\exp\left[\int_0^t \left(\alpha_s - \frac{{\sigma_s}^2}{2}\right) ds + \int_0^t \sigma_s dB(s)\right]$$

If we denote the length of a day by dt we can model the changing of the stock price during a day by

$$X(t+dt) = X(t)\exp\left[\left(\alpha_t - \frac{{\sigma_t}^2}{2}\right)dt + \sigma_t dB(t)\right]$$

We can also present $dB(t) = \sqrt{dt}\,\mathbb{Z}$, where $\mathbb{Z} \sim N(0,1)$.

And since the average annual return of a stock is

$$\frac{\log\left(\frac{X(t)}{X(0)}\right)}{t} \sim \frac{\int_0^t \left(\alpha_s - \frac{{\sigma_s}^2}{2}\right)ds}{t} \quad \text{as } t \to \infty$$

while the expectation

$$\frac{\log\left[E\left(\frac{X(t)}{X(0)}\right)\right]}{t} = \frac{\log\left[E\left(\exp\left\{\int_0^t \alpha_s ds\right\}\right)\right]}{t}$$

then we see if we will only consider the drift α while we are managing the portfolio, we take a huge risk that is caused by ignoring the volatility. In other words, the expectation itself is not the most important factor here.

Next we will deal with the dynamics of Z_t, the process that represents the value of the portfolio at time t. We write

$$Z = w_1 Z + w_2 Z + \cdots + w_N Z + w_0 Z$$

where w_i, i=1,…,N, is the weight put in i-th stock, and w_0 is the weight kept in money market. The self-finance assumption implies (we drop the time notation)

$$d(w_i Z) = \frac{w_i Z}{X_i} dX_i \,,\, i = 0,1,\ldots,N$$

Then the dynamics of Z are given by SDE (Stochastic Differential Equation)

$$\frac{dZ}{Z} = \sum_{i=0}^{N} w_i \frac{dX_i}{X_i} = \sum_{i=1}^{N} w_i(\,\alpha_i dt + \sigma_i dB_i) + w_0 r dt$$

It can be solved, based on the Ito's lemma, as

$$Z(t) = Z(0)\exp\{\sum_{i=1}^{N}[\int_{s=0}^{t} w_i(s)\alpha_i(s)ds + \int_{s=0}^{t} w_i(s)\sigma_i(s)\,dB_i(s)] + \int_{s=0}^{t} w_0(s)r(s)ds - \frac{1}{2}\int_{s=0}^{t}\left[\frac{dZ(s)}{Z(s)}\right]ds\}$$

where for any time s,

$$\left[\frac{dZ}{Z}\right] = \left(\sum_{i=1}^{N} w_i \frac{dX_i}{X_i} + w_0 r dt\right)^2 = \left(\sum_{i=1}^{N} w_i \frac{dX_i}{X_i}\right)^2 = \sum_{i,j=1}^{N} w_i w_j \sigma_{ij} dt$$

and

$$\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$$

Assuming mild restrictions on the volatilities (uniformly bounded) we get by the Law of Large Numbers

$$\frac{\sum_{i=1}^{N}\left[\int_{s=0}^{t} w_i(s)\sigma_i(s)dB_i(s)\right]}{t} \to 0 \text{ as } t \to \infty.$$

So by considering a long term, $t \to \infty$, we get

$$\frac{\log\left(\frac{Z(t)}{Z(0)}\right)}{t} \approx \frac{\int_{s=0}^{t}\left[\sum_{i=1}^{N} w_i(s)\alpha_i(s) + w_0(s)r(s) - \frac{1}{2}\sum_{i,j=1}^{N} w_i(s)w_j(s)\sigma_{ij}(s)\right] ds}{t}$$

The instant growth rate of Z at time s is given by

$$r_Z(s) = \sum_{i=1}^{N} w_i(s)\alpha_i(s) + w_0(s)r(s) - \frac{1}{2}\sum_{i,j=1}^{N} w_i(s)w_j(s)\sigma_{ij}(s)$$

Therefore, the difference between the growth rate of Z and the riskless rate, so called excess growth, is given by

$$r_Z^*(s) = r_Z(s) - r(s)$$
$$= \sum_{i=1}^{N} w_i(s)[\alpha_i(s) - r(s)] - \frac{1}{2}\sum_{i,j=1}^{N} w_i(s)w_j(s)\sigma_{ij}(s)$$

All the arguments above lead to the following theorem which is compatible with Kelly principle [9]

Theorem. The portfolio that maximizes the long term average annual return is given at time s by

$$(w_i(s))_{i=1,\dots,N} = \mathrm{argmax}_{(w_i \geq 0, \sum_1^N w_i \leq 1)}\{\sum_{i=1}^{N} w_i[\alpha_i(s) - r(s)] - 12i,j=1Nwiwj\sigma ijs\}$$

(1)

The rest of the section is devoted to solve Equation (1). The solution is based on

Karush–Kuhn–Tucker (KKT) Theorem which is an extension of Lagrange Multipliers. Here is the theorem[19]:

Theorem (KKT). Suppose that the objective function $f: \mathbb{R}^n \to \mathbb{R}$ and the constraint functions $g_i : \mathbb{R}^n \to \mathbb{R}$ and $h_j : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable at x^*. If x^* is a local maximum that satisfies conditions

$$g_i(x^*) \le 0, \text{for all } i = 1, \dots, m$$

$$h_j(x^*) = 0, \text{for all } j = 1, \dots, l$$

then there exist constants KKT multipliers $\mu_i \ge 0$ and Lagrange multipliers λ_j, such that

$$\nabla f(x^*) = \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(x^*)$$

$$\mu_i g_i(x^*) = 0 \text{ for all } i = 1, \dots, m$$

Following this theorem, we start with a system

$$A \cdot w = \alpha - r$$

which implies

$$w = A^{-1} \cdot (\alpha - r) \tag{2}$$

where

$w = (w_1, \dots, w_N, w_0)^t$,

$A = \left(a_{i,j}\right)$ is a $(N + 1) * (N + 1)$ matrix defined by

$$a_{i,j} = \begin{cases} \sigma_{i,j}, & \text{if } 1 \le i, j \le n \\ 1, & \text{if } i = n+1 \\ 0, & \text{if } i < n+1 \text{ and } j = n+1 \end{cases},$$

$$\alpha - r = (\alpha_1 - r, \dots, \alpha_n - r, 1)^t$$

If the weights given by (2) satisfy $w_i > 0, \ \forall i$, then we are done. Assume now that $I^C = \{0 \le j \le N : \ w_j \le 0\}$ is not empty. In that case we still work with the equation $A \cdot w = \alpha - r$, but we have to modify it. Start by modifying the vector $w = (w_1, \dots, w_N, w_0)^t$. Do it by replacing w_j by $\lambda_j, \ j \in I^C$ (λ_j is called a Lagrange multiplier). Then modify the matrix $A = (a_{i,j})$ as follows:

If $w_j \le 0, \ 1 \le j \le N$ then replace the j-th column of A by $(0, \dots, 0, -1, 0, \dots, 0)^t$. In other words after the modification we get $a_{j,j} = -1$ and $a_{i,j} = 0, i \ne j$. Finally, if $w_0 \le 0$ then replace the (n+1)th column by $(1, \dots, 1, 0)^t$, namely after the modification we get $a_{N+1,N+1} = 0$ and $a_{i,N+1} = 0, 1 \le i \le N$.

One thing that should get attention is that the solution is KKT solution if $\lambda_j \ge 0, j \in I^C$. Also KKT is only a necessary condition for optimal solution, i.e. it isn't sufficient. This means that in theory we need to check all the KKT solutions and select the best of them.

4. Estimations of Drifts, Volatilities and Correlations

Now the problem turns to estimation/prediction of the vector α and the matrix of σ of the next day for the stocks in the portfolio. When we look at the daily return, we see that

$$E\left(\frac{dX}{X}\right) = \alpha dt, \quad s.d.\left(\frac{dX}{X}\right) = \sigma\sqrt{dt}$$

Since $\sqrt{dt}$ is much larger than dt, namely the volatility in a day is much larger than the drift. This makes that estimating the drift a very challenging assignment. On the other hand, estimating the drift is crucial since it is important component of Formula (1). In other words, after many years when the noise is less important, the current value of the portfolio will depend on the drift estimates that the portfolio manager has used many years before the current time. The conclusion is that we must estimate the drift to the best of our ability even though statistically speaking it is almost “mission impossible”.

Another problem with the drift is that even though it looks that the drift should be basically constant for long duration, there is a possibility that the drift will change by huge number throughout relatively short time before it will go back to more “normal” level. For example, Ford’s share price dropped from $ 12.83 to $ 9.68 in 10 days, which is almost 25% loss. The estimated standard deviation of the relative price during the 10 day is around 50% larger than usual, but the thing that catches attention is that a reasonable estimate of the drift during the 10 days is -6.8348, which in absolute value is almost 100 times larger than the stock market (S&P 500) drift. In this case, predicting the drift and volatility of next day is difficult but extremely important, so we need to carefully pick up the useful historic data and select suitable estimations.

4.1 Basic Estimation

One simple way is to take the average, standard deviations and correlations of the daily returns of the stocks in the portfolio of last n days to estimate the current day's α and σ. In what follows $R_{n-i} = \frac{dX(t-idt)}{X(t-idt)}$ denotes the daily return of the i-th day before the current day. We get the following estimates (the first 2 are the drift and volatility for each stock, while the third one is the covariance estimate for stocks k and l).

$$\begin{cases} \widehat{\alpha_t} = \frac{1}{n*dt}\sum_{i=1}^{n} R_i \\ \widehat{\sigma_t} = \sqrt{\frac{1}{n*dt}\sum_{i=1}^{n}(R_i - \widehat{\alpha_t}dt)^2} \\ \widehat{\sigma_t^{k,l}} = \frac{1}{n*dt}\sum_{i=1}^{n}(R_i^k - \widehat{\alpha_t^k}dt)(R_i^l - \widehat{\alpha_t^l}dt) \end{cases}$$

(3)

where $dt = \frac{1}{250}$, which is the time proportion of one day over the trading days in one year.

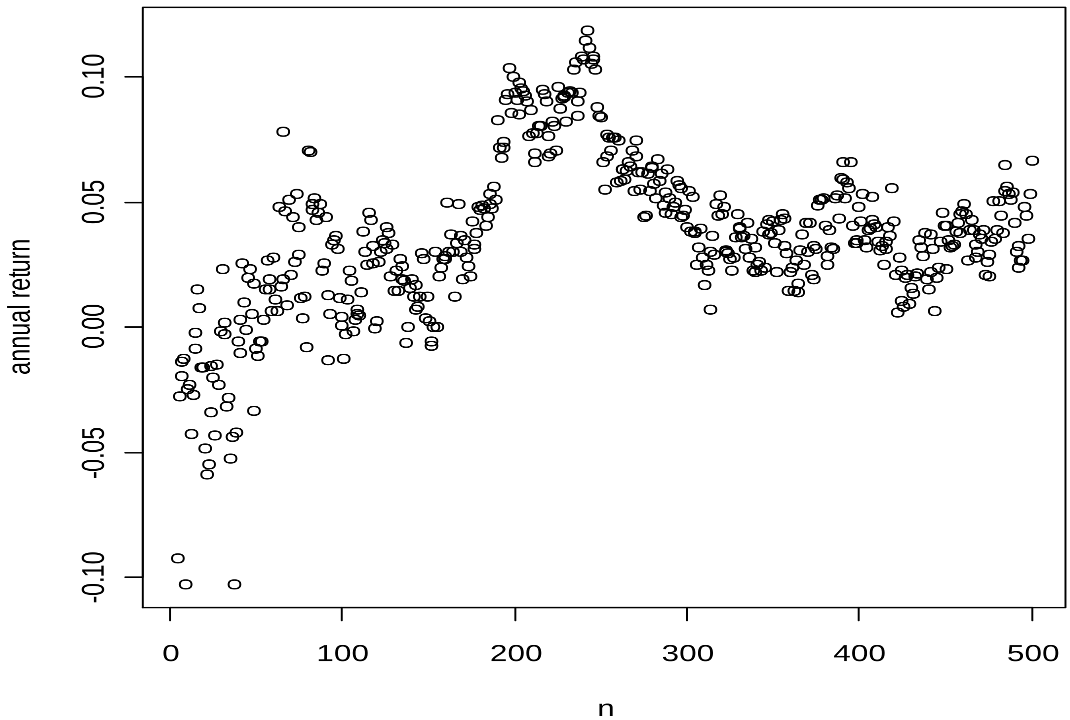
When we tested the effect of estimates (3) on managing a portfolio with historic real data, using different number of days, the return of the portfolio changed a lot. Taking the HP, Kodak and money market portfolio as an example, the annual return is quite different with different n (shown in Figure 2b), and the range is from -0.10300 to 0.11890. The maximized annual return 0.11890 happens when n is 242. While when consider a portfolio of IBM and Ford, without using money market, the maximized annual return, 0.2033101, can be achieved by n=16, which is surprisingly small number. These results mean that, although the stochastic portfolio with estimate (3) sometimes has an amazing return, i.e. the annual return of the portfolio of IBM and Ford with n=16 is two times of the annual return of the on-line portfolio, estimate (3) is not good enough because we do not know how many days to use, and the result of managing the portfolio is based to a large extent on the number of days that we use. To sum up, the question is: Which n should we use? We will talk about it later.

One simple improvement that some people may like is to replace the average used in estimate (3) by weighted average, which emphasizes the importance of recent days over a more distant past. Specifically, take $0 < \beta < 1$ and use

$$\begin{cases} \widehat{\alpha_t} = \dfrac{\sum_{i=1}^{n} \beta^i R_{n-i}}{dt \sum_{i=1}^{n} \beta^i} \\ \widehat{\sigma_t} = \sqrt{\dfrac{\sum_{i=1}^{n} \beta^i (R_{n-i} - \widehat{\alpha_t} dt)^2}{dt \sum_{i=1}^{n} \beta^i}} \\ \widehat{\sigma_t^{k,l}} = \dfrac{\sum_{i=1}^{n} \beta^i \left(R_{n-i}^k - \widehat{\alpha_t^k} dt\right)\left(R_{n-i}^l - \widehat{\alpha_t^l} dt\right)}{dt \sum_{i=1}^{n} \beta^i} \end{cases} \tag{4}$$

We do not believe that (4) will make a big improvement over (3) with regard of managing the portfolio.

Figure 2b



Annual return of the portfolio using different number of days

Another improvement of the drift estimate in (3) will be to replace the model of simple averaging by a model of linear regression. We assume that

$$E(R_i) = (a + b * i)dt, \quad i = 0, \dots, n-1$$

where a and b are unknown parameters. Then we use the standard linear regression estimate to find $\hat{a}$ and $\hat{b}$, and the prediction of $\widehat{\alpha_t} = \hat{a} + \hat{b} * n$. We recognize that the classical assumptions of linear regression model do not necessarily hold here, however due to the great difficulty in estimating the drift in a reasonable way (explained above) it makes sense to try a linear regression approach and test how the estimates achieved in this way will influence the results of portfolio management.

Remark. A test on a real data showed that a portfolio that was using the linear regression to estimate the drift didn't get a better annual return than estimates (3) or (4).

4.2 Advanced Estimation

After managing portfolio with real data based on the estimate method in Section 4.1, we found that the best number of days to use in the estimates, n, is many times (but not always) around 250, in portfolios that use money markets as one of the possible investments. On the other hand it looks that estimating the volatility is an easier task than estimating the drift and, in fact, we can estimate the volatility in a reasonable way when we keep the number of days that we use in the estimate relatively small and fixed.

Let us denote by n_σ and n_α the number of days that we use to estimate σ and α, respectively. So based on what we said above, it makes sense to let n_σ be fixed, while n_α will be dynamic and will be based on the data that comes from the market.

We tried the idea of making n_α dynamic in real data using the HP, Kodak and money market portfolio as example. We wanted to figure out each day whether to use small number of days or large number of days to estimate the drift.

Specifically we used a statistical test to decide if there is a significant difference between using $n_\alpha = 250$ and $n_\alpha = 20$ to estimate the drift using (3). If the significant difference exists, it means that α changes to a new level, and it will not make sense to use a drift estimate based on the historical data originated 20 days ago. In order to solve the significant difference problems, we came up with more advanced estimates.

Estimate of the volatility: From the SDE of the share price, we get

$$\mathrm{dlog}(X_t) = (\alpha_t - \frac{\sigma_t^2}{2})\mathrm{dt} + \sigma_t \mathrm{dB}_t$$

By subtracting $\mathrm{dlog}(X_t)$ from the equation of $\frac{dX_t}{X_t}$ we have

$$\frac{dX_t}{X_t} - \mathrm{dlog}(X_t) = \frac{\sigma_t^2}{2}\mathrm{dt}$$

Then we get an estimate of σ_t by using (X_t , X_{t+1}) which are the share price in days number $t-1$ and t respectively.

$$\widehat{\sigma_t} = \sqrt{\frac{2}{\mathrm{dt}}[\frac{X_{t+1}}{X_t} - \log\left(\frac{X_{t+1}}{X_t}\right) - 1]}$$

As a result, we can predict σ_{t+1} by $\widehat{\sigma_t}$ or more generally we may use a moving average $\widehat{\sigma_{t+1}}^2 = \frac{\sum_{j=0}^{n_\sigma - 1} \widehat{\sigma_{t-j}}^2}{n_\sigma * dt}$ of our estimates of the last n_σ days. We can even use weighted average

$$\widehat{\sigma_{t+1}}^2 = \frac{\sum_{j=0}^{n_\sigma-1} \beta^j \widehat{\sigma_{t-j}}^2}{dt \sum_{j=0}^{n_\sigma-1} \beta^j} = \frac{\sum_{j=0}^{n_\sigma-1} (1-\beta)\beta^j \widehat{\sigma_{t-j}}^2}{dt(1-\beta^{n_\sigma})}$$

where $0 < \beta < 1$.

Estimate of the drift: After we estimate σ_t , we can treat it as known to estimate α_t. We assume that α is a constant in the n_α days that we are working on, so we have

$$\frac{dX_t}{X_t} \sim N(\alpha dt, \sigma_t^2 dt)$$

The likelihood function of $(\frac{dX_1}{X_1}, \dots, \frac{dX_{n_\alpha}}{X_{n_\alpha}})$ is given by

$$L\left(\alpha; (\frac{dX_t}{X_t})\right) = \prod_{t=1}^{n_\alpha} \frac{1}{\sqrt{2\pi\widehat{\sigma_t^2}}} \exp\{\frac{\left(\frac{dX_t}{X_t} - \alpha dt\right)^2}{2\widehat{\sigma_t^2} dt}\}$$

The maximum of $L(\alpha)$ is achieved at (solve for α in $\frac{\partial \log(L(\alpha))}{\partial \alpha} = 0$)

$$\widehat{\alpha} = (\frac{1}{dt}) \frac{\sum_{t=1}^{n_\alpha} (\frac{dX_t}{X_t}) \widehat{\sigma_t^{-2}}}{\sum_{t=1}^{n_\alpha} \widehat{\sigma_t^{-2}}}$$

We get (at least in theory) that $\widehat{\alpha}$ is a better estimate of α than $(\frac{1}{dt}) \frac{\sum_{t=1}^{n_\alpha} \frac{dX_t}{X_t}}{n_\alpha}$, which we got in (3) . Actually, the formula that we got for estimating α can be explained also by a regression between $\frac{dX}{X}$ and dt :

$$\frac{dX}{X} * \frac{1}{\hat{\sigma}\sqrt{dt}} = \frac{\alpha\sqrt{dt}}{\hat{\sigma}} + \varepsilon$$

where ε is i.i.d standard normal distribution.

Estimate of the covariance: Estimating $\sigma_{i,j}$ for $i \neq j$ where i, j denote two different stocks, is based on the estimated α_t. We have: $\sigma_{i,j} = \sigma_i \sigma_j \rho_{i,j}$, where $\rho_{i,j} dt = dW_i dW_j$

Since $\frac{dX_i}{X_i} = \alpha_i dt + \sigma_i dW_i$ and similarly $\frac{dX_j}{X_j} = \alpha_j dt + \sigma_j dW_j$ we can simply estimate

$$\widehat{\sigma_{i,j}} = \left(\frac{1}{dt}\right)\left(\frac{dX_i}{X_i} - \widehat{\alpha_i} dt\right)\left(\frac{dX_j}{X_j} - \widehat{\alpha_j} dt\right)$$

We can check if what we do makes sense by observing that we should get

$$|\widehat{\sigma_{i,j}}| \leq \widehat{\sigma_i}\widehat{\sigma_j}$$

Finally we should probably use a moving average $\widehat{\sigma_{i,j}}$ by averaging over n_σ days:

$$\widehat{\sigma_{t+1}^{i,j}} = \frac{\sum_{k=0}^{n_\sigma - 1} \widehat{\sigma_{t-k}^{i,j}}}{n_\sigma * dt}$$

Now we go back to the problem of designing an adaptive procedure of n_α. We can calculate the estimated drift, $\widehat{\alpha^{(1)}}$, based on the last n_α days, e.g. $n_\alpha = 20$, and the estimated drift, $\widehat{\alpha^{(2)}}$, based on the 230 days that preceded the last 20 days:

$$\widehat{\alpha^{(1)}} = (\frac{1}{dt})\frac{\sum_{s=t-20}^{t-1}(\frac{dX_s}{X_s})\widehat{\sigma_s^{-2}}}{\sum_{s=t-20}^{t-1}\widehat{\sigma_s^{-2}}}$$

$$\widehat{\alpha^{(2)}} = (\frac{1}{dt})\frac{\sum_{s=t-250}^{t-21}(\frac{dX_s}{X_s})\widehat{\sigma_s^{-2}}}{\sum_{s=t-250}^{t-21}\widehat{\sigma_s^{-2}}}$$

$$E\left(\widehat{\alpha^{(2)}} - \widehat{\alpha^{(1)}}\right) = \alpha^{(2)} - \alpha^{(1)}$$

Since $\frac{dX_t}{X_t} \sim N((\alpha - \frac{\sigma^2}{2})dt, \sigma^2 dt)$, and we have independence, we have the variance of the difference

$$var\left(\widehat{\alpha^{(2)}} - \widehat{\alpha^{(1)}}\right) = var\left(\widehat{\alpha^{(1)}}\right) + var(\widehat{\alpha^{(2)}})$$

where

$$var\left(\widehat{\alpha^{(1)}}\right) = \frac{1}{(dt)^2} var\left(\frac{\sum_{s=t-20}^{t-1}\left(\frac{dX_s}{X_s}\right)\widehat{\sigma_s^{-2}}}{\sum_{s=t-20}^{t-1}\widehat{\sigma_s^{-2}}}\right)$$

$$= \frac{1}{(dt)^2}\sum_{s=t-20}^{t-1}\left(\frac{\widehat{\sigma_s^{-2}}}{\sum_{s=t-20}^{t-1}\widehat{\sigma_s^{-2}}}\right)^2 * \widehat{\sigma_s^2} dt$$

$$= \frac{1}{dt} * \frac{1}{\sum_{s=t-20}^{t-1}\widehat{\sigma_s^{-2}}}$$

$$var\left(\widehat{\alpha^{(2)}}\right) = \frac{1}{dt} * \frac{1}{\sum_{s=t-250}^{t-21}\widehat{\sigma_s^{-2}}}$$

Then we get

$$var\left(\widehat{\alpha^{(2)}} - \widehat{\alpha^{(1)}}\right) = \frac{1}{dt}(\frac{1}{\sum_{s=t-20}^{t-1}\widehat{\sigma_s^{-2}}} + \frac{1}{\sum_{s=t-250}^{t-21}\widehat{\sigma_s^{-2}}})$$

and

$$\widehat{\alpha^{(2)}} - \widehat{\alpha^{(1)}} \sim N(\alpha^{(2)} - \alpha^{(1)}, \frac{1}{dt}\left(\frac{1}{\sum_{s=t-20}^{t-1}\widehat{\sigma_s^{-2}}} + \frac{1}{\sum_{s=t-250}^{t-21}\widehat{\sigma_s^{-2}}}\right))$$

when we assume σ_s are fixed for each time s. With a hypothesis $H_0: \alpha^{(2)} = \alpha^{(1)}$, we test, in 95% confidence level, whether

$$\frac{\left(\widehat{\alpha^{(2)}} - \widehat{\alpha^{(1)}}\right)\sqrt{dt}}{\sqrt{\frac{1}{\sum_{s=t-20}^{t-1}\widehat{\sigma_s^{-2}}} + \frac{1}{\sum_{s=t-250}^{t-21}\widehat{\sigma_s^{-2}}}}} \sim N(0,1).$$

If the data significantly reject H_0, it means that right now the drift jumps to another level, so that using the last 250 days to estimate the drift is no longer a good possibility. From now on, we should estimate the drift based on the data created as of 20 days ago.

We applied the updated method on historical stock market data of the last 20 years. With $n_\alpha = 27$ and $n_\sigma = 11$, the annual return of the HP, Kodak and money market portfolio reaches to 0.1149204. Recall: The annual returns of HP, Kodak and money market are 0.08698516, -0.1855094 and 0.03463049, respectively. Using $n_\alpha = 30$ and $n_\sigma = 20$, a portfolio of IBM, Ford and money market has annual return 0.135471 with advanced estimate method. Recall: The annual returns of IBM, Ford and money market are 0.1181272, 0.05999252 and 0.03463049, respectively. It is interesting to observe that the annual return of the portfolio based on the basic estimation from Section 4.1 is 0.1438097 when the number of days used to estimate was 241. But 241 days is the choice which is the best for the historical data that we used. When one tries other choices of number of days the results of the basic method can be much worse so the practicality of the basic estimate success is questionable at best. It looks that the advanced method of dynamic managing of n_α is more robust.

Finally we discuss the problem of selecting n_σ. Although the situation here is less sensitive than in the case the drift, it is still important to know how to select an appropriate n_σ. Let us look at some examples. In all the following examples, the best choice of n_α is around 30, however the optimal choice of n_σ varies. We have 3 examples: (i) For the HP, Kodak and money market portfolio, the best return is achieved when n_α is around 10; (ii) The best IBM, Ford and money market portfolio picks n_σ around 20; (iii) We also looked at a portfolio of IBM, Kodak and money market and we saw that the best n_σ is around 200.

The results above tell us that although the updated method reduces the model's sensitivity, in terms of managing the portfolio, to the number of days used for estimating the drifts, the sensitivity of estimating the volatilities to the number of days used, still exists.

4.3 A Bayesian Approach

The following is a version of Rogers [16]. The assumptions are that σ, the volatility $N \times N$ matrix, is known(in fact we know how to estimate it daily) and the N-dimension drift α has to be estimated. The prior on α is chosen to be multivariate normal distribution denoted by $N(\widehat{\alpha_0}, {\tau_0}^{-1})$ where τ_0 is a non-singular matrix. We recommend that the prior will be selected with care. Specifically we will observe the market for a while and use the estimates of Sections 4.2 as a basis, so that $\widehat{\alpha_0}$ will be the last estimated drift (it is α_n in the formula below) and ${\tau_0}^{-1}$ will be the empirical covariance matrix of the α estimates produced during the n days that we observed the market, i.e.

$${\tau_0}^{-1} = \frac{\sum_{i=1}^{n}(\alpha_i - \widehat{\alpha_0})(\alpha_i - \widehat{\alpha_0})^t}{n}$$

Using Girsanov formula one can calculate the posterior distribution of α, which is also multivariate normal $N(\widehat{\alpha_1}, {\tau_1}^{-1})$, where

$$\widehat{\alpha_1} = {\tau_1}^{-1}(\tau_0\widehat{\alpha_0} + (Y_0)_{N\times 1})$$

$$\tau_1 = \tau_0 + dt * I$$

$$(\log\left(\frac{X_1}{X_0}\right))_{N\times 1} = (V_0)_{N\times N}(Y_0)_{N\times 1} - \frac{1}{2}(\widehat{\sigma_I^2})_{N\times 1}dt$$

$$\widehat{\sigma_0} = V_0 \cdot {V_0}^t$$

where X_1 is a column price vector of the N stocks on the first day and , $(\widehat{\sigma_I^2})$ is a column vector of price variances of the N stocks on the first day.

This calculation can now continue daily and we get the following formula how to proceed from the t-th day to (t+1)th day:

$$\widehat{\alpha_{t+1}} = \tau_{t+1}{}^{-1}(\tau_t \widehat{\alpha_t} + Y_t)$$

$$\tau_{t+1} = \tau_t + dt * I$$

$$(\log\left(\frac{X_{t+1}}{X_t}\right)) = V_t Y_t - \frac{1}{2}(\widehat{\sigma_\imath^2})_t dt$$

$$\widehat{\sigma_t} = V_t \cdot V_t{}^t$$

The Bayesian methodology seems to be a good improvement on what we have up to now. However, when tested on the real data, the Bayesian approach didn't perform as we expected, in the sense that the portfolio in which the Bayesian estimates were used did not get an impressive annual rate of return. The annual return of the HP, Kodak and money market portfolio was only 0.08815587, and although IBM is a brilliant choice to invest in, the IBM, Ford and money market portfolio with the Bayesian estimation performs even worse than the individual Ford stock.

5 Summary

We went over two methods of managing a portfolio: The on-line method and the stochastic portfolio method. Each of them is advantages and disadvantages. The big advantage of the on-line portfolio is that one does not need to estimate parameters and in fact there are no distribution assumptions. However, in many portfolios when implemented with historical data, the on-line portfolio did not perform that well. There is a gap between the theory and the actual results. The problem seems to be that it takes very long time to achieve even partial results of what the theory promises.

When it comes to stochastic portfolio the big problem is to estimate the parameters which look like a very difficult assignment. However with a good dynamic estimation of drifts and volatilities, investments following the stochastic portfolio can lead to much better annual returns in comparison with the on-line portfolio. Still the estimation procedure leaves many issues unanswered. For example it is not very clear how many days one should look back in order to make the estimates useful. Since the outcome seems to be sensitive to the exact method that one is using in the estimates of parameters, there are still many issues that have to be worked on.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] P.H. Algoet. Universal schemes for prediction, gambling, and portfolio selection. Annals of Probability, 20:901-941, 1992.

[2] R. Bell and T.M. Cover. Competitive optimality of logarithmic investment. Mathematics of Operations Research, 5:161-166, 1980

[3] T.M. Cover. An algorithm for maximizing expected log investment return. IEEE Transactions in Information Theory, 30:369-373, 1984

[4] T.M. Cover and D. Gluss. Empirical Bayses stock market portfolios. Advances in Applied Mathematics, 7, 1986

[5] T.M. Cover. University portfolio. Mathematical Finance, 1(1):1-29, 1991

[6] R. Fernholz, and B. Shay. Stochastic portfolio theory and stock market equilibrium. The Journal of Finance, 37(2):615-624, 1981

[7] D.P. Helmbold, J. Kivinen, and M.K. Warmuth. Worst-case loss bounds for sigmoided linear neurons. In advances in Neural Information Processing Systems, 8, 1996

[8] D.P. Helmbold, R.E. Schapire, Y. Singer, and M.K. Warmuth. On-line portfolio selection using multiplicative updates. Mathematical Finance, 8(4):325-347, 1998

[9] J.I. Kelly. A new interpretation of information rate. Bell Systems Technical Journal, 35:917-926, 1956

[10] J. Kivinen, and M.K. Warmuth. Additive versus exponentiated gradient updates for linear prediction. Information and Computation, 132(1):1-64, 1997

[11] J. Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics, 47:13-37, 1965a

[12] H.M. Markowitz. Portfolio selection. Journal of Finance, 7:77-91, 1952

[13] H.M. Markowitz. The optimization of a quadratic function subject to linear constraints. Naval Research and Logistics Quarterly, 3:111-133, 1956

[14] H.M. Markowitz. Portfolio Selection: Efficient Diversification of Investments. John Wiley and Sons, New York, 1959

[15] R.C. Merton. An intertemporal capital asset pricing model. Econometrica, 41:867-887, 1973

[16] L.C.G. Rogers. Optimal Investment, SpringerBriefs in Quantitative Finance, 2013

[17] B. Rosenber, and J.A. Onlson. The stationary distribution of returns and portfolio separation in capital markets: a fundamental contradiction. Journal of Financial and Quantitative Analysis, 11:393-401, 1976

[18] W.F. Sharpe. Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance, 19:425-442, 1964

[19] Wikipedia. Karush–Kuhn–Tucker conditions http://en.wikipedia.org/wiki/Karush%E2%80%93Kuhn%E2%80%93Tucker_conditions#cite_note-4