Market Models in Stochastic Portfolio Theory

Yunxian Zhang

21st June 2012

St Catherine's College
University of Oxford

MSc Thesis

Contents

1 Introduction **5**

2 A review of Stochastic Portfolio Theory **5**

2.1 The basic model 5

2.2 The market portfolio and relative return 9

3 Diversity **12**

4 Rank processes **14**

5 Empirical survey in the UK market **14**

5.1 Further discussion 16

6 Abstract market models **19**

6.1 Rank-based model 20

6.2 Volatility-stabilised markets 21

6.3 The model with bounded range 22

6.3.1 Deduction of formulae 22

6.3.2 Simulation 25

7 Summary **28**

Acknowledgement

I would like to express my thanks to my supervisor, Dr. Johannes Ruf for his excellent guidance and support during my research on this dissertation. I sincerely believe that I would not complete this dissertation of this quality, without the ideas and methods he provided me.

Abstract

Stochastic Portfolio Theory is the concept of studying and revealing the structure of stock markets. Under such a mathematical framework, the long-term performance and equilibrium of the stock market can be analysed systematically. In the first several sections, the main concepts and results of Stochastic Portfolio Theory will be reviewed, mainly based on the literatures contributed by Fernholz and Karatzas (2009), and Fernholz (2002). Later we will provide some econometric survey by using the data from the UK equity market to reveal some features of the relationship between the ranks and weights in the market. In the last section, a new model with bounded range will be introduced. Thorough discussion around the development and implementation of the model will be demonstrated.

1 Introduction

Under the framework of Stochastic Portfolio Theory, the properties of the long-run performance of a portfolio are analysed by introducing the growth and return of the stocks. We will start from introducing the fundamental concepts, and in the following sections more detailed properties related to the long-run behaviour will be discussed. Motivated by the ideas used in the construction of rank-based models, the empirical research implementation in the UK market was done and discussed following the review. In order to get more insights from the behaviour of the weights of the stocks, a new model was generated to deal with the weight distribution, given that all the stocks are constrained all the way in the evolution. Although as we will see, some of the intuition of this model was compromised, since this research was not meant to go that far, the main results from the implementation were impressive and inspiring.

2 A review of Stochastic Portfolio Theory

In this section, the most fundamental market models, the portfolios and their properties will be introduced. The main objective of this section is to demonstrate and recall the theories, rather than explaining the approaches in detail. Hence most theorems will be briefly explained and further insightful information should refer to Fernholz and Karatzas (2009), and Fernholz (2002), which also contain the main resources of this whole section.

2.1 The basic model

Assume in the market the price processes are Itô processes on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with the filtration $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}$ generated by a d-dimensional Brownian motion $B_t = (B_t^1, B_t^2, \dots, B_t^d)'$. Then we have the market model $\mathcal{M}$ of the form

$$\begin{aligned} \mathrm{d}S_t^0 &= r_t S_t^0 \mathrm{d}t,\ S_0^0 = 1 \\ \mathrm{d}S_t^i &= S_t^i \left(b_t^i \mathrm{d}t + \sum_{j=1}^d \sigma_t^{ij} \mathrm{d}B_t^j \right)\ ,\ i = 1, \dots, n \\ S_0^i &= s_i,\ i = 1, \dots, n \end{aligned} \tag{2.1}$$

where S^0 is the riskless cash account, r_t is the interest rate process, and $S^1, S^2, \dots, S^n$ are n risk assets (stocks), driven by the d-dimensional Brownian motion (normally $d \geq n$).

The rate of return process $b_t = (b_t^1, b_t^2, \dots, b_t^n)$ and the entries σ_t^{ij} of the $n \times d$ volatility matrix are $\mathbb{F}$-adapted processes satisfying the integrability conditions

$$\int_0^T \|r_t\| \, \mathrm{d}t + \sum_{i=1}^n \int_0^T \left(\left\|b_t^i\right\| + \sum_{j=1}^n \left(\sigma^{ij}\right)^2 \right) \mathrm{d}t < \infty, \; a.s. \tag{2.2}$$

We can also define the semi-positive definite covariance matrix by

$$a_t^{ij} = \sum_{k=1}^d \sigma_t^{ik} \sigma_t^{jk} = \left(\sigma_t \sigma_t'\right)_{ij} = \frac{\mathrm{d}}{\mathrm{d}t} \left\langle \log S^i, \log S^j \right\rangle_t \tag{2.3}$$

Then the market model (2.1) can be presented in the form

$$\begin{gathered} \mathrm{d} \log S_t^i = \gamma_t^i \mathrm{d}t + \textstyle\sum_{j=1}^d \sigma_t^{ij} \mathrm{d}B_t^j \\ \text{where } \gamma_t^i = b_t^i - \frac{1}{2} a_t^{ii} \end{gathered} \tag{2.4}$$

We also have the **logarithmic representation** of the stock processes

$$S_t^i = s_i \exp \left\{ \int_0^t \gamma_u^i \mathrm{d}u + \sum_{j=1}^d \int_0^t \sigma_u^{ij} \mathrm{d}B_u^j \right\}, \; 0 \le t < \infty \tag{2.5}$$

γ_t^i is called the ***growth rate*** of the ith stock, justified through the following argument:

If none of the individual stock variances increases too quickly, i.e., if

$$\lim_{T \to \infty} \left(\frac{\log \log T}{T^2} \int_0^T a_t^{ii} \mathrm{d}t \right) = 0, \; \text{a.s.} \tag{2.6}$$

then,

$$\lim_{T \to \infty} \frac{1}{T} \left(\log S_T^i - \int_0^T \gamma_t^i \mathrm{d}t \right) = 0, \; \text{a.s.} \tag{2.7}$$

Definition 2.1. A portfolio $\pi_t = \left(\pi_t^1, \pi_t^2, \cdots, \pi_t^n\right)'$ is an $\mathbb{F}$-progressively measurable process, bounded uniformly in (t, ω), with values in the set

$$\bigcup_{k \in \mathbb{N}} \left\{ \left(\pi^1, \pi^2, \cdots, \pi^n\right) \in \mathbb{R}^n \middle| \left(\pi^1\right)^2 + \left(\pi^2\right)^2 + \cdots + \left(\pi^n\right)^2 \le k^2, \pi^1 + \pi^2 + \cdots + \pi^n = 1 \right\}$$

A long-only portfolio $\pi_t = \left(\pi_t^1, \pi_t^2, \cdots, \pi_t^n\right)'$ is a portfolio that takes the values in the set

$$\triangle^n = \left\{ \left(\pi^1, \pi^2, \cdots, \pi^n \right) \in \mathbb{R}^n | \pi^1 \geq 0, \pi^2 \geq 0, \cdots, \pi^n \geq 0 \, and \, \pi^1 + \pi^2 + \cdots + \pi^n = 1 \right\}$$

Hence denote

$$\triangle_+^n = \left\{ \left(\pi^1, \pi^2, \cdots, \pi^n \right) \in \triangle^n | \pi^1 > 0, \pi^2 > 0, \cdots, \pi^n > 0 \right\} \tag{2.8}$$

Now we denote h_t^i as the money invested in S_t^i, and π_t^i to be the proportion of the wealth $V_t^{\omega,\pi}$ at time t, then

$$\pi_t^i = \frac{h_t^i}{V_t^{\omega,\pi}}, \; i = 1, \ldots, n$$

Then the wealth process $V_t^{\omega,\pi}$ satisfies the stochastic equation

$$\frac{\mathrm{d}V_t^{\omega,\pi}}{V_t^{\omega,\pi}} = \sum_{i=1}^{n} \pi_t^i \frac{\mathrm{d}S_t^i}{S_t^i} = \pi_t' \left[b_t \mathrm{d}t + \sigma_t \mathrm{d}B_t \right] = b_t^\pi \mathrm{d}t + \sum_{j=1}^{d} \sigma_t^{\pi j} \mathrm{d}B_t^j, \; V_0^{\omega,\pi} = \omega \tag{2.9}$$

where

$$b_t^\pi = \sum_{i=1}^{n} \pi_t^i b_t^i, \; \sigma_t^{\pi j} = \sum_{i=1}^{n} \pi_t^i \sigma_t^{ij}. \; j = 1, \ldots, d \tag{2.10}$$

Hence by (2.4), this can be transformed to

$$\mathrm{d} \log V_t^{\omega,\pi} = \gamma_t^\pi \mathrm{d}t + \sum_{j=1}^{d} \sigma_t^{\pi j} \mathrm{d}B_t^j, \; V_0^{\omega,\pi} = \omega \tag{2.11}$$

or in the logarithmic representation

$$V_t^{\omega,\pi} = \omega \exp \left\{ \int_0^t \gamma_u^\pi \mathrm{d}u + \sum_{j=1}^{d} \int_0^t \sigma_u^{\pi j} \mathrm{d}B_u^j \right\}, \; 0 \leq t < \infty \tag{2.12}$$

where the *growth rate* of the portfolio

$$\gamma_t^\pi = \sum_{i=1}^{n} \pi_t^i \gamma_t^i + \gamma_t^{\pi *} \tag{2.13}$$

and the *excess growth rate* of the portfolio

$$\gamma_t^{\pi *} = \frac{1}{2}\left(\sum_{i=1}^{n} \pi_t^i a_t^{ii} - \sum_{i=1}^{n}\sum_{j=1}^{n} \pi_t^i a_t^{ij} \pi_t^j\right) \tag{2.14}$$

As similar argument in (2.6) and (2.7), we have that, if

$$\lim_{T\to\infty}\left(\frac{\log\log T}{T^2}\int_0^T \|a_t\|\,\mathrm{d}t\right) = 0,\ a.s. \tag{2.15}$$

then,

$$\lim_{T\to\infty}\frac{1}{T}\left(\log V_T^{\omega,\pi} - \int_0^T \gamma_t^{\pi}\mathrm{d}t\right) = 0,\ a.s. \tag{2.16}$$

From linear algebra, the sufficient condition of (2.15) is that all the eigenvalues of the covariance matrix, as defined in (2.3), are uniformly bounded, i.e.

$$\begin{gathered} x' a_t x = x'\sigma_t \sigma_t' x \le K\|x\|^2, \forall t \in [0,\infty)\ and\ x \in \mathbb{R}^n\ a.s. \\ \text{for some constant } K \in [0,\infty) \end{gathered} \tag{2.17}$$

This is the ***uniform boundedness*** condition on the volatility structure $\mathcal{M}$.

Definition 2.2. If σ_t is nonsingular for all $t \in [0,\infty)$, a.s., then the market $\mathcal{M}$ is *nondegenerate* if there exists a number $\varepsilon > 0$ such that

$$x a_t x' \ge \varepsilon\|x\|^2,\ x \in \mathbb{R}^n,\ t \in [0,\infty),\ a.s. \tag{2.18}$$

The market has *bounded variance* if there exists a number $K > 0$ such that

$$x a_t x' \le K\|x\|^2,\ x \in \mathbb{R}^n,\ t \in [0,\infty),\ a.s. \tag{2.19}$$

From now on, denote $V_t^{\pi} = V_t^{1,\pi}$ when initial wealth $\omega = 1$. Then (2.11) can be transformed to

$$\mathrm{d}\log V_t^{\pi} = \gamma_t^{\pi *}\mathrm{d}t + \sum_{i=1}^{n}\pi_t^i \mathrm{d}\log S_t^i \tag{2.20}$$

Now rearrange the weights of a portfolio in descending order

$$\max_{1\le i\le n}\pi_t^i = \pi_t^{(1)} \ge \pi_t^{(2)} \ge \cdots \ge \pi_t^{(n)} = \min_{1\le i\le n}\pi_t^i \tag{2.21}$$

Let e_i denote the ith unit vector in $\mathbb{R}^n$, we introduce the new parameters

$$a_t^{\pi i} = \sum_{j=1}^{n} \pi_t^j a_t^{ij},\ a_t^{\pi\pi} = \sum_{i=1}^{n}\sum_{j=1}^{n} \pi_t^i a_t^{ij} \pi_t^j = \sum_{k=1}^{d} \left(\sigma_t^{\pi k}\right)^2 \tag{2.22}$$

$$\tau_t^{\pi,ij} = \sum_{k=1}^{d} (\sigma_t^{ik} - \sigma_t^{\pi k})(\sigma_t^{jk} - \sigma_t^{\pi k}) = (\pi_t - e_i)' a_t (\pi_t - e_j) = a_t^{ij} - a_t^{\pi i} - a_t^{\pi j} + a_t^{\pi\pi} \tag{2.23}$$

Clearly, $a_t^{\pi\pi}$ is the variance of the portfolio.

Remark 2.1. The process $\tau_t^{\pi} = \left(\tau_t^{\pi,ij}\right)_{1\le i,j\le n}$ satisfies the property

$$\sum_{j=1}^{n} \tau_t^{\pi,ij} \pi_t^j = 0, \text{ for } i = 1, 2, \cdots, n \tag{2.24}$$

2.2 The market portfolio and relative return

Now let the share prices S_t^i to be the whole capitalisation of each stock. Then we have (Fernholz 2002 p.14-15)

Definition 2.3. Let $S_t = S_t^1 + S_t^2 + \cdots + S_t^n$ be the total capitalisation of the market. The portfolio μ with weights $\mu^1, \mu^2, \dots, \mu^n$ defined by

$$\mu_t^i = \frac{S_t^i}{S_t}\ i = 1, 2, \dots, n \tag{2.25}$$

is called the *market portfolio*, and the weights μ^i are called the *market weights*.

Then we can verify the portfolio criteria are satisfied, i.e. $0 < \mu_t^i < 1$ and $\sum_{i=1}^{n} \mu_t^i = 1$. Recall the result from (2.9), the wealth process $V_t^{\omega,\mu}$ satisfies

$$\frac{\mathrm{d}V_t^{\omega,\mu}}{V_t^{\omega,\mu}} = \sum_{i=1}^{n} \mu_t^i \frac{\mathrm{d}S_t^i}{S_t^i} = \sum_{i=1}^{n} \frac{\mathrm{d}S_t^i}{S_t} = \frac{\mathrm{d}S_t}{S_t} \tag{2.26}$$

Hence we have

$$V_t^{\omega,\mu} = \frac{\omega}{S_0} S_t \tag{2.27}$$

The solution of (2.26) can be represented by the result discussed in (2.11)

$$\mathrm{d}\log V_t^{\omega,\mu} = \gamma_t^{\mu}\mathrm{d}t + \sum_{j=1}^{d}\sigma_t^{\mu j}\mathrm{d}B_t^j,\ V_0^{\omega,\mu} = \omega \tag{2.28}$$

Recall the dynamic of each single stock (2.4), and the notations appeared in (2.10) and (2.13), the market weights can be represented as

$$\mathrm{d}\log\mu_t^i = \left(\gamma_t^i - \gamma_t^{\mu}\right)\mathrm{d}t + \sum_{j=1}^{d}\left(\sigma_t^{ij} - \sigma_t^{\mu j}\right)\mathrm{d}B_t^j \tag{2.29}$$

Furthermore, from the notions introduced in (2.23), we have

$$\tau_t^{\mu,ij} = \sum_{k=1}^{d}(\sigma_t^{ik} - \sigma_t^{\mu k})(\sigma_t^{jk} - \sigma_t^{\mu k}) = \frac{\mathrm{d}\left\langle \mu^i, \mu^j\right\rangle_t}{\mu_t^i\mu_t^j\mathrm{d}t},\ 1 \le i,j \le n \tag{2.30}$$

$$\text{and } \frac{\mathrm{d}\mu_t^i}{\mu_t^i} = \left(\gamma_t^i - \gamma_t^{\mu} + \frac{1}{2}\tau_t^{\mu,ii}\right)\mathrm{d}t + \sum_{j=1}^{d}\left(\sigma_t^{ij} - \sigma_t^{\mu j}\right)\mathrm{d}B_t^j \tag{2.31}$$

where the quantity $\tau_t^{\mu,ij}$ is called the covariance of the individual stocks relative to the entire market.

Definition 2.4. For a stock S_t^i, $1 \le i \le n$ and portfolio π_t, the process

$$R_t^{\pi,i} = \log\left(\frac{S_t^i}{V_t^{\omega,\pi}}\right)\bigg|_{\omega = S_0^i},\ t \in [0,\infty) \tag{2.32}$$

is called the *relative return* process of S_t^i versus π_t. Then the *relative return* process of a portfolio π_t versus another portfolio ρ_t can be defined in a similar manner.

$$R_t^{\pi,\rho} = \log\left(\frac{V_t^{\rho}}{V_t^{\pi}}\right),\ t \in [0,\infty) \tag{2.33}$$

Now we shall establish some of the nice properties if the relative return process. The proofs will be omitted and can be found in Fernholz and Karatzas (2009) and Fernholz (2002).

Lemma 2.1. *For the relative covariance defined in (2.23), and any portfolio π_t, we have*

$$\tau_t^{\pi,ij} = \frac{\mathrm{d}}{\mathrm{d}t}\left\langle R^{\pi,i}, R^{\pi,j}\right\rangle_t,\ \textit{and}\ \tau_t^{\pi,ii} = \frac{\mathrm{d}}{\mathrm{d}t}\left\langle R^{\pi,i}\right\rangle_t \ge 0\ \textit{a.s.} \tag{2.34}$$

and the matrix $\tau_t^\pi = \left(\tau_t^{\pi,ij}\right)_{1\le i,j\le n}$ *is almost surely non-negative definite. In addition, if the covariance matrix* a_t *(as defined in (2.3)) is positive definite, the matrix* τ_t^π *has rank* $n-1$*, and the null space of* τ_t^π *is spanned by the vector* π_t *a.s..*

Lemma 2.2. *For any two portfolios* π *and* ρ *in the market* $\mathcal{M}$*, we have*

$$\mathrm{d}\log\left(\frac{V_t^\pi}{V_t^\rho}\right) = \gamma_t^{\pi*}\mathrm{d}t + \sum_{i=1}^n \pi_t^i \mathrm{d}\log\left(\frac{S_t^i}{V_t^\rho}\right) \tag{2.35}$$

a.s., for $t \in [0,\infty)$*.*

Remark 2.2. Apply this lemma in conjunction with the market portfolio, we have, for any portfolio π in the market $\mathcal{M}$,

$$\mathrm{d}\log\left(\frac{V_t^\pi}{V_t^\mu}\right) = \gamma_t^{\pi*}\mathrm{d}t + \sum_{i=1}^n \pi_t^i \mathrm{d}\log\mu_t^i = (\gamma_t^{\pi*} - \gamma_t^{\mu*})\mathrm{d}t + \sum_{i=1}^n \left(\pi_t^i - \mu_t^i\right)\mathrm{d}\log\mu_t^i \tag{2.36}$$

This representation demonstrated that the relative return of a portfolio versus the market portfolio can be represented by using the changes in the market weights.

Lemma 2.3. *For any two portfolios* π *and* ρ*, we have*

$$\gamma_t^{\pi*} = \frac{1}{2}\left(\sum_{i=1}^n \pi_t^i \tau_t^{\rho,ii} - \sum_{i=1}^n\sum_{j=1}^n \pi_t^i\pi_t^j\tau_t^{\rho,ij}\right) \tag{2.37}$$

Furthermore, by recalling the representation in 2.24, we have

$$\gamma_t^{\pi*} = \frac{1}{2}\sum_{i=1}^n \pi_t^i \tau_t^{\pi,ii} \tag{2.38}$$

This represents the excess growth rate in terms of a weighted average of the relative variances of stocks. In particular, for a long-only portfolio, we have $\gamma_t^{\pi*} \ge 0$ *(recall (2.34)).*

Now combine this lemma with (2.36) and (2.30), we have for the relative return

$$\mathrm{d}\log\left(\frac{V_t^\pi}{V_t^\mu}\right) = \sum_{i=1}^n \frac{\pi_t^i}{\mu_t^i}\mathrm{d}\mu_t^i - \frac{1}{2}\left(\sum_{i=1}^n\sum_{j=1}^n \pi_t^i\pi_t^j\tau_t^{\mu,ij}\right) \tag{2.39}$$

3 Diversity

The concept of market diversity was introduced in Fernholz (1999), and further developed in Fernholz (2002) and Fernholz and Karatzas (2009). Informally, a market is not diverse if the capital is heavily focused on one single stock and is diverse if the capital is distributed uniformly (more or less) among a large number of stocks. It is shown in Fernholz (2002) that the capital distribution is related to the excess growth rate of the stocks, and surprisingly, the market with all stocks have the same growth rate will not tend to be diverse in the long term. It was also shown that the weak diversity will lead to existence of arbitrage. In this section we review these concepts and the properties.

Now let us start with the investigation of the log-term behaviour of the stocks. Here we always consider the time-average values of the stock. For example, for the growth rate γ_t^i, we consider its log-term structure $\lim_{T\to\infty} \frac{1}{T}\int_0^T \gamma_t^i \mathrm{d}t$.

Definition 3.1. The market $\mathcal{M}$ is coherent if

$$\lim_{t\to\infty} \frac{1}{t} \log \mu_t^i = 0, \text{ for } i = 1, \dots, n, \ a.s. \tag{3.1}$$

From the definition of market weight (2.25), this can be expressed as

$$\lim_{t\to\infty} \frac{1}{t} \left(\log S_t^i - \log S_t\right) = 0, \text{ for } i = 1, \dots, n, \ a.s. \tag{3.2}$$

Lemma 3.1. *The followings are equivalent*

(1). $\mathcal{M}$ is coherent

(2). for $i = 1, \dots, n$, $\lim_{T\to\infty} \frac{1}{T}\int_0^T \left(\gamma_t^i - \gamma_t^\mu\right) \mathrm{d}t = 0$, a.s.

(3). for $i, j = 1, \dots, n$, $\lim_{T\to\infty} \frac{1}{T}\int_0^T \left(\gamma_t^i - \gamma_t^j\right) \mathrm{d}t = 0$, a.s.

Hence from the condition (3) above, we can conclude that if all the stocks have the same growth rate process, then $\mathcal{M}$ is coherent.

Lemma 3.2. *Suppose that all the stocks have constant growth rate. Then the market $\mathcal{M}$ is coherent if and only if the growth rates are all the same.*

Lemma 3.3. *For any portfolio π in a nondegenerate market, there exists a constant $\varepsilon \in (0, \infty)$, such that for $t \in [0, \infty)$*

$$\tau_t^{\pi, ii} \geq \varepsilon \left(1 - \pi_t^i\right)^2, \ i = 1, \dots, n \ a.s. \tag{3.3}$$

In particular for a long-only portfolio, we have

$$\gamma_t^{\pi *} \geq \frac{\varepsilon}{2}\left(1 - \pi_t^{(1)}\right) \tag{3.4}$$

where $\pi_t^{(1)}$ is the maximum of the weights, as defined in (2.21).

Lemma 3.4. *For a long-only portfolio π in a market with bounded variance, there exists a constant $\varepsilon \in (0, \infty)$ such that for $t \in [0, \infty)$*

$$\pi_t^{(1)} \leq 1 - \varepsilon \gamma_t^{\pi *}, \ i = 1, \dots, n \ a.s. \tag{3.5}$$

Lemma (3.3) tells us that with a nondegenerate market, if π_{max} is bounded away from 1, then $\gamma_t^{\pi *}$ is bounded away from 0. While Lemma (3.4) shows that within a market with bounded variance, if $\gamma_t^{\pi *}$ is bounded away from 0, then π_{max} is bounded away from 1.

Lemma 3.5. *Consider a long-only portfolio π in a nondegenerate market $\mathcal{M}$ with bounded variance, if all the weights are constant with at least two positive weight. Then*

$$\liminf_{T \to \infty} \frac{1}{T} \log \left(\frac{V_T^{\pi}}{V_T^{\mu}} \right) > 0, \ a.s. \tag{3.6}$$

Now we show the formal definition of diversity and the development in the relationship with excess growth rate.

Definition 3.2. Recall the order statistics (2.21) used to define the ordering weights. The market $\mathcal{M}$ is *diverse* if there exists a number $\delta > 0$ such that

$$\mu_t^{(1)} \leq 1 - \delta, \ t \in [0, \infty), \ a.s. \tag{3.7}$$

Similarly, $\mathcal{M}$ is *weakly diverse* if there exists a number $\delta > 0$ such that

$$\frac{1}{T} \int_0^T \mu_t^{(1)} \mathrm{d}t \leq 1 - \delta, t \in [0, \infty), \ a.s. \tag{3.8}$$

From this definition, we can see that a market is diverse if the capital is not distributed heavily on one single stock over the whole time horizon, and if this holds in average it is weakly diverse.

From Lemma 3.3 and Lemma 3.4, we can deduce the following relationship between diversity and bounded variance/nondegenerate property.

Proposition 3.1. *Suppose the market $\mathcal{M}$ has bounded variance, then it is diverse if*

$$\gamma_t^{\mu *} \geq \delta, \ t \in [0, \infty), \ a.s. \tag{3.9}$$

and it is weakly diverse if

$$\frac{1}{T} \int_0^T \gamma_t^{\mu *} dt \geq \delta, t \in [0, \infty), \ a.s. \tag{3.10}$$

Conversely, if the market is nondegenerate and diverse (weakly diverse), then (3.9) (respectively, (3.10)) holds.

4 Rank processes

Generally we study the stocks by using their names, i.e. $S^1, S^2, \cdots, S^n$. However, sometimes when studying on the capital, it is more convenient to arrange the stocks by their ranks (size of capitalisation in each stock). Now we introduce the tools required to develop new models based on the rank processes.

Definition 4.1. Let $X_t^1, X_t^2, \ldots, X_t^n$ be processes. Then for $k = 1, 2, \ldots, n$,

$$X_t^{(k)} = \max_{1 \leq i_1 < \cdots < i_k < n} \min\left(X_t^{i_1}, \ldots, X_t^{i_k}\right),\ t \in [0, T] \tag{4.1}$$

is the kth *rank process* of $\left\{X_t^1, X_t^2, \ldots, X_t^n\right\}$.

Remark 4.1. Clearly, $\max\left(X^1, \ldots, X^n\right) = X^{(1)} \geq X^{(2)} \geq \cdots \geq X^{(n)} = \min\left(X^1, \ldots, X^n\right)$

Definition 4.2. If we consider the processes of market weights in stocks $\mu_t^1, \mu_t^2, \ldots, \mu_t^n$, the ranked family $\left\{\mu_t^{(1)}, \mu_t^{(2)}, \ldots, \mu_t^{(n)}\right\}$ is the *capital distribution* of the market at time t.

5 Empirical survey in the UK market

The capital distribution demonstrates the size of the companies considered in the market. Hence it is reasonable to investigate the models based on the ranks of firms. Such model has been introduced in Fernholz (2002) and further discussed in Fernholz and Karatzas (2009). Recall the *capital distribution* defined in Definition 4.2, the *capital distribution curve* is the log-log plot of the market weight versus their corresponding ranks (i.e. $\log \mu_t^{(k)}$ v.s. $\log k$ for $k = 1, 2, \ldots, n$). According to Fernholz (2002), if the firms with smallest ranks are ideally replaced in a constant rate, the capital distribution curve will be a straight line. Under such assumption and motivation, some implementation was done by using actual market data. Firstly we shall define the mathematical term 'market' in the real world. Here we use all the companies included in the FTSE350 index. Using such a large group of companies, we believe that the overall behaviour of these companies should be stable enough to investigate the features of the real equity market. The following graph demonstrated the capital distribution curve of all the companies in FTSE350, from 2003 to 2012. The data was the total capitalisation of the companies, taken monthly on the 15th of each month, until 15th May 2012.

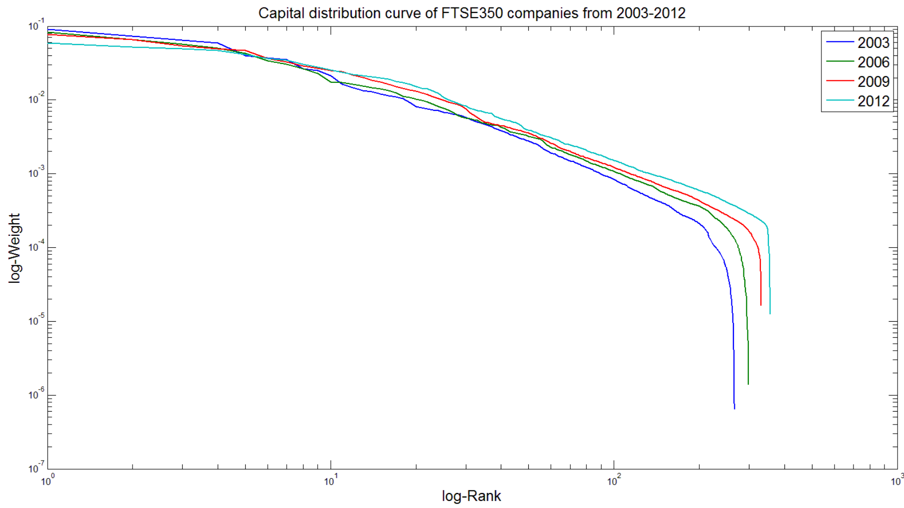

Figure 1. The capital distribution curve of all the companies in FTSE350, from 2003 to 2012. The data was the total capitalisation of the companies, taken monthly on the 15th of each month, until 15th May 2012.

The details of how to get the data and the codes of plotting the graph are displayed in Appendix 1. Due to some technical restriction, the component companies used in our research were not dynamic, but were from a snapshot list of FTSE350 on the date the data was retrieved (15th May 2012). This means that the data was not as so real as expected, because it only contains the companies on 15th May 2012. From this dataset, we could not tell which companies have entered or left the market. Nevertheless, we still believe that this incomplete market data is still able to develop the behaviour of the firms. The reasons to such brave assumption are that, firstly the group is large enough to represent the major tends of the curves, and secondly most of the companies which have ever left or entered the markets were the smallest companies. The second point is rather straightforward by observing the historical data - all the large companies remained large over a long period, even they fluctuated in the high ranks. Indeed, the curves are quite stable in the high ranks part and have steep tails in the smallest weights. The tails do make sense to us because it means that the capitalisation between small companies are quite large (with comparison to their ranks). Despite the flaws of the data, the steep tail is not to be expected to reduce for any improved data, since the small companies which drops dramatically does not exit the index immediately in the real cases. Now to investigate the real capital distribution further, we use a different dataset, which contains all the UK companies which are currently being traded in the market. This company list is also a snapshot on the 15th May 2012, but it contains 550 companies. Hence this result should be more convincing because it includes much more companies than the previous one, and should be more stable. Furthermore, the curves were plotted for the last five years, which significantly removed the effect of the change in the market.

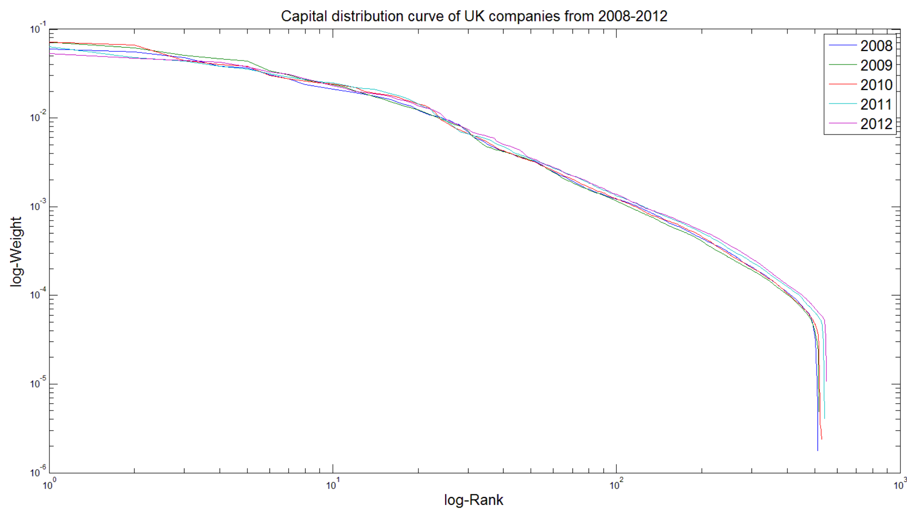

Figure 2. The capital distribution curve of 550 companies in the UK, from 2008 to 2012.

The shape of this graph is similar to the previous one with less companies, and is fairly stable. Although these are not straight lines as under the ideal conditions, one should be convinced that the actual behaviour has been developed here. Indeed, the stability of the model was discussed in Fernholz (2002) Chapter 5.

5.1 Further discussion

Besides of the research on the academic side, it is significantly valuable to investigate the data from the real market. Rather than to develop a sophisticated market model, the main objective of this part is to implement the market data and attempt to demonstrate a different approach to develop a new model based on such analysis. Despite of the incompleteness of the data we retrieved, some of other properties could still be demonstrated by various ways of analysis.

Now we focus on the change and stability of the market ranks.The following graphs shows how the market ranks in FTSE350 evolved in the succeeding year, from 2002 to 2012. Formally, it is the position of the ranks in the next year versus current ranks (i.e. $\text{rank}((X_t^{(k)})_{t+1})$ v.s. k). For example, for $k = 1$, there are only three points $1, 2,$ and 3 appeared on the graph. This means that for each year over the 11, the biggest company in the market only remained the biggest or went to second or third in the next year.

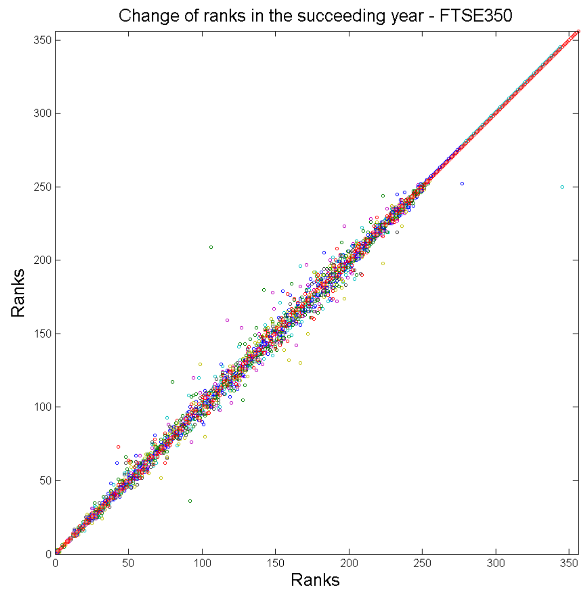

Figure 3. How the market ranks in FTSE350 evolved in the succeeding year, from 2002 to 2012. Please note that it looks like red straight lines joint in the end of the curve, because the dots in that area are too dense.

We can see that, surprisingly, small companies tend to be stable in the market. This result seems to contradict with the result from previous graph and the assumption in the volatility-stabilised model, under which small companies have large volatility. In fact, the misuse of the term 'stable' might be misleading here. The small companies are indeed stable in terms of their ranks, even they have large volatilities. This is resulted by the fact that in the low ranks group, the capitalisation between any two adjacent companies (say $X_t^{(k)}$ and $X_t^{(k+1)}$) are fairly large. In consequences, even a small company changes dramatically with respect to its (small) capitalisation, the rank of it does not move notably in the market. Indeed, this property has been revealed in Figure 2. Now we compute the mean of each rank and connecting them.

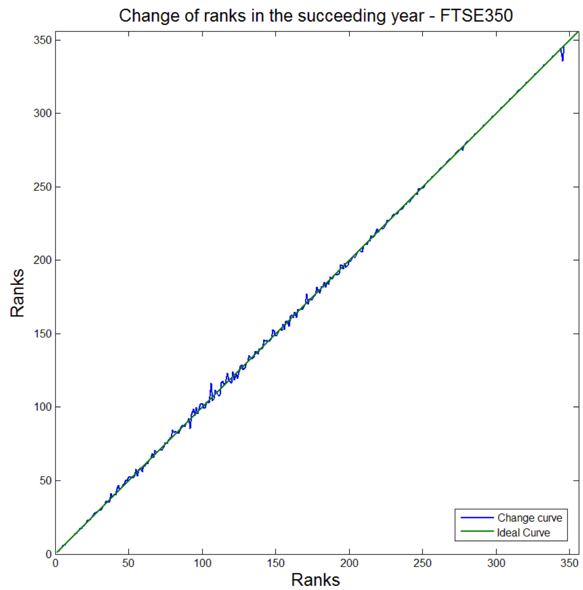

Figure 4. The curve of the mean of the change in each rank, with plotting a straight line which is the extreme case that the companies never change ranks in average.

The change is significantly more dramatic in the middle ranks, while the extent of the curve going up and down around the straight line are more or less the same. This is an apparent result, since if some companies moves in the rank, there have to be others going the opposite way. We could also observe some spikes in the tail. As explained earlier, small companies tend to move slightly in the rank, consequently, some of the 'outliers', which were caused by some of the extremely growing small companies, may influence the average value greatly. The reason why there is no spike shape in the opposite way will be explained later. Instead of the FTSE350 companies, now let us try the same analysis for the whole UK market introduced earlier.

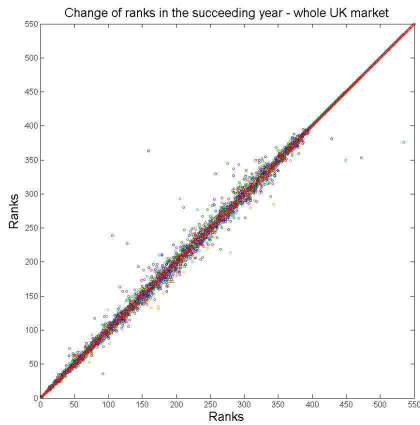

Figure 5. The points show how the market ranks in the UK evolved in the succeeding year.

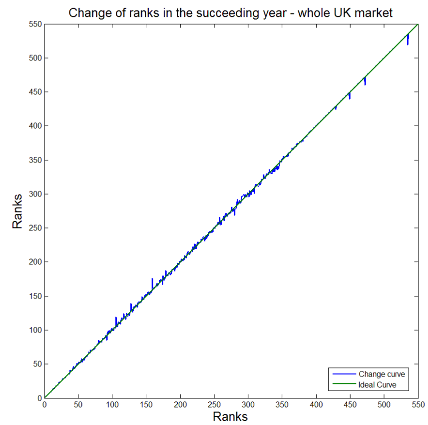

Figure 6. The curve of the mean of last graph, with the extreme line as reference.

Similar results can be seen from these two graphs. Additionally, as there are much more companies included, more spikes appeared in the tail of the curve. One odd observation is that, all the spikes are below the straight line, which means they are all caused by the booming of small companies. In fact some spikes above the straight line should be expected because some small companies are expected to decline dramatically, as well as to grow fast. The absence of these is due to the absence of the corresponding data from the dataset, because all the previously bankrupted companies have already been removed from the list we used.

6 Abstract market models

The basic market model and its properties have been introduced in Section 1. However, the study in this general model is not sufficient for us to learn from the actual market. Hence, the abstract market models, which have some more specific properties, are introduced in order to deal with some of the certain aspects of the actual markets. In this section, two of the particular models will be discussed, the volatility-stabilized markets and the rank-based models. Furthermore, a new model with bounded range will be introduced and carefully developped. Similar to the discussion in Section 1, this section is provided to review the nature of the first two models, while both of the two models were thoroughly developed and discussed in Fernolz and Karatzas (2009).

6.1 Rank-based model

In the last section, the empirical survey was motivated by the idea to reveal the relationship between the ranks and weights. Previous research published in Fernholz (2002) and Banner, Fernholz and Karatzas (2005) have suggested that it is sensible to name the stocks by their ranks rather than the actual names, and a model could be done based on this. Now let's show the construction of the Atlas model based on the ranks in a formulated way.

The Atlas model was firstly introduced in Fernholz (2002) and then further discussed in Banner, Fernholz and Karatzas (2005). One feature of this model is that it assigns constant coefficients to each stock. All the stocks have the same constant volatility term, while only the smallest stock has positive growth rate and others are zero.

Suppose we have n stocks with capitalisation $S_t^1, S_t^2, \dots S_t^n$ at time t, then we have the dynamics for $i = 1, 2, \dots, n$

$$\mathrm{d} \log S_t^i = \gamma_t^i \mathrm{d}t + \sigma_t^i \mathrm{d}B_t^i \tag{6.1}$$

where $B^1, B^2, \dots, B^n$ are independent Brownian motions, and the growth rates and volatilities are

$$\gamma_t^i = ng\mathbf{1}_{\left\{S_t^i = S_t^{p_t^n}\right\}}, \ \sigma_t^i = \sigma \tag{6.2}$$

where g and σ are positive constants, and p_t^n is specified by the random permutation $\left(p_t^1, p_t^2, \dots, p_t^n\right)$ of $(1, 2, \dots, n)$, for which

$$S^{p_t^k} = S_t^{(k)}, \ p_t^k < p_t^{k+1} \text{ if } S_t^{(k)} = S_t^{(k+1)} \text{ for } k = 1, 2, \dots, n \tag{6.3}$$

Hence p_t^k is the k-th largest stock at time t.

Based on the concept above, now a more general case can be introduced. Suppose we have constants γ, $g_1, g_2, \dots, g_n$ and $\sigma_1, \sigma_2, \dots, \sigma_n > 0$ such that

$$\begin{gathered} g_1 < 0, \ g_1 + g_2 < 0, \ g_2 + g_3 < 0 \ \dots \\ g_1 + g_2 + \cdots g_n = 0 \end{gathered} \tag{6.4}$$

Then the growth rates and volatilities are given by

$$\gamma_t^i = \gamma + \sum_{k=1}^{n} g_k \mathbf{1}_{\left\{S_t^i = S_t^{p_t^n}\right\}}, \ \sigma_t^i = \sum_{t=1}^{n} \sigma_k \mathbf{1}_{\left\{S_t^i = S_t^{p_t^n}\right\}} \tag{6.5}$$

This general model then assigns the growth rate $\gamma + g_k$ and volatility γ to the k-th biggest stock. Compare the general form with the one as shown in (6.2), we have

$$\gamma = g > 0,\ g_k = -g \text{ for } k = 1, \dots, n-1 \text{ and } g_n = (n-1)\, g \tag{6.6}$$

Hence the smallest stock contributes all the growth to the market, as it is the only one with non-zero growth rate. The further discussion includes the uniqueness of the solutions under such framework, and the long-term behaviours of the stocks. These additional study was provided in Banner, Fernholz, and Karatzas (2005).

6.2 Volatility-stabilised markets

Compared to the basic model (2.1), the most extraordinary improvement of these models is that, from the big picture the market is designed as expected to be a exponential Brownian motion with drift, while the market will not eventually focused on one single stock. Further more, as ideally constructed, the small stocks have relatively large volatility and big ones have small volatility.

Now construct the abstract market model $\mathcal{M}$ of the form,

$$\mathrm{d}\log S_t^i = \frac{\alpha}{2\mu_t^i}\mathrm{d}t + \frac{1}{\sqrt{\mu_t^i}}\mathrm{d}B_t^i,\ i = 1, 2, \cdots, n \tag{6.7}$$

where α is a non-negative constant. According to Fernholz and Karatzas (2009), this is equivalent to solving the following stochastic differential equation system, for $i = 1, 2, \cdots, n$

$$\mathrm{d}S_t^i = \frac{1+\alpha}{2}\left(S_t^1 + S_t^2 + \cdots + S_t^n\right)\mathrm{d}t + \sqrt{S_t^i\left(S_t^1 + S_t^2 + \cdots + S_t^n\right)}\mathrm{d}B_t^i \tag{6.8}$$

Furthermore, this system determines the distribution in Δ_+^n uniquely. The integrability conditions (2.2) and the market price of risk are satisfied with

$$\begin{array}{l} b_t^i = \frac{1+\alpha}{2\mu_t^i},\ \sigma_t^{ij} = \frac{\delta^{ij}}{\sqrt{\mu_t^i}} \\ r_t \equiv 0,\ \theta_t^j = \frac{1+\alpha}{2\sqrt{\mu_t^j}} \end{array} \quad 1 \le i, j \le n \tag{6.9}$$

Note that in the model (6.7), the stocks have covariances $a_t^{ij} = 0$, for $j \neq i$, and variances $a_t^{ii} = \frac{1}{\mu_t^i}$, which represented the property that smaller stocks have relatively larger variances and vice versa. Hence, the individual stock have volatile behaviours while the market as whole evolve in a stable manner (refer to Fernholz and Karatzas (2009) for details). This special nature is called *stabilization by volatility* when $\alpha = 0$, and *stabilization by both volatility and drift* when $\alpha > 0$.

Similar to (2.13), (2.14), and (2.22), the growth rate, excess growth rate and the portfolio variance are computed as

$$\gamma_t^{\mu} = \gamma = \frac{(1+\alpha)\,n-1}{2} > 0,\ \gamma_t^{\mu*} = \gamma^{\mu*} = \frac{n-1}{2} > 0,\ a_t^{\mu\mu} = 1 \tag{6.10}$$

6.3 The model with bounded range

In this section, a new model with the bounded range of the stocks value is introduced. To calibrate this model in a sophisticated way is far beyond the aim and the reach of this paper. However, under some assumptions, the model could be derived in a rather nice form. From this point of view, it is believed that this model can be improved in the way we look for, and features of the new model could be anticipated through our work. The intuition of this model is straightforward while it often requires large scale of computations. Nevertheless, the actual implementation of the model may not be as complicated as expected. Hence following the analytical derivation, a possible solution of implementing the model will be displayed, and some discussion of the efficiency and accuracy will be showed.

6.3.1 Deduction of formulae

The intuition of this model is to study the behaviour of the weight of one single asset among a group of assets. To simplify the model, firstly assume that the asset behaves as standard Brownian motions. We would like to constrain the value of any asset between 0 and 1.

Now denote

$$\overline{X}_t = \max_{0\le s\le t} X_s \ \text{ and } \ \underline{X}_t = \min_{0\le s\le t} X_s. \tag{6.11}$$

We are interested in the joint distribution of X_t^i and X_t^j, conditioned on their constraints. That is, the distributions of

$$X_t^i, X_t^j \,\Big|\, 0 < \underline{X_t^i} < \overline{X_t^i} < 1, 0 < \underline{X_t^i} < \overline{X_t^i} < 1, 0 < (\underline{X^i + X^j})_t < (\overline{X^i + X^j})_t < 1 \tag{6.12}$$

Then the major problem arises here. Since X_t^i and X_t^j are not independent, the work taken to derive the joint distribution of all these terms is too complex to be included in this paper. Therefore, to avoid this problem, we work on a different approach where all the assets are independent, and we will derive the model under such assumption.

We are interested in the density of the weights $\frac{X_t^i}{X_t^1+X_t^2+\cdots+X_t^n}$, conditioned on the constraints. Since all the assets are standard Brownian motions, for $i = 1, 2, \ldots, n$ we have (A. N. Borodin and P. Salminen (2002))

$$\mathrm{P}\left(0 < \underline{X_t^i} < \overline{X_t^i} < 1, X_t \in \mathrm{d}z\right) = \frac{1}{\sqrt{2\pi t}} \sum_{k=-\infty}^{\infty} \left(e^{-\frac{(z-x+2k)^2}{2t}} - e^{-\frac{(z+x+2k)^2}{2t}}\right) \mathrm{d}z, \tag{6.13}$$

$$\mathrm{P}\left(0 < \underline{X_t^i} < \overline{X_t^i} < 1\right) = \frac{1}{\sqrt{2\pi t}} \sum_{k=-\infty}^{\infty} \int_0^1 \left(e^{-\frac{(z-x+2k)^2}{2t}} - e^{-\frac{(z+x+2k)^2}{2t}}\right) \mathrm{d}z \tag{6.14}$$

where x is the value of X_t^i at time $t = 0$. Then it can be derived that

$$f_{X_t|0<\underline{X_t^i}<\overline{X_t^i}<1}(z) = \frac{\sum_{k=-\infty}^{\infty} \left(e^{-\frac{(z-x+2k)^2}{2t}} - e^{-\frac{(z+x+2k)^2}{2t}}\right)}{\sum_{k=-\infty}^{\infty} \int_0^1 \left(e^{-\frac{(y-x+2k)^2}{2t}} - e^{-\frac{(y+x+2k)^2}{2t}}\right) \mathrm{d}y} \tag{6.15}$$

In order to compute more efficiently in the next part, the error function is used here instead of calculating the integral of exponential functions directly. Recall

$$\mathrm{erf}\,(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \mathrm{d}t$$

The cumulative distribution function of a normal distribution can be expressed as

$$\Phi\,(x) = \frac{1}{2} + \frac{1}{2} \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)$$

Hence

$$\begin{aligned}\int_0^1 e^{-\frac{(y-x+2k)^2}{2t}} \mathrm{d}y &= \sqrt{2\pi t}\left(\Phi\left(\frac{1-x+2k}{\sqrt{t}}\right) - \Phi\left(\frac{-x+2k}{\sqrt{t}}\right)\right) \\ &= \frac{\sqrt{2\pi t}}{2}\left(\mathrm{erf}\left(\frac{1-x+2k}{\sqrt{2t}}\right) - \mathrm{erf}\left(\frac{-x+2k}{\sqrt{2t}}\right)\right)\end{aligned}$$

Similarly, $\displaystyle\int_0^1 e^{-\frac{(y+x+2k)^2}{2t}} \mathrm{d}y = \frac{\sqrt{2\pi t}}{2}\left(\mathrm{erf}\left(\frac{1+x+2k}{\sqrt{2t}}\right) - \mathrm{erf}\left(\frac{x+2k}{\sqrt{2t}}\right)\right)$

Hence

$$f_{X_t|0<\underline{X}_t^i<\overline{X}_t^i<1}(z) = \frac{2\sum_{k=-\infty}^{\infty}\left(e^{-\frac{(z-x+2k)^2}{2t}} - e^{-\frac{(z+x+2k)^2}{2t}}\right)}{\sqrt{2\pi t}\sum_{k=-\infty}^{\infty}\left(\operatorname{erf}\left(\frac{1-x+2k}{\sqrt{2t}}\right) - \operatorname{erf}\left(\frac{-x+2k}{\sqrt{2t}}\right) - \operatorname{erf}\left(\frac{1+x+2k}{\sqrt{2t}}\right) + \operatorname{erf}\left(\frac{x+2k}{\sqrt{2t}}\right)\right)}. \tag{6.16}$$

Now denote $f_{\eta_i}(z) = f_{X_t^i|0<\underline{X}_t^i<\overline{X}_t^i<1}(z)$.

Lemma 6.1. *The distribution of the sum of the weight is*

$$f_\eta(z) := f_{\eta_1+\eta_2+\cdots+\eta_n}(z) =$$
$$\int_{\max(0,z-1)}^{\min(n-1,z)} \cdots \int_{\max(0,z_3-1)}^{\min(2,z_3)} \int_{\max(0,z_2-1)}^{\min(1,z_2)} f_{\eta_1}(z_1)\, f_{\eta_2}(z_2-z_1)\,\mathrm{d}z_1 f_{\eta_3}(z_3-z_2)\,\mathrm{d}z_2$$
$$\cdots f_{\eta_n}(z-z_{n-1})\,\mathrm{d}z_{n-1}$$

Proof. We prove this by induction, the joint density of η_1 and η_2 is $f_{\eta_1,\eta_2}(z_1,z_2) = f_{\eta_1}(z_1)\, f_{\eta_2}(z_2)$ as they are independent.

Then the joint density of η_1 and $\eta = \eta_1+\eta_2$ is $f_{\eta_1,\eta}(z_1,z) = f_{\eta_1}(z_1)\, f_{\eta_2}(z-z_1)$

As $0<\eta_1<1, 0<\eta_2<1$ and $0<\eta_1+\eta_2<2$, we have

$$\max(0,\eta-1) < \eta_1 < \min(1,\eta)$$

Hence

$$f_\eta(z) = \int_{\max(0,z-1)}^{\min(1,z)} f_{\eta_1}(z_1)\, f_{\eta_2}(z-z_1)\,\mathrm{d}z_1 \tag{6.17}$$

As $\eta_1+\eta_2$ and η_3 are independent, the density of $\eta_1+\eta_2+\eta_3$ is derived in the similar manner. Eventually we get the formula for $f_{\eta_1+\eta_2+\cdots+\eta_n}(z)$. □

Since $\eta_1+\eta_2+\cdots+\eta_n$ and η_i are not independent, more steps are used to derive the density of $\frac{\eta_i}{\eta}$.

Lemma 6.2. *The distribution of the weight of the ith asset in the market at time t is*

$$f_{\frac{\eta_i}{\eta}}(z) = \int_0^{min\left(1,\frac{z(n-1)}{1-z}\right)} y f_{\eta_i}(y)\, f_{\eta^i}\left(\frac{(1-z)\,y}{z}\right)\mathrm{d}y \tag{6.18}$$

where $f_{\eta_i}(\cdot)$ and $f_{\eta^i}(\cdot)$ *are the density functions for the ith asset and the sum of other assets without the ith, respectively.*

Proof. Denote $\eta^i = \eta_1 + \eta_2 + \ldots + \eta_{i-1} + \eta_{i+1} + \ldots + \eta_n$ the sum without the i-th term. From the joint density of η_i and η^i, we have,

$$f_{\eta^i, \eta_i}(x, y) = f_{\eta^i}(x) f_{\eta_i}(y), \text{ as they are independent} \tag{6.19}$$

Hence $f_{\frac{\eta^i}{\eta_i}, \eta_i}(z, y) = y f_{\eta^i, \eta_i}(yz, y) = y f_{\eta_i}(y) f_{\eta^i}(zy)$. As $0 < y < 1$, and $0 < yz < n - 1$, we have $0 < y < min\left(1, \frac{n-1}{z}\right)$.

$$f_{\frac{\eta^i}{\eta_i}}(z) = \int_0^{min\left(1, \frac{n-1}{z}\right)} y f_{\eta_i}(y) f_{\eta^i}(zy) \, \mathrm{d}y \tag{6.20}$$

As $\frac{\eta}{\eta_i} = \frac{\eta^i}{\eta_i} + 1$,

$$f_{\frac{\eta}{\eta_i}}(z) = f_{\frac{\eta^i}{\eta_i}}(z - 1) \tag{6.21}$$

Finally,

$$f_{\frac{\eta_i}{\eta}}(z) = \frac{1}{z^2} f_{\frac{\eta}{\eta_i}}\left(\frac{1}{z}\right) = \frac{1}{z^2} f_{\frac{\eta^i}{\eta_i}}\left(\frac{1}{z} - 1\right) = \int_0^{min\left(1, \frac{z(n-1)}{1-z}\right)} y f_{\eta_i}(y) f_{\eta^i}\left(\frac{(1-z)\, y}{z}\right) \mathrm{d}y \tag{6.22}$$

□

6.3.2 Simulation

The result above does not involve any complex concept, but it is not easy to express it in a neat form. Now let us see how we can simulate the distributions when dealing with the data.

Firstly, in order to compute the infinite sum efficiently, set a level of tolerance $\left(10^{-30} \text{ here}\right)$. Start summing for $k = 0, 1, -1, 2, -2 \ldots l, -l$, and the summing does not stop until the value for $k = l$ or $-l$ is smaller than the tolerance level.

The value of time t matters when computing the value of the density. It changes the value of the density in scale, so it does not change the trend of the density. The natural choice of t could be $t = 0, 1, 2, \ldots$. Therefore, we use $t = 1$ throughout the simulation.

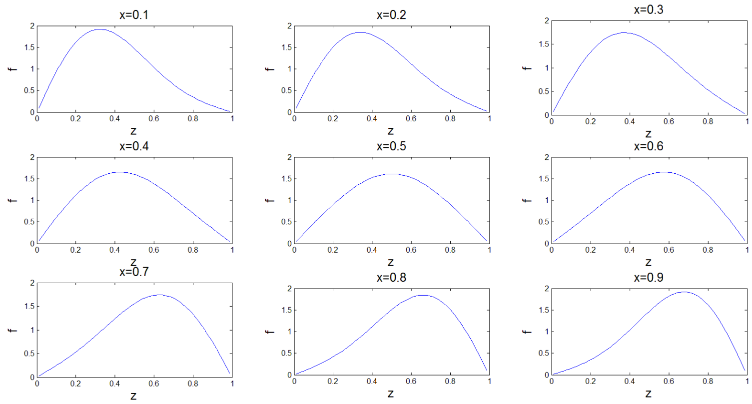

Figure 7. the different density functions $f_\eta(z)$ for $x = 0.1, 0.2, \ldots, 0.9$.

The mean of each density is the value of x, which is the initial value of the stock, this has been clearly showed in the graph. In addition, the range of the asset value z is strictly from 0 to 1.

To simulate the density of the sum, the start points x are chosen to be five stocks from the market, the 1st, 6th, 11th, 16th, and 21st by ranks in FTSE350. For the ease of computation, x is rescaled by their sum - this does not affect the purpose of this simulation as we are looking for the weights among this small group of stocks anyway. For instance $x = [0.5562\ 0.2297\ 0.1045\ 0.0630\ 0.0465]$. To simulate the density of the sum, which includes complicated integrals, we use the method of interpolation to find the function which interpolate the density functions and then integrate directly.

In fact, this method shows great scale of accuracy here. The result obtained here is stable and efficient to demonstrate major properties of the density function. As a consequence of various summing form in the computation, the most number of asset we could simulate here is up to 5. More research and attempt could be useful to reduce the work needed in the simulation and to improve the whole efficiency.

Below is the graph of the density function of the sum of 5 assets, start points $x = [0.5562\ 0.2297\ 0.1045\ 0.0630\ 0.0465]$

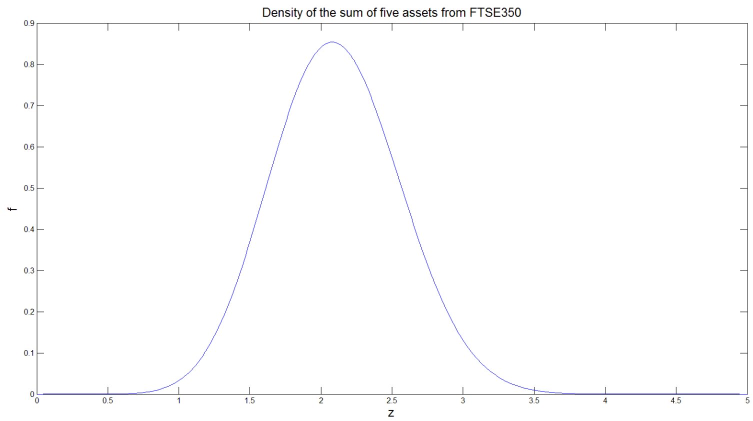

Figure 8. The density function of the sum of 5 assets from FTSE350, at time $t = 1$. The curve is nicely bell shaped, as expected.

Continue with the computation, the curve of the density $f_{\frac{\eta_i}{\eta}}$ is obtained.

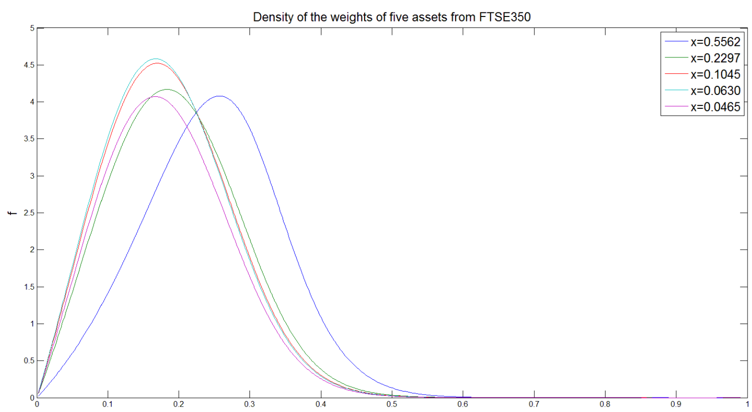

Figure 9. Density of five stocks from FTSE350, at time $t = 1$. Different colours represent different asset weights.

As we can see, the bigger companies are expected to be smaller, while the small ones tend to grow. Now let us verify this result by computing the expectations of the density functions at time $t = 1$. We firstly interpolate the density functions by polynomials up to degree 40, and then compute the expectations directly by integration.

The expected weight of the five stocks with initial value $x = [0.5562\ 0.2297\ 0.1045\ 0.0630\ 0.0465]$, at time $t = 1$, are

$$w = [0.2430\ 0.1924\ 0.1967\ 0.1911\ 0.1682]$$

Clearly, with the time going, the companies are expected to come closer in our model. This is not true in real cases, and may be caused by the inefficiency of the model. The stocks are not independent in real market, and as mentioned earlier, more work could be involved to improve this model to a great extent. In addition, by noticing that the model is based on the condition related to time, the model is changed when the time varies. Nevertheless, when we construct the simulation at one time, it will be straightforward to do the same for each time horizon required.

7 Summary

Throughout this paper, the development of stochastic portfolio theory has been demonstrated briefly in the first sections. Under these theoretical concepts, some implementations were used to reveal the actual practice. Although the ideas which led to the implementations were straightforward, the results from them are impressive, and the process of doing this could provide more thoughts on the theoretical side. During the empirical survey in the UK market, some errors in the data was tolerated. Nevertheless, the results obtained from the data were efficient to reveal the main properties we expected. The model with bounded range, which was newly developed, was generated by simple concept. Even the original idea was compromised as explained in the beginning of the section, the implementations shown later proved that the model is well-constructed and could be improved in a more sophisticated way in the future.

Appendix

1. The data used in Figure 1 was acquired from *Thomson Reuters Datastream*. The market values were retrieved on 15th of every month from May 2001 to May 2012. The mnemonic used in *Datastream* to get the value of FTSE350 is FTSE350 and it is LFTSE350 for the constituents of FTSE350. Because the constituents changes over time, there are 356 companies in total. For those which did not consist in the index is the past, the values presented on the specific day were zero.

Obtained from *Datastream*, the data was then imported from *Excel* to *Matlab*. The codes used to plot was:

```
load('matlab.mat')
[m,n]=size(ftse350c); %ftse350c stores the monthly constituents
of FTSE350
ftse350co=zeros(m,n); %create the matrix storing the ordered
constituents
for i=1:n
ftse350co(:,i)=sort(ftse350c(:,i),'descend')/ftse350(i);
end
f350year=zeros(356,11);
for i=1:11
f350year(:,i)=mean(ftse350co(:,(i-1)*12+1:i*12),2);
end
loglog(1:m,f350year(1:m,2:3:11),'LineWidth',1.5)
xlabel('\fontsize{16}log-Rank');
ylabel('\fontsize{16}log-Weight');
le=legend('\fontsize{16}2003','\fontsize{16}2006'
,'\fontsize{16}2009','\fontsize{16}2012');
title('\fontsize{16}Capital distribution curve of FTSE350
companies from 2003-2012');
```

2. The similar graph has been plotted for the whole UK market, shown in Figure 2. The mnemonics in *Datastream* for the whole UK market capitalisation and the market constituents are TOTMKUK and LTOTMKUK respectively. There are 550 companies in total.

3. The codes of plotting the graphs in Figure 3 and Figure 4 are shown below

```
load('matlab.mat')
f350year=zeros(356,11);
for i=1:11
f350year(:,i)=mean(ftse350c(:,(i-1)*12+1:i*12),2);
end
n=356;
nexti=zeros(n,10);
[ftse350co,order]=sort(ftse350c,'descend');
for j=1:n
for i=1:10
nexti(j,i)=find(order(:,i+1)==order(j,i));
end
end
plot([1:n]',nexti','o','MarkerSize',3);
hold all;
frank=mean(nexti,2);
%plot(1:n,frank, 'LineWidth',1.5);
%hold all;
m=max(max(nexti));
%plot(1:n,1:n,'LineWidth',1.5);
%lgd=legend('Change curve','Ideal Curve');
set(gca,'PlotBoxAspectRatio',[1 1 1],'DataAspectRatio',[1 1
1],'XLim',[0,m],'YLim',[0,m]);
xlabel('\fontsize{16}Ranks'); ylabel('\fontsize{16}Ranks');
title('\fontsize{16}Change of ranks in the succeeding year -
FTSE350');
```

The parameter n is important here, as it is introduced to control the number of ranks considered in the plot. The detailed plot could be observed by changing this parameter.

4. In order to compute the densities efficiently, by noticing that the numerator part can be expressed as functions of $g(y) = e^{-\frac{(y+2k)^2}{2t}}$, we make a table of values by taking large number of values of y from -1 to 2. Then we interpolate the values to obtain a polynomial of order 40, as the density function. Below is the codes of making the table

```
function res=make_table(t)
scale=100; step=1/scale;
res=zeros(1,3*scale);
for i=1:3*scale
z=i*step-1;
sum=exp(-(z^2)/(2*t));
for k=1:200
sum=sum+exp(-(z+2*k)^2/(2*t))+exp(-(z-2*k)^2/(2*t));
end
res(1,i)=sum;
end
```

To fit the points:

```
function [res,fitvalue]=fit_table(table)
step=3/length(table);
domain=-1+step:step:2;
degree=40;
res=polyfit(domain,table,degree);
fitvalue=polyval(res,domain);
end
```

The accuracy achieved by this method is impressive, shown by the graph below.

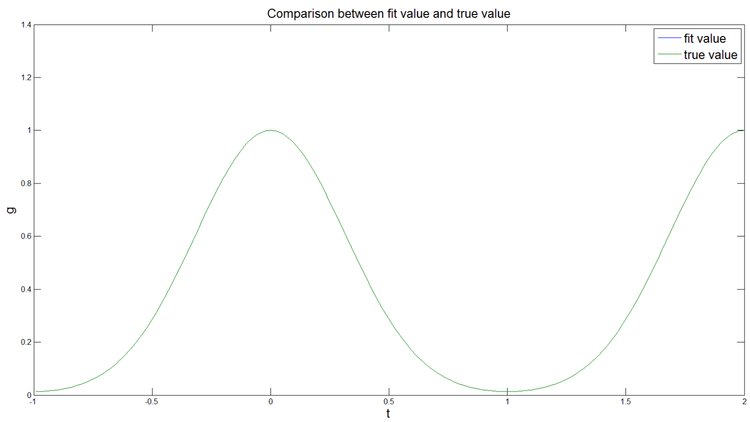

Figure 10. The comparison between the true value and the fit value of the exponential function. The two lines indeed overlapped perfectly.

The denominator part of the density is computed separately, as in the actual computation we divide this value after computing the numerator.

```
function res=int_value(x,t)
res=erf((1-x)/sqrt(2*t))-erf(-x/sqrt(2*t))-...
erf((1+x)/sqrt(2*t))+erf(x/sqrt(2*t));
for p=1:100
res=res+erf((1-x+2*p)/sqrt(2*t))-erf((-x+2*p)/sqrt(2*t))-...
erf((1+x+2*p)/sqrt(2*t))+erf((x+2*p)/sqrt(2*t))+...
erf((1-x-2*p)/sqrt(2*t))-erf((-x-2*p)/sqrt(2*t))-...
erf((1+x-2*p)/sqrt(2*t))+erf((x-2*p)/sqrt(2*t));
end
res=res*sqrt(2*pi*t)/2;
end
```

5. The joint density function of the sum of the assets $f_\eta(z)$ is written as a Matlab function, to make convenient use of function features.

```
function res=joint_v(w,t)
n=length(w); scale=100; table=make_table(t);
joint=zeros(1,n*scale);
deno=int_value(w(1,1),t);
for i=1:scale
v1=round(i-scale*w(1,1))+scale;
v2=round(i+scale*w(1,1))+scale;
joint(1,i)=(table(1,v1)-table(1,v2))/deno;
end
for i=2:n
deno=int_value(w(1,i),t);
temp=joint(1,1:(i-1)*scale);
for k=1:i*scale
z2=k/scale; low=round(100*max(0,z2-1));
up=round(100*min(i-1,z2)); sum=0;
for j=low+1:up
pos1=round(k-j-w(1,i)*100+100);
pos2=round(k-j+w(1,i)*100+100);
sum=sum+temp(1,j)*(table(1,pos1)-table(pos2));
end
joint(1,k)=sum/deno/scale;
end
end
res=joint;
end
```

6. The following codes are for simulating the weights, and plotting the graph in Figure 9

```
n=length(w); table=make_table(t);
[co_table b]=fit_table(table);
scale=100; degree=20; joint=zeros(n,(n-1)*scale);
co=zeros(n,degree+1); ww=zeros(1,n-1); res=zeros(n,1500);
for i=1:n
for j=1:i-1
ww(1,j)=w(1,j);
end
for j=i+1:n
ww(1,j-1)=w(1,j);
end
joint(i,:)=joint_v(ww,t);
co(i,:)=polyfit(1/scale:1/scale:n-1,joint(i,:),degree);
end
%vw=zeros(1,20*scale);
for k=1:n
deno=int_value(w(1,k),t);
for i=1:10*100 %z=i/scale;
if i>(n-1)*100
up=(n-1)*100/i;
else up=1;
end summ=0;
for j=up/100:up/100:up
summ=summ+j*(polyval(co_table,j-w(1,k))-...
polyval(co_table,j+w(1,k)))*polyval(co(k,:),i/100*j);
end
res(k,i)=summ*(up/100)*(1+i/100)^2/deno;
end
for i=11:510
up=(n-1)/i; summ=0;
for j=up/100:up/100:up
summ=summ+j*(polyval(co_table,j-w(1,k))-...
polyval(co_table,j+w(1,k)))*polyval(co(k,:),i*j);
end
res(k,i-10+1000)=summ*(up/100)*(1+i)^2/deno;
end
end
domain=zeros(1,1500);
for i=1:1000
domain(1,i)=1/(1+i/100);
end
for i=1001:1500
domain(1,i)=1/(1+i-990);
end
plot(domain,res);
set(gca,’YLim’,[0,5]);
title(’\fontsize{16}Density of the weights of five...
assets from FTSE350’);
xlabel(’\fontsize{16}z’); ylabel(’\fontsize{16}f’);
legend(’\fontsize{16}x=0.5562’,’\fontsize{16}x=0.2297’,...
’\fontsize{16}x=0.1045’,’\fontsize{16}x=0.0630’,...
’\fontsize{16}x=0.0465’);
```

7. To compute the expectations, we use the following codes:

```
for k=1:n
co_weight(k,:)=polyfit(domain(1,:),res(k,:),80);
end
syms x y;
for j=1:n
y=0;
for i=1:41
y=y+co_weight(j,i)*x^(42-i);
end
int_value(j,1)=int(y,x,0,0.8);
end
```

References

[1] E. Robert Fernholz, I. Karatzas, 2009. Mathematical Modelling and Numerical Methods in Finance, Volume 15: Special Volume (Handbook of Numerical Analysis). North Holland.

[2] E. Robert Fernholz, 2002. Stochastic Portfolio Theory . Springer.

[3] I. Karatzas, 1991. Brownian Motion and Stochastic Calculus (Graduate Texts in Mathematics). 2nd Edition. Springer.

[4] E. Robert Fernholz, I. Karatzas, 2005. Relative arbitrage in volatility-stabilized markets, Annals of Finance, Springer, Volume 1, Number 2, pages 149-177.

[5] E. Robert Fernholz, 2005. Diversity and relative arbitrage in equity markets. Finance and Stochastics, Volume 9, Number 1, pages 1-27.

[6] A. Banner, D. Fernholz, I. Karatzas, 2005. On Atlas models of equity markets. Annals of Applied Probability, Volume 15, pages 2296-2330.

[7] A. N. Borodin, P. Salminen, 2002. Handbook of Brownian Motion - Facts and Formulae (Probability and its Applications). 2nd Edition. Birkhäuser Basel.

[8] S. E. Shreve, 2004. Stochastic Calculus for Finance II: Continuous-Time Models (Springer Finance). Springer.