Truncated Ordered Stick Breaking Financial Market Model and Corresponding Bayesian Estimation

TRUNCATED ORDERED STICK BREAKING FINANCIAL MARKET MODEL AND CORRESPONDING BAYESIAN ESTIMATION

BY

MU HE, B.Sc.

A THESIS

SUBMITTED TO THE DEPARTMENT OF MATHEMATICS & STATISTICS

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Master of Science (2016) McMaster University

(Mathematics & Statistics) Hamilton, Ontario, Canada

TITLE:	Truncated Ordered Stick Breaking Financial Market Model and Corresponding Bayesian Estimation
AUTHOR:	Mu He B.Sc. Applied Mathematics The Hong Kong Polytechnic University
SUPERVISOR:	Prof. Shui Feng
NUMBER OF PAGES:	x, 61

To my family

Abstract

The Poisson-Dirichlet distribution is a probability distribution on the infinite-dimensional simplex, which has applications in population genetics, Bayesian statistics and finance & economics partition theories. Recently the Poisson-Dirichlet distribution has been used in modeling market structure and portfolio behaviours. The financial market model applying the GEM distribution, which is named after Griffiths, Engen and McCloskey, has just been introduced and some improvements can be discussed.

This thesis focuses on the introduction and development of a two-parameter Poisson-Dirichlet distribution and GEM distribution, modification of the financial market model by the truncated ordered stick breaking process and Bayesian estimation of the new models.

To summarize, the two new truncated ordered stick breaking model introduced give restrictions on the ranks of the markets weights and show better fitting results for real data sets.

Acknowledgements

I will take this opportunity to thank my supervisor Dr. Shui Feng. It was a nice research experience to work under his supervision. Also, I want to thank my supervisory committee members, Dr. Viveros and Dr. Hoppe. Further, I am grateful for Dr. Balakrishnan's recommendations in my work, my friend Benedict Min-Oo for his kind help, Dr. Xiaojun Zhu for his suggestions, Tian Feng, Dejing Kong, Xinyang Wu, Yanling Jin, Chengwei Qin and all the other researchers and friends of mine for their kind assistance.

Notation and abbreviations

CRP ... Chinese Restaurant Process

GEM Distribution ... Griffiths, Engen and McCloskey Distribution

INID ... Independent but not Identically Distributed

ISBP ... Invariant under Size Biased Permutation

MCMC ... Markov Chain Monte Carlo

PD Distribution... Poisson-Dirichlet Distribution

SDE ... Stochastic Differential Equation

Contents

List of Figures

Chapter 1

Introduction

Poisson-Dirichlet Distribution and Financial Market Model. The Dirichlet process was introduced by Ferguson (1973), whose aim was to find a prior of distributions in the infinite dimensional case. Hence, it could be applied as a prior for Bayesian non-parametrics statistics. Kingman (1975) introduced the Poisson-Dirichlet distribution, which is a probability measure on the infinite-dimensional simplex of ranked weights. The two-parameter Poisson-Dirichlet process was introduced by Perman *et al.* (1992) in the context of studying ranked jumps of subordinators. Related sampling formulas were developed by Ewens (1972) and Pitman and Yor (1997). The two-parameter Ewens' sampling formula of Poisson-Dirichlet distribution has been widely applied towards the species diversity in the field of biology study. Definitions and properties of the family are demonstrated in Chapter 2. The review is mainly based on Johnson *et al.* (1997), Feng (2010), Sibuya (2014) and Carlton (1999).

The Capital Distribution Curve. Stochastic Portfolio Theory is a financial framework modeling the behavior of portfolios and the capitalization market structure. The Rank-based model is one of the results derived from Stochastic Portfolio Theory, which was introduced by Fernholz (2002). The distribution curve, defined as the log-log plots of capital weights and ranks, gives a stable pattern for modeling the stock markets.

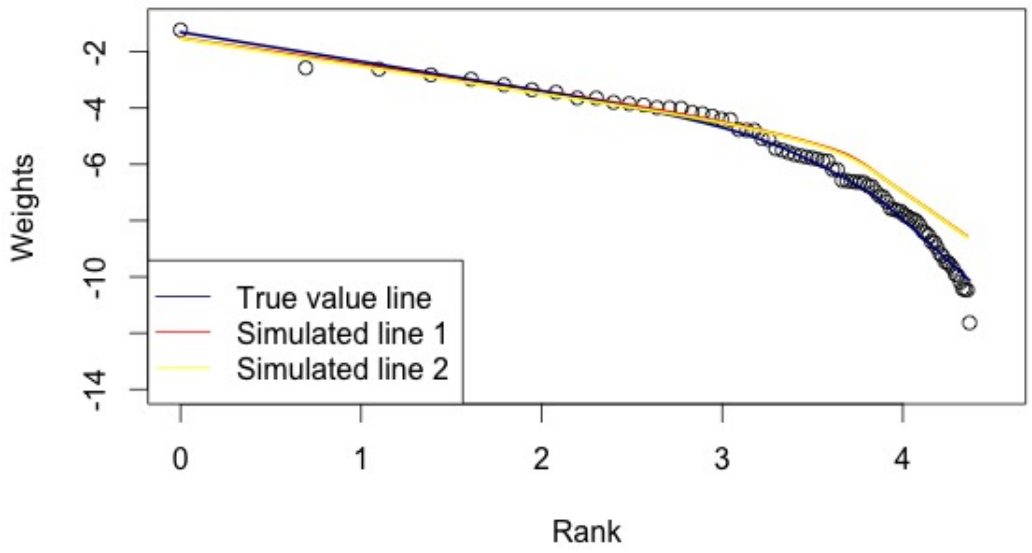

Figure 1.1: Log-log plot: The world stock market weights versus ranks

Chatterjee and Pal (2010) proved that the point process generated by the market weights converges weakly to a Poisson-Dirichlet Process under certain circumstances. This theorem supported the idea of applying the Poisson-Dirichlet distribution in financal applications. This makes sense in reality if we imagine that the capitalizations

are the set to be randomly partitioned into different stocks in different countries. Figure 1.1 gives an intuitive view, in which data is from the World Federation of Exchanges monthly report in 2014.

Sosnovskiy (2015) gave a model using the two-parameter GEM distribution for modeling the market weights based on the results of Feng and Wang (2007) and in this paper we modify the model for a better fitting. The details are reviewed in Section 2.2, mainly based on Fernholz (2002). New models are presented in Chapter 3.

Permanent Approximation. Some important results in order statsitics for independent but not identical (INID) variables are used in our work. Permanent and INID order statistics are reviewed in Section 2.3, mainly based on Balakrishnan (2007). To conclude, computing the permanent is a research bottleneck without effective solutions. The method for computing the exact value of the permanent most efficiently is Ryser's algorithm by Ryser (1963). Valiant (1979) gave an explanation that the exact computation can not finish in polynomial time. Jerrum *et al.* (2004) gave a polynomial-time approximation algorithm using Markov Chains to sample from weighted permutations. Huber and Law's alogrithm by Huber and Law (2008) is applied in our inference, which gives a faster computational time. In Section 2.4, we briefly go over the algorithms above.

The results and simulations are shown in Chapter 4, and Chapter 5 presents the benefits, limitations and future work.

Chapter 2

Literature Review

2.1 Stochastic Portfolio Theory

The main objective of this section is to review the work of Stochastic Portfolio Theory. For more details refer to the book of Fernholz (2002).

2.1.1 Basic Model

To begin with, we introduce a few assumptions for the theory:

1. In the market, the number of companies is a bounded constant, which means no new shares are issued and no companies collapse or merge.

2. There are no transaction costs or taxes and the value of shares is infinitely divisible.

3. Assume that there are no dividends.

Under the above assumptions, we suppose that there are n stocks considered. The stock price $\{S_i(t), i = 1, ..., n\}$ follows random process, which is defined on a probability space $\{\Omega, \mathcal{F}, P\}$. The randomness of the model is from n-dimensional Brownian motion:

$$\mathbb{W} = \{W(t) = (W_1(t), ..., W_n(t), \mathcal{F}_t, t \in [0, \infty)\} \tag{2.1}$$

where the filtration $\mathcal{F}_t$ is the augmentation of the natural filtration $\{\mathcal{F}_t^W = \sigma(W(s, 0 < s < t))\}$ under P.

The markets' stocks are modeled by the logarithmic model for simplicity. We can assume every company has only one share, therefore the price and capitalization process are equal.

Stock Price Process is defined as $S_i(t), i = 1, ..., n$ that satisfies the stochastic system:

$$d(\log S_i(t)) = \gamma_i(t)dt + \sum_{v=1}^{n} \xi_{iv}(t)dW_v(t), \quad t \in [0, \infty) \tag{2.2}$$

where the growth rate process γ_i satisfies:

$$\int_0^T |\gamma_i(t)|dt < \infty, \quad \text{for all } t \in [0, \infty) \quad \text{a.s.} \tag{2.3}$$

where the variance process ξ_{iv} is measurable, adapted and satisfies:

$$\begin{aligned} &\int_0^T (\xi_{i1}^2(t) + ... + \xi_{in}^2(t))dt < \infty, \quad t \in [0, \infty) \quad a.s. \\ &\lim_{t\to\infty} t^{-1}(\xi_{i1}^2(t) + ... + \xi_{in}^2(t)) \log\log(t) = 0 \quad a.s. \\ &\xi_{i1}^2(t) + ... + \xi_{in}^2(t) > 0 \quad t \in [0, \infty) \quad a.s. \end{aligned} \tag{2.4}$$

where a process $\{S(t), \mathcal{F}_t, t \in [0, \infty)\}$ is adapted means $S(t)$ is $\mathcal{F}_t$ measurable for $t \in [0, \infty)$.

(2.2) can be modified into a normal form:

$$dS_i(t) = (\gamma_i(t) + \frac{1}{2}\sum_{v=1}^{n} \xi_v^2(t))S_i(t)dt + S_i(t)\sum_{v=1}^{n} \xi_{iv}(t)dW_v(t), \qquad t \in [0, \infty) \quad (2.5)$$

Accordingly, the variance and covariance process of the stock are defined as:

$$\sigma_{ii}(t) = \xi_{i1}^2(t) + ... + \xi_{in}^2(t), \qquad \sigma_{ij}(t) = \sum_{v=1}^{n} \xi_{iv}(t)\xi_{jv}(t), \qquad t \in [0, \infty) \quad (2.6)$$

The return process α_i is related to the growth rate process γ_i for the i-th stock:

$$\alpha_i(t) = \gamma_i(t) + \frac{\sigma_{ii}(t)}{2}, \qquad t \in [0, \infty) \quad (2.7)$$

A market M is defined as a family of stocks $\{S_1, ..., S_n\}$ and a portfolio Π is defined as the weights of each stock in the portfolio $\pi_1, ..., \pi_n$.

Naturally, the market portfolio is defined as $\mathbf{X}$ which contains $X_1, X_2, ..., X_n$:

$$X_i = \frac{S_i(t)}{S_1(t) + S_2(t) + ... + S_n(t)}, \quad t \in [0, \infty) \quad (2.8)$$

where $i = 1, ..., n$. The sum of stock or capitalization is defined as:

$$T_X(t) = S_1(t) + S_2(t) + ... + S_n(t), \quad t \in [0, \infty) \quad (2.9)$$

which represents the value of the total market at time t. Suppose then we have a $T_\Pi(t)$, which represents the value of investment in the portfolio Π at time t. $T_{\pi_i}(t)$

represents the value of investment for the i-th stock in the portfolio Π at time t.

$$T_\Pi(t) = \sum_{i=1}^{n} T_{\pi_i}(t), \qquad T_{\pi_i}(t) = \pi_i T_\Pi(t), \quad \sum_{i=1}^{n} \pi_i = 1, \quad t \in [0, \infty) \tag{2.10}$$

Our main concern is on the relative performance with the market:

$$d\big(\log\big(\frac{T_\Pi(t)}{T_X(t)}\big)\big) = \sum_{i=1}^{n} \pi_i(t) d(\log(X_i(t))) + \gamma_\pi^*(t)dt, \quad t \in [0, \infty) \tag{2.11}$$

where,

$$\gamma_\pi^*(t) = \frac{1}{2}(\sum_{i=1}^{n} \pi_i(t)\sigma_{ii}(t) - \sum_{i,j=1}^{n} \pi_i(t)\pi_j(t)\sigma_{ij}(t)), \qquad t \in [0, \infty) \tag{2.12}$$

Therefore, our portfolio's perfomance depends on the changing of market weights X_i and excess growth rate γ_π^*, leading to the importance of understanding the market weight performance.

2.1.2 Rank Based Model and Capital Distribution Curve

Further development of the theory gives extension over capital distribution and capital distribution curve. The capital distribution of the market is defined to be the family of ranked markets weights:

$$X_{(1)}(t) \geq X_{(2)}(t) \geq \cdots \geq X_{(n)}(t) \tag{2.13}$$

Meanwhile, capital distribution curve is defined as the log-log plot of the market weights ranked in descending order.

In research of the capital distribution curve, many assumptions and results come

out. One of the theorems is proposed by Chatterjee and Pal (2010) that the limiting point process representing the markets weights converges weakly to a one-parameter Poisson-Dirichlet Process with parameter 2η if and only if $\eta \in (0, 1/2)$, where η satisfies:

$$\begin{aligned} &\lim_{n\to\infty} (\bar{\gamma}(n) - \gamma_i(n)) = \eta \quad \mathrm{i} \geq 1 \\ &\lim_{n\to\infty} \sup \max(\bar{\gamma}(n) - \gamma_i(n)) \leq \eta \qquad \mathrm{i} \geq 1 \end{aligned} \tag{2.14}$$

where $\bar{\gamma}(n)$ means the average of the growth rate.

To understand the theorem in an intuitive way, we may consider that under the assumptions, the market can be considered as a closed system. When investors invest in the market, the flow of capitalizations is similar to assigning partitions into different stocks. Hence, this theorem has a good explanation in reality. Later, the two-paremeter Poisson-Dirichlet distribution financial market model is introduced, which will be reviewed in the next section in detail.

2.2 Dirichlet Distribution Family

2.2.1 Dirichlet Distribution

To define Dirichlet distribution, the term simplex is needed. Simplex is a surface in $\mathbb{R}^n$ space, which is the generalization of a tetrahedral region space to n dimensions. A set of n components vector $\mathbf{y}$ is defined as $\{y_1, y_2, ..., y_n\}$, which usually denoted by Δ_n, satisfying $\Delta_n = \{\mathbf{y} \in \mathbb{R}^n | \sum_{j=1}^n y_j = 1, y_j \geq 0, \text{for } j = 1, 2, 3, ...n\}$. Then a probability mass function can be defined on a $(n-1)$ dimensional probability simplex. Dirichlet distribution can be considered as a probability distribution over the $(n-1)$ dimensinal probability simplex Δ_n. The distribution density function is given as:

$$P_{Y_1,...,Y_n}(y_1, ..., y_n) = \frac{\Gamma(\sum_{j=1}^n \theta_j)}{\prod_{j=1}^n \Gamma(\theta_j)} \prod_{j=1}^n y_j^{\theta_j - 1} \tag{2.15}$$

Usually, we denote $\theta_0 = \sum_{j=1}^n \theta_j$.

If $\mathbf{P} \sim D(\theta_1, \theta_2, ..., \theta_n)$, then the marginal distribution P_i follows Beta$(\theta_i, \theta_0 - \theta_i)$. This indicates Dirichlet distribution is the general case of beta distribution in higher dimensions.

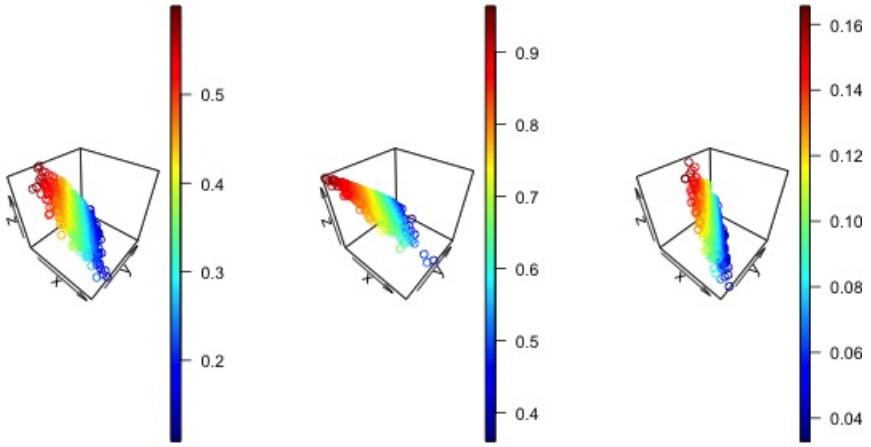

Figure 2.1: Examples of Dirichlet distribution with various parameters

Figure 2.1 shows a 1000 sample drawn from Dirichet(10,10,10), Dirichlet(1,15,100) and Dirichlet(100,50,15).

To generalize the Dirichlet distribution into an infinite case, Ferguson (1973) gave the definition of Dirichlet process by defining a random probability measure $\mathfrak{DP} \sim DP(\theta, \mathfrak{B})$ on a measuable space ($\mathcal{X}$,$\mathcal{B}$). The measure $\mathfrak{DP}$ with a scale paramter θ and a measure $\mathfrak{B}$ is a Dirichlet process if for any finite measurable partitions of the set $\mathcal{X}$: $(B_1, B_2..., B_n)$

$$\mathfrak{DP}(B_1), \mathfrak{DP}(B_2), ..., \mathfrak{DP}(B_n) \sim \text{Dirichlet}(\theta\mathfrak{B}(B_1), \theta\mathfrak{B}(B_2), ..., \theta\mathfrak{B}(B_n)) \tag{2.16}$$

Usually, there are three ways to generate samples from Dirichlet distribution or Dirichlet process.

Generating dirichlet distribution from random gamma variables is the most computationally efficient method. First, we generate n gamma random variables G_i, $i = 1, ..., n$, from Gamma$(\theta, 1)$. Then, the normalized pmf $q_i = \frac{G_i}{\sum_{j=1}^{n} G_i}$ follows Dirichlet(θ_i) distribution. This could be proved by finding the joint distribution by the Jacobian directly.

Another possible method is known as P*ó*lya urn model or Chinese Restaurant Process(CRP). Suppose there are infinite tables in a Chinese restaurant with infinitely many customers who can sit at one table, and each table initially offers one dish. The first customer X_1 always sits at 1st table. After the n-th customer comes in, we have k tables $U_1, ..., U_k$ with each table having customers $n_1, ..., n_k$. While the $n+1$-th customer has two options: to get a new table or sit at the k-th table that already having n_k people. The probability is set as:

$$\begin{cases} P(X_{n+1} \text{ into a new table } U_{n+1}) = \frac{\theta}{\theta+n} \\ P(X_{n+1} \text{ into an existing table } U_i) = \frac{n_i}{\theta+n} \end{cases} \qquad i = 1, ..., k \tag{2.17}$$

where θ is the dispersion value of DP and n is the total number of customers in the restaurant at a given time. Therefore, we could also generate Dirichlet samples through CRP.

The last way to generate dirichlet distribution or dirichlet process raised from a stick breaking process. Assuming that we have a stick with length 1, we break the stick by a piece of propotion and keep breaking for infinite steps. More formally, a random probability measure $\mathfrak{DP}$ defined on ($\mathcal{X}$,$\mathcal{B}$) by

$$\mathfrak{DP} = \sum_{n=1}^{\infty} \Pi_n \sigma_{Y_n} \tag{2.18}$$

is a Dirichlet process $\mathrm{DP}(\theta, \mathfrak{B})$, where σ_{Y_n} is a delta function that takes one if $Y = Y_n$. Then we set

$$\begin{aligned} &\beta_k \sim \mathrm{Beta}(1, \theta), \quad Y_n \sim \mathfrak{B} \\ &\pi_k = \beta_k \prod_{l=1}^{k-1}(1-\beta_l), \quad \mathfrak{DP} = \sum_{n=1}^{\infty} \Pi_n \sigma_{Y_n} \end{aligned} \tag{2.19}$$

Hence we get a stick breaking process for generating Dirichlet samples.

The stick breaking distribution over π is sometimes written as $\mathrm{GEM}(0, \theta)$ distribution. This leads to the importance of understanding GEM distribution, which is named after Griffiths, Engen and McCloskey who contributed the research of this distribution, see Ewens (1990).

GEM distribution has a remarkable property. Given a probability sequence (P_n) representing proportions for a population Z_+ of infinite distinct types of objects, if we take samples from the population, it is the same we make a permutation of (P_n) respective to the order in which the different types are observed. This procedure is equivalent to a reordering of (P_n), getting a $(\tilde{P_n})$: $\tilde{P_1} = P_n$ if the first component of the sample is of type n, $n \in Z_+$, and $\tilde{P_2} = P_m$ if the following member of the sample not from n is from type m, $m \in Z_+|n$ and so forth. Mathematical definition is given as:

$$\begin{aligned} &P(\tilde{P_1} = P_n | P_1, P_2, ...) = P_n \\ &P(\tilde{P_{j+1}} = P_n | \tilde{P_1}, ..., \tilde{P_j}, P_1, P_2, ...) = \frac{P_n}{1 - \tilde{P_1} - ... - \tilde{P_j}} I(P_n \neq \tilde{P_1}, ..., \tilde{P_j}) \end{aligned} \tag{2.20}$$

Then, the new sequence $(\tilde{P_n})$ is a size biased permutation of (P_n).

If $(\tilde{P_n}) \xrightarrow{d} (P_n)$, then (P_n) is invariant under size-biased permutation(ISBP). In our case, GEM distribution is ISBP.

The ordered distribution of Dirichlet distribution or above GEM distribution is known as Poisson-Dirichlet distribution(PD) $\sim$ PD$(0, \theta)$. Generally this is also called one-parameter Poisson-Dirichlet distribution. As the one-parameter case is not the focus of our work, we will skip it and talk about a more general two-parameter case in the next section.

2.2.2 Two-Parameter Poisson-Dirichlet Distribution

Poisson-Dirichlet distribution is the infinite and ranked version of Dirichlet distribution. It is an infinitely dimensional simplex whose sides length are sorted. A formal definition is given by stick breaking process, see Feng (2010): Suppose $PD(\alpha, \theta)$ is the two-parameter Poisson-Dirichlet distribution with $\alpha \in [0, 1)$, $\theta \in (-\alpha, \infty)$. The $PD(\alpha, \theta)$ has the law of descending order $(V_1^{\alpha,\theta}, V_2^{\alpha,\theta}, ...)$. The size-biased random permutation $(V_1, V_2, ...)$ of $PD(\alpha, \theta)$ is a stick breaking process:

$$V_1 = U_1, \quad V_2 = (1 - U_1)U_2, \quad V_3 = (1 - U_1)(1 - U_2)U_3, \quad ... \tag{2.21}$$

where $\{U_k : k = 1, 2, ...\}$ are independent $Beta(1 - \alpha, \theta + k\alpha)$ random variables.
When Poisson-Dirichlet distribution is discussed, it is always worth mentioning GEM distribution and the result can be easily deduced from above definition that the two-paramter Poisson-Dirichlet distribution is the ranked distribution of GEM(α, θ) distribution.
Similar to the Dirichlet distribution, we can also fomulate an urn model to describe the two-parameter Poisson-Dirichlet distribution. The two-parameter Ewens' sampling formula is a typical formula for two parameter random partitions family. Suppose there are balls $B_1, B_2, ...$ randomly distributed into urns $U_1, U_2, ...$

The rule of distribution is B_1 goes to U_1 with probability 1, while $B_2, B_3..., B_n$ are put into $U_1, ..., U_k$ with probability:

$$\begin{cases} P(B_{n+1} \text{ into a new } U_{n+1}) = \frac{\theta+k\alpha}{\theta+n} \\ P(B_{n+1} \text{ into an old } U_j) = \frac{c_j-\alpha}{\theta+n} \end{cases} \tag{2.22}$$

where c_j are the balls in the existing urns U_j, $j = 1, ...k$, also $\sum_{j=1}^{k} c_j = n$.
Thus, we can model it with an ordered partition set $a_{(1)}, .., a_{(k)}$, satisfying:

$$P(\mathbf{a}, \theta, \alpha) = \frac{1}{\theta^{[n]}} \prod_{j=1}^{k} (\theta + (j-1)\alpha)(1-\alpha)^{c_j - 1} \tag{2.23}$$

where $0 \leq \alpha \leq 1$ and $\theta \geq -\alpha$ or $\alpha < 0$ and $\theta = -m\alpha$. In most circumstances, we consider the first condition. $c_j = |a_{(j)}|$, which is the number of components in each partition $a_{(j)}$ while $x^{[n]}$ is the upper factorial. Here, the probability has nothing with the partition components.
Moreover, we can define a size index $\mathbf{S}$ satisfying:

$$s_j = |i : c_i = j, i = 1, ..., k|, \quad j = 1, ..., n \quad \sum_{j=1}^{n} s_j = k \quad \sum_{j=1}^{n} j s_j = n \tag{2.24}$$

Similarly, an unordered partition set $a_1, ..., a_k$ satisfies the probability:

$$P(\mathbf{a}, \theta, \alpha) = \frac{1}{\theta^{[n]}} \prod_{i=1}^{k} (\theta + (i-1)\alpha) \prod_{j=1}^{n} ((1-\alpha)^{[j-1]})^{s_j} \tag{2.25}$$

Considering another Urn model where both balls and urns are undistinguished, which is similar to Ewens' sampling formula, generally known as Pitman's Sampling Formula

or two-parameter Ewens' sampling formula:

$$P(\mathbf{S}, \theta, \alpha) = \frac{n!}{s_j! \theta^{[n]}} \prod_{i=1}^{k} (\theta + (i-1)\alpha) \prod_{j=1}^{n} (\frac{(1-\alpha)^{[j-1]}}{j!})^{s_j} \tag{2.26}$$

2.2.3 Inference of Poisson-Dirichlet Distribution

As for Poisson-Dirichlet distribution, both one-parameter and two-parameter case require infinitely dimensional data. Even there exists finite version of probability density function for Poisson-Dirichlet distribution, it still contains infinite sum. Hence, the inference of the parameters is conducted based on the sampling formulas. The most common ways are least squares estimator and maximum likelihood estimator. Besides, there are alternative methods and we will introduce some of them.

Least Squares Fit

Sosnovskiy (2015) used the least squares methods to estimate parameters for two-parameter Poisson-Dirichlet Distribution, which is an empirical method averaging the simulated samples to fit the least squares function. By finding the least sum of squares, he gave an intuitive result that Poisson-Dirichlet distribution is a good model in capitalization market and portfolio analysis.

More generally, Sibuya (2014) showed a least square estimator. In the urn model, least sum of squares estimator is deduced as:

$$(\hat{\theta}, \hat{\alpha}) = \arg\min || \frac{jS_j}{n} - E\big(\frac{jS_j}{n}\big) || \tag{2.27}$$

where

$$E(\frac{jS_j}{n}) = \begin{pmatrix} n-1 \\ j-1 \end{pmatrix} \frac{(\theta+\alpha)^{[n-j]}(1-\alpha)^{[j-1]}}{(\theta+1)^{[n-1]}} \tag{2.28}$$

Here, j is the group with exact j representives and s_j indicates the size index of the group with size j.

Maximum Likelihood Estimation

Carlton (1999) gave a detailed view of estimation using the maximum likelihood estimator. When both estimators are unknown, the necessary but not sufficient condition for MLEs$(\hat{\alpha_n}, \hat{\theta}_n)$ to exist is that both first derivatives equal zero and the Hessian matrix satisfies certain conditions. The log-likelihood function is calculated as:

$$\begin{aligned} l(\alpha,\theta) &= \log(P_n(\mathbf{S}|\alpha,\theta)) \\ &= \text{Constant} - \sum_{l=1}^{n-1}\log(\theta+l) + \sum_{m=1}^{k-1}\log(\theta+m\alpha) + \sum_{j=2}^{n} s_j \sum_{i=2}^{j-1}\log(i-\alpha) \end{aligned} \tag{2.29}$$

Then the corresponding first derivatives are:

$$l_\alpha(\alpha,\theta) = \sum_{m=1}^{k-1}\frac{m}{\theta+m\alpha} - \sum_{j=2}^{n} s_j \sum_{i=1}^{j-1}\frac{1}{i-\alpha} \tag{2.30}$$

$$l_\theta(\alpha,\theta) = -\sum_{l=1}^{n-1}\frac{1}{\theta+l} + \sum_{m=1}^{k-1}\frac{1}{\theta+m\alpha} \tag{2.31}$$

Meanwhile, to find the Hessian matrix, we need to find the second derivatives:

$$H(\alpha,\theta) = \begin{bmatrix} l_{\alpha\alpha} & l_{\theta\alpha} \\ l_{\alpha\theta} & l_{\theta\theta} \end{bmatrix} \tag{2.32}$$

where,

$$l_{\alpha\alpha} = -\sum_{m=1}^{k-1} \frac{m^2}{(\theta + m\alpha)^2} - \sum_{j=2}^{n} s_j \sum_{i}^{j-1} \frac{1}{(i-\alpha)^2} \tag{2.33}$$

$$l_{\theta\theta} = \sum_{i=1}^{n-1} \frac{1}{(\theta + l)^2} - \sum_{m=1}^{k-1} \frac{1}{(\theta + m\alpha)^2} \tag{2.34}$$

$$l_{\alpha\theta} = -\sum_{m}^{k-1} \frac{m}{(\theta + m\alpha)^2} \tag{2.35}$$

When $\alpha > 0$, the analytic solution of MLE can not be found. In addition, $\hat{\theta}$ is always not a consistent estimator of θ, even if we can get a numeric solution.

Other Possible Estimators

Another possible estimator is also given by Sibuya (2014). Denote R_i is a function of $\mathbf{S}$:

$$R_i = \sum_{j=1}^{n} \frac{j^{[i]} s_j}{n^{[i]}}, \quad i = 0, 1, ... \quad (R_0 = k, R_1 = n) \tag{2.36}$$

Then, the simplest estimation is

$$\hat{\alpha} = \frac{s_1}{R_0}, \qquad \hat{\theta} = \frac{(1 - \hat{\alpha})}{R_2 - 1} \tag{2.37}$$

While some complicated estimators are represented by higher Rs, such as R_2 and R_3:

$$\hat{\alpha} = \frac{\frac{R_3}{R_2} - 2R_2 + R_3}{\frac{R_3}{R_2} - R_2} \tag{2.38}$$

$$\hat{\theta} = \frac{1 + R_2 - \frac{2R_3}{R_2}}{\frac{R_3}{R_2} - R_2} \tag{2.39}$$

These methods are similar to method of moment.

2.2.4 Poisson-Dirichlet Market Model

Prior to the introduction of the Market Model, a few definitions are to be reviewed first.

A diffusion Z={Z(t): $t > 0$ } is a continuous-time Markov process with value in an interval (l,r) if:

1. the sample paths of Z are almost surely continuous
2. for every Z in (l,r) and every $\epsilon > 0$, the following limits exist:

$$\begin{aligned} b(z) &= \lim_{j \to 0} \frac{1}{h} \mathbb{E}(Z(t+h) - Z(t) | Z(t) = z) \\ a(z) &= \lim_{h \to 0} \frac{1}{h} \mathbb{E}((Z(t+h) - Z(t))^2 | Z(t) = z) \end{aligned} \tag{2.40}$$

where the functions b and a on (l,r) $\to \mathbb{R}$ are named infinitesimal drift and the infinitesimal variance of Z.

The infinitesimal drift $b(z)$ determines the expected change in a small increment of Z starting at z:

$$\mathbb{E}(Z(t) - Z(0) | Z(0) = z) = b(z)t + o(t) \tag{2.41}$$

Similarly, The infinitesimal variance $a(z)$ determines the variance of a small increment of Z starting at z:

$$\mathbb{E}((Z(t) - Z(0))^2 | Z(0) = z) = a(z)t + o(t) \tag{2.42}$$

A probability distribution π is called stationary for a continuous-time stochastic process Z if at every time $t \geq 0$, π is the distribution of the process at that time.
A Wright-Fisher diffusion process Z(t) driven by stochastic differential equation(SDE) is defined as:

$$dZ(t) = \frac{1}{2}(\alpha(1 - Z(t)) - \theta Z(t))dt + \sqrt{Z(t)(1 - Z(t))}dW(t) \tag{2.43}$$

where $Z(t)$ has stationary beta(α, θ) distribution.
Hence, we can model diffusion process with stationary Poisson-Dirichlet distribution by applying the stick breaking methods. The market model is based on the work of Feng and Wang (2007), in which reversibility of corresponding infinite-dimensional process was proved. The application was first introduced by Sosnovskiy (2015).
We consider a Wright-Fisher diffusion process Z(t). Accordingly, we can set $X_n(t)$ as the market weights of the n-th stock at time t, the stock weights are driven by stick breaking process

$$X_1(t) = Z_1(t), \quad ... \quad X_n(t) = Z_n(t)(1 - \sum_{i=1}^{n-1} X_i(t)) \tag{2.44}$$

where processes $Z_n(t)$ are determined by independent SDEs

$$dZ_n(t) = \frac{1}{2}((1 - \alpha)(1 - Z_n(t)) - (\theta + \alpha n)Z_n(t))dt + \sqrt{Z_n(t)(1 - Z_n(t))}dW_n(t) \tag{2.45}$$

with stationary beta distributions of size biased stick breaking:

$$Z_n(t) \sim \text{Beta}(1 - \alpha, \theta + n\alpha) \tag{2.46}$$

Initial values of processes $Z_n(0)$ are determined by

$$Z_1(0) = X_1(0), \quad ... \quad Z_n(0) = \frac{X_n(0)}{1 - \sum_{i=1}^{n-1} X_i(0)} \tag{2.47}$$

The market capitalization M(t) can be written as:

$$dM(t) = \frac{1}{2}(\theta - cM(t))d(t) + \sqrt{M(t)}dW(t) \tag{2.48}$$

where M(t) has a stationary gamma distribution Gamma(θ,c). The c could be calculated by M(0) = E(Gamma(θ,c)). Hence the local behaviour of the stock prices should be represented as:

$$P_n(t) = \frac{M(t) \times X_n(t)}{q_n}, \tag{2.49}$$

where q_n is defined as the number of shares of n-th company.

Some simulation results could be found in Chapter 3, where we modify the model and give some new pictures illustrating the stock weights and the market.

2.3 Order Statistics

2.3.1 Permanent

Order statistics for independent and identically distributed (IID) random variables have been studied intensively. While in most cases, the data or sample is not ideally IID. In our work, we need the joint distribution of a group of variables $X_1, ..., X_k$ to be independent Beta distributions with different parameters. Thus, the order statistics for independent but not identically distributed (INID) variables are needed. To find it, we need to define the permanent first.

The permanent of an n by n matrix A with components $A_{i,j}$, $i, j = 1, ..., n$ is defined as:

$$\text{Per}(A) = \sum_{\sigma} \prod_{i=1}^{n} A_{i,\sigma(i)} \tag{2.50}$$

where $\sum_{\sigma}$ is the sum over all permuations $\sigma(1), ..., \sigma(n)$ of $1, ..., n$.

From the definition, we have a direct comparision between determinant and permanent. The difference is that the permanent's operations do not have alternating signs, postive ones only.

Two important properties of the permanent are useful.

Property 1 Permanent of A does not change with the permutations of rows or columns.

Property 2 Suppose $A_{-i,-j}$ denotes the $n-1$ dimensional sub-matrix of A by deleting the i-th row and j-th column, then

$$\begin{aligned} \text{Per}(A) &= \sum_{i=1}^{n} a_{i,j}\text{Per}(A_{-i,-j}) \\ &= \sum_{j=1}^{n} a_{i,j}\text{Per}(A_{-i,-j}), \quad j = 1, 2, ..., n \end{aligned} \tag{2.51}$$

This property shows that the permanents can be expanded in a similar way to that of determinants.

2.3.2 Joint Distribution of INID Order Statistics

Let $X_1, X_2, ..., X_n$ be independent but not identically distributed random variables with cumulative distribution functions $F_{r_1}(x), F_{r_2}(x), ...F_{r_n}(X)$. Then, the joint CDF for the order statistics $X_{(r_1)}, X_{(r_2)}, ..., X_{(r_n)}$:

$$F_{X_{(r_1)},...,X_{(r_n)}} = \sum_{i_n=n}^{n} \cdots \sum_{i_2=2}^{i_3} \sum_{i_1=1}^{i_2} \frac{P_{i_1,...,i_n}(x_1, ..., x_n)}{i_1!(i_2 - i_1)! \ldots (n - i_n)!} \tag{2.52}$$

where

$$
\begin{aligned}
&P_{i_1,\ldots,i_n}(x_1,..,x_n) \\
&= \mathrm{Per}\begin{pmatrix}
F_1(x_1) & F_2(x_1) & \ldots & F_n(x_1) \\
\vdots & \vdots & & \vdots \\
F_1(x_1) & F_2(x_1) & \ldots & F_n(x_1) \\
F_1(x_2)-F_1(x_1) & F_2(x_2)-F_2(x_1) & \ldots & F_n(x_2)-F_n(x_1) \\
\vdots & \vdots & & \vdots \\
F_1(x_2)-F_1(x_1) & F_2(x_2)-F_2(x_1) & \ldots & F_n(x_2)-F_n(x_1) \\
\vdots & \vdots & & \vdots \\
1-F_1(x_n) & 1-F_2(x_n) & \ldots & 1-F_n(x_n) \\
\vdots & \vdots & & \vdots \\
1-F_1(x_n) & 1-F_2(x_n) & \ldots & 1-F_n(x_n)
\end{pmatrix}
\end{aligned}
\tag{2.53}
$$

where the j-th group, $j = 1, ..., k+1$, contains $i_j - i_{j-1}$ repetitions of the same row. Similarly, the joint density function of k variables $X_{(1)}, X_{(2)}, ..., X_{(k)}$ chosen from the sample $X_1, X_2, ..., X_n$ is

$$
\begin{aligned}
&f_{r_1,r_2,\ldots,r_k}(x_1,x_2,...,x_k) \\
&= \frac{1}{(r_1-1)!(r_2-r_1-1)!\ldots(r_k-r_{k-1}-1)!(n-r_k)!}\mathrm{Per}(A_k)
\end{aligned}
\tag{2.54}
$$

where $-\infty < x_1 < x_2 < ... < x_k < \infty$ and

$$A_k = \begin{pmatrix} F_1(x_1) & F_2(x_1) & \dots & F_n(x_1) \\ \vdots & \vdots & & \vdots \\ F_1(x_1) & F_2(x_1) & \dots & F_n(x_1) \\ f_1(x_1) & f_2(x_1) & \dots & f_n(x_1) \\ F_1(x_2) - F_1(x_1) & F_2(x_2) - F_2(x_1) & \dots & F_n(x_2) - F_n(x_1) \\ \vdots & \vdots & & \vdots \\ F_1(x_2) - F_1(x_1) & F_2(x_2) - F_2(x_1) & \dots & F_n(x_2) - F_n(x_1) \\ f_1(x_2) & f_2(x_2) & \dots & f_n(x_2) \\ \vdots & \vdots & & \vdots \\ F_1(x_n) - F_1(x_{k-1}) & F_2(x_k) - F_2(x_{k-1}) & \dots & F_n(x_k) - F_n(x_{k-1}) \\ \vdots & \vdots & & \vdots \\ F_1(x_k) - F_1(x_{k-1}) & F_2(x_k) - F_2(x_{k-1}) & \dots & F_n(x_k) - F_n(x_{k-1}) \\ f_1(x_k) & f_2(x_k) & \dots & f_n(x_k) \\ 1 - F_1(x_k) & 1 - F_2(x_k) & \dots & 1 - F_n(x_k) \\ \vdots & \vdots & & \vdots \\ 1 - F_1(x_k) & 1 - F_2(x_k) & \dots & 1 - F_n(x_k) \end{pmatrix} \tag{2.55}$$

where the row with entries $F_i(x_k), ...,$ where $i = 1, ..., n$ and $k = 1, ..., n$ contains $r_k - r_{k-1} - 1$ rows. Similarly, $f_i(x_k)$, where $i = 1, ..., n$ and $k = 1, ..., n$ contains 1 row. Lastly, the last group $1 - F_i(x_k)$ contains $n - r_k$ rows.

2.3.3 Ryser's Exact Algorithm

Ryser (1963) gave an algorithm to calculate the permanent value:

$$\mathrm{Per}(A) = \sum_{t=0}^{n-1}(-1)^t \sum_{X\in\Gamma_{n-t}} r_1(X)r_2(X)...r_n(X) \tag{2.56}$$

where $\Gamma_k = \{X \in \mathcal{R}^{n\times k} | \text{X consists of columns of A}\}$ is the set of all $n\times k$ submatrices of A and $r_i(X) =$ sum of row i-th of matrix X is the i-th row sum of X.
It could be finished in $O((n^2)(2^n))$ time. Later, Valiant (1979) showed the complexity of computing exact value of the permanent in a polynomial time, so researchers started working on approximation methods.

2.3.4 HL-algorithm for Permanent Approximation

Various methods in approximation of the permanent of a matrix has been developed, due to the complexity of calculating the exact value of the permanent. In the work of Huber and Law (2008), an efficient approximate algorithm solving the permanent for a non-negative matrix with expected running time $\mathrm{O}(n^4\log(n))$ is given.
They scale the matrix first and then find a counting method for approximation.

Scaling of Matrix

An $n \times n$ matrix is called a doubly stochastic matrix if its components $a_{i,j}$ satisfy:

$$\sum_{i=1}^{n} a_{i,j} = \sum_{j=1}^{m} a_{i,j} = 1 \tag{2.57}$$

For a matrix with entries of random non-negative values, we first scale it into a doubly stochastic matrix for calculation simpilicity. The algorithm is as follows:

1 Set n is the length of a $n \times n$ matrix, and a constraint parameter ϵ for error, and each row sums (rowsum(i)), $i = 1, ..., n$. Initialize the row scaling and column scaling parameters x and y as unit vectors.

2 While $\max(|\text{rowsums} - 1|) > \epsilon$, keep the iteration:

Set D_r as a diagonal matrix with entries of reciprocal of each row sums.

Update $A = D_r A$ to scale rows.

Update row scaling paramter $x = x \times$ reciprocal of each row sums.

Set D_c as a diagonal matrix with entries of reciprocal of each columns.

Update $A = D_c A$ to scale columns.

Update the column scaling parameter $y = y \times$ reciprocal of each row columns.

3 End, output the updated matrix B, x and y with error ϵ.

Hence, we get an updated matrix A with columns sums equal to 1 while row sums is near 1 within an error restriction ϵ.

After the first scaling, we get a nearly doubly stochastic matrix A. Then, we need to do another scaling work by dividing the largest entry of each row max(i). Thus, we get a new updated matrix and save updated A as A_{save} for later calculation. Thus, the range of the row sums r(i) are in the interval: $(\frac{1-\epsilon}{\max(i)}, \frac{1+\epsilon}{max(i)})$.

HL-Factor

After we get the manipulated matrix A, we now can get the row sums $r(i)$. A HL-factor is defined to find an upper bound of the matrix.

$$h(r) = \begin{cases} r + \frac{1}{2}\log(r) + e - 1, & r \geq 1 \\ 1 + (e-1)r, & r \in [0,1] \end{cases} \tag{2.58}$$

For any $n \times n$ matrix A with entries between 0 and 1, the row sums of the i-th row r(i), $i = 1, ..., n$.

$$\text{Per}(A) \leq \prod_{i=1}^{n} \frac{h(r(i))}{e} \tag{2.59}$$

Usually an accept/reject algorithm for a function is:

1 Find a $g(\alpha) \geq f(\alpha|\mathbf{X}, \theta)$ for all α. Calculating

$$\infty > c = \int_{-\infty}^{\infty} g(\alpha)d\alpha \geq 1 \qquad h(\alpha) = \frac{g(\alpha)}{c} \tag{2.60}$$

2 Generate a proposal from $h(\alpha)$ and a u = Uniform(0,1) random variable

3 Accept y as a simulation result if $u \leq \frac{f(\alpha|\mathbf{X},\theta)}{g(\alpha)}$, else repeat the enumerate.

Here, we have the similar pattern to the idea of the above method. First, we find an upper bound of the value of the permanent as in (2.59). Then we find a typical 'characteristic' function for the permanent based on random self-reducibility property of permanents, i.e. a typical feature that can represent the permanent. Last, we repeat the iterations and count the number of success, just like dropping needles and seeing how many are in the area (satisfying the features) of the target permanent. Details and a rigorous proof can be found in Huber and Law (2008).

Here, we present a brief idea. The idea of sampling is to generate variates W over Ω satisfying:

$$P(W = x) = \frac{w(x)}{Z} \tag{2.61}$$

where w(x) is a nonnegative weight for all x, and $Z = \sum_{x\in\Omega} w(x)$. Then, we assume Ω can be partitioned into $\Omega_1, ...\Omega_n$. For every Z_i, $Z_i = \sum_{x\in\Omega_i} w(x)$, and there is an upper bound: $Z_i \leq U(\Omega_i)$ and the upper bounds for partitions satisfying $\sum_{i=1}^{n} U(\Omega_i) \leq U(\Omega)$.

After the above settings, we can begin the algorithm. We first consider a random variable I generated with probability:

$$\begin{aligned} P(I = i) &= \frac{U(\Omega_i)}{U(\Omega)} \quad i = 1, ..., n \\ P(I = 0) &= 1 - \sum_{i=1}^{n} \frac{U(\Omega_i)}{U(\Omega)} \end{aligned} \tag{2.62}$$

Then, if we get any non-zero I, we remove $\Omega_{I=i}$ from Ω and set the reduced set Ω_{-I} as the new Ω.

If at some step, we generated an $I = 0$, which means this iteration fails, then we stop and start again. Else if there is no $I = 0$ and we finish the reducing process, we success in getting a random variable W, which equals the last single element.

In the permanent approximation, the set of permutations can be partitioned into n pieces by the choice of which row is assigned to the first colomn. Accordingly, we set the weight of a permuatation for A to be $w(\sigma) = \prod_{\sigma^{-1}(j)=1}^{n} A_{\sigma^{-1}(j),j}$ and $\Omega = \{\sigma : w(\sigma) > 0\}$. Then the steps follow as:

1 Set the iterations: k, input matrix A equals A_{save}.

2 For iterations from 1 to k,

Let $A = A_{save}$ for every new iteration:

For column j from 1 to n, we choose a row and reduce the matrix to $A_{-i,-j}$.

When the first $j-1$ columns have been reduced to A_{j-1}, we choose a row at column j: $\sigma_{(j)}^{-1} = i$ with probability for any row i:

$$P(\text{choose row i at column j}) = \begin{cases} \frac{A_{i,j}M(f(A_{j-1},i,j))}{M(A_{j-1})} & i = 1, ..., n \\ 1 - \sum_1^n \frac{A_{i,j}M(f(A_{i,j},i,j))}{M(A_{j-1})} & i = 0 \end{cases} \tag{2.63}$$

where

$$M(A) = \prod_{i=1}^{n} \frac{h(r(i))}{e} \tag{2.64}$$

$$f(A, i, j) = A_{-i,-j} \tag{2.65}$$

where $A_{-i,-j}$ is a matrix A with entry $i, j = 1$ and all other entries in row i, column j equal 0.

Then, if $i > 0$, comparing the cumulative distribution of choosing row i at column j with a uniform(0,1) random number. Assigning $\sigma_{(j)}^{-1} = i$, i.e. choosing a row i and A reduce to $f(A, i, j)$. If we can get a complete permutation from above algorithm, we say it is a success and add one to count of success.

3 Now get the approximate permanent value for scaled matrix, we can do the transformation:

$$s = \prod((diag(D_r) \times diag(D_c))) \times \prod_{i=1}^{n}(max(i))^{-1} \tag{2.66}$$

$$\text{Per}(A^*) = \frac{M(A) \times \text{count of success}}{\text{iterations} \times s} \tag{2.67}$$

Last, we can get an approximate value for permanent of a matrix A.

Chapter 3

Methodologies

In this chapter, we will introduce three truncated models applying the stick breaking process. The discussion is based on the assumptions that the process has converged at a time $t = N$ and $N \to \infty$. The samples of the ordered weights are chosen from the stationary distribution after that time point. We make such assumptions as our data sets are rather stable from time to time and we are interested in the value of parameters for the process.

3.1 Truncated Ordered Stick Breaking Models

3.1.1 Direct Model

To avoid the infinite dimension for the Market Model in Section 2.2.4, we can construct truncated models for real data set. One possible method is to define $X_1, X_2, ...X_n$ to

be the n-th largest weights directly and conduct the truncation:

$$\begin{aligned} &X_1 = Z_1 \quad X_2 = Z_2(1-Z_1) \quad \dots \quad X_n = Z_n(1-Z_1)...(1-Z_{n-1}) \\ &X_0 = 1 - \sum_{i=1}^{n} X_i \end{aligned} \tag{3.1}$$

where X_0 can be considered as a small amount left after finite steps $\lim_{n\to\infty} X_0 \to 0$. Also

$$Z_i \sim \text{Beta}(1-\alpha, \theta + i\alpha) \qquad i = 1, 2, ..., n \tag{3.2}$$

Hence, the joint distribution of $X_1, X_2, ...X_n$ can be calculated as

$$\begin{aligned} f(X_1, X_2, ..., X_n) &= f(Z_1, Z_2, ..., Z_n)|J| \\ &= f(Z_1)f(Z_2)...f(Z_n)|J| \\ &= \prod_{i=1}^{n} \big(\frac{1}{\text{B}(1-\alpha, \theta+i\alpha)}\big) \times {X_1}^{-\alpha}(1-X_1)^{\theta+\alpha-1} \\ &\times \prod_{j=2}^{n} (\frac{X_i}{1-\sum_{k=1}^{j-1} X_k})^{-\alpha} (\frac{1-\sum_{k=1}^{j} X_k}{1-\sum_{k=1}^{j-1} X_k})^{\theta+j\alpha-1} \times |J| \\ &= \prod_{i=1}^{n} (\frac{1}{\text{B}(1-\alpha, \theta+i\alpha)}) \times \prod_{j=1}^{n} X_j^{-\alpha} \\ &\times (1 - X_1 - X_2 - \cdots - X_n)^{\theta+n\alpha-1} \times |J| \end{aligned} \tag{3.3}$$

where

$$
\begin{aligned}
J &= \begin{pmatrix} \frac{\partial Z_1}{\partial X_1} & \frac{\partial Z_1}{\partial X_2} & \cdots & \frac{\partial Z_1}{\partial X_n} \\ \frac{\partial Z_2}{\partial X_1} & \frac{\partial Z_2}{\partial X_2} & \cdots & \frac{\partial Z_2}{\partial X_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial Z_n}{\partial X_1} & \frac{\partial Z_n}{\partial X_2} & \cdots & \frac{\partial Z_n}{\partial X_n} \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 \\ \frac{X_2}{(1-X_1)^2} & \frac{1}{1-X_1} & 0 & 0 & \cdots & 0 \\ \frac{X_3}{(1-X_1-X_2)^2} & \frac{X_3}{(1-X_1-X_2)^2} & \frac{1}{1-X_1-X_2} & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{X_i}{(1-X_1-X_2-X_{i-1})^2} & \cdots & \frac{X_i}{(1-X_1-X_2-X_{i-1})^2} & \frac{1}{1-X_1-X_2-X_{i-1}} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{X_n}{(1-X_1-X_2-X_{n-1})^2} & \cdots & \cdots & \cdots & \cdots & \frac{1}{1-X_1-X_2-X_{n-1}} \end{pmatrix} \\
&= \frac{1}{(1-X_1)(1-X_1-X_2)\ldots(1-X_1-X_2-...-X_{n-1})}
\end{aligned} \tag{3.4}
$$

and

$$
\mathrm{B}(1-\alpha, \theta+i\alpha) = \frac{\Gamma(1-\alpha)\Gamma(\theta+i\alpha)}{\Gamma(1-\alpha+\theta+i\alpha)} \tag{3.5}
$$

This model is the most direct way to find the truncated approximation. And we call it **Direct Truncated Stick Breaking Model**.

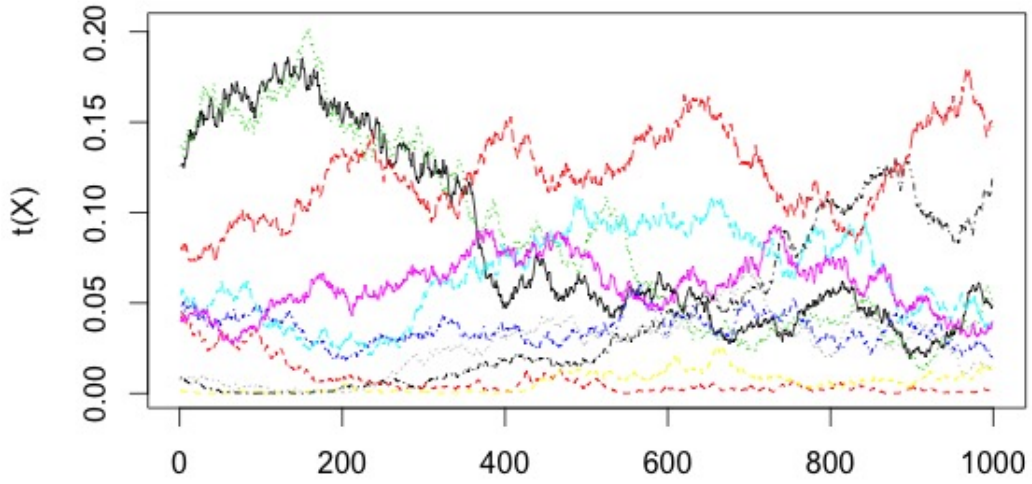

Figure 3.1: Simulation of stock weights for Direct Model

Figure 3.1 illustrates our simulation work of the model. There are 10 stocks and we can see the ranks of the stocks change a lot when time passes, i.e. the largest stock at time 0 becomes the 4-th largest in the end. This is irregular in the real market, as a dominating company's market capitalizations will not vary a lot in a short period. Hence, it is natural that we seek some better models.

3.1.2 Decreasing Model

Since we have various choices to formulate the model, a truncated ordered stick breaking process $Z_{(1)} \geq Z_{(2)} \geq Z_{(3)} \geq ... \geq Z_{(n)}$ can be considered to model the market weights, it can be:

$$\begin{aligned} &X_1 = Z_{(1)} \quad X_2 = Z_{(2)}(1 - Z_{(1)}) \quad \dots \quad X_n = Z_{(n)}(1 - Z_{(1)})...(1 - Z_{(n-1)}) \\ &X_0 = 1 - \sum_{i=1}^{n} X_i \end{aligned} \tag{3.6}$$

Accordingly, we have the joint distribution:

$$\begin{aligned} f(X_1, X_2...X_n) &= f(Z_{(1)}, Z_{(2)}, ..., Z_{(n)}) \times |J| \\ &= \mathrm{Per}(A_n) \times |J| \end{aligned} \tag{3.7}$$

where

$$
\begin{aligned}
A_n &= \begin{pmatrix}
f_1(Z_{(1)}) & f_2(Z_{(1)}) & \dots & f_n(Z_{(1)}) \\
f_1(Z_{(2)}) & f_2(Z_{(2)}) & \dots & f_n(Z_{(2)}) \\
\vdots & \vdots & & \vdots \\
f_1(Z_{(k)}) & f_2(Z_{(k)}) & \dots & f_n(Z_{(k)}) \\
\vdots & \vdots & & \vdots \\
f_1(Z_{(n)}) & f_2(Z_{(n)}) & \dots & f_n(Z_{(n)})
\end{pmatrix} \\
&= \begin{pmatrix}
f_1(X_1) & f_2(X_1) & \dots & f_n(X_1) \\
f_1(\frac{X_2}{1-X_1}) & f_2(\frac{X_2}{1-X_1}) & \dots & f_n(\frac{X_2}{1-X_1}) \\
\vdots & \vdots & & \vdots \\
f_1(\frac{X_k}{1-\sum_{j=1}^{k-1} X_j}) & f_2(\frac{X_k}{1-\sum_{j=1}^{k-1} X_j}) & \dots & f_n(\frac{X_k}{1-\sum_{j=1}^{k-1} X_j}) \\
\vdots & \vdots & & \vdots \\
f_1(\frac{X_n}{1-\sum_{j=1}^{n-1} X_j}) & f_2(\frac{X_n}{1-\sum_{j=1}^{n-1} X_j}) & \dots & f_n(\frac{X_n}{1-\sum_{j=1}^{n-1} X_j})
\end{pmatrix}
\end{aligned}
\tag{3.8}
$$

$$
\begin{aligned}
J &= \begin{pmatrix}
\frac{\partial Z_{(1)}}{\partial X_1} & \frac{\partial Z_{(1)}}{\partial X_2} & \cdots & \frac{\partial Z_{(1)}}{\partial X_n} \\
\frac{\partial Z_{(2)}}{\partial X_1} & \frac{\partial Z_{(2)}}{\partial X_2} & \cdots & \frac{\partial Z_{(2)}}{\partial X_n} \\
\cdots & \cdots & \cdots & \cdots \\
\frac{\partial Z_{(n)}}{\partial X_1} & \frac{\partial Z_{(n)}}{\partial X_2} & \cdots & \frac{\partial Z_{(n)}}{\partial X_n}
\end{pmatrix} \\
&= \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
\frac{X_2}{(1-X_1)^2} & \frac{1}{1-X_1} & 0 & 0 & \cdots & 0 \\
\frac{X_3}{(1-X_1-X_2)^2} & \frac{X_3}{(1-X_1-X_2)^2} & \frac{1}{1-X_1-X_2} & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{X_i}{(1-X_1-X_2-X_{i-1})^2} & \cdots & \frac{X_i}{(1-X_1-X_2-X_{i-1})^2} & \frac{1}{1-X_1-X_2-X_{i-1}} & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{X_n}{(1-X_1-X_2-X_{n-1})^2} & \cdots & \cdots & \cdots & \cdots & \frac{1}{1-X_1-X_2-X_{n-1}}
\end{pmatrix} \\
&= \frac{1}{(1-X_1)(1-X_1-X_2)\ldots(1-X_1-X_2-...-X_{n-1})}
\end{aligned}
\tag{3.9}
$$

This model is built on the decreasing ordered stick breaking of INID Beta distributions, named **Decreasing Ordered Truncated Stick Breaking Model**.

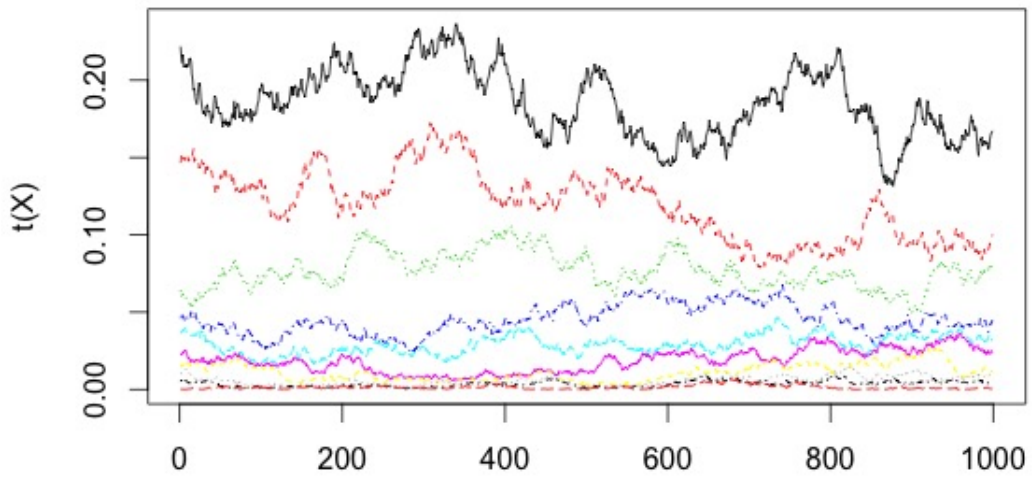

Figure 3.2: Simulation of stock weights for Decreasing Model

The dynamic motivation is that the ranks of weights under this model won't change with time. This means that the market weights will remain in the same order throughout.

Figure 3.2 gives an illustration of the Decreasing Model. This model is not very satisfying again as it is unlike the real world, where the ranks remain straight.

3.1.3 Increasing Model

There is another possible model we can use, which focuses more on ranks. Define the weights accordingly, named **Increasing Ordered Truncated Stick Breaking Model**:

$$\begin{aligned} &X_n = Z_{(n)} \quad X_{n-1} = Z_{(n-1)}(1 - Z_{(n)}) \quad \dots \quad X_1 = Z_{(1)}(1 - Z_{(n)})...(1 - Z_{(2)}) \\ &X_0 = 1 - \sum_{i=1}^{n} X_i \end{aligned} \tag{3.10}$$

$$\begin{aligned} f(X_1, X_2...X_n) &= f(Z_{(1)}, Z_{(2)}, ..., Z_{(n)}) \times |J| \\ &= \mathrm{Per}(A_n) \times |J| \end{aligned} \tag{3.11}$$

where

$$\begin{aligned} A_n &= \begin{pmatrix} f_1(Z_{(1)}) & f_2(Z_{(1)}) & \dots & f_n(Z_{(1)}) \\ f_1(Z_{(2)}) & f_2(Z_{(2)}) & \dots & f_n(Z_{(2)}) \\ \vdots & \vdots & & \vdots \\ f_1(Z_{(k)}) & f_2(Z_{(k)}) & \dots & f_n(Z_{(k)}) \\ \vdots & \vdots & & \vdots \\ f_1(Z_{(n)}) & f_2(Z_{(n)}) & \dots & f_n(Z_{(n)}) \end{pmatrix} \\ &= \begin{pmatrix} f_1(\frac{X_1}{1-\sum_{j=2}^{n} X_j}) & f_2(\frac{X_1}{1-\sum_{j=2}^{n} X_j}) & \dots & f_n(\frac{X_1}{1-\sum_{j=2}^{n} X_j}) \\ f_1(\frac{X_2}{1-\sum_{j=3}^{n} X_j}) & f_2(\frac{X_2}{1-\sum_{j=3}^{n} X_j}) & \dots & f_n(\frac{X_2}{1-\sum_{j=3}^{n} X_j}) \\ \vdots & \vdots & & \vdots \\ f_1(\frac{X_k}{1-\sum_{j=k+1}^{n} X_j}) & f_2(\frac{X_k}{1-\sum_{j=k+1}^{n} X_j}) & \dots & f_n(\frac{X_k}{1-\sum_{j=k+1}^{n} X_j}) \\ \vdots & \vdots & & \vdots \\ f_1(X_n) & f_2(X_n) & \dots & f_n(X_n) \end{pmatrix} \end{aligned} \tag{3.12}$$

$$
\begin{aligned}
J &= \begin{pmatrix} \frac{\partial Z_{(1)}}{\partial X_1} & \frac{\partial Z_{(1)}}{\partial X_2} & \cdots & \frac{\partial Z_{(1)}}{\partial X_n} \\ \frac{\partial Z_{(2)}}{\partial X_1} & \frac{\partial Z_{(2)}}{\partial X_2} & \cdots & \frac{\partial Z_{(2)}}{\partial X_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial Z_{(n)}}{\partial X_1} & \frac{\partial Z_{(n)}}{\partial X_2} & \cdots & \frac{\partial Z_{(n)}}{\partial X_n} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{1-X_1-X_2-\ldots-X_{n-1}} & \cdots & \frac{X_1}{(1-X_1-X_2-\ldots-X_{n-1})^2} \\ \cdots & \cdots & \cdots \\ 0\ldots & \frac{1}{1-X_1-X_2-\ldots-X_{i-1}} & \cdots \frac{X_i}{(1-X_1-X_2-\ldots-X_{i-1})^2} \\ \cdots & \cdots & \cdots \\ 0 & \ldots & 1 \end{pmatrix} \\
&= \frac{1}{(1-X_n)(1-X_n-X_{n-1})\ldots(1-X_2-X_3-\ldots-X_n)}
\end{aligned}
\tag{3.13}
$$

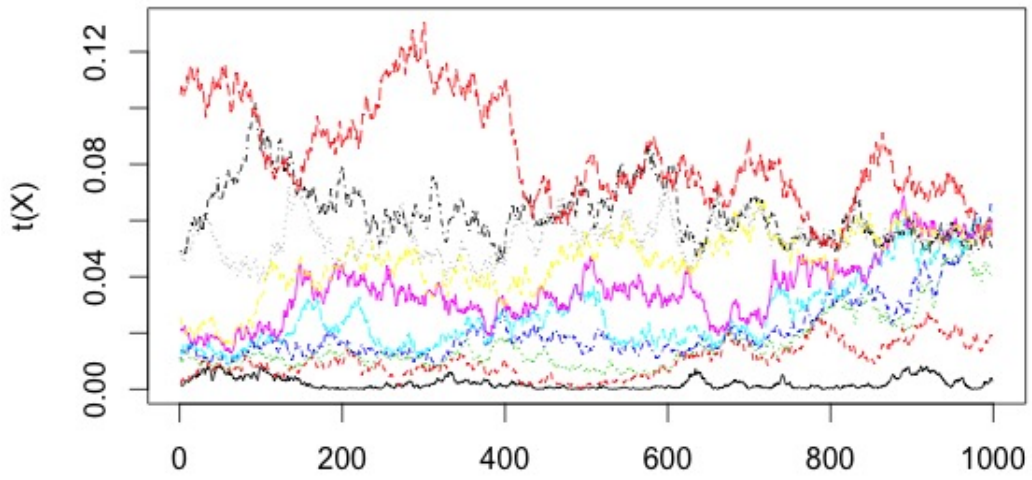

Figure 3.3: Simulation of stock weights for Increasing Model

To simulate the dynamic changes, Figure 3.3 gives a satisfying pattern. As time passes, there are a few interactions among the weights. However, the general ranks stay stable.

This can be explained that we introduce an adjustment value in our model. For exmaple, at time t= t_1: $X_{n-1} - X_n = Z_{(n-1)} - Z_{(n)} - Z_{(n-1)}Z_{(n)}$, where $Z_{(n-1)}Z_{(n)}$ is considered as an adjustment term for the ranks.

3.2 Bayesian Estimation

As the complexity of the joint distribution increases, traditional inference methods like the MLE are not applicable. As a consequence, the Markov Chain Monte Carlo estimation is used. Markov Chain Monte Carlo (MCMC) estimation is a Bayesian method for inference. To set up an inference method using it, first we need to get a likelihood of variables and assume appropriate priors for the target parameters. Details of the method and the settings in our case are introduced in the latter paragraphs. Metropolis Hasting is one of the most commonly used MCMC methods and it is applicable for our estimations. Suppose we have a likelihood $f(\mathbf{X}|\alpha,\theta)$, then the posterior is proportional to the likelihood times the prior: $f(\alpha,\theta|\mathbf{X}) \propto f(\mathbf{X}|\alpha,\theta)f(\alpha,\theta)$.

Independent Metropolis Hasting

1 Initialize $\alpha^{(0)}$, given $\theta^{(0)}$

2 for iteration $i = 1, 2, ...$

Generate a random proposal α^*, θ^* near $\alpha^{(i-1)}, \theta^{(i-1)}$ by a jumping distribution $G_t(\alpha^*, \theta^*)$.

3 Calculate the ratio:

$$r = \frac{f(\alpha^*, \theta^*|\mathbf{X})/G_t(\alpha^*, \theta^*)}{f(\alpha^{(t-1)}, \theta^{(t-1)}|\mathbf{X})/G_t(\alpha^{(t-1)}, \theta^{(t-1)})} \tag{3.14}$$

4 Accept the proposal if the ratio is larger than 1 or a uniform(0,1) random variable.

In our estimation, details will be discussed in the next Chapter. Briefly, Metropolis Hasting alogrithm applies, assuming α and θ are independent:

$$f(\alpha, \theta|\mathbf{X}, \theta) \propto f(\mathbf{X}|\alpha, \theta)f(\alpha, \theta) = f(\mathbf{X}|\alpha, \theta)f(\alpha)f(\theta) \tag{3.15}$$

$$f(\mathbf{X}|\alpha, \theta) \sim f_{(1),(2),...,(n)}(x_1, x_2, ..., x_n|\alpha, \theta) \tag{3.16}$$

$$f(\alpha) \sim \text{Beta}(a, b) \tag{3.17}$$

$$f(\theta) \sim \text{Gamma}(m, n) \tag{3.18}$$

The jump function for α is

$$G_\alpha(t) \sim \text{Uniform}(-w_\alpha, w_\alpha) \tag{3.19}$$

The jump function for θ is

$$G_\theta(t) \sim \text{Uniform}(-w_\theta, w_\theta) \tag{3.20}$$

Chapter 4

Estimation Results and Simulation Work

This chapter mainly presents the results of Metropolis Hasting MCMC inference. The data is the world stocks market capitalizations in 2014, i.e. Nasdaq, London Securities Exchanges and Canadian Securities Exchange, etc. There are totally 79 stock markets included, which could be considered as the 78 steps in breaking a 'money' stick. The pieces broke are the money invested in the different stock markets.

The algorithm we use is the Metropolis Hasting introduced in Chapter 3. There are a few assumptions: firstly, we assume the parameters α and θ are independent, given support $0 < \alpha < 1$ and $\theta > 0$. Secondly, we consider the proir for α is Beta(a,b) distribution and for θ is a Gamma(m,n) distribution. Lastly, the jump functions we choose for α and θ are uniform distributions with appropriate scales. We run 5000 iterations and take a burn-in for the first 200.

In section 4.1, we present the results of the direct truncated stick breaking model. In section 4.2 and section 4.3, decreasing ordered truncated stick breaking model and

increasing ordered truncated stick breaking model are shown accordingly.

4.1 Direct Truncated Stick Breaking Model

Figure 4.1 shows the log-log plot of market weights and their ranks. The blue line is lowess line of the data set of Nov. 2014. The yellow line is the simulation result of least squares estimator for two-parameter Poisson-Dirichlet market model. The red line is the simulation result of MCMC estimator for direct truncated model. For the simulation, we generate 100 samples and take the average of them.

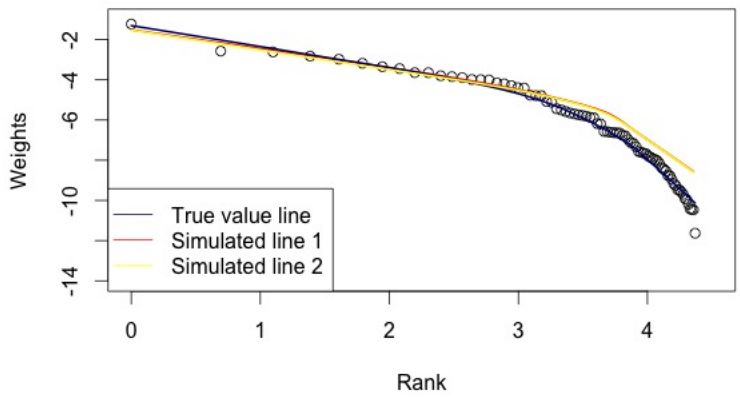

Figure 4.1: Nov. 2014 Lowess line comparision (Direct Model)

It is shown that the fitting is not satisfying in the tail for both models.

Table 4.1 summarizes estimations and intervals of α and θ.

Date	mean of α	95% density interval of α	mean of θ	95% density interval of θ
Jan. 2014	0.47	(0.38,0.56)	22.64	(15.44,30.39)
Jun. 2014	0.47	(0.37,0.55)	23.89	(16.30, 32.55)
Nov. 2014	0.47	(0.38,0.56)	22.96	(16.16, 30.32)

Table 4.1: Selected estimation of Direct Model

4.2 Decreasing Ordered Truncated Stick Breaking Model

Figure 4.2 illustrates the simulation result of Decreasing Model over the data on Nov. 2014.

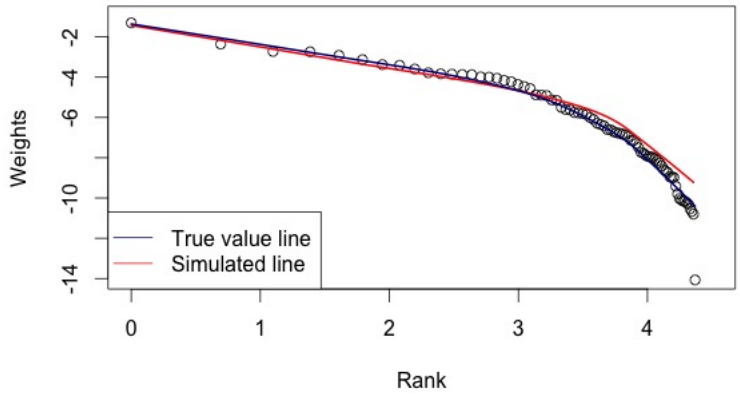

Figure 4.2: Nov. 2014 Lowess line comparision (Decreasing Model)

Date	mean of α	95% density interval of α	mean of θ	95% density interval of θ
Jan. 2014	0.52	(0.42,0.62)	14.70	(9.07,23.79)
Jun. 2014	0.50	(0.41,0.59)	14.64	(9.07,20.85)
Nov. 2014	0.53	(0.42,0.66)	16.29	(8.95,24.13)

Table 4.2: Selected estimation of Decreasing Model

Table 4.2 summarizes the estimations and intervals of α and θ for different dates. Similar to Direct Model, the graph shows that the fit has bias in the tail.

4.3 Increasing Ordered Truncated Stick Breaking Model

The tail problem can be solved by the Increasing Model as presented in Figure 4.3. This indicates if one concerns more on small individual companies, Increasing Model is an appropriate choice.

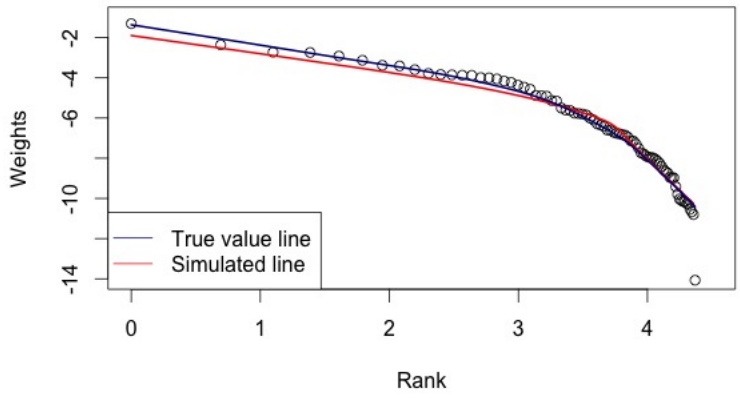

Figure 4.3: Nov. 2014 Lowess line comparision (Increasing Model)

Date	mean of α	95% density interval of α	mean of θ	95% density interval of θ
Jan. 2014	0.62	(0.54,0.69)	20.00	(13.87,28.25)
Jun. 2014	0.61	(0.54,0.69)	19.51	(13.21,25.70)
Nov. 2014	0.60	(0.51,0.68)	20.92	(12.94,27.82)

Table 4.3: Sample estimation of Increasing Model

In addition, we can draw the conlusion that the market capitalization curve is stable from time to time as the estimations are almost the same for different time.

To prove our estimation is powerful, we simulate data by the stick breaking process and use the MCMC to estimate known parameters. Here, we set $\alpha = 0.3$ and $\theta = 30$ for the generated samples, then we take 100 samples' average and estimate the parameters.

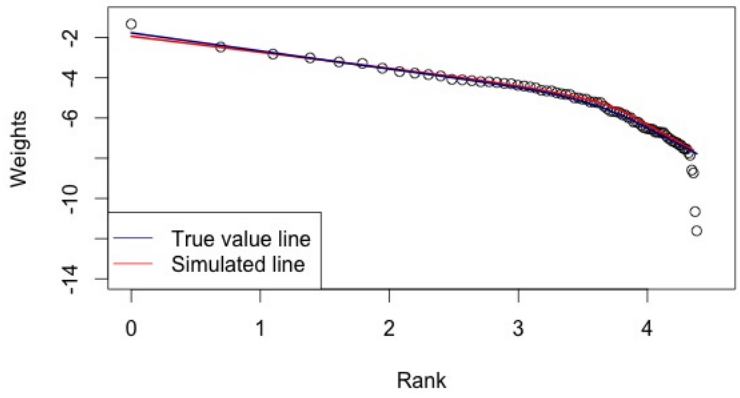

Figure 4.4: Simulated lowess line comparision

n	mean of α	95% density interval of α	mean of θ	95% density interval of θ
50	0.33	(0.16,0.48)	29.25	(20.64,39.33)
80	0.30	(0.19,0.41)	30.65	(22.43,39.71)

Table 4.4: Simulated estimation of Increasing Model

Figure 4.4 gives the illustration of simulated data. Table 4.4 gives the numerical results, showing that our method has a very good estimation outcome.

Chapter 5

Discussion

To summarize, our work is motivated by using Poisson-Dirichlet distribution to model market weights. Assuming the stock markets weights processes are driven by modified versions of stick breaking process, we take samples after a certain time $t = N$, $N \to \infty$, in which the process has been stable. Then, through MCMC we can get the estimators of the parameters α and θ for the process. To interpret the parameters, we conclude that the larger α and smaller θ we have, the greater gaps among the weights. This is an empirical result. Similar to the Poisson-Dirichlet parameters, we can consider they affect the diversity of the system in the same way.

In conclusion, this thesis comes up with a potential stochastic processes which can be established to model the market portfolios.

5.1 Benefits

All three models present nice patterns in fitting real data sets. Especially, the Decreasing Model and the Increasing Model give different restrictions on the ranks and

this makes the models more applicable in reality.

Different models also provide various choices for head concerned and tail concentrated users.

5.2 Limitations

The computational methods are restricted by the machine, as the largest number R can represent is 1.797693e+308. Thus, when n goes larger than 100, the MCMC alogrithm does not always work fine. We need to give some scaling adjustments in the middle steps. However, the good news is that size 80 seems good enough for most cases.

5.3 Future Work

Since the Poisson-Dirichlet model is widely used in biological and economic studies, our truncated model could also be applied to those areas and interesting outcomes may be found.

Furthermore, when we conduct the estimation, we assume that we sample from stationary distributions. However, it takes some time before the process converges. This means we can calculate the general distribution of the stochastic process before converging.

Appendix A

Appendix

A.1 Selected MCMC Results Plot

As there are many MCMC results, we list some selected plots here as a representation. The Figures A.1-A.4 give the Direct Model results and diagnostics for the MCMC estimation.

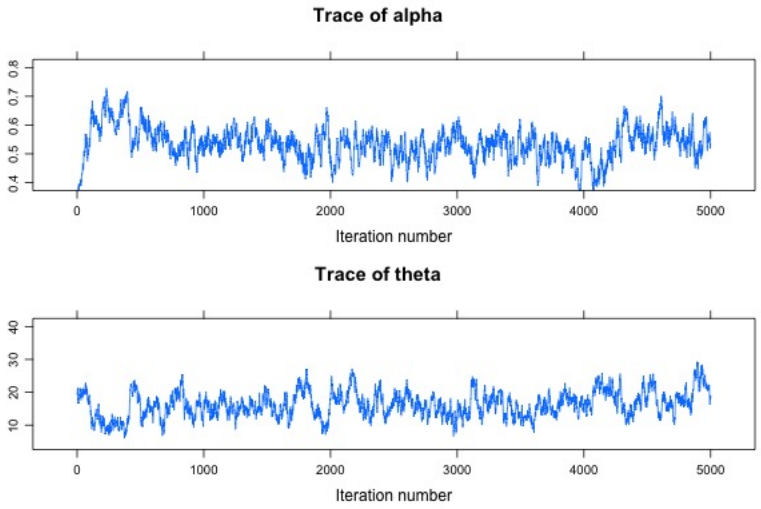

Figure A.1: Trace plot of parameters (Direct Model)

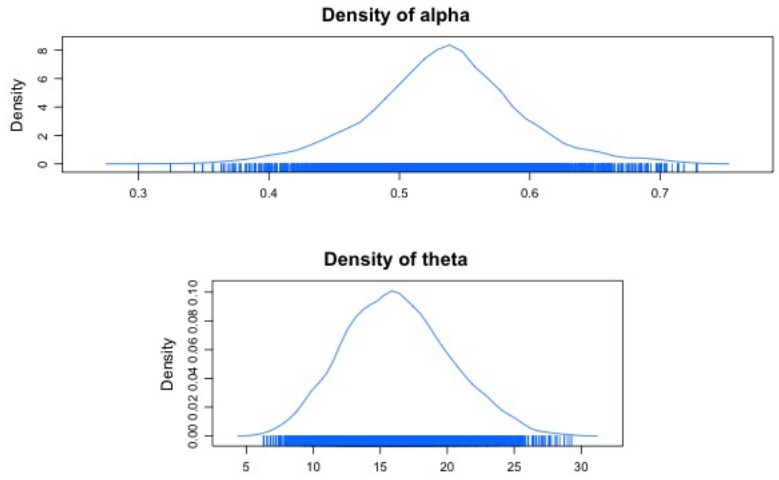

Figure A.2: Density plot of parameters (Direct Model)

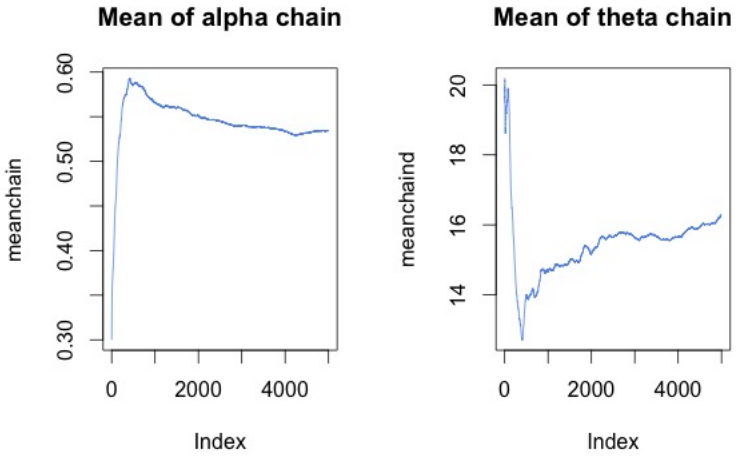

Figure A.3: Trace plot of mean for parameters (Direct Model)

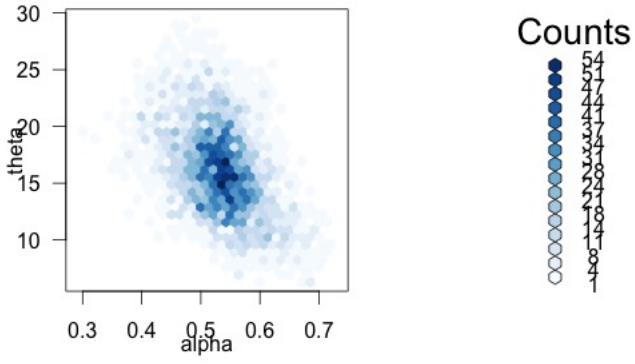

Figure A.4: Most frequent points (Direct Model)

The Figures A.5-A.8 give the Decreasing Model results and diagnostics for the MCMC estimation.

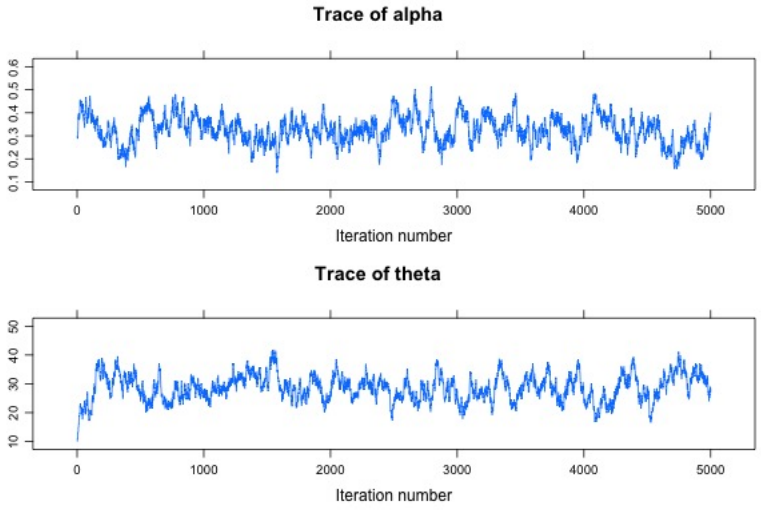

Figure A.5: Trace plot of parameters (Decreasing Model)

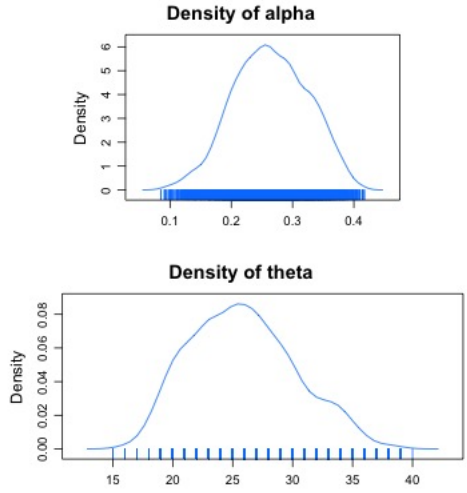

Figure A.6: Density plot of parameters (Decreasing Model)

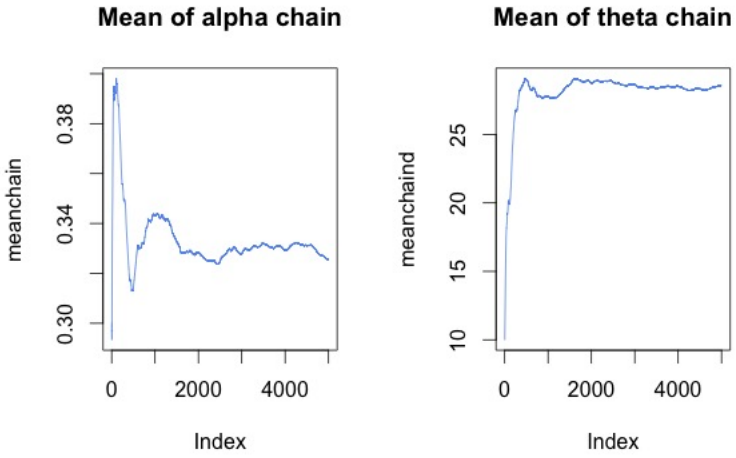

Figure A.7: Trace plot of mean for parameters (Decreasing Model)

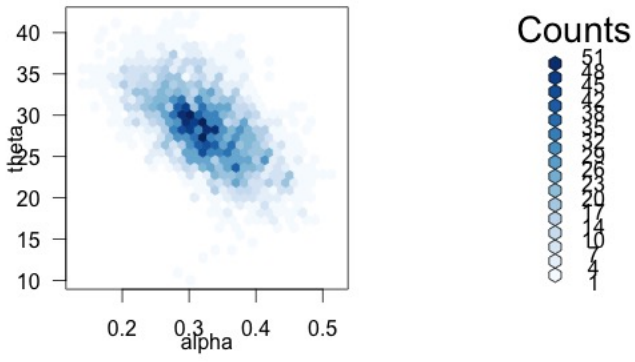

Figure A.8: Most frequent points (Decreasing Model)

The Figures A.9-A.12 give the Increasing Model results and diagnostics for the MCMC estimation.

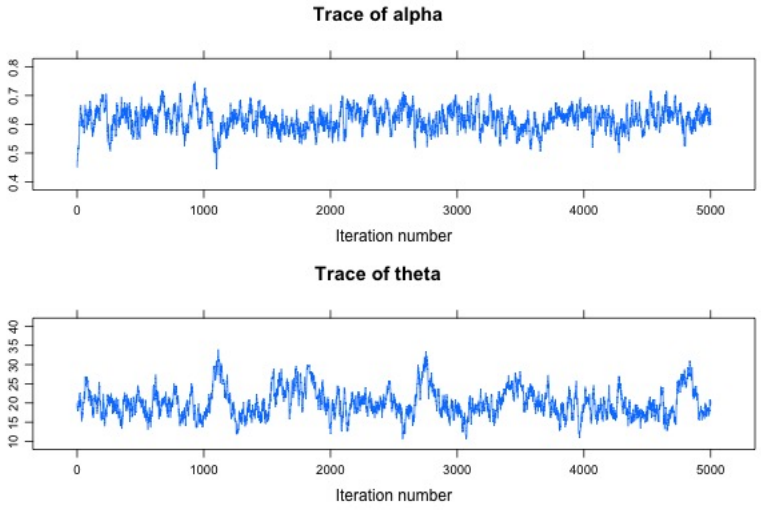

Figure A.9: Trace plot of parameters (Increasing Model)

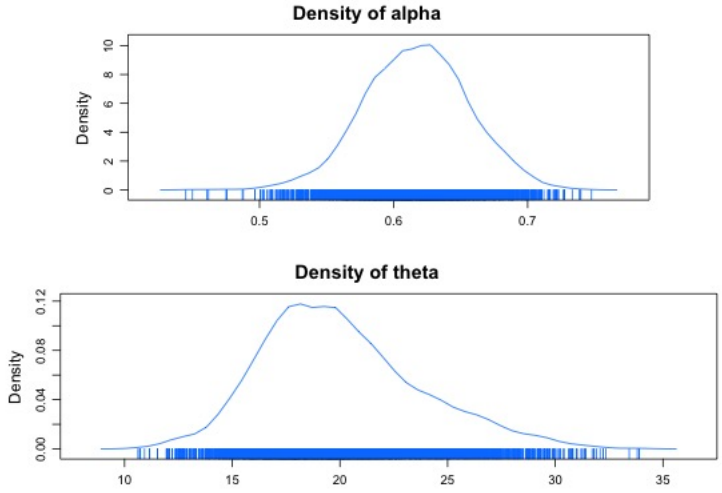

Figure A.10: Density plot of parameters (Increasing Model)

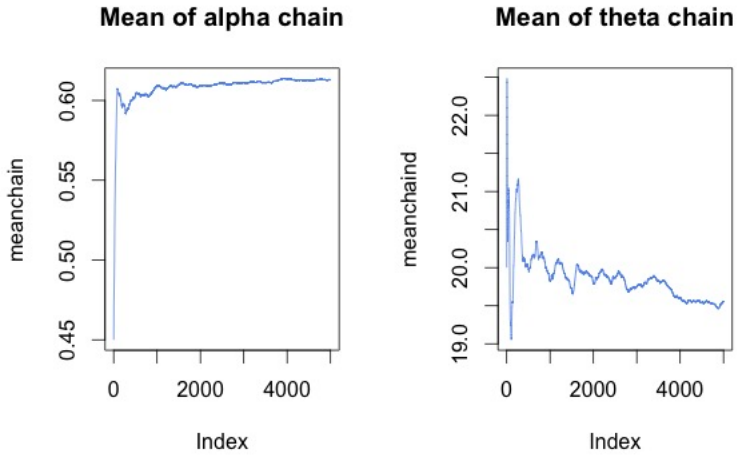

Figure A.11: Trace plot of mean for parameters (Increasing Model)

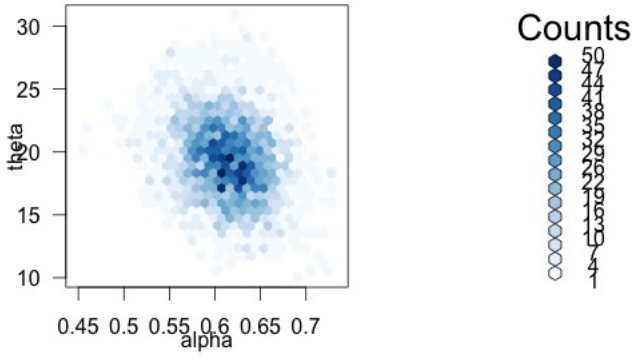

Figure A.12: Most frequent points (Increasing Model)

Bibliography

Balakrishnan, N. (2007). Permanents, order statistics, outliers, and robustness. *Revista Matemática Complutense*, **20**, 7–107.

Carlton, M. (1999). *Applications of the two-parameter Poisson-Dirichlet distribution*. Ph.D. thesis, University of California, Los Angeles.

Chatterjee, S. and Pal, S. (2010). A phase transition behavior for Brownian motions interacting through their ranks. *Probability Theory and Related Fields*, **147**, 123–159.

Ewens, W. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**, 87–112.

Ewens, W. (1990). Population genetics theory-the past and the future. *Mathematical and Statistical Developments of Evolutionary Theory*, **299**, 177–227.

Feng, S. (2010). *The Poisson-Dirichlet distribution and related topics: models and asymptotic behaviors*. Springer-Verlag Berlin Heidelberg.

Feng, S. and Wang, F. (2007). A class of infinite-dimensional diffusion processes with connection to population genetics. *Journal of Applied Probability*, **44**, 938–949.

Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1**, 209–230.

Fernholz, E. (2002). *Stochastic portfolio theory*. Springer New York.

Huber, M. and Law, J. (2008). Fast approximation of the permanent for very dense problems. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 681–689. Society for Industrial and Applied Mathematics.

Jerrum, M., Sinclair, A., and Vigoda, E. (2004). A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM*, **51**, 671–697.

Johnson, N., Kotz, S., and Balakrishnan, N. (1997). *Discrete multivariate distributions*. John Wiley and Sons, New York.

Kingman, J. (1975). Random discrete distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, **37**, 1–22.

Perman, M., Pitman, J., and Yor, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probability Theory and Related Fields*, **92**, 21–39.

Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, **25**, 855–900.

Ryser, H. (1963). *Combinatorial mathematics*. The Mathematical Association of America.

Sibuya, M. (2014). Prediction in Ewens-Pitman sampling formula and random samples from number partitions. *Annals of the Institute of Statistical Mathematics*, **66**, 833–864.

Sosnovskiy, S. (2015). On financial applications of the two-parameter Poisson-Dirichlet distribution. https://arxiv.org/pdf/1501.01954v3.pdf. Research Note.

Valiant, L. (1979). The complexity of computing the permanent. *Theoretical Computer Science*, **8**, 189–201.