Multivariate Extremes, Max-Stable Process Estimation and Dynamic Financial Modeling

by

Zhengjun Zhang

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics.

Chapel Hill

2002

Approved by

Advisor: Professor Richard L. Smith

Reader: Professor A. Ronald Gallant

Reader: Professor Harry Hurd

Reader: Professor Chuanshu Ji

Reader: Professor Pranab K. Sen

Reader: Professor Gordon Simons

ABSTRACT

ZHENGJUN ZHANG: Multivariate Extremes, Max-Stable Process Estimation and Dynamic Financial Modeling.
(Under the direction of Professor Richard L. Smith)

Studies have shown time series data from finance, insurance and environment etc. are fat tailed and clustered when extremal events occur. In order to characterize such extremal processes, max-stable processes or min-stable processes have been proposed since the 1980s and some probabilistic properties have been obtained, but the applications are very limited due to lack of efficient statistical estimation methods.

In this work, some probabilistic properties of the processes are proved and a series of estimation procedures to estimate the underlying max-stable processes are proposed, i.e. multivariate maxima of moving maxima processes. The first proposed method is purely probabilistic. It is designed for the time series with only one signature pattern, which can be regarded as a clustering pattern. It gives true parameter values if the model is correct. The second proposed method is a two step estimating method. At the first step, the method gives exact parameter values within each signature pattern, then it estimates the proportions of different signature patterns in the process. Consistency and asymptotic properties for the estimators are proved. The third proposed method is a generalized version of the second one but is not tied with the data, i.e. the data are not assumed to follow the model exactly. It is practically applicable. Three variants of the third method are proposed. They are designed to provide more specific estimators for special cases of the model, such as symmetric, monotone and asymmetric data structure respectively. All the estimators have been proved to be consistent and asymptotically normal.

To date, the exceedance over threshold approach which uses a generalized Pareto distribution(GPD) has been advocated. Assuming the population distribution belongs to the multivariate domains of attraction of multivariate extreme value distributions we develop threshold methods to estimate the parameters of the underlying max-stable process from the observed data. All previously developed six methods have their corresponding version under threshold methods.

How to manage a portfolio efficiently, with the highest expected return for a given level of risk, or equivalently, the least risk for a given level of expected return, is a key to the success or failure of a financial system. As an application of max-stable processes, financial time series data are standardized and transformed. The new time series are modelled as max-stable processes. The VaR (Value at Risk), maximal possible losses of portfolios under given confidence level, of portfolios are calculated and portfolio optimizations under VaR constraints are then studied.

ACKNOWLEDGEMENTS

I am deeply indebted to Prof. Richard L. Smith, my advisor, for his stimulating suggestions, encouragement and complete support through all my time and research here at Chapel Hill.

I would like to express my gratitude to my committee members, Professor A. Ronald Gallant, Professor Harry Hurd, Professor Chuanshu Ji, Professor Pranab K. Sen, Professor Gordon Simons, who gave me their full support in completing this thesis. A series of presentations given to a study group organized by Professor Smith and Professor Gallant increased my, along with other students', knowledge base at a time when I was beginning to investigate applications of my research. Professor Gallant suggested studying stock return data as an application of max-stable process modeling. Professor Sen suggested several extreme value theory references and put forth two theoretical considerations which, once I proved, formed the basis of several important theorems and examples in this thesis. The discussions I had with Professor Ji, Professor Simons and Professor Hurd helped me to better understand the problems I was studying and helped me make progress.

I would also like to express my deep thanks to Professor Jianqing Fan, Professor Edward Carlstein, Professor Amarjit Budhiraja, Professor Paul Embrechts from ETH-Zentrum, Professor Ruey S. Tsay from the University of Chicago, Professor Robert Stine from the University of Pennsylvania and Dr. Dougal Goodman from the Foundation for Science and Technology of the United Kingdom for their comments and suggestions. Given that part of results were developed in the Newton Institute for Mathematical Science at Cambridge University, I would like to thank Professor H. K. Moffatt and the institute for their supporting and assisting me in participating the *Managing Uncertainty Program.*

I have taken courses from almost all the professors in this department. I want to thank them for teaching me and helping me develop my career. Professor Fan has given me constant supports in many ways. Professor Carlstein guided me in becoming a good teacher throughout the undergraduate courses I have taught while at Chapel Hill. I enjoyed my three years as computer coordinator, a job which gave me opportunities to help our faculty, staff and graduate students. Everything was beautiful at Chapel Hill.

Especially, I would like to give my special thanks to my wife, Dr. Chunming Zhang, for her endless support and love.

Contents

List of Figures

Chapter 1

Introduction

1.1 General introduction

Extreme events or rare events have major impacts (bad or good) on the real world. Imagine the major damage caused by a disaster hurricane or the impact of winning three million dollars in the lottery. Such rare events are part of our life. We must face, understand and study those phenomenon and problems caused by rare events. Indeed, the study of extreme events has become very important and drawn major attention in probability and statistical research.

The extreme type theorems play a central role of the study of extreme value theory. In the literature, Fisher and Tippett (1928) were the first who discovered the extreme type theorems and later these results were proved in complete generality by Gnedenko (1943). Galambos (1987), Leadbetter, Lindgren and Rootzén (1983), and Resnick (1987) are excellent reference books on the probabilistic aspect. Smith (1990) gives a comprehensive account of statistical aspects, especially maximum likelihood methods in parameter estimation. A recent book by Embrechts, Klüppelberg, and Mikosch (1997) gives an excellent viewpoint of modeling extremal events. The extreme type theorems say that for a sequence of i.i.d. random variables with suitable normalizing constants, the limiting distribution of maximum statistics, if it exists, follows one of three types of extreme value distributions. In the multivariate context, the maximum is taken componentwise and there is no specific parametric type of limiting distribution. However, there have been many attempts to characterize the possible limits, such as de Haan and Resnick (1977), de Haan (1985) and Resnick's (1987) point process approach, and Pickands's (1981) representation theorem for multivariate extreme value distribution with unit Fréchet margins. Some efforts have been devoted to extending the i.i.d classical results to dependent sequences under some conditions such as stationary, mixing conditions etc. For instance, Leadbetter, *et al* (1983) contains an abundant account of the theory of extreme values for dependent sequences, both stationary and non-stationary, as well as for stationary continuous time processes at a rigorous mathematical level. The extremal index, originated by Cartwright (1958), Newell (1964), Loynes (1965), O'Brien (1974) and Leadbetter (1983), is a quantity which allows one to associate the limiting distribution of a dependent sequence to the extreme value distributions. In a multivariate analog, Nandagopalan (1990, 1994) introduced the multivariate extreme index and derived some elementary properties.

In the statistical aspects, the focus is on extremal events modeling, parameter estimation and testing of hypotheses. There are many applications to real problems. Among them, extreme value theory has been largely applied to environmental problems such as river flow, wind speed, sea level, temperature and rainfall, and insurance and finance (cf. Smith 1990 and Embrechts, Klüppelberg, and Mikosch 1997). To model extremal events in a univariate context, usually the generalized extreme value distribution is adopted. To date, the exceedance over threshold approach which uses a generalized Pareto distribution(GPD) (Pickands 1975, Davison and Smith 1990) and a fixed number of extreme order statistics approach (Weissman 1978, Smith 1986, Tawn 1988, Robinson and Tawn 1995, Smith 1997) have been advocated.

Although there are well-developed approaches to model univariate extremal processes, problems concerning the environment, finance and insurance etc. are multivariate in character: for example, floods may occur at different sites along a coastline; the failure of a portfolio management may be caused by a single extreme price movement or multiple movements. Here multivariate extreme modeling is essential for risk management and precision of modeling. What one needs is to choose or develop appropriate multivariate extreme value distributions which can be used in modeling multivariate extremal processes. As we mentioned earlier, multivariate extreme value distributions have no specific parametric form. Fortunately several models have been developed that are multivariate extreme value distributions. Among them, it is worth mentioning models such as Bivariate Logistic model by Marshall and Olkin (1983), Bivariate Exponential model by Mardia (1970), Asymmetric Logistic model by Tawn (1990), Negative Asymmetric Logistic model by Joe (1989), Dirichlet model by Coles and Tawn (1991), Bilogistic model by Smith (1990), Nested Logistic model by McFadden (1978) and Time Series Logistic model by Coles and Tawn (1991). These models are listed in Coles and Tawn (1991). Coles and Tawn (1994) also demonstrate how statistical methods for multivariate extremes may be applied to a very practical problem of data analysis.

In general, an multivariate distribution function characterizes the dependence structure within the random vector. It does not show the time dependent structure of the vector time series. Since multivariate extreme value distributions are in fact max-stable distributions (Resnick 1987), and the extreme values of an multivariate stationary process may be characterized in terms of a limiting max-stable process under quite general conditions (Smith and Weissman 1996), it is very natural to model extreme processes by max-stable processes.

In this work, I mainly focus on proving probabilistic properties of a certain class of max-stable processes and proposing a series of estimation procedures of estimating the underlying max-stable process. The consistency and asymptotic properties of all estimators are proved. Applications of max-stable process in finance will be addressed.

In the rest of this chapter we will give some background results. We will discuss extreme value theory for multivariate random variables in section 1.2. Also in section 1.2 extreme value theory for univariate random variables will be briefly reviewed since the marginal distributions of MEVD have to be univariate extreme value distributions. And multivariate maxima of moving maxima processes will be discussed in section 1.3.

In chapter 2, we will study the properties of multivariate maxima of moving maxima processes, extend probability properties, propose statistical estimation methods for the parameters and prove consistency and asymptotics. Some examples will be given in this chapter to demonstrate the processes and the statistical aspects.

In chapter 3, we will consider modeling multivariate maxima of moving maxima processes by using the bivariate joint distribution of the sequence of dependent random variables. We shall study estimation of parameters and consistent properties as well as asymptotics. Examples will be illustrated.

In chapter 4, we will consider the parametric structure for multivariate maxima of moving maxima processes. Estimation of parameters, consistency and asymptotic properties will be addressed.

In chapter 5, we first review literature on the applications of extreme value theory to finance and insurance. We will briefly review the definition of VaR and some typical calculation methods. Then extreme value approaches will be discussed. Finally we model multivariate extreme value distributions to multivariate financial time series data and illustrate VaR calculation as well as portfolio optimizations. Comparison among extreme value approaches and other approaches will be illustrated also.

1.2 Multivariate extreme value theory

1.2.1 Extreme value theory for univariate random variables

Suppose $X_1, X_2, \ldots, X_n$ are an i.i.d. sequence with distribution function $F(x)$ and let

$$M_n = \max(X_1, X_2, \ldots, X_n). \tag{1.1}$$

Then M_n has the distribution function

$$\Pr\{M_n \le x\} = \Pr\{X_1 \le x, \cdots, X_n \le x\} = F^n(x). \tag{1.2}$$

It is clear that the maximum of a sample simply tends to the right-hand endpoint of the distribution almost surely, no matter whether it is finite or infinite. Let X_F be the right endpoint, since

$$\sum_{n=1}^{\infty} \Pr\{|M_n - X_F| > \epsilon\} = \sum_{n=1}^{\infty} \Pr\{M_n < X_F - \epsilon\} = \sum_{n=1}^{\infty} \Pr\{X_1 < X_F - \epsilon\}^n$$
$$= \frac{\Pr\{X_1 < X_F - \epsilon\}}{1 - \Pr\{X_1 < X_F - \epsilon\}} < \infty$$

and this shows $M_n \xrightarrow{a.s.} X_F$. What we are interested in is the form of limits

$$\lim_{n\to\infty} F^n(a_n x + b_n) = \lim_{n\to\infty} \Pr\Big(\frac{(M_n - b_n)}{a_n} \le x\Big) = H(x) \tag{1.3}$$

for suitable normalizing constants $a_n > 0$, and b_n.

If (1.3) holds, we say F (or X) belongs to the (maximum) domain of attraction of H and write $F \in MDA(H)$ (or $X \in MDA(H)$). H has one of the following three parametric forms (which are generally called extreme value distributions)

$$\begin{aligned}
&Type \quad I: && H(x) = \exp\{-\exp(-x)\} \quad (-\infty < x < \infty)\\
&Type \quad II: && H(x) = \begin{cases} 0 & \text{if } x \le 0, \\ \exp(-x^{-\alpha}) & \text{if } x > 0, \end{cases}\\
&Type \quad III: && H(x) = \begin{cases} \exp(-(-x)^{\alpha}) & \text{if } x < 0, \\ 1 & \text{if } x \ge 0. \end{cases}
\end{aligned}$$

In II and III, α is any positive number. The three types are also often called the Gumbel, Fréchet and Weibull types respectively.

The following theorems are very useful in finding the $MDA(H)$ of F and the suitable normalizing constants. The proofs of the theorems can be found in Leadbetter et al. (1983), Resnick (1987), Galambos (1987), etc..

Theorem 1.1 *Let* $0 \le \tau \le \infty$ *and suppose that for suitable normalizing constants* $a_n > 0$ *and* b_n*,* $u_n = u_n(x) = \frac{x}{a_n} + b_n$ *such that*

$$n(1 - F(u_n)) \to \tau \qquad \text{as } n \to \infty \tag{1.4}$$

then

$$P(M_n \le u_n) \to e^{-\tau} \qquad \text{as } n \to \infty \tag{1.5}$$

Conversely, if (1.5) holds for some τ*,* $0 \le \tau \le \infty$*, then (1.4) holds.*

Theorem 1.2 *Necessary and sufficient conditions for the distribution* F *belongs to the* MDA *of*

Type I*:* $\int_0^\infty (1 - F(u))du < \infty$*,*

$$\lim_{t \uparrow X_F} \frac{1 - F(t + xg(t))}{1 - F(t)} = e^{-x}$$

for all real x*, where*

$$g(t) = \frac{\int_t^{X_F} (1 - F(u))du}{1 - F(u)}$$

for $t < X_F$*.*

Type II*:* $X_F = \infty$ *and*

$$\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}$$

$\alpha > 0$*, for each* $x > 0$*.*

Type III*:* $X_F < \infty$ *and*

$$\lim_{h \downarrow 0} \frac{1 - F(X_F - xh)}{1 - F(X_F - h)} = x^{\alpha}$$

$\alpha > 0$*, for each* $x > 0$*.*

Some other theoretical results may be very useful for finding the $MDA(H)$ of F and finding the normalizing constants. Those results and examples whose distributions belong to each of the three domains of attraction can be found in Leadbetter et al. (1983), Resnick (1987), Galambos (1987), etc.. As a simple example, we consider now the Pareto distribution

$$F(x) = 1 - \kappa x^{-\alpha}, \;\; \alpha > 0, \kappa > 0, x \ge \kappa^{1/\alpha}.$$

We have

$$\frac{1-F(tx)}{1-F(t)} = \frac{(tx)^{-\alpha}}{t^{-\alpha}} = x^{-\alpha}$$

so F belongs to MDA of a Type II extreme value distribution. By setting

$$n(1-F(u_n)) = \tau$$

we have

$$u_n = (\kappa n/\tau)^{1/\alpha}.$$

By putting $\tau = x^{-\alpha}$ for $x \geq 0$, we have

$$P\{(\kappa n)^{-1/\alpha} M_n \leq x\} \to \exp(-x^{-\alpha})$$

so

$$a_n = (\kappa n)^{-1/\alpha},\ b_n = 0.$$

The extreme value distributions are *max-stable* distributions. We say a non-degenerate distribution H is *max-stable*, if $H^n(a_n x + b_n) = H(x)$ holds for some constants $a_n > 0$, and b_n for each $n = 2, 3, \ldots$. The next result (Theorem 1.4.1 in Leadbetter et al. 1983) shows the relation.

Theorem 1.3 *Every max-stable distribution is of extreme value type, i.e. equal to $H(ax+b)$ for some $a > 0$ and b; Conversely, each distribution of extreme value type is max-stable.*

The three types of extreme value distributions can be written into a generalized extreme value (GEV) distribution form (which is very useful for statistical purposes)

$$H(x;\mu,\sigma,\xi) = \exp\{-[1+\frac{\xi(x-\mu)}{\sigma}]^{-1/\xi}\} \tag{1.6}$$

where $1+\xi(x-\mu)/\sigma > 0$, $\sigma > 0$ and μ, ξ arbitrary. The case $\xi = 0$ is interpreted as the limit $\xi \to 0$, that is

$$H(x;\mu,\sigma,0) = \exp\{-\exp[-\frac{(x-\mu)}{\sigma}]\} \tag{1.7}$$

Type II and III correspond to $\xi > 0$ ($\xi = \frac{1}{\alpha}$) and $\xi < 0$ ($\xi = -\frac{1}{\alpha}$) respectively. Smith (1990) has a detailed review of statistical treatments, applications and estimations, of GEV.

Suppose now $\{X_i, i = 1, 2, \ldots, \}$ is a stationary sequence with a continuous marginal distribution function $F(x)$ and $\{\widehat{X}_i, i = 1, 2, \ldots, \}$ is the so-called associated sequence of i.i.d. random variables with the same marginal distribution function F. M_n stands for the maximum as usual, defined by (1.1), while $\widehat{M_n}$ denotes the corresponding maximum of $\{\widehat{X}_1, \cdots, \widehat{X}_n\}$. The limiting distribution of M_n can be related to the limiting distribution of $\widehat{M_n}$ via a quantity θ defined below.

If for every $\tau > 0$ there exists a sequence of thresholds $\{u_n\}$ such that

$$\Pr\{\widehat{M_n} \le u_n\} \to e^{-\tau} \tag{1.8}$$

and under quite mild additional conditions,

$$\Pr\{M_n \le u_n\} \to e^{-\theta\tau} \tag{1.9}$$

Then θ is called the *extremal index* of the sequence $\{X_n\}$. This concept originated in papers by Cartwright (1958), Newell (1964), Loynes (1965), O'Brien (1974). Leadbetter (1983) gave a formal definition.

The index θ can take any values in [0,1] and $\frac{1}{\theta}$ is interpreted as mean cluster size of exceedance over some high threshold. When $\theta = 0$, it corresponds to a strong dependence (infinite cluster sizes) but not so strong that all the values can be the same. While $\theta = 1$ is a form of asymptotic independence of extremes, but it does not mean that the original sequence is independent.

If (1.9) holds for some τ and corresponding $\{u_n\}$, then it holds for all τ' (equal or not equal to τ) and its corresponding $\{u'_n\}$. Estimators of the extremal index have been proposed by Leadbetter, Weissman, de Haan, and Rootzén (1989), Nandagopalan (1990), Hsing (1993). Smith and Weissman (1994) gave a review of estimating the extreme index and proposed two estimating methods, i.e., blocks method and runs method. Other references include chapter 8 in the book by Embrechts et al. (1997).

1.2.2 Limit laws of multivariate extremes

Suppose $\{\mathbf{X}_i = (X_{i1}, \cdots, X_{iD}), i = 1, 2, ...\}$ is a D-dimensional i.i.d. random process with distribution $F(\mathbf{x}) = F(x_1, \ldots, x_D) = \Pr\{X_{id} \le x_d, d = 1, \ldots, D\}$ and marginal distributions $F_d(x) = \Pr\{X_{id} \le x_d\}, d = 1, \ldots, D$. Let $\mathbf{M}_n = (M_{n1}, \cdots, M_{nD})$ denote the vector of pointwise maxima, where $M_{nd} = \max\{X_{id}, 1 \le i \le n\}$. If there exist normalizing constants $\mathbf{a}_n > 0, \mathbf{b}_n$ such that

$$\begin{aligned}
\Pr\{\mathbf{M}_n \leq \mathbf{a}_n\mathbf{x} + \mathbf{b}_n\} &= \Pr\{M_{nd} \leq a_{nd}x_d + b_{nd}, d = 1, \ldots, D\} \\
&= F^n(a_{n1}x_1 + b_{n1}, a_{n2}x_2 + b_{n2}, \ldots, a_{nD}x_D + b_{nD}) \quad (1.10)\\
&= F^n(\mathbf{a}_n\mathbf{x} + \mathbf{b}_n) \to H(\mathbf{x})
\end{aligned}$$

as $n \to \infty$ and for the limit distribution H being non-degenerate such that each $H_i, i = 1, \ldots, D$ is non-degenerate and must be in the GEV family, then the distribution H is called a D-dimensional multivariate extreme value distribution and F is said to belong to the domain of attraction of H, which we write $F \in D(H)$.

These distributions have received theoretical consideration in works by de Haan and Resnick (1977), de Haan (1985), Pickands (1981), and Resnick (1987). In the characterization of the multivariate extreme distribution, *max-stable (or min-stable)* distributions play a central role. We say a distribution $H(\mathbf{x})$ is *max-stable* if for every $t > 0$ there exist functions $\boldsymbol{\alpha}(t) > 0, \boldsymbol{\beta}(t)$ such that

$$H^t(\mathbf{x}) = H(\boldsymbol{\alpha}(t)\mathbf{x} + \boldsymbol{\beta}(t)) = H(\alpha_1(t)x_1 + \beta_1(t), \ldots, \alpha_D(t)x_D + \beta_D(t)). \quad (1.11)$$

The following theorem describes the equivalence between multivariate extreme value distributions and max-stable distributions.

Theorem 1.4 *The class of multivariate extreme value distributions is precisely the class of max-stable distribution functions with non-degenerate marginals.*

This is Proposition 5.9 in Resnick (1987). After slight modification of Pickands' representation of a min-stable multivariate exponential into a representation for a max-stable multivariate Fréchet distribution, we have

Theorem 1.5 *Suppose $H(\mathbf{x})$ is a limit distribution satisfying (1.10), then*

$$H(\mathbf{x}) = \exp\{-\int_{S_D} \max_{1\leq i\leq D}(\frac{w_i}{x_i})dG(w)\} \quad (1.12)$$

where G is a positive finite measure on the unit simplex

$$S_D = \{(w_1, \ldots, w_D) : \sum_{i=1}^{D} w_i = 1, w_i \geq 0, i = 1, \ldots, D\}$$

and G satisfies

$$\int_{S_D} w_i dG(w) = 1, i = 1, \ldots, D \quad (1.13)$$

Note $v(\mathbf{x}) = \int_{S_D} \max_{1 \le i \le D} (\frac{w_i}{x_i}) dG(w)$ is called the exponent measure by de Haan and Resnick (1977). So to model a multivariate extreme value distribution function is in fact to model the measure function G. De Haan (1985) gave a simple nonparametric procedure for modeling the measure function G. Coles and Tawn (1991) argued that parametric models are preferable when one wants simultaneously to estimate the exponent measure and the dependence structure.

In section 1.2.1, we looked at the limit distribution of a dependent sequence of univariate random variables, and some of the results can be extended in the multivariate context. Suppose now $\{\mathbf{X}_i = (X_{i1}, \cdots, X_{iD}), i = 1, 2, ...\}$ is a D-dimensional stationary stochastic processes with distribution function F and marginals F_d. Also let $\{\widehat{\mathbf{X}}_i\}$ be the associated sequence of i.i.d. random vectors having the same distribution function F. $\mathbf{M}_n$ and $\widehat{\mathbf{M}}_n$ are both pointwise maxima of $\{\mathbf{X}_i\}$ and $\{\widehat{\mathbf{X}}_i\}$ respectively. Suppose

$$\begin{aligned} \lim_{n\to\infty} \Pr\{M_{n1} \le u_{n1}, \ldots, M_{nD} \le u_{nD}\} = H(\boldsymbol{\tau}) \\ \lim_{n\to\infty} \Pr\{\widehat{M}_{n1} \le u_{n1}, \ldots, \widehat{M}_{nD} \le u_{nD}\} = \widehat{H}(\boldsymbol{\tau}) \end{aligned} \tag{1.14}$$

both exist and are nonzero, then a quantity that Nandagopalan (1990, 1994) called the *multivariate extremal index* can relate the extreme value properties of a stationary process to those of i.i.d. sequence. The *multivariate extremal index* is defined by

$$H(\boldsymbol{\tau}) = \widehat{H}(\boldsymbol{\tau})^{\theta(\boldsymbol{\tau})} \tag{1.15}$$

where $\theta(\boldsymbol{\tau})$ satisfies

(i) $0 \le \theta(\boldsymbol{\tau}) \le 1$ for all $\boldsymbol{\tau}$,

(ii) $\theta(0, \ldots, 0, \tau_d, 0, \ldots, 0) = \theta_d$ for $\tau_d > 0$, where θ_d is the extremal index of the d^{th} component process.

(iii) $\theta(c\boldsymbol{\tau}) = \theta(\boldsymbol{\tau})$ for all $c > 0$(Theorem 1.1 of Nandagopalan 1994).

Smith and Weissman (1996) pointed out that these properties are not sufficient to characterize the function $\theta(\boldsymbol{\tau})$. They also argued two reasons why one needs to obtain a more precise characterization to cover a much broader range of processes and to correspond to real stochastic processes, for instance, multivariate maxima of moving maxima processes which we are going to address in this work. The first reason is that "the number of examples for which the multivariate extreme index has been calculated is currently very small (Nandagopalan 1994, Weissman 1994) and it is important to be able to extend this class to cover a much broader range of processes". The second reason is that "why we need a characterization is statistical: crude estimators of $\theta(\tau)$ are easy to construct, but would not correspond to multivariate extreme index of any real stochastic process."

1.2.3 Basic properties of multivariate extreme value distributions

In this subsection, we study some basic properties of multivariate extreme value distribution functions. The following two lemmas are very general, not restricted to MEV, they are theorems 5.1.1 and lemma 5.2.1 in Galambos (1987).

Lemma 1.6 *Let $F(\mathbf{x})$ be a D-dimensional distribution function with marginals $F_d(x)$, $1 \le d \le D$. Then, for all $x_1, x_2, \ldots, x_D$,*

$$\max(0, \sum_{d=1}^{D} F_d(x_d) - D + 1) \le F(x_1, x_2, \ldots, x_D) \le \min(F_1(x_1), F_2(x_2), \ldots, F_D(x_D)).$$

Lemma 1.7 *Let $F_n(\mathbf{x})$ be a sequence of D-dimensional distribution functions, $F_{nd}(x_d)$ be the dth univariate marginal of $F_n(\mathbf{x})$. If $F_n(\mathbf{x})$ converges weakly to a nondegenerate continuous distribution function $F(\mathbf{x})$, then, for each d with $1 \le d \le D$, $F_{nd}(x_d)$ converges weakly to dth marginal $F_d(x_d)$ of $F(\mathbf{x})$.*

The **Copula**, or **dependence function**, is a very useful concept in the investigation of limit distributions for normalized extremes. It is an multivariate distribution with all marginals being uniform $U(0, 1)$.

Definition 1.1 *Let $F(\mathbf{x})$ be a D-dimensional distribution function, with dth univariate margin F_d. The copula associated with F, is a distribution function $C : [0, 1]^m \to [0, 1]$ that satisfies*

$$F(x_1, x_2, \ldots, x_D) = C[F_1(x_1), F_2(x_2), \ldots, F_D(x_D)].$$

Write $C_F = C_F(\mathbf{y}) = C(\mathbf{y})$ over the unit cube $0 \le y_d \le 1,\ 1 \le d \le D$.

Based on the function $C(\mathbf{y})$, we now re-state theorems which relate the univariate marginals and the multivariate or dependence structure of the limit distributions.

Theorem 1.8 *If (1.10) holds, then the dependence function C_H of the limit $H(\mathbf{x})$ satisfies*

$$C_H^k(y_1^{1/k}, y_2^{1/k}, \ldots, y_D^{1/k}) = C_H(y_1, y_2, \ldots, y_D)$$

where $k \geq 1$ is an arbitrary integer. (This is Theorem 5.2.1 of Galambos 1987).

Theorem 1.9 *A D-dimensional distribution function $H(\mathbf{x})$ is a limit of (1.10) if and only if its univariate marginals are of the same type as one of three type distributions and if its copula C_H satisfies the condition of Theorem 1.8. (This is theorem 5.2.4 of Galambos 1987).*

Theorem 1.9 tells in principle that if we want to determine $\mathbf{a}_n$ and $\mathbf{b}_n$ we just need to determine the components from the marginal limit convergence forms. Let's look at a simple example to illustrate how Theorem 1.9 works.

Example 1.1 *Let (X, Y) have a bivariate exponential distribution function $F(x, y)$. If $\frac{\mathbf{M}_n - \mathbf{a}_n}{\mathbf{b}_n}$ converges weakly to a nondegenerate distribution function $H(x, y)$, we can choose*

$$\mathbf{a}_n = (\log n, \log n) \quad \textit{and } \mathbf{b}_n = (1, 1).$$

1.3 Subclasses of max-stable processes

Davis and Resnick (1989) studied what they called the max-autoregressive moving average (MARMA(p,q)) process of a stationary process $\{X_n\}$ which satisfy the MARMA recursion,

$$X_n = \phi_1 X_{n-1} \vee \cdots \vee \phi_p X_{n-p} \vee Z_n \vee \theta_1 Z_{n-1} \vee \cdots \vee \theta_q Z_{n-q}$$

for all n where $\phi_i, \theta_j \geq 0, 1 \leq i \leq p, 1 \leq j \leq q$ and $\{Z_n\}$ is i.i.d. with common distribution function $F(x) = \exp\{-\sigma x^{-1}\}$. For any given $\{\phi_i\}, \{\theta_j\}$, the corresponding process is a max-stable process. They have argued "it is unlikely that another subclass of the max-stable processes can be found which is as broad and tractable as the MARMA class". Some basic properties of the MARMA processes have been shown and the prediction of a max-stable process has been studied relatively completely. However, much less is known about estimation of MARMA process. For prediction, see also Davis and Resnick (1993). A naive estimation procedure for ϕ_i, θ_j's when the order $q = 1$ is given in Davis and Resnick(1989).

Deheuvels (1983) defined what he called the moving minimum(MM) corresponding process as

$$T_i = \min\{\delta_k Z_{i-k}, -\infty < k < \infty\}, -\infty < i < \infty,$$

where $\delta_k > 0$, and $\{Z_k\}$ are i.i.d. exponential 1. The main theorem of Deheuvels (1983) is exactly stated as the following theorem.

Theorem 1.10 *If $(T_0, \ldots, T_m)$ follows a joint multivariate extreme value distribution for minima with exponentially $E(1)$ distributed margins, then there exist $m+1$ sequences $\{a_k^i(n),\ -\infty < k < \infty\}$ depending on $n = 1, 2, \ldots,$ of positive numbers, such that, if $T_i(n) = \min\{a_k^i(n) Z_{-k},\ -\infty < k < \infty\}$, $i = 0, \ldots, m$, then $(T_0(n), \ldots, T_m(n))$ converges in distribution to $(T_0, \ldots, T_m)$ as $n \to \infty$.*

The results of Deheuvels (1983) are very strong, but the model itself is still not easily tractable for the estimation of parameters. Notice that the reciprocal of $\frac{1}{T_i}$ gives the moving maximum processes as

$$\frac{1}{T_i} = \max\{\frac{1}{\delta_k} Z'_{i-k},\ -\infty < k < \infty\},\ -\infty < i < \infty$$

where $\{Z'_k\}$ are i.i.d unit Fréchet random variables. Smith and Weissman (1996) extended this definition to a more general framework which is more realistic and is called multivariate maxima of moving maxima (henceforth M4) process. The definition is

$$Y_{id} = \max_l \max_k a_{l,k,d} Z_{l,i-k}, \quad d = 1, \cdots, D, \tag{1.16}$$

where $\{Z_{li}, l \geq 1, -\infty < i < \infty\}$ are an array of independent unit Fréchet random variables. The constants $\{a_{l,k,d}, l \geq 1, -\infty < k < \infty, 1 \leq d \leq D\}$ are nonnegative constants satisfying

$$\sum_{l=1}^{\infty} \sum_{k=-\infty}^{\infty} a_{l,k,d} = 1 \text{ for } d = 1, \ldots, D \tag{1.17}$$

As we have seen that $M4$ processes deal with D dimensional random processes whereas MM processes deal with univariate processes ($D = 1$). Under the model (1.16), Smith and Weissman (1996) have shown very attractive results. Some are parallel to the results of Deheuvels (1983). Although MM processes are only specified over one index there are possibilities to easily extend to over two indexes. The extension of MM processes to $M4$ processes results in hopes to estimate model parameters easily.

Following de Haan(1984), (1.16) defines max-stable processes because for any finite number r and positive constants $\{y_{id}\}$ we have

$$
\begin{aligned}
&\Pr\{Y_{id} \le y_{id}, 1 \le i \le r, 1 \le d \le D\} \\
&\quad = \Pr\{Z_{l,i-k} \le \tfrac{y_{id}}{a_{l,k,d}} \text{ for } l \ge 1, -\infty < k < \infty, 1 \le i \le r, 1 \le d \le D\} \\
&\quad = \Pr\{Z_{l,m} \le \min_{1-m \le k \le r-m} \min_{1 \le d \le D} \tfrac{y_{m+k,d}}{a_{l,k,d}}, l \ge 1, -\infty < m < \infty\} \\
&\quad = \exp[-\textstyle\sum_{l=1}^{\infty} \sum_{m=-\infty}^{\infty} \max_{1-m \le k \le r-m} \max_{1 \le d \le D} \tfrac{a_{l,k,d}}{y_{m+k,d}}]
\end{aligned}
\tag{1.18}
$$

This is (2.5) of Smith and Weissman (1996) and we have

$$\Pr^n\{Y_{id} \le ny_{id}, 1 \le i \le r, 1 \le d \le D\} = \Pr\{Y_{id} \le y_{id}, 1 \le i \le r, 1 \le d \le D\}$$

which tells that $\{\mathbf{Y}_i\}$ are max-stable. They have argued that the extreme values of a multivariate stationary process may be characterized in terms of a limiting max-stable process under quite general conditions. They also showed that a very large class of max-stable processes may be approximated by the M4 processes mainly because those processes have the same multivariate extremal index (Theorem 2.3 in Smith and Weissman 1996). The theorem and conditions appear below.

Now fix $\tau = \{\tau_1, \ldots, \tau_D\}$ with $0 \le \tau_d < \infty$, $d = 1, \ldots, D$. Let $\{u_{nd},\ n \ge 1\}$ be a sequence of thresholds such that $n\{1 - F_d(u_{nd})\} \to \tau_d$ under the model assumption. Since Z_{lk} is unit Fréchet we can take $u_{nd} = \frac{n}{\tau_d}$. Denote $\mathbf{u}_n = (u_{n1}, \ldots, u_{nd})$ and $\mathcal{B}_j^k(\mathbf{u}_n)$ the σ-field generated by the events $\{X_{id} \le u_{nd},\ j \le i \le k, 1 \le d \le D\}$ for $1 \le j \le k \le n$. Define

$$\alpha_{nt} = \sup\{|P(A \cap B) - P(A)P(B)| :\ A \in \mathcal{B}_1^k(\mathbf{u}_n),\ B \in \mathcal{B}_{k+t}^n(\mathbf{u}_n)\} \tag{1.19}$$

where the supremum is taken over $1 \le k \le n - t$ and two respective σ-fields. If there exists a sequence $\{t_n, n \ge 1\}$ such that

$$t_n \to \infty,\ t_n/n \to 0,\ \alpha_{n,t_n} \to 0 \ \text{ as } n \to \infty \tag{1.20}$$

the mixing condition $\triangle(\mathbf{u}_n)$ is said to hold (Nandagopalan 1994, Smith and Weissman 1996). And further, there exists a sequence $\{k_n,\ n \ge 1\}$ such that

$$k_n \to \infty,\ k_n t_n/n \to 0,\ k_n \alpha_{n,t_n} \to 0 \ \text{ as } n \to \infty. \tag{1.21}$$

Let $r_n = [n/k_n]$ the integer part of n/k_n. We now exactly state a lemma and a theorem (Lemma 2.2 and their main theorem Theorem 2.3 of Smith and Weissman 1996).

Lemma 1.11 *Suppose (1.19)-(1.21) hold. Then*

$$\theta(\boldsymbol{\tau}) = \lim_{n\to\infty} \Pr\{Y_{id} \le u_{nd},\ 2 \le i \le r_n, 1 \le d \le D | \max_d (\frac{Y_{1d}}{u_{nd}}) > 1\}. \tag{1.22}$$

Alternatively, if we assume

$$\lim_{r\to\infty} \lim_{n\to\infty} \sum_{i=r}^{r_n} \sum_{d=1}^{D} \Pr\{Y_{id} > u_{nd} | \max_d (\frac{Y_{1d}}{u_{nd}}) > 1\} = 0, \tag{1.23}$$

then (1.22) is equivalent to

$$\theta(\boldsymbol{\tau}) = \lim_{r\to\infty} \lim_{n\to\infty} \Pr\{Y_{id} \le u_{nd},\ 2 \le i \le r, 1 \le d \le D | \max_d (\frac{Y_{1d}}{u_{nd}}) > 1\}. \tag{1.24}$$

This lemma is basically a restatement of results of O'Brien, for example O'Brien (1987).

Theorem 1.12 *Suppose $\triangle(\mathbf{u}_n)$ and (1.23) hold for $\{\mathbf{Y}_i\}$, so that the multivariate extremal index $\theta^{\mathbf{Y}}(\boldsymbol{\tau})$ is given by (1.24). Suppose also the same assumptions hold for $\{\mathbf{X}_i\}$ (with the same t_n, k_n sequences). So the multivariate extremal index $\theta^{\mathbf{X}}(\boldsymbol{\tau})$ is also given by (1.24) with X_{id} replacing Y_{id} everywhere. Then $\theta^{\mathbf{Y}}(\boldsymbol{\tau}) = \theta^{\mathbf{X}}(\boldsymbol{\tau})$.*

The extremal index of the process defined by (1.16) is

$$\theta(\boldsymbol{\tau}) = \frac{\sum_l \max_k \max_d a_{l,k,d}\tau_d}{\sum_l \sum_k \max_d a_{l,k,d}\tau_d}. \tag{1.25}$$

Although theoretical results have been obtained, the estimation of parameters in both MARMA(p,q) and M4 processes are not well developed and the applications of the two processes are very limited. Recently Hall, Peng and Yao (2001) discussed moving maximum models

$$Y_i = \sup\{a_{j-i}Z_i,\ -\infty < i < \infty\}$$

where the distribution of Z_i is assumed either $F(z|\theta) = \exp(-z^{-\theta})$ or the generalized Pareto distribution $F(z|\theta) = 1 - (1+z)^{-\theta}$. Then for a finite number of parameters, they chose $(\theta, a_{(m)})$ to minimize

$$\begin{aligned} D_m(\theta, a_{(m)}) = \textstyle\int (\widehat{G}(y) - \prod_{i=2-m}^{k} & F[\min\{a_{j-i}^{-1} y_j, \\ & \max(i,1) \le j \le \min(i+m,k)\}|\theta])^2 w(y)dy, \end{aligned} \tag{1.26}$$

where the integral is over $y = (y_1, \ldots, y_k) \in \mathbb{R}_+^k$ and

$$\widehat{G}(y) = (n-k)^{-1} \sum_{i=1}^{n-k} I_{(Y_{i+j-1} \le y_j \text{ for } 1 \le j \le k)}, \tag{1.27}$$

and w is a nonnegative weight function. We state their main theorem as follows.

Theorem 1.13 *Under conditions*

- *F has support on the positive half-line, and is in the domain of attraction of a Type II extreme value distribution.*
- *each a_i is nonnegative and, for some $\epsilon \in (0, r)$, $0 < \sum_i a_i^{r-\epsilon} < \infty$.*

then

$$\sup_{-\infty < y_1, \ldots, y_k < \infty} |\Pr(Y_1^* \leq y_1, \ldots, Y_k^* \leq y_k | Y_1, \ldots, Y_n) - \Pr(Y_1 \leq y_1, \ldots, Y_k \leq y_k)| \to 0 \tag{1.28}$$

where Y_j^* is defined by

$$Y_j^* = \sup\{\widehat{a}_{j-i} Z_i^*, \ -\infty < i < \infty\}$$

$\widehat{a}_{j-i}$ and $\widehat{\theta}$ are solutions of (1.26) and Z_i^* has distribution function $F(.|\widehat{\theta})$. Moreover, if $m \geq C_4 (\log n)^2$ for C_4 sufficiently large, the rate of convergence in (1.28) is $O_p(n^{-(1/2)+\delta})$ for all $\delta > 0$.

Our present work on the estimation of $M4$ processes is somewhat parallel to Hall et al. (2001)'s work. In contrast to the bootstrapped processes which Hall et al. (2001) used to construct confidence intervals and prediction intervals, we directly construct parameter estimators and prove their asymptotic properties. We will systematically solve the estimation problems of M4 processes in this work.

Chapter 2

Probabilistic Properties of Multivariate Maxima of Moving Maxima Processes and Basic Estimation of Parameters

2.1 Introduction

In this chapter we consider the model specified in Smith and Weissman (1996), extend some properties and propose estimating procedures which determine max moving range, signature patterns and estimation of parameters.

Let $\{Z_{lk}, l \geq 1, -\infty < k < \infty\}$ be an array of independent unit Fréchet random variables. Smith and Weissman (1996) studied the following process which they call $M4$ process.

$$Y_{id} = \max_{l} \max_{k} a_{l,k,d} Z_{l,i-k}, \quad d = 1, \cdots, D,$$

for nonnegative constants $\{a_{l,k,d}, l \geq 1, -\infty < k < \infty\}$ satisfying $\sum_{l=1}^{\infty} \sum_{k=-\infty}^{\infty} a_{l,k,d} = 1$ for $d = 1, \cdots, D$.

Since in practice we will not have infinitely many parameters, usually we have $l = 1, \cdots, L$ and $-K_1 \leq k \leq K_2$ for some finite numbers L, K_1 and K_2. Here L corresponds to the maximum number of moving patterns. And K_1 and K_2 characterize the range of the sequence dependence. We will focus on the finite dimensional $M4$ process which we state as

$$Y_{id} = \max_{1 \leq l \leq L} \max_{-K_1 \leq k \leq K_2} a_{l,k,d} Z_{l,i-k}, \quad d = 1, \cdots, D, \tag{2.1}$$

where $\sum_{l=1}^{L} \sum_{k=-K_1}^{K_2} a_{l,k,d} = 1$ for $d = 1, \cdots, D$.

The model assumptions made here can be related to some real motivations in insurance and finance as well as environmental engineering. For example, insurance claims result from different factors (or patterns) and claims are usually made within a certain period. Stock market variation results from an internal or external big market price movement and will last a certain period. As we mentioned in section 1.1, the exceedance over threshold approaches have been advocated in modern extreme value theory applications. The exceedances over a high threshold of the observed process are modeled as the exceedances over threshold of an $M4$ process. We are not modeling the whole process as $M4$.

Under model (2.1), it is easily to obtain the finite distribution of $\{Y_{id},\ 1 \le i \le r,\ 1 \le d \le D\}$ as a consequence of (1.18). The distribution has the following form

$$\Pr\{Y_{id} \le y_{id}, 1 \le i \le r, 1 \le d \le D\} = \exp[-\sum_{l=1}^{L} \sum_{m=1-K_2}^{r+K_1} \max_{1-m \le k \le r-m} \max_{1 \le d \le D} \frac{a_{l,k,d}}{y_{m+k,d}}]$$

where $a_{l,k,d} = 0$ for $k < -K_1$ or $k > K_2$. The goal is to estimate all parameters $\{a_{l,k,d}\}$ under the constraints that all parameters are nonnegative and the summation is equal to 1 for each $d = 1, \dots, D$. Due to the singularity that appears in the distribution function, maximum likelihood method is not directly applicable because of the singularities. One way to avoid this problem is to use a grouped likelihood approach, which has been advocated in similar circumstances by Barnard (1965) and Kempthorne (1966), and developed in detail by Giesbrecht and Kempthorne (1976) for the particular case of a three parameter log-normal distribution. But this is not so easy to apply in a multivariate context, so we consider alternative approaches.

In this work, first we study the structure of model (2.1) and prove probabilistic properties which can be used to construct estimating procedures. Second, we study empirical distribution functions of the finite number of random variables. Guaranteed by the strong law of large numbers or ergodicity, we are able to construct estimators of all parameters and prove the consistency and asymptotics of the proposed estimators. We will start from simple examples which help us to understand the model structure and easily construct some basic estimating procedure which is based on the probabilistic properties of the $M4$ process. The related results are illustrated in section 2.2. Then in section 2.4 we study more general case and develop an estimating procedure which first estimates the weights within the same pattern and then estimates the weights among patterns.

2.2 Extended properties

Under model (2.1), it is possible that a big value of Z_{lk} (unobservable) dominates all other values within a certain period of length $K_2 + K_1 + 1$ and there is a strong dependence among the big values (this is part of motivation for the model (2.1)). Actually this occurs infinitely many times of the whole process. It will be clear after looking at some examples and some theoretical results later on.

Consider now a simplified model,

$$Y_{id} = \max_{-K_1 \le k \le K_2} a_{kd} Z_{i-k}, \tag{2.2}$$

which is corresponding to $L = 1$ (single pattern).

Define

$$\begin{aligned} A_t^d = [a_{-K_1 d} Z_{t+K_1} &\geq \max_{\substack{-K_1 \le k \le K_2 \\ k \ne -K_1}} a_{kd} Z_{t-k}, \\ a_{-K_1+1,d} Z_{t+K_1} &\geq \max_{\substack{-K_1 \le k \le K_2 \\ k \ne -K_1+1}} a_{kd} Z_{t+1-k}, \\ &\vdots \\ a_{K_2 d} Z_{t+K_1} &\geq \max_{\substack{-K_1 \le k \le K_2 \\ k \ne K_2}} a_{kd} Z_{t+K_1+K_2-k}] \end{aligned} \tag{2.3}$$

We have $\Pr(A_t^d) > 0, t \geq 1$. We now derive the explicit form of $\Pr(A_t^d)$. Denote

$$\begin{aligned} A_t^{dz} = [a_{-K_1 d} z &\geq \max_{\substack{-K_1 \le k \le K_2 \\ k \ne -K_1}} a_{kd} Z_{t-k}, \\ a_{-K_1+1,d} z &\geq \max_{\substack{-K_1 \le k \le K_2 \\ k \ne -K_1+1}} a_{kd} Z_{t+1-k}, \\ &\vdots \\ a_{K_2 d} z &\geq \max_{\substack{-K_1 \le k \le K_2 \\ k \ne K_2}} a_{kd} Z_{t+K_1+K_2-k}] \end{aligned} \tag{2.4}$$

Based on (2.3), we can draw the following diagram:

$$\begin{array}{ccccccccc|c|ccc} Z_{t-K_2} & \cdots & Z_{t-2} & Z_{t-1} & Z_t & Z_{t+1} & \cdots & Z_{t+K_1-1} & & Z_{t+K_1} & Z_{t+K_1+1} & Z_{t+K_1+2} & \cdots \\ a_{K_2 d} & \cdots & a_{2d} & a_{1d} & a_{0d} & a_{-1d} & \cdots & a_{-K_1+1,d} & & a_{-K_1 d} & & & \\ & \cdots & a_{3d} & a_{2d} & a_{1d} & a_{0d} & \cdots & a_{-K_1+2,d} & & a_{-K_1+1,d} & a_{-K_1 d} & & \\ & & & & & & & & & \vdots & & & \\ & & & & & & & & & a_{K_2 d} & a_{K_2-1,d} & a_{K_2-2,d} & \cdots \end{array}$$

then

$$
\begin{aligned}
\Pr(A_t^{zd}) &= \Pr(Z_{t-K_2} \le \tfrac{a_{-K_1d}}{a_{K_2d}} z, \\
&\qquad Z_{t-K_2+1} \le \min(\tfrac{a_{-K_1d}}{a_{K_2-1,d}}, \tfrac{a_{-K_1+1,d}}{a_{K_2d}}) z \\
&\qquad \dots \\
&\qquad Z_{t+K_1-1} \le \min(\tfrac{a_{-K_1d}}{a_{-K_1+2,d}}, \tfrac{a_{-K_1+1,d}}{a_{-K_1+2,d}}, \dots, \tfrac{a_{K_2-2,d}}{a_{K_2d}}) z \\
&\qquad Z_{t+K_1+1} \le \min(\tfrac{a_{-K_1+1,d}}{a_{-K_1d}}, \tfrac{a_{-K_1+2,d}}{a_{-K_1+1,d}}, \dots, \tfrac{a_{K_2d}}{a_{K_2-1,d}}) z \\
&\qquad Z_{t+K_1+2} \le \min(\tfrac{a_{-K_1+2,d}}{a_{-K_1d}}, \tfrac{a_{-K_1+3,d}}{a_{-K_1+1,d}}, \dots, \tfrac{a_{K_2d}}{a_{K_2-2,d}}) z \\
&\qquad \dots \\
&\qquad Z_{t+2K_1+K_2-1} \le \min(\tfrac{a_{K_2-1,d}}{a_{-K_1d}}, \tfrac{a_{K_2,d}}{a_{-K_1+1,d}}) z \\
&\qquad Z_{t+2K_1+K_2} \le \tfrac{a_{K_2d}}{a_{-K_1d}} z) \\
&= \exp[-\tfrac{1}{z}\{\sum_{j=1}^{K_1+K_2} (\bigwedge_{i=1}^{j} \tfrac{a_{-K_1+i-1,d}}{a_{K_2-j+i,d}} + \bigwedge_{i=1}^{j} \tfrac{a_{K_2-i+1,d}}{a_{-K_1+j-i,d}})\}] \\
&= \exp[-\tfrac{\Delta_d}{z}]
\end{aligned}
$$

so

$$\Pr(A_t^d) = \int_0^\infty \Pr(A_t^{zd}) \frac{1}{z^2} e^{-\frac{1}{z}} dz = \int_0^\infty \frac{1}{z^2} e^{-\frac{1+\Delta_d}{z}} dz = \frac{1}{(1+\Delta_d)^2}$$

thus

$$\Pr(A_t^d) = \frac{1}{[1 + \sum_{j=1}^{K_1+K_2} (\bigwedge_{i=1}^{j} \frac{a_{-K_1+i-1,d}}{a_{K_2-j+i,d}} + \bigwedge_{i=1}^{j} \frac{a_{K_2-i+1,d}}{a_{-K_1+j-i,d}})]^2}. \tag{2.5}$$

For $P(A_t^d A_{t+m}^d)$, it is clear $P(A_t^d A_{t+m}^d) = (P(A_t^d))^2$ if $m > K_1 + K_2$. Suppose $1 \le m \le K_1 + K_2$, then
from A_t we get

$$a_{-K_1+m,d} Z_{t+K_1} \ge a_{-K_1d} Z_{t+m+K_1}, \tag{2.6}$$

from A_{t+m}^d we get

$$a_{-K_1d} Z_{t+m+K_1} \ge a_{-K_1+m,d} Z_{t+K_1}. \tag{2.7}$$

So (2.6) and (2.7) imply $a_{-K_1+m,d} Z_{t+K_1} = a_{-K_1d} Z_{t+m+K_1}$, thus

$$P(A_t^d A_{t+m}^d) = \begin{cases} (P(A_t^d))^2 & \text{if } m > K_1 + K_2, \\ 0 & \text{if } 1 \le m \le K_1 + K_2. \end{cases} \tag{2.8}$$

We have the following lemma.

Lemma 2.1 *Under the model (2.2), for each d we have*

$$\Pr(A_t^d,\ i.o.) = 1$$

Proof. Let $\{t_1, t_2, \cdots, \}$ be subsequence of $\{t \geq 1\}$ such that $t_{i+1} - t_i > K_1 + K_2, i \geq 1$, then $\{A^d_{t_i}\}$ is an independent sequence of events. By the Borel-Cantelli lemma for independent events we get

$$\Pr(\bigcap_{i=1}^{\infty}\bigcup_{n=i}^{\infty} A^d_{t_n}) = 1$$

since $P(A^d_{t_i}) > 0$. But

$$\bigcap_{t=1}^{\infty}\bigcup_{n=t}^{\infty} A^d_n \supseteq \bigcap_{i=1}^{\infty}\bigcup_{n=i}^{\infty} A^d_{t_n},$$

so

$$\Pr(A^d_t,\ i.o.) = \Pr(\bigcap_{t=1}^{\infty}\bigcup_{n=t}^{\infty} A^d_n) = 1$$

and this completes the proof. □

This theorem tells there are an infinite number of time periods within which the process is driven by a single extreme jump. For example, a real-world interpretation might be that a flood in a certain region and a certain time period is caused by a specific hurricane. The strengths of different hurricanes are different and the costs are different. Or we say they follow different patterns.

We have following theorems.

Corollary 2.2 *Under the model (2.2), for each d we have*

$$P(Y_{td} = a_{-K_1d}Z_{t+K_1}, Y_{t+1,d} = a_{-K_1+1,d}Z_{t+K_1}, \cdots, Y_{t+K_2+K_1,d} = a_{K_2d}Z_{t+K_1}, i.o.) = 1$$

or

$$P(\frac{Y_{td}}{Y_{td} + Y_{t+1,d} + \cdots + Y_{t+K_2+K_1,d}} = a_{-K_1d}, \quad i.o.) = 1, \tag{2.9}$$

and equivalently

$$P(\frac{Y_{t+m,d}}{Y_{td} + Y_{t+1,d} + \cdots + Y_{t+K_2+K_1,d}} = a_{-K_1+m,d}, \quad i.o.) = 1 \qquad m = 0, \cdots, K_1 + K_2 \tag{2.10}$$

Proof. The condition defining the set A^d_t implies

$$Y_{td} = a_{-K_1d}Z_{t+K_1},\ Y_{t+1,d} = a_{-K_1+1,d}Z_{t+K_1}, \cdots, Y_{t+K_2+K_1,d} = a_{K_2d}Z_{t+K_1},$$

and hence by the theorem we have proved the corollary. □

Theorem 2.3 *Under the model (2.2), if* $P(\frac{Y_{t+m,d}}{Y_{td}+Y_{t+1,d}+\cdots+Y_{t+K_2+K_1,d}} = c_{md}, \quad i.o.) = 1$ *for* $m = 0, \cdots, K_1 + K_2$*, then* $c_{md} = a_{K_1+m,d}$ *and those* $\{Y_{td}, \cdots, Y_{t+K_1+K_2,d}\}$ *form the events* A_t^d *if* $\frac{Y_{t+m,d}}{Y_{td}+Y_{t+1,d}+\cdots+Y_{t+K_2+K_1,d}} = c_{md}$.

Remark: The theorem says, for example when $m = 0$, there is only one constant $c_{0d} = a_{-K_1 d}$ such that (2.9) is true. And if $\frac{Y_{t+m,d}}{Y_{td}+Y_{t+1,d}+\cdots+Y_{t+K_2+K_1,d}} = c_{md}$ is true for one m, it is true for all m.

Proof. We only prove the case when $m = 0$. Define random variables

$$T_{td} = sI_{(Y_{td}=a_{t-s,d}Z_s)}.$$

Notice that $t - K_2 \le s \le t + K_1$ and T_{td} is uniquely defined for each t and hence for all t because the Z_ts have an absolutely continuous distribution. The event A_t^d corresponds to

$$T_{td} = T_{t+1,d} = \cdots = T_{t+K_1+K_2,d} = t + K_1.$$

Suppose now we have that

$$\frac{Y_{td}}{Y_{td} + Y_{t+1,d} + \cdots + Y_{t+K_2+K_1,d}} = p_d \tag{2.11}$$

occurs infinitely many times for $p_d \neq a_{-K_1 d}$, then $T_{td}, T_{t+1,d}, \cdots, T_{t+K_1+K_2,d}$ must follow one of the following two cases.

(1) $T_{td} = T_{t+1,d} = \cdots = T_{t+K_1+K_2,d} = t + K$, where $K \neq K_1, K \in \{-K_2, -K_2 + 1, \cdots, K_1 - 1\}$

(2) $T_{td}, T_{t+1,d}, \cdots, T_{t+K_1+K_2,d}$ contain at least two different values.

For case 1, $Y_{td} = a_{-Kd}Z_{t+K}$, $Y_{t+1,d} = a_{1-K,d}Z_{t+K}, \cdots$, $Y_{t+K_1+K_2,d} = a_{K_1+K_2-K,d}Z_{t+K} = 0$ since $K_1 + K_2 - K > K_2$, and $a_{K_2+1,d} = a_{K_2+2,d} = \cdots = 0$. This is a contradiction to $Y_{td} > 0$ all t.

For case 2, this means the LHS of (2.11) is a function of at least two different Z_t's, because if they are the same, the value must be equal to $t + K_1$ which corresponds to $p_d = a_{-K_1 d}$, otherwise it is the case (i). (2.11) can be written into

$$\frac{a_{t-s_1,d}Z_{s_1}}{a_{t-s_1,d}Z_{s_1} + a_{t+1-s_2,d}Z_{s_2} + \cdots + a_{t+K_1+K_2-s_{K_1+K2+1},d}Z_{s_{K_1+K_2+1}}} = p_d \tag{2.12}$$

where $s_1,\ s_2, \ldots, s_{K_1+K_2+1}$ depend on t. Since the range of $t - s_1,\ t + 1 - s_2, \ldots, t + K_1 + K_2 - s_{K_1+K_2+1}$ is finite under the assumption of (2.2), there are fixed numbers

$h_0, h_1, \ldots, h_{K_1+K_2}$ such that

$$\frac{a_{h_0,d}Z_{s_1}}{a_{h_0,d}Z_{s_1} + a_{h_1,d}Z_{s_2} + \cdots + a_{h_{K_1+K2},d}Z_{s_{K_1+K_2+1}}} = p_d \tag{2.13}$$

occurs infinitely many times. This implies

$$\begin{aligned}\Pr(&\tfrac{a_{h_0,d}Z_{s_1}}{a_{h_0,d}Z_{s_1}+a_{h_1,d}Z_{s_2}+\cdots+a_{h_{K_1+K2},d}Z_{s_{K_1+K_2+1}}} \\ &= \tfrac{a_{h_0,d}Z_{s'_1}}{a_{h_0,d}Z_{s'_1}+a_{h_1,d}Z_{s'_2}+\cdots+a_{h_{K_1+K2},d}Z_{s'_{K_1+K_2+1}}}) > 0\end{aligned} \tag{2.14}$$

for some $s_1,\ s_2, \ldots, s_{K_1+K_2+1}$ and $s'_1,\ s'_2, \ldots, s'_{K_1+K_2+1}$ such that

$$\max(s_1,\ s_2, \ldots, s_{K_1+K_2+1}) < \min(s'_1,\ s'_2, \ldots, s'_{K_1+K_2+1}).$$

But (2.14) is not possible since all Z_t's are continuous random variables and the quantity of the LHS in the bracket in (2.14) is independent of the quantity of the RHS in the bracket in (2.14). This shows that (2.11) can not be true.

Suppose now (2.11) occurs at $t = t_1$ and t_2, i.e.

$$\begin{aligned}\Pr(&\tfrac{a_{h_0,d}Z_{s_1}}{a_{h_0,d}Z_{s_1}+a_{h_1,d}Z_{s_2}+\cdots+a_{h_{K_1+K2},d}Z_{s_{K_1+K_2+1}}} \\ &= \tfrac{a_{h'_0,d}Z_{s'_1}}{a_{h'_0,d}Z_{s'_1}+a_{h'_1,d}Z_{s'_2}+\cdots+a_{h'_{K_1+K2},d}Z_{s'_{K_1+K_2+1}}} = p_d) > 0\end{aligned} \tag{2.15}$$

then $s_1,\ s_2, \ldots, s_{K_1+K_2+1}$ and $s'_1,\ s'_2, \ldots, s'_{K_1+K_2+1}$ must have some common values, otherwise (2.15) cannot be true. But (2.15) implies (2.13) occurs infinitely often, and (2.13) implies (2.14) and so we have (2.11) cannot occur even twice.

Both cases have shown contradictions for $p_d \neq a_{K_1}$. So $c_{0d} = a_{-K_1 d}$and those $\{Y_{td}, \cdots, Y_{t+K_1+K_2,d}\}$ form events A_t^d, therefore the proof is completed. □

The following theorem tells that the range cannot be over $K_2 + K_1 + 1$ numbers in order to get infinitely many ratios which are equal to a constant.

Theorem 2.4 *Under the model (2.2), for each d*

$$P(\frac{Y_{td}}{Y_{td} + Y_{t+1,d} + \cdots + Y_{t+K_2+K_1+1,d}} = c_d, \quad i.o.) = 0$$

for any constant c_d.

Proof. Because Y_{td} and $Y_{t+K_1+K_2+1,d}$ cannot be written as functions of just one Z_t, $\frac{Y_{td}}{Y_{td}+Y_{t+1,d}+\cdots+Y_{t+K_2+K_1+1,d}}$ is a function of at least two different Z_t's. The proof then follows by the same arguments as in Theorem 2.3. □

Theorems 2.3 and 2.4 tell that in order to estimate the parameter a_{-K_1+m}, we only need to observe two equal values of $\frac{Y_{t_1+m,d}}{Y_{t_1d}+\cdots+Y_{t_1+K_1+K_2,d}}$, $\frac{Y_{t_2+m,d}}{Y_{t_2d}+\cdots+Y_{t_2+K_1+K_2,d}}$ at time t_1 and t_2, where $t_2 - t_1 > K_1 + K_2$. Naturally one wants to know how fast a sequence defined by $\frac{Y_{t+m,d}}{Y_{td}+\cdots+Y_{t+K_1+K_2,d}}$, $t = 1, 2, ...$, reaches the desired value $a_{-K_1+m,d}$. We use discrete time Markov chain method to study this problem in section 2.5.

2.3 Examples of $M4$ processes

In order to have insight into what the theorems have shown, now we turn to give some examples to demonstrate applications of the theorems.

2.3.1 Case $L = 1$

Example 2.1 *Consider the model*

$$Y_{td} = \max(a_{1d}Z_{t-1}, a_{0d}Z_t), \ d = 1, \ldots, D$$

Define

$$A_t^d = [a_{1d}Z_{t-1} \leq a_{0d}Z_t, \quad a_{1d}Z_t \leq a_{0d}Z_{t+1}]$$

We have $P(A_t^d) > 0$, so by Theorem 2.3,

$$P(Y_{td} = a_{0d}Z_t, Y_{t+1,d} = a_{1d}Z_t, \quad i.o.) = 1$$

or

$$P(\frac{Y_{t+1,d}}{Y_{td}} = \frac{a_{1d}}{a_{0d}} = \frac{1 - a_{0d}}{a_{0d}} = \frac{1}{a_{0d}} - 1, \quad i.o.) = 1.$$

We seek all those ratios of $\frac{Y_{t+1,d}}{Y_{td}}$ such that the ratios are close to a constant. Or equivalently consider

$$P(\frac{Y_{td}}{Y_{td} + Y_{t+1,d}} = a_{0d}, \quad i.o) = 1$$

The ratios are bounded and in [0,1], we have $a_{1d} = 1 - a_{0d}$. In this way we can find accurate estimates for parameters. Figure 2.1 plots the partial points from the original sequence and points from (2.9). Those points falling onto a horizontal line (in the right figure) correspond to those spikes with same shape (in the left figure). Those spikes follow the same moving pattern. The intercept of the line to vertical axis gives the value of a_{0d}.

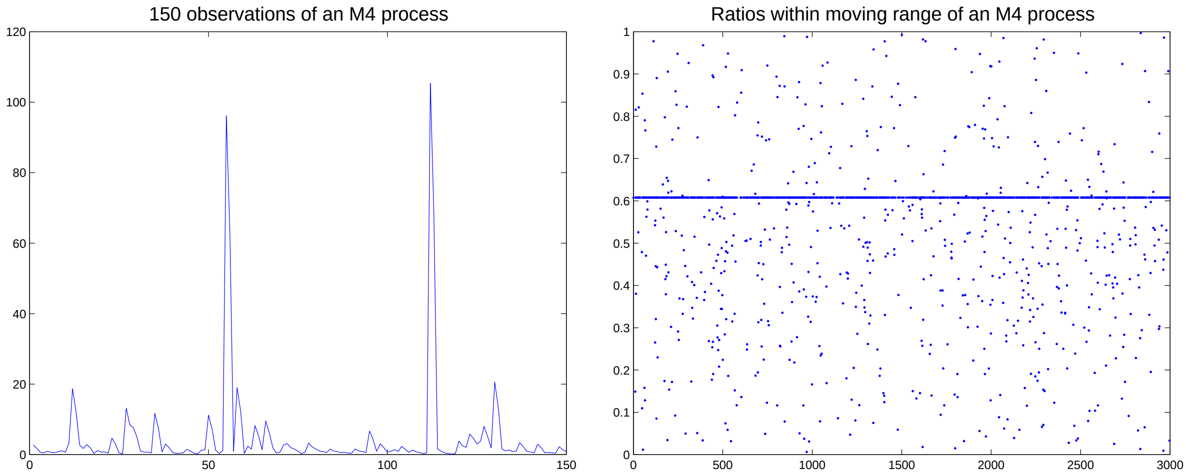

Figure 2.1: Left figure is a time series plot of 150 observations of process $Y_{td} = \max(a_{1d}Z_{t-1}, a_{0d}Z_t)$ for some d. Right figure is a time series plot of 3000 observations of ratios $\frac{Y_{td}}{Y_{td}+Y_{t+1,d}}$. The value of a_{0d} can be read from the right figure.

Example 2.2

$$Y_{td} = \max(a_{1d}Z_{t-1}, a_{0d}Z_t, a_{-1d}Z_{t+1})$$

Define

$$\begin{aligned} A_t^d = [a_{-1d}Z_{t+1} &\geq \max(a_{1d}Z_{t-1}, a_{0d}Z_t), \\ a_{0d}Z_{t+1} &\geq \max(a_{1d}Z_t, a_{-1d}Z_{t+2}), \\ a_{1d}Z_{t+1} &\geq \max(a_{0d}Z_{t+2}, a_{-1d}Z_{t+3})] \end{aligned}$$

we have $P(A_t^d) > 0$, so by Theorem 2.3,

$$P(Y_{td} = a_{-1d}Z_{t+1}, \quad Y_{t+1,d} = a_{0d}Z_{t+1}, \quad Y_{t+2,d} = a_{1d}Z_{t+1}, \quad i.o.) = 1$$

or

$$P(\frac{Y_{td}}{Y_{td} + Y_{t+1,d} + Y_{t+2,d}} = a_{-1d}, \quad i.o.) = 1.$$

Equivalently,

$$P(\frac{Y_{t+1,d}}{Y_{td} + Y_{t+1,d} + Y_{t+2,d}} = a_{0d}, \quad i.o.) = 1$$

and

$$P(\frac{Y_{t+2,d}}{Y_{td} + Y_{t+1,d} + Y_{t+2,d}} = a_{1d}, \quad i.o.) = 1.$$

Figure 2.2 again shows the significant pattern of value a_{-1d}.

Examples 2.1 and 2.1 have shown how to get the moving coefficients in each individual process, but we are mainly interested in multivariate processes. In other words,

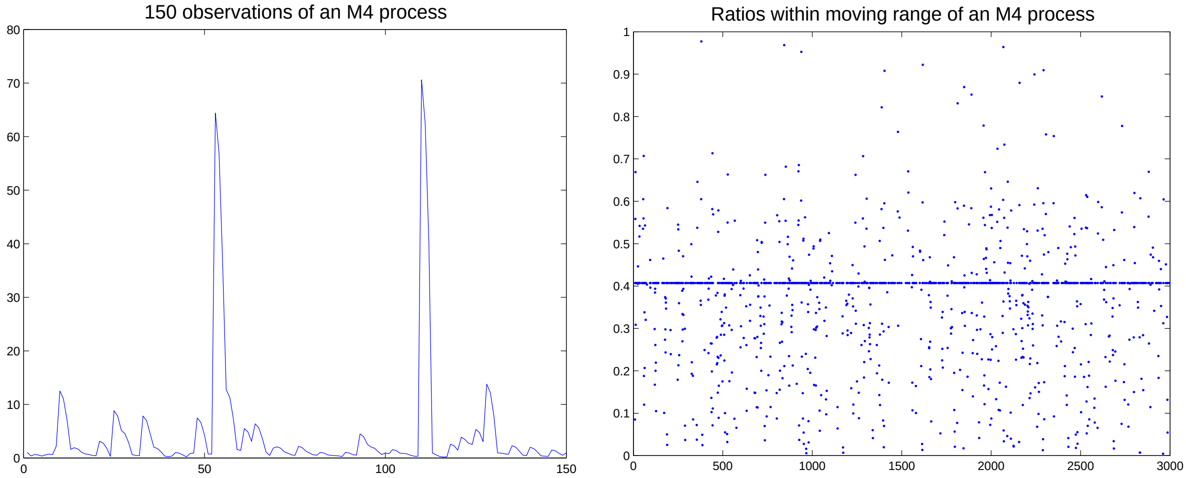

Figure 2.2: Left figure is a time series plot of 150 observations of process $Y_{td} = \max(a_{1d}Z_{t-1}, a_{0d}Z_t, a_{-1d}Z_{t+1})$ for some d. Right figure is a time series plot of 3000 observations of ratios $\frac{Y_{td}}{Y_{td}+Y_{t+1,d}+Y_{t+2,d}}$. The value of a_{-1d} can be read from the right figure.

we need to know how to distinguish different processes. For example we have two bivariate processes

$$\begin{cases} Y_{i1} = \frac{1}{2}\max(Z_{1,i-1},\ Z_{1,i}) \\ Y_{i2} = \frac{1}{3}\max(Z_{1,i-1},\ Z_{1,i},\ Z_{1,i+1}) \end{cases} \tag{2.16}$$

$$\begin{cases} Y_{i1} = \frac{1}{2}\max(Z_{1,i-1},\ Z_{1,i}) \\ Y_{i2} = \frac{1}{3}\max(Z_{1,i},\ Z_{1,i+1},\ Z_{1,i+2}) \end{cases} \tag{2.17}$$

By plotting

$$\frac{Y_{i1}}{Y_{i1}+Y_{i+1,1}},\ \frac{Y_{i+1,1}}{Y_{i1}+Y_{i+1,1}}$$

or

$$\frac{Y_{i2}}{Y_{i2}+Y_{i+1,2}+Y_{i+2,2}},\ \frac{Y_{i+1,2}}{Y_{i2}+Y_{i+1,2}+Y_{i+2,2}},\ \frac{Y_{i+2,2}}{Y_{i2}+Y_{i+1,2}+Y_{i+2,2}}$$

we can get all coefficients $\frac{1}{2}$ and $\frac{1}{3}$, which can be read off from the pictures. But we need to know which model, (2.16) or (2.17), is the true model. We now study this.

When

$$\frac{Y_{i1}}{Y_{i1}+Y_{i+1,1}} = \frac{1}{2},\ \frac{Y_{i+1,1}}{Y_{i1}+Y_{i+1,1}} = \frac{1}{2}$$

were from $Z_{1,i}$, then

$$\frac{Y_{i-1,2}}{Y_{i-1,2}+Y_{i,2}+Y_{i+1,2}} = \frac{1}{3},\ \frac{Y_{i,2}}{Y_{i-1,2}+Y_{i,2}+Y_{i+1,2}} = \frac{1}{3},\ \frac{Y_{i+1,2}}{Y_{i-1,2}+Y_{i,2}+Y_{i+1,2}} = \frac{1}{3} \tag{2.18}$$

for model (2.16), but

$$\frac{Y_{i-2,2}}{Y_{i-2,2}+Y_{i-1,2}+Y_{i,2}}=\frac{1}{3},\ \frac{Y_{i-1,2}}{Y_{i-2,2}+Y_{i-1,2}+Y_{i,2}}=\frac{1}{3},\frac{Y_{i,2}}{Y_{i-2,2}+Y_{i-1,2}+Y_{i,2}}=\frac{1}{3} \quad (2.19)$$

for model (2.17). So if (2.18) is the case we conclude model (2.16), otherwise it's model (2.17). Some other comparison also can be done in order to distinguish the models.

2.3.2 Case $L > 1$

Now consider $L > 1$, where the model is (2.1). Define for each l

$$\begin{aligned} A_t^{ld} = [a_{l,-K_1,d}Z_{l,t+K_1} &\geq \max_{\substack{-K_1\le k\le K_2\\ k\ne -K_1}} a_{l,k,d}Z_{l,t-k}, \\ a_{l,-K_1+1,d}Z_{l,t+K_1} &\geq \max_{\substack{-K_1\le k\le K_2\\ k\ne -K_1+1}} a_{l,k,d}Z_{l,t+1-k}, \\ &\vdots \\ a_{l,K_2,d}Z_{l,t+K_1} &\geq \max_{\substack{-K_1\le k\le K_2\\ k\ne K_2}} a_{l,k,d}Z_{l,t+K_1+K_2-k}]. \end{aligned} \qquad (2.20)$$

Remark: we can define such event for all l simultaneously, but we don't need here. Notice $P(A_t^{ld}) > 0$, so by Theorem 2.3, for each $m = 0, 1, \cdots, K_1 + K_2$, we have

$$P(\tfrac{Y_{t+m,d}}{Y_{td}+Y_{t+1,d}+\cdots+Y_{t+K_2+K_1,d}} = \tfrac{a_{l,-K_1+m,d}}{a_{l,-K_1,d}+a_{l,-K_1+1,d}+\cdots+a_{l,K_2,d}}, \quad i.o.) = 1 \qquad (2.21)$$

We expect to have L signature patterns on the plot of $\frac{Y_{t+m,d}}{Y_{td}+Y_{t+1,d}+\cdots+Y_{t+K_2+K_1,d}}$, and these patterns give estimates of $\frac{a_{l,-K_1+m,d}}{a_{l,-K_1,d}+a_{l,-K_1+1,d}+\cdots+a_{l,K_2,d}}$, $1 \le l \le L$. Figure 2.3 shows three different signature patterns (points fall onto 3 horizontal lines) which correspond to $L = 3$. As we have already seen, the plots give accurate estimates of the ratios. When $L = 1$, we can exactly get all the values of a_{kd}. But for $L > 1$ we cannot. Even for $L = 1$, we have assumed that the model assumptions are exactly satisfied, not something we would expect to use in practice. Also, the whole method presupposes that the margins are transformed into unit Fréchet margin and this wouldn't be exact in practice, either. We need to develop estimation procedures to obtain estimates of $a_{l,k,d}$ in a more practical way. We study this in Chapter 3. In the next section we still assume that the model assumptions are exactly satisfied but L is greater than 1.

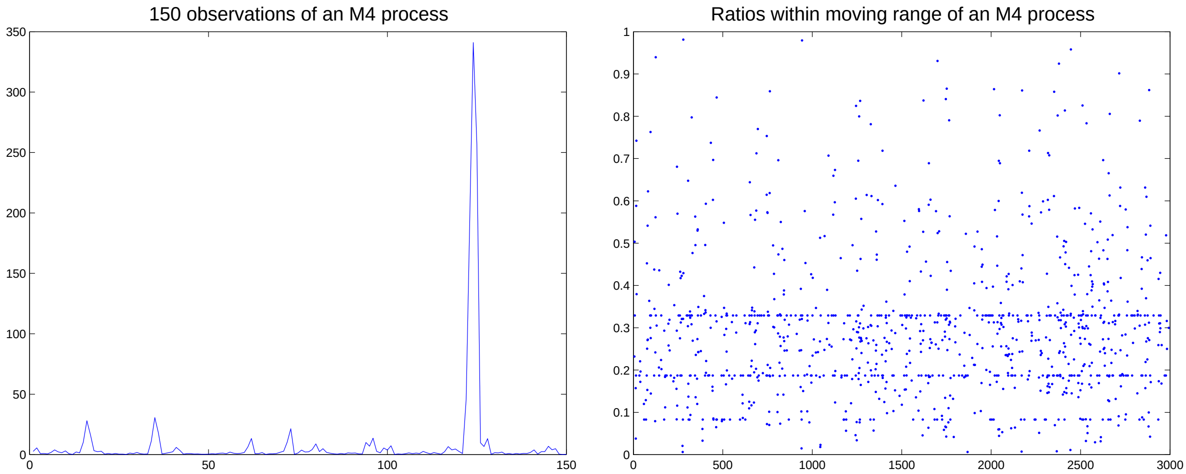

Figure 2.3: A demo of multiple signature patterns.

2.4 Estimation of weight parameters

In the previous section, (2.21) gives estimates of $\frac{a_{l,-K_1+m,d}}{a_{l,-K_1,d}+a_{l,-K_1+1,d}+\cdots+a_{l,K_2,d}}$, not the parameters themselves. We solve this problem in this section. Rewrite the model as

$$\begin{aligned} Y_{id} &= \max_{1\le l\le L}\max_{-K_1\le k\le K_2} a_{l,k,d}Z_{l,i-k} \\ &= \max_{1\le l\le L} b_{ld}\max_{-K_1\le k\le K_2} c_{l,k,d}Z_{l,i-k} \end{aligned} \tag{2.22}$$

where b_{ld} is the weight of l's signature pattern and such that $\sum_l b_{ld} = 1$ and $\sum_k c_{l,k,d} = 1$ for each l and d.

It is easy to show $P(Y_{1d} \le y_{1d}) = e^{-\frac{1}{y_{1d}}}$ from (1.18). As mentioned in section 2.1 our goal is to approximate the distribution function and from the approximation we obtain estimates of all parameters. Since the univariate distribution studied here does not relate any parameters to the distribution function, we seek a jointly k-variate distribution function approximation. Throughout this work we will consider $k = 2$ only because the cases of $k > 2$ can be generalized from the case $k = 2$. First we have,

$$P(Y_{1d} \le y_{1d}, Y_{2d} \le y_{2d}) = \exp[-\sum_{l=1}^{L} b_{ld}\sum_{m=1-K_2}^{2+K_1}\max(\frac{c_{l,1-m,d}}{y_{1d}}, \frac{c_{l,2-m,d}}{y_{2d}})] \tag{2.23}$$

where $c_{l,K_2+1,d} = 0, c_{l,-K_1-1,d} = 0$

Under the model (2.22) all $c_{l,k,d}$ can be estimated by looking into

$$P(\frac{Y_{t+m,d}}{Y_{td} + Y_{t+1,d} + \cdots + Y_{t+K_2+K_1,d}} = c_{l,-K_1+m,d}, \quad i.o.) = 1 \tag{2.24}$$

and therefore we only need to estimate b_{ld}.

It is known that an empirical process approximates the random process from which observations are obtained. Let A be any subset of $\mathbb{R}_+^2 = (0,\infty)\times(0,\infty)$ and define

$$X_{id} = I_A(Y_{id}, Y_{i+1,d}).$$

Let μ_d be the mean of X_{id}, then

$$E(X_{id}) = P((Y_{id}, Y_{i+1,d}) \in A) = \mu_d$$

$$Var(X_{id}) = E(X_{id}^2) - (E(X_{id}))^2 = \mu_d - \mu_d^2$$

By appropriately choosing A, we can construct parameter estimators.

2.4.1 Estimation using independent observed values from a dependent sequence

Now let $A_{1d} = (0, x_{1d}) \times (0, x'_{1d}), \cdots, A_{L-1,d} = (0, x_{L-1,d}) \times (0, x'_{L-1,d})$ be different and define

$$\bar{X}_{A_{jd}} = \frac{1}{n}\sum_{i=1}^{n} I_{A_{jd}}(Y'_{id}, Y'_{i+1,d}) \tag{2.25}$$

where $(Y'_{id}, Y'_{i+1,d})$ are i.i.d pairs taken from an M-dependent time series, which we study in this work. Then SLLN implies

$$\bar{X}_{A_{jd}} \xrightarrow{a.s.} P(A_{jd}) = P(Y_{1d} \le x_{jd}, Y_{2d} \le x'_{jd}). \tag{2.26}$$

Now let

$$\exp[-\sum_{l=1}^{L}\widehat{b}_{ld}\sum_{m=1-K_2}^{2+K_1}\max(\frac{c_{l,1-m,d}}{x_{jd}}, \frac{c_{l,2-m}}{x'_{jd}})] = \bar{X}_{A_{jd}}, \quad j = 1, \cdots, L-1 \tag{2.27}$$

then we can construct parameter estimators from solving D systems of linear equations

$$\begin{cases} \sum_{l=1}^{L}\widehat{b}_{ld}\sum_{m=1-K_2}^{2+K_1}\max(\frac{c_{l,1-m,d}}{x_{1d}}, \frac{c_{l,2-m,d}}{x'_{1d}}) & = & -\log(\bar{X}_{A_{1d}}) \\ & \vdots & \\ \sum_{l=1}^{L}\widehat{b}_{ld}\sum_{m=1-K_2}^{2+K_1}\max(\frac{c_{l,1-m,d}}{x_{L-1,d}}, \frac{c_{l,2-m,d}}{x'_{L-1,d}}) & = & -\log(\bar{X}_{A_{L-1,d}}) \\ \sum_{l=1}^{L}\widehat{b}_{ld} = 1 & & \end{cases} \tag{2.28}$$

We choose values of $x_{1d}, x'_{1d}, \cdots, x_{L-1,d}, x'_{L-1,d}$ such that this system of linear equations has unique solution. Since now $c_{l,k,d}$'s are known, we are able to choose values of

$x_{1d}, x'_{1d}, \cdots, x_{L-1,d}, x'_{L-1,d}$ such that the determinant of the system of linear equations is not zero. This may not be true for some special cases, for example when all coefficients are identical we get $\sum_{l=1}^{L} \widehat{b}_{ld} = 1$ for each equation of those L equations in (2.28).

We say two signature patterns, lth and l'th, are identical when the coefficients can be written as

$$b_{ld}(c_{l,-K_1,d}, c_{l,-K_1+1,d}, \cdots, c_{l,K_2,d}),$$

$$b_{l'd}(c_{l,-K_1,d}, c_{l,-K_1+1,d}, \cdots, c_{l,K_2,d}).$$

The identical patterns have the following property:

$$\max(b_{ld} \max_{-K_1 \le k \le K_2} c_{l,k,d} Z_{l,i-k},\ b_{l'd} \max_{-K_1 \le k \le K_2} c_{l',k,d} Z_{l',i-k}) \stackrel{d}{=} (b_{ld} + b_{l'd}) \max_{-K_1 \le k \le K_2} c_{l',k,d} Z'_{l,i-k}$$

where $Z'_{l,i-k}$'s are unit Fréchet. We assume there are no identical patterns in models considered in this section. The results may apply to some cases when there exist identical patterns.

For a specific d, define

$$\Delta_d = \begin{bmatrix} \sum\limits_{m=1-K_2}^{2+K_1} \max(\frac{c_{1,1-m,d}}{x_{1d}}, \frac{c_{1,2-m,d}}{x'_{1d}}) & \cdots & \sum\limits_{m=1-K_2}^{2+K_1} \max(\frac{c_{L,1-m,d}}{x_{1d}}, \frac{c_{L,2-m,d}}{x'_{1d}}) \\ \vdots & \ddots & \vdots \\ \sum\limits_{m=1-K_2}^{2+K_1} \max(\frac{c_{1,1-m,d}}{x_{L-1,d}}, \frac{c_{1,2-m,d}}{x'_{L-1,d}}) & \cdots & \sum\limits_{m=1-K_2}^{2+K_1} \max(\frac{c_{L,1-m,d}}{x_{L-1,d}}, \frac{c_{L,2-m,d}}{x'_{L-1,d}}) \\ 1 & \cdots & 1 \end{bmatrix}.$$

i.e. $|\Delta_d|$ is the determinant of the system of linear equations. Assume now the L determinants of the $(L-1) \times (L-1)$ matrices formed from the bottom $L-1$ rows are not all zero. For fixed x'_{1d}, since $c_{l,k,d}$ are known and $\sum_{m=1-K_2}^{2+K_1} c_{l,i-m,d} = 1,\ i = 1, 2,$ then there exist $x_{\min,d}$ and $x_{\max,d}$ such that when $x_{1d} < x_{\min,d}$ or $x_{1d} > x_{\max,d}$, all elements of first row in Δ_d are $\frac{1}{x_{1d}}$ or $\frac{1}{x'_{1d}}$ respectively. And so when $x_{1d} < x_{\min,d}$ or $x_{1d} > x_{\max,d}$, $|\Delta_d| = 0$. When x_{1d} varies in $[x_{\min,d}, x_{\max,d}]$, denote Δ_d by $\Delta_d(x_{1d})$, then

$$|\Delta_d(x_{1d})| = \frac{1}{x_{1d}} \sum c_{ijd} |\Delta_d|_{1j} + \frac{1}{x'_{1d}} \sum c_{i'j'd} |\Delta_d|_{1j'} \tag{2.29}$$

where $|\Delta_d|_{1j} \neq 0$, $|\Delta_d|_{1j'} \neq 0$ are the $(1, j)$ or $(1, j')$ minors of Δ_d. Both summations are over all non-zero minors of the first row of Δ_d and the corresponding $\frac{c_{ijd}}{x_{1d}}$ or $\frac{c_{i'j'd}}{x'_{1d}}$. If $|\Delta_d(x_{1d})| = 0$, by varying x_{1d} in $[x_{\min,d}, x_{\max,d}]$, at some point x, some $\frac{1}{x_{1d}} c_{ijd} |\Delta_d|_{1j}$ of the summation $\frac{1}{x_{1d}} \sum c_{ijd} |\Delta_d|_{1j}$ change to $\frac{1}{x'_{1d}} c_{i'j'd} |\Delta_d|_{1j'}$ and add to $\frac{1}{x'_{1d}} \sum c_{i'j'd} |\Delta_d|_{1j'}$, or *vice versa*, and this change results in $|\Delta_d(x)| \neq 0$. Hence it cannot be true that

$|\Delta_d| = 0$ for all x_{1d}. This argument can be applied to lower dimension matrices. On the other hand, we can start from a 2×2 matrix and extend it to $L \times L$ matrix such that the determinant is not zero as required. Therefore, there exist constants $x_{1d}, x'_{1d}, \cdots, x_{L-1,d}, x'_{L-1,d}$ such that each system of linear equations (2.28) has a unique solution.

So we get

$$\begin{array}{rcl} \widehat{b}_{ld} & = & \sum_{j=1}^{L-1} \theta_{ljd} \log(\bar{X}_{A_{jd}}) + \text{constant}_{ld} \\ & \xrightarrow{a.s.} & \sum_{j=1}^{L-1} \theta_{ljd} \log(\Pr(A_{jd})) + \text{constant}_{ld} = b_{ld} \end{array} \tag{2.30}$$

for suitable constants θ_{ljd} which are the elements of the inverse of Δ_d . Let

$$\mu_{jd} = E(I_{A_{jd}}(Y'_{1d}, Y'_{2d})) = \Pr(A_{jd})$$

$$\mu_{ijd} = E(I_{A_{id}}(Y'_{1d}, Y'_{2d}) I_{A_{jd}}(Y'_{1d}, Y'_{2d})) = E(I_{A_{id}A_{jd}}(Y'_{1d}, Y'_{2d})) = \Pr(A_{id}A_{jd})$$

then we have well-known asymptotic normality properties.

Proposition 2.5 *If $(Y'_{id}, Y'_{i+1,d})$ are i.i.d pairs, and $\bar{X}_{A_{jd}}$ is defined as in (2.25), then*

(i) $\sqrt{n}(\bar{X}_{A_{jd}} - \mu_{jd}) \xrightarrow{d} N(0, \mu_{jd} - \mu_{jd}^2)$

(ii) $\sqrt{n}(\log(\bar{X}_{A_{jd}}) - \log(\mu_{jd})) \xrightarrow{d} N(0, \frac{\mu_{jd} - \mu_{jd}^2}{\mu_{jd}^2})$

(iii) $\sqrt{n}\left(\begin{bmatrix} \bar{X}_{A_{1d}} \\ \vdots \\ \bar{X}_{A_{L-1,d}} \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_{L-1} \end{bmatrix}\right) \xrightarrow{d} N(0, \Sigma_d)$,

where $\begin{cases} \sigma_{ijd} = \mu_{ijd} - \mu_{id}\mu_{jd} & \text{if } i \neq j, \\ \sigma_{ijd} = \mu_{id} - \mu_{id}^2 & \text{if } i = j. \end{cases}$

The proofs are trivial and can be found in most theoretical statistical books, for example Arnold (1990), Chen (1981). □

Theorem 2.6 *For each l,*

$$\sqrt{n}(\widehat{b}_{ld} - b_{ld}) \xrightarrow{d} N(0, \sigma_{ld}^2)$$

where

$$\sigma_{ld}^2 = (\frac{\theta_{l1d}}{\mu_{1d}}, \cdots, \frac{\theta_{l,L-1,d}}{\mu_{L-1,d}}) \Sigma_d \begin{bmatrix} \frac{\theta_{l1d}}{\mu_{1d}} \\ \vdots \\ \frac{\theta_{l,L-1,d}}{\mu_{L-1,d}} \end{bmatrix}$$

Proof. By the mean-value theorem,

$$\sqrt{n}(\widehat{b}_{ld} - b_{ld}) = \sqrt{n}[\frac{dh}{d\xi}]' \left(\begin{bmatrix} \bar{X}_{A_{1d}} \\ \vdots \\ \bar{X}_{A_{L-1,d}} \end{bmatrix} - \begin{bmatrix} \mu_{1d} \\ \vdots \\ \mu_{L-1,d} \end{bmatrix} \right)$$

where $h(\mu_d) = \sum_{j=1}^{L-1} \theta_{ljd} \log(\mu_{jd})$, $[\frac{dh}{d\mu_d}]' = (\frac{\theta_{l1d}}{\mu_{1d}}, \cdots, \frac{\theta_{l,L-1,d}}{\mu_{L-1,d}})$ and ξ' is between $(\bar{X}_{A_{1d}}, \cdots, \bar{X}_{A_{L-1,d}})$ and $(\mu_{1d}, \cdots, \mu_{L-1,d})$. By the proposition and Slutsky theorem,

$$\begin{aligned} \sqrt{n}h'(\xi)^T \left(\begin{bmatrix} \bar{X}_{A_{1d}} \\ \vdots \\ \bar{X}_{A_{L-1,d}} \end{bmatrix} - \begin{bmatrix} \mu_{1d} \\ \vdots \\ \mu_{L-1,d} \end{bmatrix} \right) & \xrightarrow{d} \ [\tfrac{dh}{d\mu_d}]'Z, \quad Z \sim N(0, \Sigma_d) \\ & \sim \ N(0, [\tfrac{dh}{d\mu_d}]'\Sigma[\tfrac{dh}{d\mu_d}]) \end{aligned}$$

and this completes the proof. □

A generalization of the theorem to the joint distribution of $\widehat{b}_{1d}$, ..., $\widehat{b}_{Ld}$ can be obtained and the proof arguments are similar. We have

Theorem 2.7 $\sqrt{n}(\widehat{\mathbf{b}_d} - \mathbf{b}_d) \xrightarrow{d} N(0, \Theta_d\Sigma_d\Theta_d')$,
where

$$\widehat{\mathbf{b}_d} = \begin{bmatrix} \widehat{b}_{1d} \\ \widehat{b}_{2d} \\ \vdots \\ \widehat{b}_{Ld} \end{bmatrix}, \mathbf{b}_d = \begin{bmatrix} b_{1d} \\ b_{2d} \\ \vdots \\ b_{Ld} \end{bmatrix}, \Theta_d = \begin{bmatrix} \theta_{11d}/\mu_{1d} & \theta_{12d}/\mu_{2d} & \cdots & \theta_{1,L-1,d}/\mu_{L-1,d} \\ \theta_{21d}/\mu_{1d} & \theta_{22d}/\mu_{2d} & \cdots & \theta_{2,L-1,d}/\mu_{L-1,d} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{L,1d}/\mu_{1d} & \theta_{L,2d}/\mu_{2d} & \cdots & \theta_{L,L-1,d}/\mu_{L-1,d} \end{bmatrix}.$$

Note: $\Theta_d\Sigma_d\Theta_d'$ is singular because $\sum_{l=1}^{L} \widehat{b}_{ld} = 1$.

We still need to specify the asymptotic joint distribution of all b_{ld}. Now let

$$\mu_{ijdd'} = E(I_{A_{id}}(Y_{1d}', Y_{2d}')I_{A_{jd'}}(Y_{1d'}', Y_{2d'}')).$$

$$\mathbf{\Sigma} = (\Sigma_{dd'})$$

where each component of $\mathbf{\Sigma}$ is a covariance matrix. $\Sigma_{dd} = \Sigma_d$, $\Sigma_{dd'}(ij) = \mu_{ijdd'} - \mu_{ijd}\mu_{ijd'}$. Then we have the following generalization.

Corollary 2.8 $\sqrt{n}(\widehat{\mathbf{b}} - \mathbf{b}) \xrightarrow{d} N(0, \mathbf{\Theta\Sigma\Theta}')$,
where

$$\widehat{\mathbf{b}} = \begin{bmatrix} \widehat{\mathbf{b}_1} \\ \widehat{\mathbf{b}_2} \\ \vdots \\ \widehat{\mathbf{b}_D} \end{bmatrix}, \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_D \end{bmatrix}, \mathbf{\Theta} = \begin{bmatrix} \Theta_1 & & & \\ & \Theta_2 & & \\ & & \ddots & \\ & & & \Theta_D \end{bmatrix}.$$

2.4.2 Estimation using the whole dependent sequence

In subsection 2.4.1, we were using independent draws from the bivariate distribution of $(Y_{td}, Y_{t+1,d})$ for each d. Since the observed process is a dependent process, independent draws do not contain all information from the data. The estimates may not be accurate. If we use the entire observed process to estimate all parameters, the efficiency of the estimators may be higher and of course it is a more realistic scenario for practical applications.

Like previous subsection, we estimate parameters associated to dth series by the observed values of that series. We now drop the sub-index d from $Y_{td}, a_{l,k,d}$, etc. We write Y_t, a_{lk} only.

We use the same notations (without index d) as in section 2.4.1, for A_js and define

$$\bar{X}_{A_j} = \frac{1}{n}\sum_{i=1}^{n} I_{A_j}(Y_i, Y_{i+1})$$

which means we use the original observations which are an M-dependent sequence ($M = K_1 + K_2 + 1$ here). In order to derive similar results as in section 2.4.1 without loss of data information, we will apply ergodicity. We quote an ergodic theorem here, whose proof can be found in Billingsley (1995).

Let $(\Omega, \mathcal{F}, \mu)$ be a complete probability space and $T : \Omega \rightarrow \Omega$ be a one-to-one onto map such that T and T^{-1} are both measurable: $T^{-1}\mathcal{F} = T\mathcal{F} = \mathcal{F}$. Assume further that $\mu(T^{-1}E) = \mu(E)$ for all $E \in \mathcal{F}$. A map T satisfying these conditions is called a measure-preserving transformation (or m.p.t. for short). The $\mathcal{F}$-set A is *invariant* under T if $T^{-1}A = A$; it is a *nontrivial* invariant set if $0 < \mu(A) < 1$. And T is said *ergodic* if there are no nontrivial invariant sets in $\mathcal{F}$. A measurable function f is invariant if $f(T\omega) = f(\omega)$ for all ω.

Theorem 2.9 The ergodic theorem *Suppose that T is a m.p.t. on $(\Omega, \mathcal{F}, \mu)$ and that f is measurable and integrable. Then*

$$\lim_n \frac{1}{n}\sum_{k=1}^{n} f(T^{k-1}\omega) = \widehat{f}(\omega)$$

with probability 1, where $\widehat{f}$ is invariant and integrable and $E[\widehat{f}] = E[f]$. If T is ergodic, then $\widehat{f} = E[f]$ with probability 1.

If $f = I_A$ and T is ergodic, we have

$$\lim_n \frac{1}{n} \sum_{k=1}^{n} I_A(T^{k-1}\omega) = P(A)$$

with probability 1.

By the mean-value ergodic theorem

$$\bar{X}_{A_j} \xrightarrow{a.s.} E(I_{A_j}(Y_i, Y_{i+1})) = \Pr(A_j)$$

where T is taken to be a Bernoulli shift. And again we apply the same method we had in section 2.4.1, we get

$$\begin{aligned} \widehat{b}_l &= \textstyle\sum_{j=1}^{L-1} \theta_{lj} \log(\bar{X}_{A_j}) + \text{constant}_l \\ &\xrightarrow{a.s.} \textstyle\sum_{j=1}^{L-1} \theta_{lj} \log(\Pr(A_j)) + \text{constant}_l = b_l \end{aligned}$$

In order to study asymptotic normality, we introduce the following proposition which is Theorem 27.4 in Billingsley (1995). First we introduce the so-called α-mixing condition.

For a sequence $Y_1, Y_2, ...$ of random variables, let α_n be a number such that

$$\mid P(A \cap B) - P(A)P(B) \mid \leq \alpha_n$$

for $A \in \sigma(Y_1, ..., Y_k)$, $B \in \sigma(Y_{k+n}, Y_{k+n+1}, ...)$, and $k \geq 1, n \geq 1$. When $\alpha_n \to 0$, the sequence $\{Y_n\}$ is said to be α-mixing. This means that Y_k and Y_{k+n} are approximately independent.

Proposition 2.10 *Suppose that $X_1, X_2, \cdots,$ is stationary and α-mixing with $\alpha_n = O(n^{-5})$ and that $E[X_n] = 0$ and $E[X_n^{12}] < \infty$. If $S_n = X_1 + \cdots + X_n$, then*

$$n^{-1}\, Var[S_n] \to \sigma^2 = E[X_1^2] + 2 \sum_{k=1}^{\infty} E[X_1 X_{1+k}],$$

where the series converges absolutely. If $\sigma > 0$, then $S_n/\sigma\sqrt{n} \xrightarrow{d} N(0,1)$.

Remark: The constants $\alpha_n = O(n^{-5})$ and $E[X_n^{12}] < \infty$ are stronger than necessary as stated in the remark followed Theorem 27.4 in Billingsley (1995) to avoid technical complication in the proof.

Proposition 2.11 *If $\sigma_j > 0$, $\sqrt{n}(\bar{X}_{A_j} - \mu_j) \xrightarrow{d} N(0, \sigma_j^2)$,*
where

$$\sigma_j^2 = \mu_j - \mu_j^2 + 2 \sum_{k=1}^{K_1+K_2+1} (\Pr(Y_1 \leq x_j, Y_2 \leq x_j', Y_{1+k} \leq x_j, Y_{2+k} \leq x_j') - \mu_j^2)$$

Proof. Let $X_n = I_{A_j}(Y_n, Y_{n+1}) - \mu_j$, then $E[X_n] = 0$ and $E[X_n^{12}] < \infty$ because X_n is bounded. And the α-mixing condition is satisfied since Y_n's are M-dependent. So the conditions of Proposition 2.10 are satisfied. Then the proof follows after calculating the following values and applying Proposition 2.10.

$$X_1^2 = I_{A_j}(Y_1, Y_2) - 2\mu_j I_{A_j}(Y_1, Y_2) + \mu_j^2$$

$$EX_1^2 = \mu_j - 2\mu_j^2 + \mu_j^2 = \mu_j - \mu_j^2$$

$$\begin{aligned} X_1 X_{1+k} &= (I_{A_j}(Y_1, Y_2) - \mu_j)(I_{A_j}(Y_{1+k}, Y_{2+k}) - \mu_j) \\ &= I_{A_j}(Y_1, Y_2) I_{A_j}(Y_{1+k}, Y_{2+k}) - \mu_j I_{A_j}(Y_1, Y_2) - \mu_j I_{A_j}(Y_{1+k}, Y_{2+k}) + \mu_j^2 \end{aligned}$$

and

$$E(X_1 X_{1+k}) = \Pr(Y_1 \leq x_j, Y_2 \leq x_j', Y_{1+k} \leq x_j, Y_{2+k} \leq x_j') - \mu_j^2$$

□

Lemma 2.12 $\sqrt{n}\left(\begin{bmatrix} \bar{X}_{A_1} \\ \vdots \\ \bar{X}_{A_{L-1}} \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_{L-1} \end{bmatrix}\right) \xrightarrow{d} N(0, \Sigma + \sum_{k=1}^{K_1+K_2+1}\{W_k + W_k'\})$ *where*
$\sigma_{ij} = \mu_{ij} - \mu_i\mu_j$*, the matrix* W_k *has entries* $w_k^{ij} = \Pr(Y_1 \leq x_i, Y_2 \leq x_i', Y_{1+k} \leq x_j, Y_{2+k} \leq x_j') - \mu_i\mu_j$*,* $\mu_{ii} = \mu_i$*.*

Proof. Let

$$U_1 = (I_{A_1}(Y_1, Y_2) - \mu_1, \cdots, I_{A_{L-1}}(Y_1, Y_2) - \mu_{L-1})',$$

$$U_{1+k} = (I_{A_1}(Y_{1+k}, Y_{2+k}) - \mu_1, \cdots, I_{A_{L-1}}(Y_{1+k}, Y_{2+k}) - \mu_{L-1})',$$

and $\alpha = (\alpha_1, \cdots, \alpha_{L-1})' \neq 0$ be an arbitrary vector.

Let $X_1 = \alpha' U_1, X_2 = \alpha' U_2, \cdots$, then $E[X_n] = 0$ and $E[X_n^{12}] < \infty$. And so Proposition 2.10 can apply. We say expectation are applied on all elements if expectation is applied on a random matrix. But $E[X_1^2] = \alpha' E[U_1 U_1']\alpha = \alpha'\Sigma\alpha$, $E[X_1 X_{1+k}] = \alpha' E[U_1 U_{1+k}']\alpha = \alpha' W_k \alpha$ where

$$E[(I_{A_i}(Y_1, Y_2) - \mu_i)(I_{A_j}(Y_1, Y_2) - \mu_j)] = \mu_{ij} - \mu_i\mu_j$$

$$\begin{aligned} &E[(I_{A_i}(Y_1, Y_2) - \mu_i)(I_{A_j}(Y_{1+k}, Y_{2+k}) - \mu_j)] \\ &\quad = \Pr(Y_1 \leq x_i, Y_2 \leq x_i', Y_{1+k} \leq x_j, Y_{2+k} \leq x_j') - \mu_i\mu_j \end{aligned}$$

So the proof is completed by applying the Cramér-Wold device (see below). □

Cramér-Wold device: (Cambanis and Leadbetter 1994) *Let* $\xi = (\xi_1, \ldots, \xi_D)$, $\xi_n = (\xi_{n1}, \ldots, \xi_{nD})$, $n = 1, 2, \ldots,$ *be random vectors. Then*

$$\xi_n \xrightarrow{d} \xi \quad \text{as } n \to \infty$$

if and only if

$$\alpha_1 \xi_{n1} + \cdots + \alpha_D \xi_{nD} \xrightarrow{d} \alpha_1 \xi_1 + \cdots + \alpha_D \xi_D \quad \text{as } n \to \infty$$

for all $\alpha_1, \ldots, \alpha_D \in \mathbf{R}$.

Theorem 2.13 *For each* l,

$$\sqrt{n}(\widehat{b}_l - b_l) \xrightarrow{d} N(0, \sigma_l^2)$$

where

$$\sigma_l^2 = (\frac{d_{l1}}{\mu_1}, \cdots, \frac{d_{l,L-1}}{\mu_{L-1}})(\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\}) \begin{bmatrix} \frac{d_{l1}}{\mu_1} \\ \vdots \\ \frac{d_{l,L-1}}{\mu_{L-1}} \end{bmatrix}$$

Proof. This follows from lemma 2.12 using the same argument as in the proof of Theorem 2.6. □

A generalization of the theorem to the joint distribution of $\widehat{b}_1$, ..., $\widehat{b}_L$ can be obtained and the proof arguments are similar. We have

Theorem 2.14

$$\sqrt{n}(\widehat{\mathbf{b}} - \mathbf{b}) \xrightarrow{d} N(0, \Theta(\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\})\Theta'),$$

where $\widehat{\mathbf{b}}$, $\mathbf{b}$ *and* Θ *are defined as the same as in Theorem 2.7.*

For all $\widehat{b}_{ld}$, a similar result of Corollary 2.8 can be obtained.

2.5 Results from discrete time Markov chain theory

In order to simplify the notation, we only consider the case of $D = 1$ and $L = 1$ in this section. Now define

$$X_t = I_{A_t},\ t = 0, \pm 1, \pm 2, ...$$

and suppose the sequence $\{X_t\}$ starts at $t = 0$ with $X_{t-k} = 0,\ k = 1, \dots, K_1 + K_2$. If $X_t = 0$, then

$$\begin{aligned}\Pr(X_{t+1} = 0|X_t = 0) &= \Pr(A^c_{t+1}|A^c_t) = \tfrac{\Pr(A^c_t A^c_{t+1})}{\Pr(A^c_t)} = \tfrac{1-\Pr(A_t)-\Pr(A_{t+1})}{1-\Pr(A_t)} \\ &= \tfrac{1-2\Pr(A_t)}{1-\Pr(A_t)} = p\end{aligned}$$

If $X_t = 1$, then

$$\begin{aligned}&\Pr(X_{t+1} = 0|X_t = 1) = 1,\ \Pr(X_{t+2} = 0|X_t = 0) = 1, \dots, \\ &\Pr(X_{t+K_1+K_2+1} = 0|X_{t+K_1+K_2} = 0) = 1.\end{aligned}$$

If we consider all the 0's before the sequence $\{X_t\}$ reaches 1 are different, and similarly for the 0's after reaching 1, then we can construct a Markov chain which has the following transition diagram. Where $b1$ corresponds to A^c_t, $b2$ corresponds to A^c_{t+1},

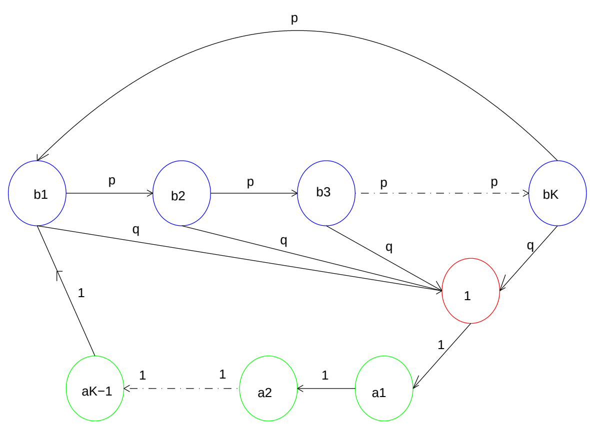

Figure 2.4: Transition Diagram of the Constructed Markov Chain.

and similarly for $b3, \dots, b_K, K = K_1 + K_2 + 1$. Once the chain reaches the state b_K, it moves either to 1 or to b_1. Because $A_{t+K_1+K_2+2}$ is independent of A_t, so we can think the chain return to b_1 and 'restart' again. Once the chain reaches the state 1, it must move at least $K_1 + K_2 + 1$ steps to reach the state 1 again. Or after $K_1 + K_2$ steps, the chain 'restart' again. We have the following transition probability matrix of

$\{b_1, b_2, \ldots, b_K, 1, a_1, a_2, \ldots, a_{K-1}\}$.

$$P = \begin{bmatrix} 0 & p & 0 & \cdots & 0 & q & 0 & 0 & \cdots & 0 \\ 0 & 0 & p & \cdots & 0 & q & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & p & q & 0 & 0 & \cdots & 0 \\ p & 0 & 0 & \cdots & 0 & q & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

Define now $a_1 = 2, a_2 = 3, \ldots, a_{K-1} = K, b_1 = K+1, b_2 = K+2, \ldots, b_K = 2K+1$, and

$$T = \min\{n : X_n = 1\},$$

and

$$u_i(n) = \Pr\{T \leq n | X_0 = i\},\ m_i = E(T|X_0 = i).$$

Then we have standard DTMC results:

$$u_i(n) = P_{i1} + \sum_{j=2}^{2K+1} P_{ij} u_j(n-1),\ i \geq 2,\ n \geq 1,$$

$$u_i(1) = 0,\ i \geq 1,$$

$$m_i = 1 + \sum_{j=2}^{2K+1} P_{ij} m_j,\ i = 2, \ldots, 2K+1.$$

The proofs of these results can be found from books such as Kulkarni (1995). We give an example to show how fast we can get a desired value.

Example 2.3 *Let*

$$Y_t = \max\{.1Z_{t-2},\ .2Z_{t-1},\ .4Z_t,\ .2Z_{t+1},\ .1Z_{t+2}\}$$

then by (2.5) we can have

$$\Pr(A_t) = .0331,\ p = \frac{1 - 2\Pr(A_t)}{1 - \Pr(A_t)} = .9658,\ q = .0342.$$

The value of .0342 approximately tells among 100 independent events, we can get 3 times of the desired value. But to get 100 independent events, we need to have 600 A_ts. Since A_ts are dependent, the number of 600 can be dramatically reduced. In fact from $m_i = E(T|X_0 = i)$, 29 is the mean number of needed A_ts to get the desired value once.

2.6 Approximation of two $M4$ processes

The characterization of extreme observations of a stationary process in Smith and Weissman (1996) are infinite order $M4$ processes. But in practice, it is unrealistic to estimate infinite many parameters. It is natural to apply models with finite number of parameters to real data if the candidate model well approximates the true model.

In this section, we will create conditions under which two processes are arbitrarily close and the following two theorems show that.

Lemma 2.15 *Suppose* $\sum_{-\infty<i<\infty} \alpha_i = 1$, $\sum_{|i|>K} \alpha_i = \delta$, $X = \bigvee_{|i|>K} \alpha_i Z_i$ *and* $Y = \bigvee_{|i|\le K} \alpha_i Z_i$, *where* $\{Z_i\}$ *are i.i.d Fréchet. Let* $Z_\delta = \frac{1}{1-\delta} Y$, *then*

$$\lim_{\delta \to 0} P[|Z_\delta - X \vee Y| > \epsilon] = 0.$$

Proof. It is easy to check that $X,\ Y,\ X \vee Y,\ Z_\delta$ have the distributions:

$$X \sim e^{-\frac{\delta}{x}},\ Y \sim e^{-\frac{1-\delta}{y}},\ X \vee Y \sim e^{-\frac{1}{x}},\ Z_\delta \sim e^{-\frac{1}{z}},$$

$$\begin{aligned} P(X > Y) &= \int_0^\infty P[X > y] \tfrac{1-\delta}{y^2} e^{-\frac{1-\delta}{y}} dy \\ &= \int_0^\infty (1 - e^{-\frac{\delta}{y}}) \tfrac{1-\delta}{y^2} e^{-\frac{1-\delta}{y}} dy \\ &= 1 - (1-\delta) \int_0^\infty \tfrac{1}{y^2} e^{-\frac{1}{y}} dy \\ &= \delta, \end{aligned}$$

$$\begin{aligned} P[Z_\delta - Y > \epsilon] &= P[(\tfrac{1}{1-\delta} - 1) Y > \epsilon] = P[\tfrac{\delta}{1-\delta} Y > \epsilon] \\ &= P[Y > \tfrac{(1-\delta)\epsilon}{\delta}] = 1 - e^{-\frac{(1-\delta)\delta}{(1-\delta)\epsilon}} = 1 - e^{-\frac{\delta}{\epsilon}}. \end{aligned}$$

Now

$$\begin{aligned} P[|Z_\delta - X \vee Y| > \epsilon] &= P[Z_\delta - X \vee Y > \epsilon] + P[X \vee Y - Z_\delta > \epsilon] \\ &= P[Z_\delta - X > \epsilon,\ X > Y] + P[Z_\delta - Y > \epsilon,\ Y > X] \\ &\quad + P[X - Z_\delta > \epsilon,\ X > Y] + P[Y - Z_\delta > \epsilon,\ Y > X] \\ &\le 2P[X > Y] + P[Z_\delta - Y > \epsilon] + 0 \\ &= 2\delta + 1 - e^{-\frac{\delta}{\epsilon}} \end{aligned}$$

which proves the assertion. $\square$

Now pick δ_n so that $P[|Z_\delta - X \vee Y| > \epsilon] \leq 2^{-n}$ for $\delta \leq \delta_n$; then

$$\begin{aligned}
P[\lim_{n\to\infty} Z_{\delta_n} \neq X \vee Y] &= P[|Z_{\delta_n} - X \vee Y| > \epsilon,\ i.o.] \\
&= P[\bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} |Z_{\delta_j} - X \vee Y| > \epsilon] \\
&= \lim_{n\to\infty} P[\bigcup_{j=n}^{\infty} |Z_{\delta_j} - X \vee Y| > \epsilon] \\
&\leq \lim_{n\to\infty} \sum_{j=n}^{\infty} P[|Z_{\delta_j} - X \vee Y| > \epsilon] \\
&\leq \lim_{n\to\infty} 2^{-n+1} = 0
\end{aligned}$$

So $Z_{\delta_n} \xrightarrow{a.s.} X \vee Y$.

Apply this result to a process, we have $Z_{i\delta_n} \xrightarrow{a.s.} X_i \vee Y_i$, each i, and

$$\begin{aligned}
P[\bigcap_{i=1}^{\infty} \lim_{n\to\infty} Z_{i\delta_n} = X_i \vee Y_i] &= 1 - P[\bigcup_{i=1}^{\infty} \lim_{n\to\infty} Z_{i\delta_n} \neq X_i \vee Y_i] \\
&\geq 1 - \sum_{i=1}^{\infty} P[\lim_{n\to\infty} Z_{i\delta_n} \neq X_i \vee Y_i] = 1.
\end{aligned}$$

So

$$P[\bigcup_{i=1}^{\infty} \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} |Z_{i\delta_n} - X_i \vee Y_i| > \epsilon] = 0.$$

We now state the theorem which shows how a finite moving range model arbitrarily closely approximates an infinite range moving process. The proof is just a generalization of the arguments above.

Theorem 2.16 *Suppose* $\sum_{l=1}^{\infty} \sum_{k=-\infty}^{\infty} a_{l,k,d} = 1$, $\sum_{\{lk\} \not\subseteq K} a_{l,k,d} = \delta_d > 0$, *where* $\mathcal{K}$ *is a finite index set.*

$$Y_{id} = \max_l \ \max_k a_{l,k,d} Z_{l,i-k}, \quad d = 1, \cdots, D, \tag{2.31}$$

$$\tilde{Y}_{i\delta_d} = \max_{\{lk\} \subseteq \mathcal{K}} b_{l,k,d} Z_{l,i-k}, \quad d = 1, \cdots, D, \tag{2.32}$$

where $\sum_{l=1}^{L} \sum_{k=-K_1}^{K_2} b_{l,k,d} = 1$ *for* $d = 1, \cdots, D$. *And* $b_{l,k,d} = \frac{1}{1-\delta_d} a_{l,k,d}$ *for* $\{lk\} \subseteq \mathcal{K}$, *then there exist* $\{\delta_{md}\}$, $\delta_{md} \to 0$ *as* $m \to \infty$, *such that*

$$P[\bigcup_{d=1}^{D} \bigcup_{i=-\infty}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} |\tilde{Y}_{i\delta_{md}} - Y_{id}| > \epsilon] = 0.$$

Therefore we conclude $\{\tilde{Y}_{i\delta_d}\} \to \{Y_{id}\}$ *for all* i *and* d *with probability one.*

Chapter 3

Estimation Based on Bivariate Distribution

3.1 Introduction

The methods developed in previous chapter are idealized methods because they assume the model holds exactly. Also the $M4$ process is itself just an approximation to the general max-stable process. The process may be max-stable without being exactly $M4$. In practice we may not be able to estimate the ratios accurately as stated in sections 2.4.1 and 2.4.2, especially when the data are with error. In this chapter, we will develop methods which can be applied to estimate $a_{l,k,d}$ directly. In the previous chapter, the bivariate distribution functions were used to estimate the weight parameters. Here our intention is to determine when the bivariate distribution functions determine all the a_{id}'s, and then to construct estimators. Define $A_{1d}, A_{2d}, \cdots, A_{L\times(K_1+K_2+1)-1,d}$ similarly as we did in the previous chapter and then solve a system of nonlinear equations

$$\begin{cases} \sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1}\max(\frac{\widehat{a}_{l,1-m,d}}{x_{1d}},\frac{\widehat{a}_{l,2-m,d}}{x'_{1d}}) & = & -\log(\bar{X}_{A_{1d}}) \\ \sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1}\max(\frac{\widehat{a}_{l,1-m,d}}{x_{2d}},\frac{\widehat{a}_{l,2-m,d}}{x'_{2d}}) & = & -\log(\bar{X}_{A_{2d}}) \\ & \vdots & \\ \sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1}\max(\frac{\widehat{a}_{l,1-m,d}}{x_{L\times(K_1+K_2+1),d}},\frac{\widehat{a}_{l,2-m,d}}{x'_{L\times(K_1+K_2+1),d}}) & = & -\log(\bar{X}_{A_{L\times(K_1+K_2+1),d}}) \end{cases}$$

Under some conditions, this will give unique solutions which converge to true parameter values for each d, i.e. the bivariate distribution function determines the whole process. We will introduce such conditions and prove some theoretical results. Like Chapter 2, we drop the index d when we deal with a single process or we treat multiple processes separately.

3.2 Modeling time dependence

In this section we mainly focus on the time dependence of each single process and we will not use the index d in sub-sections 3.2.1 and 3.2.2.

3.2.1 Preliminary estimation

First we let $L = 1$ and study the structure of bivariate distribution function

$$P(Y_1 \leq y_1, Y_2 \leq y_2) = \exp[-\sum_{m=1-K_2}^{2+K_1} \max(\frac{a_{1-m}}{y_1}, \frac{a_{2-m}}{y_2})] \tag{3.1}$$

where $a_{K_2+1} = 0, a_{-K_1-1} = 0$. Now define

$$q(x) = a_{-K_1} + \sum_{j=-K_1}^{K_2-1} \max(xa_j, a_{j+1}) + xa_{K_2}, \tag{3.2}$$

then $P(Y_1 \leq 1, Y_2 \leq x) = \exp[-q(x)/x]$.
Define

$$M(x) = \{j : \frac{a_{j+1}}{a_j} > x\} \tag{3.3}$$

where we include $-K_1 - 1 \in M(x)$, $K_2 \in \bar{M}(x)$ complement of $M(x)$ for all $x \in (0, \infty)$. Note that $M(x) \uparrow$ as $x \downarrow$. Then

$$q(x) = x \sum_{j \in \bar{M}(x)} a_j + \sum_{j \in M(x)} a_{j+1}, \tag{3.4}$$

and $q'(x) = \sum_{j \in \bar{M}(x)} a_j$ everywhere except when x is one of $\frac{a_{-K_1+1}}{a_{-K_1}}, \frac{a_{-K_1+2}}{a_{-K_1+1}}, \cdots, \frac{a_{K_2}}{a_{K_2-1}}$. A typical $q(x)$ picture is shown in Figure 3.1.

As $x \to 0$, $q(x) \to \sum_{-K_1}^{K_2} a_j = 1$. For x sufficiently large, $q(x) = a_{-K_1} + x \sum a_j = a_{-K_1} + x$. So if $r_1 < r_2 < \cdots < r_p$ denote the $p \leq K_1 + K_2$ distinct values of $\frac{a_{j+1}}{a_j}$, we can deduce from $q(x)$,

(i) a_{-K_1},

(ii) the values of $r_1, r_2, \cdots, r_p$,

(iii) all sums of the form $\sum_{j \in A(i)} a_j$ and $\sum_{j \in A(i)} a_{j+1}$ where $A(i) = \{j : \frac{a_{j+1}}{a_j} = r_i\}$. This is true because $\sum_{j \in A(i)} a_j$ is just the change in $q'(x)$ at $x = r_i$, while $\sum_{j \in A(i)} a_{j+1} = r_i \sum_{j \in A(i)} a_j$.

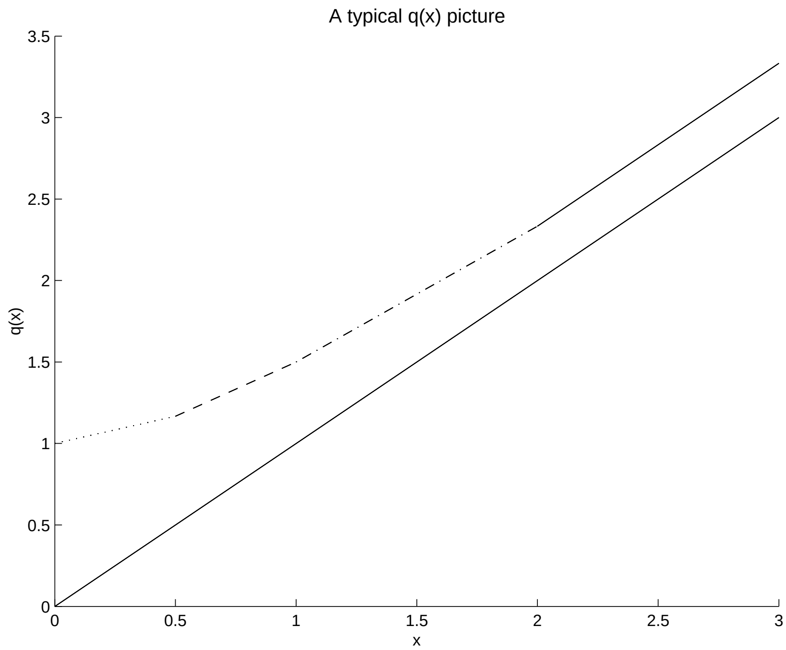

Figure 3.1: A demo of $q(x)$ and its slope $q'(x)$ and ratio change points.

In particular, when all the ratios $\frac{a_{j+1}}{a_j}$ are distinct we can deduce all the r_i's and the values of a_j and $a_{j+1} = r_i a_j$ corresponding to each r_i. However, the model may still be non-identifiable if there are non-trivial permutations of the a_j's that preserve the r_i's.

Proposition 3.1 *If all $\binom{K_1+K_2+1}{2}$ ratios $\frac{a_j}{a_{j'}}$ are distinct, the model is uniquely identified by $q(x)$.*

The reason why Proposition 3.1 is true is that in this case, any permutation of the a_j's must create a new set of values of $r_1, \cdots, r_{K_1+K_2}$.

Remark 1: This justifies statements like "for almost all (w.r.t Lebesgue measure) choices of coefficients $a_{-K_1}, \cdots, a_{K_2}$, the model is identifiable from $q(x)$".

Remark 2: The uniqueness means the values of the vector

$$(a_{K_1},\ a_{-K_1+1},\ \ldots,\ a_{K_2})$$

are uniquely determined. The reason is because we can not simply distinguish the following two processes without further analysis.

$$Y_i = \max(.2Z_{i-1},\ .3Z_i,\ .5Z_{i+1}),$$

$$Y_i' = \max(.2Z_i,\ .3Z_{i+1},\ .5Z_{i+2}).$$

But it should be no ambiguity that we treat them as one model since they have the same joint distribution functions within each sequence.

Since we use the bivariate distribution to construct estimators of parameters, from the previous arguments, if the condition of Proposition 3.1 is false then we may not be able to identify the model. But we may be able to identify the model via some higher-order joint distribution. We now construct an artificial example to demonstrate this idea.

Example 3.1 *This is a counterexample to show a process that is not identifiable via the bivariate joint distribution, but can be identifiable from the trivariate joint distribution. Let* $(a_0, \cdots, a_4) = \frac{1}{6}(1, 1, 2, 1, 1)$ *and* $(b_0, \cdots, b_4) = \frac{1}{6}(1, 2, 1, 1, 1)$*. We consider the two processes generated by the sequences* $a_0, \cdots, a_4$ *and* $b_0, \cdots, b_4$*. Then* $p = 3, r_1 = \frac{1}{2}, r_2 = 1, r_3 = 2$ *for both configurations, so* $q(x)$ *is the same and displayed in Figure 3.1 with*

$$q'(x) = \begin{cases} \frac{1}{6} & 0 < x < \frac{1}{2}, \\ \frac{1}{2} & \frac{1}{2} < x < 1, \\ \frac{5}{6} & 1 < x < 2, \\ 1 & 2 < x \end{cases}$$

i.e. we can't distinguish the a_i*'s from the* b_i*'s on the basis of* $q(x)$*. However, consider the formula*

$$\begin{aligned} \log(\Pr(Y_1 \le y_1, Y_2 \le y_2, Y_3 \le y_3)) \quad = \quad & \tfrac{a_4}{y_1} + \max(\tfrac{a_3}{y_1}, \tfrac{a_4}{y_2}) + \max(\tfrac{a_2}{y_1}, \tfrac{a_3}{y_2}, \tfrac{a_4}{y_3}) \\ & + \max(\tfrac{a_1}{y_1}, \tfrac{a_2}{y_2}, \tfrac{a_3}{y_3}) + \max(\tfrac{a_0}{y_1}, \tfrac{a_1}{y_2}, \tfrac{a_2}{y_3}) \\ & + \max(\tfrac{a_0}{y_2}, \tfrac{a_1}{y_3}) + \tfrac{a_0}{y_3} \end{aligned}$$

and let $y_1 = 1, y_2 = y_3 = c$ *where* $c > 2$*.*
With $\mathbf{a} = \frac{1}{6}(1, 1, 2, 1, 1)$*:*

$$-\log P = \frac{1}{6}[1 + 1 + 2 + 1 + 1 + \frac{1}{c} + \frac{1}{c}] = 1 + \frac{1}{3c},$$

With $\mathbf{a} = \frac{1}{6}(1, 2, 1, 1, 1)$*:*

$$-\log P = \frac{1}{6}[1 + 1 + 1 + 2 + 1 + \frac{2}{c} + \frac{1}{c}] = 1 + \frac{1}{2c},$$

So the two values of $\Pr(Y_1 \le y_1, Y_2 \le y_2, Y_3 \le y_3)$ *are distinct in this case.*

In other words, the two possible models for $\mathbf{a}$ are distinguishable from their trivariate distributions, but not bivariate. However this is a specific example where we need trivariate distribution function, in most cases bivariate distributions are enough.

Now we turn to Case $L > 1$, we have

$$P(Y_1 \leq y_1, Y_2 \leq y_2) = \exp[-\sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1} \max(\frac{a_{l,1-m}}{y_1}, \frac{a_{l,2-m}}{y_2})] \tag{3.5}$$

where $a_{l,K_2+1} = 0, a_{l,-K_1-1} = 0$. And

$$q(x) = \sum_{l=1}^{L}[a_{l,-K_1} + \sum_{j=-K_1}^{K_2-1} \max(xa_{lj}, a_{l,j+1}) + xa_{l,K_2}]. \tag{3.6}$$

And similarly, $q(x)$ is a piecewise linear function and its change points are those adjacent ratios of the coefficients. We have

Proposition 3.2 *If all* $\binom{L\times(K_1+K_2+1)}{2}$ *ratios* $\frac{a_{lj}}{a_{lj'}}$ *are distinct, the model is uniquely identified by* $q(x)$*.*

Proof. . Since all the ratios are different and are points at which $q(x)$ changes slopes or $q'(x)$ has jumps. So based on the jump points of $q(x)$, the ratios of $\frac{a_{l,j+1}}{a_{lj}}$ are uniquely determined. Let's now rewrite (3.6) as

$$q(x) = \sum_{l=1}^{L} b_l[c_{l,-K_1} + \sum_{j=-K_1}^{K_2-1} \max(xc_{lj}, a_{l,j+1}) + xc_{l,K_2}] \tag{3.7}$$

where $\sum_j c_{lj} = 1$ for each l and all c_{lj} are uniquely determine by the ratios. We also write $q(x)$ as

$$q(x) = \sum_{l=1}^{L} xb_l \sum_{m=1-K_2}^{2+K_1} \max(c_{l,1-m}, \frac{c_{l,2-m}}{x})]. \tag{3.8}$$

Suppose now $q(x)$ has a different representation, say

$$q(x) = \sum_{l=1}^{L} xb_l' \sum_{m=1-K_2}^{2+K_1} \max(c_{l,1-m}, \frac{c_{l,2-m}}{x})] \tag{3.9}$$

then

$$\sum_{l=1}^{L}(b_l - b_l') \sum_{m=1-K_2}^{2+K_1} \max(c_{l,1-m}, \frac{c_{l,2-m}}{x})] = 0 \tag{3.10}$$

for all x.

Suppose we have chosen $x_1,\ x_2,\ldots,x_{L-1}$ and formed the matrix

$$\Delta_d = \begin{bmatrix} \sum_{m=1-K_2}^{2+K_1} \max(c_{1,1-m,d}, \frac{c_{1,2-m,d}}{x_1}) & \cdots & \sum_{m=1-K_2}^{2+K_1} \max(c_{L,1-m,d}, \frac{c_{L,2-m,d}}{x_2}) \\ \vdots & \ddots & \vdots \\ \sum_{m=1-K_2}^{2+K_1} \max(c_{1,1-m,d}, \frac{c_{1,2-m,d}}{x_3}) & \cdots & \sum_{m=1-K_2}^{2+K_1} \max(c_{L,1-m,d}, \frac{c_{L,2-m,d}}{x_{L-1}}) \\ 1 & \cdots & 1 \end{bmatrix}$$

and

$$\begin{bmatrix} \sum_{m=1-K_2}^{2+K_1} \max(c_{1,1-m,d}, \frac{c_{1,2-m,d}}{x_1}) & \cdots & \sum_{m=1-K_2}^{2+K_1} \max(c_{L,1-m,d}, \frac{c_{L,2-m,d}}{x_2}) \\ \vdots & \ddots & \vdots \\ \sum_{m=1-K_2}^{2+K_1} \max(c_{1,1-m,d}, \frac{c_{1,2-m,d}}{x_3}) & \cdots & \sum_{m=1-K_2}^{2+K_1} \max(c_{L,1-m,d}, \frac{c_{L,2-m,d}}{x_{L-1}}) \\ 1 & \cdots & 1 \end{bmatrix} \begin{pmatrix} b_1 - b_1' \\ \vdots \\ \vdots \\ b_L - b_L' \end{pmatrix} = 0$$

we can follow the lines after (2.28) and show $|\Delta_d| \neq 0$, then conclude $b_l = b_l'$, all l. So $q(x)$ uniquely determine all $a_{l,j}$. □

3.2.2 Asymptotics for the case of $D = 1$

Now we define

$$\begin{aligned} \sum_{l=1}^{L} \quad [\quad & a_{l,K_2} + \max(a_{l,K_2-1}, \frac{a_{l,K_2}}{x}) + \max(a_{l,K_2-2}, \frac{a_{l,K_2-1}}{x}) \\ + \quad & \max(a_{l,K_2-3}, \frac{a_{l,K_2-2}}{x}) + \cdots + \max(a_{l,-K_1}, \frac{a_{l,-K_1+1}}{x}) + \frac{a_{l,-K_1}}{x}] = b(x) \end{aligned} \tag{3.11}$$

so we have $q(x) = xb(x)$. Let $Y_1, Y_2, \cdots, Y_n$ be observed values and

$$\widehat{b}(x) = -\log(\frac{1}{n}\sum_{i=1}^{n} I_{(Y_i \le 1, Y_{i+1} \le x)}), \tag{3.12}$$

then $\widehat{q}(x) = x\widehat{b}(x)$.

Theorem 3.3 *For each x, we have*

$$\sqrt{n}(\widehat{b}(x) - b(x)) \xrightarrow{d} N(0, \sigma^2)$$

where $\sigma^2 = \frac{\mu_x - \mu_x^2 + 2\sum_{k=1}^{K_1+K_2+1}(\Pr(Y_1 \le 1, Y_2 \le x, Y_{1+k} \le 1, Y_{2+k} \le x) - \mu_x^2)}{\mu_x^2}$, $\mu_x = \Pr(Y_1 \le 1, Y_2 \le x)$.

Proof. Directly apply Proposition 2.5 and 2.10. □

Corollary 3.4 *For each x, we have* $\sqrt{n}(\widehat{q}(x) - q(x)) \xrightarrow{d} N(0, x^2\sigma^2)$

Now suppose the functions $b(x), q(x), \widehat{b}(x), \widehat{q}(x)$ are evaluated at $x = x_1, x_2, \cdots, x_m$, then we have the following theorem.

Theorem 3.5

$$\sqrt{n}(\widehat{\mathbf{b}} - \mathbf{b}) \xrightarrow{d} N(0, \Theta(\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\})\Theta'),$$

where

$$\widehat{\mathbf{b}} = \begin{bmatrix} \widehat{b}(x_1) \\ \widehat{b}(x_2) \\ \vdots \\ \widehat{b}(x_m) \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b(x_1) \\ b(x_2) \\ \vdots \\ b(x_m) \end{bmatrix}, \Theta = \begin{bmatrix} \frac{1}{\mu_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\mu_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\mu_m} \end{bmatrix}.$$

where

$\mu_i = \Pr(Y_1 \leq 1, Y_2 \leq x_i)$, $\mu_{ij} = \Pr(Y_1 \leq 1, Y_2 \leq \min(x_i, x_j))$, $\sigma_{ij} = \mu_{ij} - \mu_i\mu_j$, $w_k^{ij} = \Pr(Y_1 \leq 1, Y_2 \leq x_i, Y_{1+k} \leq 1, Y_{2+k} \leq x_j) - \mu_i\mu_j$, $\mu_{ii} = \mu_i$.

Proof. This follows from Lemma 2.12, Proposition 2.5 and the same arguments in the proof of Theorem 2.7. □

Theorem 3.5 indicates that when the sample size n is sufficiently large, $\widehat{\mathbf{b}} - \mathbf{b}$ is an asymptotically normally distributed random vector. If the points x_is are used to get estimates of a_is, we have the following theorems.

Theorem 3.6 *If all the ratios of parameters in the model (2.22) are different, then (3.11) uniquely determine all a_{lk}.*

Proof. We only prove the case when $L = 1$ and all the ratios are different. Similar proof for $L > 1$ can be done. Now suppose we have different b_i such that

$$\begin{aligned} b_{K_2} &+ \max(b_{K_2-1}, \tfrac{b_{K_2}}{x}) + \max(b_{K_2-2}, \tfrac{b_{K_2-1}}{x}) \\ &+ \max(b_{K_2-3}, \tfrac{b_{K_2-2}}{x}) + \cdots + \max(b_{-K_1}, \tfrac{b_{-K_1+1}}{x}) + \tfrac{b_{-K_1}}{x} = b(x) \end{aligned} \tag{3.13}$$

and suppose $\{a_i\}$ are the true values in (3.11). Define

$$r_i = \frac{a_{K_2-i}}{a_{K_2-i-1}}, \quad r_i' = \frac{b_{K_2-i}}{b_{K_2-i-1}}, \ i = 0, 1, \ldots, K_1 + K_2 + 1.$$

Let $\{r_i^*\}$ be a permutation of $\{r_i\}$ such that

$$0 = r_0^* < r_1^* < r_2^* < \cdots < r_{K_1+K_2+1}^* = \infty$$

then when x, y are varying within (r_i^*, r_{i+1}^*),

$$b(y) - b(x) = \frac{a_{-K_1}}{y} - \frac{a_{-K_1}}{x} + \sum_{\frac{a_{k+1}}{a_k} > r_i^*} (\frac{a_k}{y} - \frac{a_k}{x}) \tag{3.14}$$

since $\max(a_k, \frac{a_{k+1}}{x})$ is a_k when $\frac{a_{k+1}}{a_k} < x$, and $\frac{a_{k+1}}{x}$ otherwise.

When x is in (r_i^*, r_{i+1}^*) and y is in (r_{i+1}^*, r_{i+2}^*), then

$$b(y) - b(x) = \frac{a_{-K_1}}{y} - \frac{a_{-K_1}}{x} + a_k - \frac{a_k}{x} + \sum_{\frac{a_{k+1}}{a_k} > r_{i+1}^*} (\frac{a_k}{y} - \frac{a_k}{x}) \tag{3.15}$$

for some k since $\max(a_k, \frac{a_{k+1}}{x}) = a_k$ implies $\max(a_k, \frac{a_{k+1}}{y}) = a_k$ and there exists a k such that $\max(a_k, \frac{a_{k+1}}{x}) = \frac{a_k}{x}$ and $\max(a_k, \frac{a_{k+1}}{y}) = a_k$.

Let $\{r'_{*i}\}$ be a permutation of $\{r'_i\}$ such that

$$0 = r'_{*0} < r'_{1*} < r'_{2*} < \cdots < r'_{*K_1+K_2+1} = \infty$$

and without loss of generality we assume $r'_{*0} < r_0^*$, then let $x < r_0^*, r'_{*0} < y < r_0^*$, then (3.14) and (3.15) give different $b(y) - b(x)$, and hence we must have $r'_{*0} = r_0^*$ and therefore $r'_{*i} = r_i^*$, i.e. $r'_i = r_i$, all i.

Within $(0, r_0^*)$, we have

$$b(x) = a_{K_2} + \frac{1}{x} = b_{K_2} + \frac{1}{x}$$

Within $(r_{K_1+K_2}^*, \infty)$, we have

$$b(x) = 1 + \frac{a_{-K_1}}{x} = 1 + \frac{b_{-K_1}}{x}$$

which gives $a_{-K_1} = b_{-K_1}, a_{K_2} = b_{K_2}$.

Within$(r_{K_1+K_2-1}^*, r_{K_1+K_2}^*)$, we have

$$\begin{aligned} b(y) - b(x) &= \tfrac{a_{-K_1}}{y} - \tfrac{a_{-K_1}}{x} + \sum_{\frac{a_{k+1}}{a_k} > r_{K_1+K_2-1}^*} (\tfrac{a_k}{y} - \tfrac{a_k}{x}) \\ &= \tfrac{b_{-K_1}}{y} - \tfrac{b_{-K_1}}{x} + \sum_{\frac{b_{k+1}}{b_k} > r_{K_1+K_2-1}^*} (\tfrac{b_k}{y} - \tfrac{b_k}{x}) \end{aligned}$$

which gives $a_i = b_j$ for $i \neq -K_1, j \neq -K_1$, inductively, we have $a_i = b_j = a'_i$, but by Proposition 3.1, we have all $a_i = b_i$. So the proof is completed. $\square$

Remark: the conditions of Proposition 3.1 are stronger than necessary and can be weakened.

Now let $Y_1, Y_2, \cdots, Y_n$ be observed values and

$$\widehat{b}(x) = -\log(\frac{1}{n}\sum_{i=1}^{n} I_{(Y_i \le 1, Y_{i+1} \le x)})$$

which we expect to get $(K_1 + K_2)\ \widehat{r}_i$ such that (3.14) and (3.15) are true when $b(x)$ is replaced by $\widehat{b}(x)$ and a_i are replaced by $\widehat{a}_i$.

By the ergodic theorem, $\widehat{b}(x) \xrightarrow{a.s.} b(x)$ as $n \to \infty$. Now suppose the model is identifiable from $x_1, \ldots, x_m$, where these values are different from the ratios of true parameters as stated in Corollary 3.1, and define $\mathbf{b} = (b(x_1), \ldots, b(x_m))'$, and $\widehat{\mathbf{b}} = (\widehat{b}(x_1), \ldots, \widehat{b}(x_m))'$, then $\widehat{\mathbf{b}} \xrightarrow{a.s.} \mathbf{b}$. Suppose the probability space is $(\Omega, \mathcal{F}, P)$, then there exists an $A \in \mathcal{F}$ such that $\Pr(A) = 1$ and for each $\omega \in A$, $\widehat{\mathbf{b}}_\omega \to \mathbf{b}$.

Notice that a permutation of l and l' will not change the value of $b(x)$ in (3.11). This can be easily seen in the following example.

Example 3.2 *The following two processes*

$$Y_i = \max \begin{bmatrix} 0.05Z_{1,i-1}, & 0.1Z_{1,i}, & 0.03Z_{1,i+1} \\ 0.15Z_{2,i-1}, & 0.2Z_{2,i}, & 0.02Z_{2,i+1} \\ 0.16Z_{3,i-1}, & 0.17Z_{3,i}, & 0.12Z_{3,i+1} \end{bmatrix}$$

$$Y_i' = \max \begin{bmatrix} 0.16Z_{1,i-1}, & 0.17Z_{1,i}, & 0.12Z_{1,i+1} \\ 0.05Z_{2,i-1}, & 0.1Z_{2,i}, & 0.03Z_{2,i+1} \\ 0.15Z_{3,i-1}, & 0.2Z_{3,i}, & 0.02Z_{3,i+1} \end{bmatrix}$$

have the same joint distributions.

Unless we specifically say l and l' are not permutable, otherwise we allow those l's in (3.11) are permutable.

Suppose the solutions of

$$\begin{cases} \sum_{l=1}^{L}[\widehat{a}_{l,K_2} & + \max(\widehat{a}_{l,K_2-1}, \frac{\widehat{a}_{l,K_2}}{x_1}) + \max(\widehat{a}_{l,K_2-2}, \frac{\widehat{a}_{l,K_2-1}}{x_1}) \\ & + \max(\widehat{a}_{l,K_2-3}, \frac{\widehat{a}_{l,K_2-2}}{x_1}) + \cdots + \max(\widehat{a}_{l,-K_1}, \frac{\widehat{a}_{l,-K_1+1}}{x_1}) + \frac{\widehat{a}_{l,-K_1}}{x_1}] = \widehat{b}_\omega(x_1) \\ \quad\cdots & \cdots \\ \sum_{l=1}^{L}[\widehat{a}_{l,K_2} & + \max(\widehat{a}_{l,K_2-1}, \frac{\widehat{a}_{l,K_2}}{x_m}) + \max(\widehat{a}_{l,K_2-2}, \frac{\widehat{a}_{l,K_2-1}}{x_m}) \\ & + \max(\widehat{a}_{l,K_2-3}, \frac{\widehat{a}_{l,K_2-2}}{x_m}) + \cdots + \max(\widehat{a}_{l,-K_1}, \frac{\widehat{a}_{l,-K_1+1}}{x_m}) + \frac{\widehat{a}_{l,-K_1}}{x_m}] = \widehat{b}_\omega(x_m) \end{cases} \tag{3.16}$$

are $\widehat{\mathbf{a}}_\omega$. This is equivalent to $C_{n\omega}\widehat{\mathbf{a}}_\omega = \widehat{\mathbf{b}}_\omega$ in matrix notations where $C_{n\omega}$ is uniquely determined by $\widehat{\mathbf{a}}_\omega$. The elements of $C_{n\omega}$ are either 1, $\frac{1}{x_i}$ or $1+\frac{1}{x_i}$. And the solutions of

$$\begin{cases} \sum_{l=1}^{L}[a_{l,K_2} & + \max(a_{l,K_2-1}, \frac{a_{l,K_2}}{x_1}) + \max(a_{l,K_2-2}, \frac{a_{l,K_2-1}}{x_1}) \\ & + \max(a_{l,K_2-3}, \frac{a_{l,K_2-2}}{x_1}) + \cdots + \max(a_{l,-K_1}, \frac{a_{l,-K_1+1}}{x_1}) + \frac{a_{l,-K_1}}{x_1}] = b(x_1) \\ \cdots & \cdots \\ \sum_{l=1}^{L}[a_{l,K_2} & + \max(a_{l,K_2-1}, \frac{a_{l,K_2}}{x_m}) + \max(a_{l,K_2-2}, \frac{a_{l,K_2-1}}{x_m}) \\ & + \max(a_{l,K_2-3}, \frac{a_{l,K_2-2}}{x_m}) + \cdots + \max(a_{l,-K_1}, \frac{a_{l,-K_1+1}}{x_m}) + \frac{a_{l,-K_1}}{x_m}] = b(x_m) \end{cases} \tag{3.17}$$

are $\mathbf{a}$. And similarly this is equivalent to $C\mathbf{a} = \mathbf{b}$. Since $\widehat{\mathbf{b}}_\omega \to \widehat{\mathbf{b}}$, which implies the estimated ratios converge to the true ratios, then $\widehat{\mathbf{a}}_\omega \to \widehat{\mathbf{a}}$ as $n \to \infty$ by Theorem 3.6 and the assumption that the model is identifiable from $x_1, \ldots, x_m$. But $\max(\widehat{a}_{l,K_2-i}, \frac{\widehat{a}_{l,K_2-i+1}}{x_k})$ converges to $\max(a_{l,K_2-i}, \frac{a_{l,K_2-i+1}}{x_k})$ for each i and k. In other words, x_k remains in intervals $(\frac{\widehat{a}_{l,K_2-i+1}}{\widehat{a}_{l,K_2-i}}, \frac{\widehat{a}_{l,K_2-j+1}}{\widehat{a}_{l,K_2-j}})$ for some i and j when n is sufficiently large. So $C_{n\omega} \to C$ as $n \to \infty$. And so we have proved the following theorem.

Theorem 3.7 *Suppose the model is identifiable from $x_1, \ldots, x_m$, where these values are different from the ratios of true parameters, then the solutions of (3.16) converge to the solutions of (3.17) almost surely. i.e. $\widehat{\mathbf{a}} \xrightarrow{a.s.} \mathbf{a}$ and $C_n \xrightarrow{a.s.} C$.*

Since the elements of both C_n and C are either 1, $\frac{1}{x_i}$ or $1+\frac{1}{x_i}$, then for sufficiently large n, we have $C_n = C$. Bearing this in mind, we have the following multivariate central limit theorem.

Theorem 3.8 *Suppose the model is identifiable from $x_1, \ldots, x_m$, where these values are different from the ratios of true parameters, then*

$$\sqrt{n}(\widehat{\mathbf{a}} - \mathbf{a}) \xrightarrow{d} N(0, B\Theta(\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\})\Theta' B')$$

where $B = (C'C)^{-1}C'$, Θ, Σ and W_k are defined the same as in Proposition 2.5, Theorem 2.7 and Lemma 2.12.

Proof. Since $(C_n'C_n)^{-1}C_n' \xrightarrow{a.s.} (C'C)^{-1}C'$ so

$$\begin{aligned} \sqrt{n}(\widehat{\mathbf{a}} - \mathbf{a}) &= \sqrt{n}((C_n'C_n)^{-1}C_n'\widehat{\mathbf{b}} - (C'C)^{-1}C'\mathbf{b}) \\ &= \sqrt{n}(C_n'C_n)^{-1}C_n'(\widehat{\mathbf{b}} - \mathbf{b}) + \sqrt{n}((C_n'C_n)^{-1}C_n' - (C'C)^{-1}C')\mathbf{b} \\ &\xrightarrow{d} N(0, B\Theta(\Sigma + \textstyle\sum_{k=1}^{K_1+K_2+1}\{W_k + W_k'\})\Theta' B'). \end{aligned}$$

and hence this proves the theorem. □

3.2.3 Asymptotics for a special case of $L = 1$ and $D > 1$

For all $\widehat{a}_{l,k,d}$, a similar asymptotic joint normal distribution result of Corollary 2.8 can be obtained if the permutability of index l in each d are not an issue. But this is not the case in general, so we will not give the asymptotic joint distribution here based on the estimates obtained one component at a time since we didn't simultaneously estimate $\widehat{a}_{l,k,d}$. We illustrate a special case here under which we can identify all parameters based on estimating each single process and putting all estimates together. We will discuss why this special case can not be generalized in section 3.3. Also in section 3.3 a method which can simultaneously estimate all parameters for the cases of $L > 1$ and $D > 1$ is proposed.

We now consider the model

$$Y_{id} = \max_{-K_1 \le k \le K_2} a_{k,d} Z_{i-k}, \; d = 1, \ldots, D \tag{3.18}$$

where $\sum_k a_{k,d} = 1, a_{k,d} \ge 0$ for each d.

By Proposition 3.1,

$$\begin{aligned}[a_{K_2,d} &+ \max(a_{K_2-1,d}, \frac{a_{K_2,d}}{x}) + \max(a_{K_2-2,d}, \frac{a_{K_2-1,d}}{x}) \\ &+ \max(a_{K_2-3,d}, \frac{a_{K_2-2,d}}{x}) + \cdots + \max(a_{-K_1,d}, \frac{a_{-K_1+1,d}}{x}) + \frac{a_{-K_1,d}}{x}] = b_d(x)\end{aligned} \tag{3.19}$$

uniquely determines $a_{k,d}$ for each d when the values of $b_d(x)$ are given. But we just can not simply put all values obtained from (3.19) and form (3.18) because for some d, (3.19) may give different vector values of

$$(a_{-K_1,d}, \; a_{-K_1+1,d}, \; \ldots, \; a_{K_2,d}),$$

for example when $K_1 + K_2 + 1 = 4$ we may get something like

$$(0, \; .2, \; .3, \; .5) \;\text{ or } (.2, \; .3, \; .5, \; 0).$$

But their functions in (3.18)are different and will result in a different multivariate joint distribution.

The following proposition shows that under certain conditions we can simply put the solutions of (3.19) for each d together and then those solutions uniquely determine the true model (3.18).

Proposition 3.9 *Suppose* $a_{-K_1,d} > 0,\ a_{K_2,d} > 0$ *for all* d*, and all ratios* $\frac{a_{j+1,d}}{a_{jd}}$ *are distinct for each* d*, then*

$$\begin{aligned}
[a_{K_2,d} &+ \max(a_{K_2-1,d}, \frac{a_{K_2,d}}{x}) + \max(a_{K_2-2,d}, \frac{a_{K_2-1,d}}{x}) \\
&+ \max(a_{K_2-3,d}, \frac{a_{K_2-2,d}}{x}) + \cdots + \max(a_{-K_1,d}, \frac{a_{-K_1+1,d}}{x}) + \frac{a_{-K_1,d}}{x}] = b_d(x) \\
& d = 1, \ldots, D
\end{aligned} \tag{3.20}$$

uniquely determine the matrix

$$\begin{bmatrix} a_{-K_1,1}, & a_{-K_1+1,1}, & \ldots, & a_{K_2,1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{-K_1,D}, & a_{-K_1+1,D}, & \ldots, & a_{K_2,D} \end{bmatrix} \tag{3.21}$$

Furthermore, there exist points

$$x_{1d},\ x_{2d},\ \ldots,\ x_{md},\ d = 1, \ldots, D$$

such that

$$\begin{aligned}
[a_{K_2,d} &+ \max(a_{K_2-1,d}, \frac{a_{K_2,d}}{x_{jd}}) + \max(a_{K_2-2,d}, \frac{a_{K_2-1,d}}{x_{jd}}) \\
&+ \max(a_{K_2-3,d}, \frac{a_{K_2-2,d}}{x_{jd}}) + \cdots + \max(a_{-K_1,d}, \frac{a_{-K_1+1,d}}{x_{jd}}) + \frac{a_{-K_1,d}}{x_{jd}}] = b_d(x_{jd}) \\
& d = 1, \ldots, D
\end{aligned} \tag{3.22}$$

uniquely determine the matrix in (3.21).

The proof of this proposition is obvious by noticing that $a_{-K_1,d} > 0,\ a_{K_2,d} > 0$ for all d.

Now let $A_{jd} = (0, 1) \times (0, x_{jd})$, for $j = 1, \ldots, m,\ d = 1, \ldots, D$ and define

$$\bar{X}_{A_{jd}} = \frac{1}{n} \sum_{i=1}^{n} I_{A_{jd}}(Y_{id}, Y_{i+1,d}), \tag{3.23}$$

$$\mu_{jd} = E\{I_{A_{jd}}(Y_{id},\ Y_{i+1,d})\} = P(Y_{1d} \le 1, Y_{2d} \le x_{jd}), \tag{3.24}$$

$$\begin{aligned}
\mu_{jdj'd'} &= E\{I_{A_{jd}}(Y_{id},\ Y_{i+1,d}) I_{A_{j'd'}}(Y_{id'},\ Y_{i+1,d'})\} \\
&= P(Y_{1d} \le 1, Y_{2d} \le x_{jd}, Y_{1d'} \le 1, Y_{2d'} \le x_{j'd'}),
\end{aligned} \tag{3.25}$$

Lemma 3.10 $\sqrt{n} \left(\begin{bmatrix} \bar{X}_{A_{11}} \\ \vdots \\ \bar{X}_{A_{m1}} \\ \bar{X}_{A_{12}} \\ \vdots \\ \bar{X}_{A_{mD}} \end{bmatrix} - \begin{bmatrix} \mu_{11} \\ \vdots \\ \mu_{m1} \\ \mu_{12} \\ \vdots \\ \mu_{mD} \end{bmatrix} \right) \xrightarrow{d} N(0, \Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\})$

where

Σ *has entries* $\sigma_{rs} = \mu_{jdj'd'} - \mu_{jd}\mu_{j'd'}$, *the matrix* W_k *has entries* $w_k^{rs} = P(Y_{1d} \le 1, Y_{2d} \le x_{jd}, Y_{1+k,d'} \le 1, Y_{2+k,d'} \le x_{j'd'}) - \mu_{jd}\mu_{j'd'}$ *for* $d = [\frac{r-1}{m}]$, $j = r - (d-1) \times m$, $d' = [\frac{s-1}{m}]$, $j' = s - (d-1) \times m$.

Proof. This can be done exactly as the proof of Lemma 2.12.

We now state the asymptotic joint distribution of all $\widehat{a}_{k,d}$ in the following theorem.

Theorem 3.11 *Suppose the model is identifiable from* $x_{1d}, \ldots, x_{md}$ *for each* d*, where these values are different from the ratios of true parameters, then under the conditions in Proposition 3.9*

$$\sqrt{n}(\widehat{\mathbf{a}} - \mathbf{a}) \xrightarrow{d} N(0, B\Theta(\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\})\Theta' B')$$

where C_d *is the matrix formed from (3.22) when* d*th model is considered.* $\Theta = diag\{\frac{1}{\mu_{11}}, \ldots, \frac{1}{\mu_{m1}}, \ldots, \frac{1}{\mu_{mD}}\}$, $C = diag\{C_1, C_2, \ldots, C_D\}$, $B = (C'C)^{-1}C'$.

3.2.4 Simulation examples

In this section we perform simulation studies. The first model illustrates a simulated $M4$ process with two signature patterns where each pattern has order of 2 and the second model adds Gaussian noise into the first model.

Example 3.3 *We perform two simulation experiments with the following two processes.*

$$Y_i = \max(.1Z_{1,i-1}, .4Z_{1,i}, .35Z_{2,i-1}, .15Z_{2,i}) \tag{3.26}$$

and

$$Y_i = \max(.1Z_{1,i-1}, .4Z_{1,i}, .35Z_{2,i-1}, .15Z_{2,i}) + N_i \tag{3.27}$$

where $N_i \sim N(0, .01)$ *are i.i.d.*

We plot the ratios $\frac{Y_i}{Y_i + Y_{i+1}}$ for both models. Plots in Figure 3.2 look almost exactly the same. However, when a portion of the plot is magnified, as in Figure 3.3, we can see the difference.

We now apply estimating methods developed in previous sections and list all results in the following tables.

The estimated values are based on a sample of size 10000. The standard deviations are obtained by evaluating the formula in Theorem 3.8 with the true values

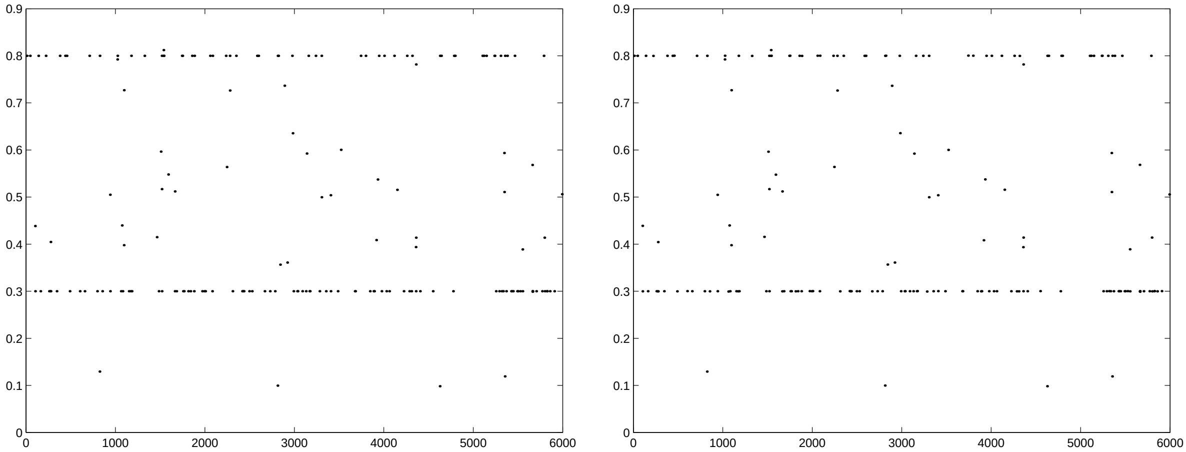

Figure 3.2: The left plot is the ratios of $\frac{Y_i}{Y_i+Y_{i+1}}$ at the threshold level 10 under the model (3.26). The right plot is the ratios of $\frac{Y_i}{Y_i+Y_{i+1}}$ at the threshold level 10 under the model (3.27).

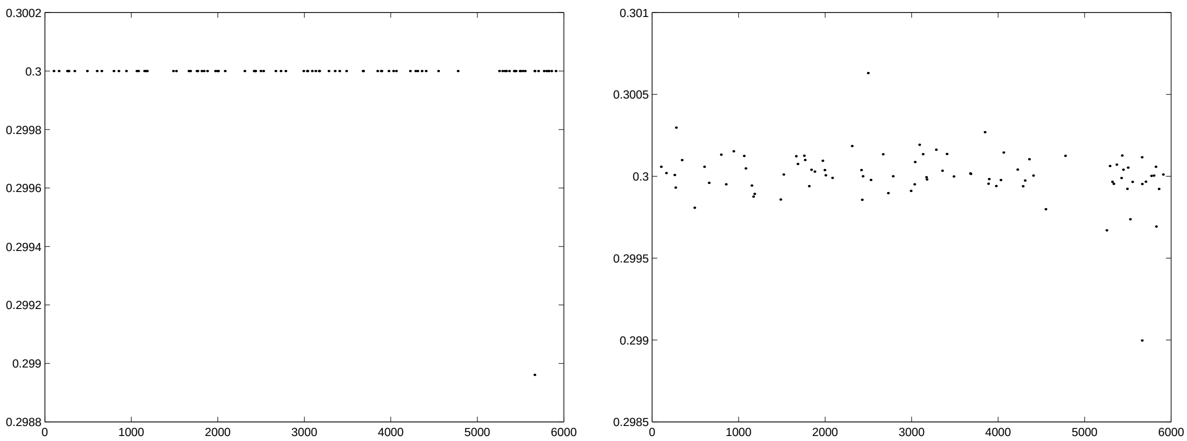

Figure 3.3: The left plot is the ratios around .3 with distance .01 at the threshold level 10 under the model (3.26). The right plot is the ratios .3 with distance .01 at the threshold level 10 under the model (3.27).

Parameter	$a_{1,-1}$	$a_{1,0}$	$a_{2,-1}$	$a_{2,0}$
True value	.1	.4	.35	.15
Estimated value	.1226	.3678	.3747	.1398
Standard Deviation.	.0145	.0469	.0513	.0181

Table 3.1: Simulation results for model (3.26). x =(0.3214, 0.6282, 1.0275, 1.3778, 1.6789, 2.4043, 3.5540, 4.5417).

Parameter	$a_{1,-1}$	$a_{1,0}$	$a_{2,-1}$	$a_{2,0}$
True value	.1	.4	.35	.15
Estimated value	.1169	.3599	.3804	.1477
Standard Deviation.	.0138	.0462	.0502	.0197

Table 3.2: Simulation results for model (3.26). x =(0.3214, 0.8279, 1.5283, 2.9791, 4.5417).

Parameter	$a_{1,-1}$	$a_{1,0}$	$a_{2,-1}$	$a_{2,0}$
True value	.1	.4	.35	.15
Estimated value	.1203	.3541	.3919	.1341
Standard Deviation.	.0144	.0466	.0509	.0180

Table 3.3: Simulation results for model (3.27). x =(0.3214, 0.6283, 1.0277, 1.3782, 1.6799, 2.4054, 3.5546, 4.5422).

Parameter	$a_{1,-1}$	$a_{1,0}$	$a_{2,-1}$	$a_{2,0}$
True value	.1	.4	.35	.15
Estimated value	.1172	.3605	.3805	.1461
Standard Deviation.	.0138	.0461	.0501	.0197

Table 3.4: Simulation results for model (3.27). x =(0.3215, 0.8277, 1.5279, 2.9773, 4.5384).

Parameter	$a_{1,-1}$	$a_{1,0}$	$a_{2,-1}$	$a_{2,0}$
True value	.1	.4	.35	.15
Estimated value	0.1401	0.3801	0.3561	0.1353
Standard Deviation.	0.0419	0.0684	0.0697	0.0493

Table 3.5: Simulation results for model (3.26)

Parameter	$a_{1,-1}$	$a_{1,0}$	$a_{2,-1}$	$a_{2,0}$
True value	.1	.4	.35	.15
Estimated value	.1404	.3791	.3572	.1351
Standard Deviation.	.0421	.0668	.0668	.0492

Table 3.6: Simulation results for model (3.27)

approximated by the empirical values. These simulation experiments show that the effectiveness of the estimating procedures proposed.

The estimated values in Tables 3.5 and 3.6 are mean values of estimates based on 100 replications of sample size 10000. The standard deviations are sample standard deviations.

3.3 Modeling temporal and inter-serial dependence

As we mentioned in section 3.2.3 we can't just estimate the coefficient one component at a time and then put them all together to derive the full model for the joint distribution of the multivariate processes, even if each single component process only has one signature pattern. Example 3.2, in section 3.2.2, showed that two different processes have the same distribution functions. The reason is because the coefficients in the second process are permutations of the coefficients of the first process. The permutations are on index l. Proposition 3.2 actually tells that all the values of $a_{l,k}$ are uniquely determined by $b(x)$ when the permutation of index l is allowed. But this is not the case when we have multivariate processes. We use the following artificial bivariate processes to illustrate why this is not the case.

Suppose we have the bivariate processes

$$
\begin{cases}
Y_{i1} & = & \max \begin{bmatrix} 0.05Z_{1,i-1}, & 0.10Z_{1,i}, & 0.03Z_{1,i+1} \\ 0.12Z_{2,i-1}, & 0.16Z_{2,i}, & 0.09Z_{2,i+1} \\ 0.16Z_{3,i-1}, & 0.17Z_{3,i}, & 0.12Z_{3,i+1} \end{bmatrix} \\
\\
Y_{i2} & = & \max \begin{bmatrix} 0.10Z_{1,i-1}, & 0.13Z_{1,i}, & 0.17Z_{1,i+1} \\ 0.07Z_{2,i-1}, & 0.04Z_{2,i}, & 0.03Z_{2,i+1} \\ 0.11Z_{3,i-1}, & 0.12Z_{3,i}, & 0.23Z_{3,i+1} \end{bmatrix}
\end{cases} \tag{3.28}
$$

Suppose we have the observed values $\{Y_{i1},\ Y_{i2}\}$ and get estimates based on the methods developed for single component process.

A possible representation or estimation of process $\{Y_{i1}\}$ could be

$$
Y'_{i1} = \max \begin{bmatrix} 0.16Z_{1,i-1}, & 0.17Z_{1,i}, & 0.12Z_{1,i+1} \\ 0.05Z_{2,i-1}, & 0.10Z_{2,i}, & 0.03Z_{2,i+1} \\ 0.12Z_{3,i-1}, & 0.16Z_{3,i}, & 0.09Z_{3,i+1} \end{bmatrix} \tag{3.29}
$$

and a possible representation or estimation of process $\{Y_{i2}\}$ could be

$$
Y'_{i2} = \max \begin{bmatrix} 0.10Z_{1,i-1}, & 0.13Z_{1,i}, & 0.17Z_{1,i+1} \\ 0.11Z_{2,i-1}, & 0.12Z_{2,i}, & 0.23Z_{2,i+1} \\ 0.07Z_{3,i-1}, & 0.04Z_{3,i}, & 0.03Z_{3,i+1} \end{bmatrix} \tag{3.30}
$$

Note: we used the exact coefficients of original processes in these two representations, in real situation this may not be the case. What we do here is just for illustration.

Now if we put the two estimated processes together, we have

$$
\begin{cases}
Y'_{i1} & = & \max \begin{bmatrix} 0.16Z_{1,i-1}, & 0.17Z_{1,i}, & 0.12Z_{1,i+1} \\ 0.05Z_{2,i-1}, & 0.10Z_{2,i}, & 0.03Z_{2,i+1} \\ 0.12Z_{3,i-1}, & 0.16Z_{3,i}, & 0.09Z_{3,i+1} \end{bmatrix} \\
\\
Y'_{i2} & = & \max \begin{bmatrix} 0.10Z_{1,i-1}, & 0.13Z_{1,i}, & 0.17Z_{1,i+1} \\ 0.11Z_{2,i-1}, & 0.12Z_{2,i}, & 0.23Z_{2,i+1} \\ 0.07Z_{3,i-1}, & 0.04Z_{3,i}, & 0.03Z_{3,i+1} \end{bmatrix}
\end{cases} \tag{3.31}
$$

It is obvious $\{Y_{i1}\}$ and $\{Y'_{i1}\}$ have the same joint distributions, $\{Y_{i2}\}$ and $\{Y'_{i2}\}$ have the same distributions, but $\{(Y_{i1},\ Y_{i2})\}$ and $\{(Y'_{i1},\ Y'_{i2})\}$ don't have the same joint distributions. This can be seen from Figure 3.4.

3.3.1 Inter-serial dependence

We now consider modeling spatial dependence of multivariate time series. We use a similar structure as we used for modeling time dependence, see section 3.2. What we

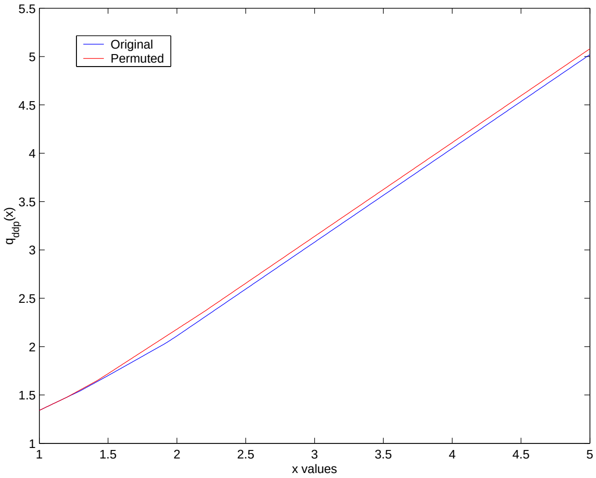

Figure 3.4: A demo of two different bivariate processes. The blue curve are drawn from the original process. The red curve are drawn from the permuted process. $q_{ddp}(x) = -x\log(P(Y_{11} \leq 1, Y_{12} \leq x))$.

do in this section is to estimate parameters based on the joint distribution of a pair of random sequence $\{(Y_{id}, Y_{id'})\}$, where $d \neq d',\ 1 \leq d, d' \leq D$.

It is easy to derive that

$$\Pr\{Y_{1d} \leq y_{1d}, Y_{1d'} \leq y_{1d'}\} = \exp[-\sum_{l=1}^{L}\sum_{m=1-K_2}^{1+K_1} \max(\frac{a_{l,1-m,d}}{y_{1,d}}, \frac{a_{l,1-m,d'}}{y_{1,d'}})]$$

and simply we have

$$\Pr\{Y_{1d} \leq 1, Y_{1d'} \leq x\} = \exp[-\sum_{l=1}^{L}\sum_{m=1-K_2}^{1+K_1} \max(a_{l,1-m,d}, \frac{a_{l,1-m,d'}}{x})].$$

Define

$$\begin{aligned} b_{dd'}(x) &= \sum_{l=1}^{L}\sum_{m=1-K_2}^{1+K_1} \max(a_{l,1-m,d}, \frac{a_{l,1-m,d'}}{x}), \\ q_{dd'}(x) &= x b_{dd'}(x), \\ \widehat{b}_{dd'}(x) &= -\log(\frac{1}{n}\sum_{i=1}^{n} I_{(Y_{id}\leq 1,\ Y_{id}\leq x)}), \\ \widehat{q}_{dd'}(x) &= x\widehat{b}_{dd'}(x), \end{aligned} \tag{3.32}$$

then it's obvious that as $n \to \infty$

$$\widehat{b}_{dd'}(x) \xrightarrow{a.s.} b_{dd'}(x),\ \widehat{q}_{dd'}(x) \xrightarrow{a.s.} q_{dd'}(x).$$

Like section 3.2, we assume now all ratios $\frac{a_{l,k,d'}}{a_{l,k,d}}$ are distinct for all l and k. Then $b_{dd'}(x)$ or $q_{dd'}(x)$ can uniquely determine all ratios $\frac{a_{l,k,d'}}{a_{l,k,d}}$ since these ratios are the jump points of piecewise linear function of $q_{dd'}(x)$, but $b_{dd'}(x)$ or $q_{dd'}(x)$ can not uniquely determine all $a_{l,k,d}$ and $a_{l,k,d'}$ since $b_{dd'}(x)$ doesn't distinguish the index k and the index l.

3.3.2 Temporal and inter-serial dependence

We now combine (3.11) and (3.32) together as a system of nonlinear equations.

$$\begin{cases} b_d(x) &= \sum_{l=1}^{L}[a_{l,K_2,d} + \max(a_{l,K_2-1,d}, \frac{a_{l,K_2,d}}{x}) + \max(a_{l,K_2-2,d}, \frac{a_{l,K_2-1,d}}{x}) \\ & \quad + \max(a_{l,K_2-3,d}, \frac{a_{l,K_2-2,d}}{x}) + \cdots + \max(a_{l,-K_1,d}, \frac{a_{l,-K_1+1,d}}{x}) + \frac{a_{l,-K_1,d}}{x}] \\ b_{d'}(x) &= \sum_{l=1}^{L}[a_{l,K_2,d'} + \max(a_{l,K_2-1,d'}, \frac{a_{l,K_2,d'}}{x}) + \max(a_{l,K_2-2,d'}, \frac{a_{l,K_2-1,d'}}{x}) \\ & \quad + \max(a_{l,K_2-3,d'}, \frac{a_{l,K_2-2,d'}}{x}) + \cdots + \max(a_{l,-K_1,d'}, \frac{a_{l,-K_1+1,d'}}{x}) + \frac{a_{l,-K_1,d'}}{x}] \\ b_{dd'}(x) &= \sum_{l=1}^{L}\sum_{m=1-K_2}^{1+K_1} \max(a_{l,1-m,d}, \frac{a_{l,1-m,d'}}{x}) \end{cases} \tag{3.33}$$

then we will show (3.33) uniquely determine all parameters $a_{l,k,d}$, $a_{l,k,d'}$.

Proposition 3.12 *Suppose all ratios $\frac{a_{l,j,d}}{a_{l,j',d}}$ for all l and $j \neq j'$ are distinct, all ratios $\frac{a_{l,j,d'}}{a_{l,j',d'}}$ for all l and $j \neq j'$ are distinct, and all ratios $\frac{a_{l,k,d}}{a_{l,k,d'}}$ for all l and k are distinct, then (3.33) uniquely determine all values of $a_{l,k,d}$ and $a_{l,k,d'}$.*

Furthermore, there exist points $x_1, x_2, \ldots, x_m$, $m \leq 3L(K_1 + K_2 + 2)$, such that

$$b_d(x_i) \ \ \textit{and} \ \ b_{dd'}(x_i), i = 1, \ldots, m$$

uniquely determine all values of $a_{l,k,d}$ and $a_{l,k,d'}$.

Proof. By Proposition 3.2, $b_d(x)$ and $b_{d'}(x)$ uniquely determine all values of parameters $a_{l,k,d}$ and $a_{l,k,d'}$ respectively. So we can get

$$(a_{l,-K_1,d},\ a_{l,-K_1+1,d}, \ldots, a_{l,K_2,d}),\ \ l = 1, \ldots, L$$

and

$$(a_{l',-K_1,d'},\ a_{l',-K_1+1,d'}, \ldots, a_{l',K_2,d'}),\ \ l' = 1, \ldots, L.$$

Since all ratios $\frac{a_{l,k,d}}{a_{l,k,d'}}$ are distinct, any permutation of index l in $a_{l,k,d'}$ will result in different ratios which will be different from the jump points of $q_{dd'}(x)$, so the jump points of $q_{dd'}(x)$ uniquely determine

$$(\frac{a_{l,-K_1,d}}{a_{l,-K_1,d'}},\ \frac{a_{l,-K_1+1,d}}{a_{l,-K_1+1,d'}},\ \cdots\ ,\ \frac{a_{l,K_2,d}}{a_{l,K_2,d'}})$$

for some l and l'. So (3.33) eventually uniquely determine all the true values of all parameters $a_{l,k,d}$ and $a_{l,k,d'}$.

The reason why $x_1, x_2, \ldots, x_m$ uniquely determine all values of $a_{l,k,d}$ and $a_{l,k,d'}$ is because $q_d(x)$, $q_{d'}(x)$ and $q_{dd'}(x)$ are piecewise linear functions which can be uniquely determined by finite number of points as long as there are at least two points between any two jump points. $\square$

Proposition 3.13 *Suppose all ratios $\frac{a_{l,j,d}}{a_{l,j',d}}$ for all l and $j \neq j'$ are distinct for each $d = 1, \ldots, D$ and $\frac{a_{l,k,1}}{a_{l',k,d'}}$ for all l, l' and k are distinct for each $d' = 2, \ldots, D$, then*

$$\begin{cases} b_d(x) &= \sum_{l=1}^{L}[a_{l,K_2,d} + \max(a_{l,K_2-1,d}, \frac{a_{l,K_2,d}}{x}) + \max(a_{l,K_2-2,d}, \frac{a_{l,K_2-1,d}}{x}) \\ & \quad + \max(a_{l,K_2-3,d}, \frac{a_{l,K_2-2,d}}{x}) + \cdots + \max(a_{l,-K_1,d}, \frac{a_{l,-K_1+1,d}}{x}) + \frac{a_{l,-K_1,d}}{x}], \\ & \qquad d = 1, \ldots, D \\ b_{1d'}(x) &= \sum_{l=1}^{L} \sum_{m=1-K_2}^{1+K_1} \max(a_{l,1-m,1}, \frac{a_{l,1-m,d'}}{x}), \ d' = 2, \ldots, D, \end{cases}$$

uniquely determine all values of $a_{l,k,d}$, $d = 1, \ldots, D$, $l = 1, \ldots, L$, $-K_1 \leq k \leq K_2$.

Furthermore, there exist points $x_1, x_2, \ldots, x_m$, $m \leq (2D-1)(K_1 + K_2 + 1) + 2D$, such that

$$b_d(x_i) \ \ \textit{and} \ \ b_{1d'}(x_i), i = 1, \ldots, m, \ d = 1, \ldots, D, \ d' = 2, \ldots, D$$

uniquely determine all values of $a_{l,k,d}$.

Proof. . This can be done by following the arguments in Proposition 3.12. $\square$

3.3.3 The estimators and asymptotics

Now for suitable choice of

$$x_{1d}, \ x_{2d}, \ \ldots, \ x_{md}, \ d = 1, \ \ldots, \ D,$$

$$x'_{1d'}, \ x'_{2d'}, \ \ldots, \ x'_{m'd'}, \ d' = 2, \ \ldots, \ D,$$

define

$$U_d(x_{jd}) = \frac{1}{n}\sum_{i=1}^{n} I_{(Y_{id} \leq 1, \ Y_{i+1,d} \leq x_{jd})}, \ j = 1, \ldots, m, \ d = 1, \ldots, D,$$

$$\widehat{b}_d(x_{jd}) = -\log(U_d(x_{jd})), \ j = 1, \ldots, m, \ d = 1, \ldots, D,$$

$$U_{1d'}(x'_{jd'}) = \frac{1}{n}\sum_{i=1}^{n} I_{(Y_{i1} \leq 1, \ Y_{id'} \leq x'_{jd'})}, \ j = 1, \ldots, m', \ d' = 2, \ldots, D,$$

$$\widehat{b}_{1d'}(x'_{jd'}) = -\log(U_{1d'}(x'_{jd'})),\ j=1,\ldots,m,\ d'=2,\ldots,D,$$

$$\mathbf{U} = \begin{bmatrix} U_1(x_{11}) \\ U_1(x_{21}) \\ \vdots \\ U_1(x_{m1}) \\ U_2(x_{12}) \\ \vdots \\ U_D(x_{mD}) \\ U_{12}(x'_{12}) \\ U_{12}(x'_{22}) \\ \vdots \\ U_{12}(x'_{m'2}) \\ U_{13}(x'_{13}) \\ \vdots \\ U_{1D}(x'_{m'D}) \end{bmatrix}, \qquad \widehat{\mathbf{b}} = \begin{bmatrix} \widehat{b}_1(x_{11}) \\ \widehat{b}_1(x_{21}) \\ \vdots \\ \widehat{b}_1(x_{m1}) \\ \widehat{b}_2(x_{12}) \\ \vdots \\ \widehat{b}_D(x_{mD}) \\ \widehat{b}_{12}(x'_{12}) \\ \widehat{b}_{12}(x'_{22}) \\ \vdots \\ \widehat{b}_{12}(x'_{m'2}) \\ \widehat{b}_{13}(x'_{13}) \\ \vdots \\ \widehat{b}_{1D}(x'_{m'D}) \end{bmatrix}.$$

Let

$$\begin{aligned} \mu_{djd} = E(U_d(x_{jd})) &= \Pr(Y_{1d} \le 1,\ Y_{2d} \le x_{jd}), \\ &\quad d=1,\ldots,D,\ j=1,\ldots,m. \end{aligned}$$

$$\begin{aligned} \mu_{1d'jd'} = E(U_{1d'}(x'_{j'd'})) &= \Pr(Y_{11} \le 1,\ Y_{1d'} \le x'_{j'd'}), \\ &\quad d'=2,\ldots,D,\ j'=1,\ldots,m'. \end{aligned}$$

$$\begin{aligned} \mu_{djd,\ d'j'd'} &= E[(I_{(Y_{1d}\le 1, Y_{2d}\le x_{jd})} - \mu_{djd})(I_{(Y_{1d'}\le 1, Y_{2d'}\le x_{j'd'})} - \mu_{d'j'd'})] \\ &= \Pr(Y_{1d} \le 1,\ Y_{2d} \le x_{jd}, Y_{1d'} \le 1,\ Y_{2d'} \le x_{j'd'}) - \mu_{djd}\mu_{d'j'd'}, \\ &\quad d,d'=1,\ldots,D,\ j,j'=1,\ldots,m. \end{aligned}$$

$$\begin{aligned} \mu_{djd,\ 1d'j'd'} &= E[(I_{(Y_{1d}\le 1, Y_{2d}\le x_{jd})} - \mu_{djd})(I_{(Y_{11}\le 1, Y_{1d'}\le x'_{j'd'})} - \mu_{1d'j'd'})] \\ &= \Pr(Y_{1d} \le 1,\ Y_{2d} \le x_{jd}, Y_{11} \le 1,\ Y_{1d'} \le x'_{j'd'}) - \mu_{djd}\mu_{1d'j'd'}, \\ &\quad d=1,\ldots,D,\ j=1,\ldots,m, \\ &\quad d'=2,\ldots,D,\ j'=1,\ldots,m'. \end{aligned}$$

$$\begin{aligned} \mu_{1d'j'd',\ djd} &= E[(I_{(Y_{11}\le 1, Y_{1d'}\le x'_{j'd'})} - \mu_{1d'j'd'})(I_{(Y_{1d}\le 1, Y_{2d}\le x_{jd})} - \mu_{djd})] \\ &= \Pr(Y_{1d} \le 1,\ Y_{2d} \le x_{jd}, Y_{11} \le 1,\ Y_{1d'} \le x'_{j'd'}) - \mu_{djd}\mu_{1d'j'd'} \\ &= \mu_{djd,\ 1d'j'd'}, \\ &\quad d=1,\ldots,D,\ j=1,\ldots,m, \\ &\quad d'=2,\ldots,D,\ j'=1,\ldots,m'. \end{aligned}$$

$$\begin{aligned} \mu_{1djd,\ 1d'j'd'} &= E[(I_{(Y_{11}\le 1, Y_{1d}\le x'_{jd})} - \mu_{1djd})(I_{(Y_{11}\le 1, Y_{1d'}\le x'_{j'd'})} - \mu_{1d'j'd'})] \\ &= \Pr(Y_{11} \le 1,\ Y_{1d} \le x'_{jd}, Y_{11} \le 1,\ Y_{1d'} \le x'_{j'd'}) - \mu_{1djd}\mu_{1d'j'd'}, \\ &\quad d=2,\ldots,D,\ j=1,\ldots,m', \\ &\quad d'=2,\ldots,D,\ j'=1,\ldots,m'. \end{aligned}$$

$$
\begin{aligned}
w^{(k)}_{djd,\ d'j'd'} &= E[(I_{(Y_{1d}\le 1, Y_{2d}\le x_{jd})} - \mu_{djd})(I_{(Y_{1+k,d'}\le 1, Y_{2+k,d'}\le x_{j'd'})} - \mu_{d'j'd'})] \\
&= \Pr(Y_{1d}\le 1,\ Y_{2d}\le x_{jd}, Y_{1+k,d'}\le 1,\ Y_{2+k,d'}\le x_{j'd'}) - \mu_{djd}\mu_{d'j'd'}, \\
&\quad d, d' = 1,\ldots,D,\ j, j' = 1,\ldots,m.
\end{aligned}
$$

$$
\begin{aligned}
w^{(k)}_{djd,\ 1d'j'd'} &= E[(I_{(Y_{1d}\le 1, Y_{2d}\le x_{jd})} - \mu_{djd})(I_{(Y_{1+k,1}\le 1, Y_{1+k,d'}\le x'_{j'd'})} - \mu_{1d'j'd'})] \\
&= \Pr(Y_{1d}\le 1,\ Y_{2d}\le x_{jd}, Y_{1+k,1}\le 1,\ Y_{1+k,d'}\le x'_{j'd'}) - \mu_{djd}\mu_{1d'j'd'}, \\
&\quad d = 1,\ldots,D,\ j = 1,\ldots,m, \\
&\quad d' = 2,\ldots,D,\ j' = 1,\ldots,m'.
\end{aligned}
$$

$$
\begin{aligned}
w^{(k)}_{1d'j'd',\ djd} &= E[(I_{(Y_{1,1}\le 1, Y_{1,d'}\le x'_{j'd'})} - \mu_{1d'j'd'})(I_{(Y_{1+k,d}\le 1, Y_{2+k,d}\le x_{jd})} - \mu_{djd})] \\
&= \Pr(Y_{1,1}\le 1,\ Y_{1,d'}\le x'_{j'd'},\ Y_{1+k,d}\le 1,\ Y_{2+k,d}\le x_{jd}) - \mu_{djd}\mu_{1d'j'd'}, \\
&\quad d = 1,\ldots,D,\ j = 1,\ldots,m, \\
&\quad d' = 2,\ldots,D,\ j' = 1,\ldots,m'.
\end{aligned}
$$

$$
\begin{aligned}
w^{(k)}_{1djd,\ 1d'j'd'} &= E[(I_{(Y_{11}\le 1, Y_{1d}\le x'_{jd})} - \mu_{1djd})(I_{(Y_{1+k,1}\le 1, Y_{1+k,d'}\le x'_{j'd'})} - \mu_{1d'j'd'})] \\
&= \Pr(Y_{11}\le 1,\ Y_{1d}\le x'_{jd}, Y_{1+k,1}\le 1,\ Y_{1+k,d'}\le x'_{j'd'}) - \mu_{1djd}\mu_{1d'j'd'}, \\
&\quad d = 2,\ldots,D,\ j = 1,\ldots,m', \\
&\quad d' = 2,\ldots,D,\ j' = 1,\ldots,m'.
\end{aligned}
$$

$$
\mu = \begin{bmatrix} \mu_{111} \\ \mu_{121} \\ \vdots \\ \mu_{1m1} \\ \mu_{212} \\ \vdots \\ \mu_{DmD} \\ \mu_{1212} \\ \mu_{1222} \\ \vdots \\ \mu_{12m'2} \\ \mu_{1313} \\ \vdots \\ \mu_{1Dm'D} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \\ \mu_{m+1} \\ \vdots \\ \mu_{D\times m} \\ \mu_{D\times m+1} \\ \mu_{D\times m+2} \\ \vdots \\ \mu_{D\times m+m'} \\ \mu_{D\times m+m'+1} \\ \vdots \\ \mu_{D\times m+(D-1)m'} \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} b_1(x_{11}) \\ b_1(x_{21}) \\ \vdots \\ b_1(x_{m1}) \\ b_2(x_{12}) \\ \vdots \\ b_D(x_{mD}) \\ b_{12}(x'_{12}) \\ b_{12}(x'_{22}) \\ \vdots \\ b_{12}(x'_{m'2}) \\ b_{13}(x'_{13}) \\ \vdots \\ b_{1D}(x'_{m'D}) \end{bmatrix}.
$$

We have the following relations

$$
\begin{cases}
\mu_{djd} \ \rightarrow\ \mu_r,\ d = [\frac{r-1}{m}] + 1,\ j = r - (d-1)\times m & \text{if } r \le D\times m, \\
\mu_{1d'jd'} \ \rightarrow\ \mu_r,\ d' = [\frac{r-D\times m-1}{m'}] + 2,\ j = r - D\times m - (d-2)\times m' & \text{if } r > D\times m,.
\end{cases}
$$

We now use the similar relations between the indexes of μ_{djd} and the indexes of μ_r define the following variables.

$$\sigma_{rs} = \begin{cases} \mu_{djd,d'j'd'} & \text{if }, r \le D \times m, s \le D \times m \\ \mu_{djd,1d'j'd'} & \text{if }, r \le D \times m, s > D \times m \\ \mu_{1djd,d'j'd'} & \text{if }, r > D \times m, s \le D \times m \\ \mu_{1djd,1d'j'd'} & \text{if }, r > D \times m, s > D \times m. \end{cases}$$

$$w_k^{rs} = \begin{cases} w_{djd,d'j'd'}^{(k)} & \text{if }, r \le D \times m, s \le D \times m \\ w_{djd,1d'j'd'}^{(k)} & \text{if }, r \le D \times m, s > D \times m \\ w_{1djd,d'j'd'}^{(k)} & \text{if }, r > D \times m, s \le D \times m \\ w_{1djd,1d'j'd'}^{(k)} & \text{if }, r > D \times m, s > D \times m. \end{cases}$$

and

$$\Sigma = (\sigma_{rs}),\ W_k = (w_k^{rs}),\ \Theta = (\text{diag}\{\mu\})^{-1}.$$

We now put everything above together. Then we obtain the following lemma. Its proof is just simply following the lines used in lemma 2.12.

Lemma 3.14 *For the choices of $x_{jd},\ x_{j'd'}$ and the definitions of each variables above, we have*

$$\sqrt{n}(\mathbf{U} - \mu) \xrightarrow{d} N(0, \Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\})$$

$$\sqrt{n}(\widehat{\mathbf{b}} - \mathbf{b}) \xrightarrow{d} N(0, \Theta(\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\})\Theta').$$

Now consider the system of non-linear equations

$$\left\{ \begin{array}{rcl} b_d(x_{jd}) & = & \sum_{l=1}^{L}[a_{l,K_2,d} + \max(a_{l,K_2-1,d}, \frac{a_{l,K_2,d}}{x_{jd}}) + \max(a_{l,K_2-2,d}, \frac{a_{l,K_2-1,d}}{x_{jd}}) \\ & & +\max(a_{l,K_2-3,d}, \frac{a_{l,K_2-2,d}}{x_{jd}}) + \cdots + \max(a_{l,-K_1,d}, \frac{a_{l,-K_1+1,d}}{x_{jd}}) + \frac{a_{l,-K_1,d}}{x_{jd}}] \\ & & \quad j = 1, \ldots, m,\ d = 1, \ldots, D \\ b_{1d'}(x'_{j'd'}) & = & \sum_{l=1}^{L} \sum_{m=1-K_2}^{1+K_1} \max(a_{l,1-m,d}, \frac{a_{l,1-m,d'}}{x'_{j'd'}}) \\ & & \quad j' = 1, \ldots, m',\ d' = 2, \ldots, D \end{array} \right. \tag{3.34}$$

and denote the left hand side of (3.34) as $\mathbf{b}$. Since (3.34)uniquely determine the values of all parameters $a_{l,k,d}$, (3.34)has the matrix representation

$$\mathbf{b} = C\mathbf{a} \tag{3.35}$$

or equivalently

$$(C'C)^{-1}C'\mathbf{b} = \mathbf{a}. \tag{3.36}$$

We now obtain our estimators by solving the system of non-linear equations

$$\begin{cases} \widehat{b}_d(x_{jd}) & = & \sum_{l=1}^{L}[\widehat{a}_{l,K_2,d} + \max(\widehat{a}_{l,K_2-1,d}, \frac{\widehat{a}_{l,K_2,d}}{x_{jd}}) + \max(\widehat{a}_{l,K_2-2,d}, \frac{\widehat{a}_{l,K_2-1,d}}{x_{jd}}) \\ & & + \max(\widehat{a}_{l,K_2-3,d}, \frac{\widehat{a}_{l,K_2-2,d}}{x_{jd}}) + \cdots + \max(\widehat{a}_{l,-K_1,d}, \frac{\widehat{a}_{l,-K_1+1,d}}{x_{jd}}) + \frac{\widehat{a}_{l,-K_1,d}}{x_{jd}}] \\ & & \quad j = 1, \ldots, m, \; d = 1, \ldots, D \\ b_{1d'}(x'_{j'd'}) & = & \sum_{l=1}^{L} \sum_{m=1-K_2}^{1+K_1} \max(\widehat{a}_{l,1-m,d}, \frac{\widehat{a}_{l,1-m,d'}}{x'_{j'd'}}) \\ & & \quad j' = 1, \ldots, m', \; d' = 2, \ldots, D \end{cases} \tag{3.37}$$

As n sufficiently large (3.37) can be written as the following matrix representation

$$(C'C)^{-1}C'\widehat{\mathbf{b}} = \widehat{\mathbf{a}}. \tag{3.38}$$

Summarize all arguments above we have obtained the following theorem

Theorem 3.15 *If all ratios $\frac{a_{l,j,d}}{a_{l,j',d}}$ for all l and $j \neq j'$ are distinct for each $d = 1, \ldots, D$ and $\frac{a_{l,k,1}}{a_{l',k,d'}}$ for all l, l' and k are distinct for each $d' = 2, \ldots, D$, of the multivariate processes $\{Y_{id}\}$, then there exist*

$$x_{1d}, \; x_{2d}, \; \ldots, \; x_{md}, \; d = 1, \; \ldots, \; D,$$

$$x'_{1d'}, \; x'_{2d'}, \; \ldots, \; x'_{m'd'}, \; d' = 2, \; \ldots, \; D,$$

such that the estimators $\widehat{\mathbf{a}}$ satisfies

$$\sqrt{n}(\widehat{\mathbf{a}} - \mathbf{a}) \xrightarrow{d} N(0, B\Theta(\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\})\Theta' B')$$

where $B = (C'C)^{-1}C'$.

3.3.4 Simulation examples

We continue the simulation examples used in section 3.2.4 and consider now the bivariate processes

$$\begin{cases} Y_{i1} & = & \max \begin{bmatrix} 0.10Z_{1,i-1}, & 0.40Z_{1,i} \\ 0.35Z_{2,i-1}, & 0.15Z_{2,i} \end{bmatrix} \\ \\ Y_{i2} & = & \max \begin{bmatrix} 0.35Z_{1,i-1}, & 0.15Z_{1,i} \\ 0.10Z_{2,i-1}, & 0.40Z_{2,i} \end{bmatrix} \end{cases} \tag{3.39}$$

We first generate data by simulating these bivariate processes, then based on the simulated data re-estimate all coefficients simultaneously and compute their asymptotic

Parameter	$a_{1,-1,1}$	$a_{1,0,1}$	$a_{2,-1,1}$	$a_{2,0,1}$	$a_{1,-1,2}$	$a_{1,0,2}$	$a_{2,-1,2}$	$a_{2,0,2}$
True value	.1	.4	.35	.15	.35	.15	.1	.4
Estimated value	.1169	.3804	.3599	.1477	.3693	.1303	.1508	.3493
Standard Deviation.	.0123	.0427	.0491	.0128	.0345	.0204	.0226	.0369

Table 3.7: Simulation results for model

covariance matrix. Notice that the coefficients in the second process are the permuted coefficients in the first process. The purpose here is to apply estimators developed in section 3.3.3 and compare the asymptotic standard deviations obtained by Theorem 3.15 with those obtained in section 3.2.4. Table 3.7 is obtained using simulated data with a sample size of 10000. If we compare Table 3.2 with Table 3.7, we find the standard deviations in Table 3.7 are smaller than those in Table 3.2. We think this is because the model in section 3.3.3 uses more data and data information than other models use.

3.4 Weighted least squares estimation

In this section, we assume the model is identifiable from the bivariate distributions. Proposition 3.2 has given sufficient conditions for this.

From (3.4) and finite number of $\{x_1, x_2, \cdots, x_{K_1+K_2}, x_{K_1+K_2+1}\}$ such that

$$x_1 < r_1 < x_2 < r_2 < \cdots < x_{K_1+K_2} < r_{K_1+K_2} < x_{K_1+K_2+1},$$

then the model is identified by $q(x_1), q(x_2), \ldots, q(x_{K_1+K_2+1})$. And we have

Corollary 3.16 *Suppose the model is identifiable w.r.t. bivariate joint distribution. Let $r_0 = 0, x_0 = 0$, $\{x_1, x_2, \cdots, x_m\}$ are m points such that $x_1 < r_1$, $x_m > r_{K_1+K_2}$, and*

$$\min(x_{i+1} - x_i, i = 0, \cdots, m) < \min(r_{i+1} - r_i, i = 0, \cdots, K_1 + K_2)$$

then the model is identified by $q(x_1), q(x_2), \ldots, q(x_m)$.

Under the assumptions in Proposition 3.1, solving a_j's iteratively is equivalent to an optimizing problem

$$\min_{\substack{\sum a_j = 1 \\ a_j > 0}} \sum_{i=1}^{m} (x_i \sum_{j \in \bar{M}(x_i)} a_j + \sum_{j \in M(x_i)} a_{j+1} - q(x_i))^2 \tag{3.40}$$

which gives least squares solutions.

In practice, what we observed is $\widehat{q}(x_i)$, or an estimation of $q(x_i)$, for each i. Based on Corollary 3.4, for sufficiently large n,

$$\widehat{q}(x_i) = q(x_i) + \epsilon_i, \quad i = 1, \cdots, m \tag{3.41}$$

where ϵ_i are normally distributed and correlated with variance-covariance matrix V, a generalization of Corollary 3.4. So the weighted least squares estimates are optimum solutions of

$$\min_{\substack{\sum a_j = 1 \\ a_j > 0}} (\widehat{\mathbf{q}} - \mathbf{q})' V^{-1} (\widehat{\mathbf{q}} - \mathbf{q}). \tag{3.42}$$

where $\mathbf{q} = (q(x_1), \cdots, q(x_m))'$, $\widehat{\mathbf{q}} = (\widehat{q}(x_1), \cdots, \widehat{q}(x_m))'$.

Suppose now $\widehat{\mathbf{a}}$ is a solution of (3.42), if we restrict the parameter space to a neighborhood of $\widehat{\mathbf{a}}$, written $\delta(\widehat{\mathbf{a}})$, then each $q(x_i)$ is a linear combination of a_j's and so there exists a matrix C which depends on $\widehat{\mathbf{a}}$ but can be exactly determined such that

$$\min_{\substack{\sum a_j = 1 \\ a_j > 0, \mathbf{a} \in \delta(\widehat{\mathbf{a}})}} (\widehat{\mathbf{q}} - C\mathbf{a})' V^{-1} (\widehat{\mathbf{q}} - C\mathbf{a}) \tag{3.43}$$

has a solution of $\widehat{\mathbf{a}} = (C'V^{-1}C)^{-1}C'V^{-1}\widehat{\mathbf{q}} = B\widehat{\mathbf{b}}$, where $B = (C'V^{-1}C)^{-1}C'V^{-1}\mathrm{diag}\{x_1, \cdots, x_m\}$, $q(x) = xb(x)$. And then we have a corollary,

Corollary 3.17

$$\sqrt{n}(\widehat{\mathbf{a}} - \mathbf{a}) \xrightarrow{d} N(0, B\Theta(\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\})\Theta' B')$$

Proof. By the same arguments as in the Theorem 3.8. □

3.5 Threshold Methods

In the real world, a time series may not follow the assumed statistical model. But in our applications the tail probability of large observations is the main concern, for example, what is the probability of a big price movement of next day given today's information on the stock market. There has been an extensive development of threshold methods in extreme value statistical research. In this section, we develop a unified procedure for modeling within-cluster behavior at extreme levels by using $M4$ processes to model the temporal dependence of exceedances. We assume those values above certain threshold value u are actually observed and follow the tail distribution of unit Fréchet. Those values below the threshold u are not observed or treated as zero. Our development is focusing on univariate time series data, but without any difficulty it can be easily extended to multivariate time series data with multiple thresholds.

$$\begin{cases} \Pr(Y_i < u + x,\ Y_{i+1} < u) = e^{-\sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1}\max(\frac{a_{l,1-m}}{u+x},\frac{a_{l,2-m}}{u})} \\ \Pr(Y_i < u,\ Y_{i+1} < u + x) = e^{-\sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1}\max(\frac{a_{l,1-m}}{u},\frac{a_{l,2-m}}{u+x})} \end{cases}$$

Let $A_1(x) = (0, u) \times (0, u + x),\ A_2(x) = (0, u + x) \times (0, u)$, and define

$$\bar{X}_{A_j(x)} = \frac{1}{n}\sum_{i=1}^{n} I_{A_j(x)}(Y_i, Y_{i+1}),\ j = 1, 2.$$

then for a fixed u, as $n \to \infty$,

$$\bar{X}_{A_1(x)} \xrightarrow{a.s.} \Pr(Y_i < u,\ Y_{i+1} < u + x) = p_1(x),$$

$$\bar{X}_{A_2(x)} \xrightarrow{a.s.} \Pr(Y_i < u + x,\ Y_{i+1} < u) = p_2(x).$$

We have

$$\begin{cases} \sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1}\max(\frac{a_{l,1-m}}{u+x_1},\frac{a_{l,2-m}}{u}) & = & -\log(p_1(x_1)) \\ \sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1}\max(\frac{a_{l,1-m}}{u},\frac{a_{l,2-m}}{u+x_1}) & = & -\log(p_2(x_1)) \\ & \vdots & \\ \sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1}\max(\frac{a_{l,1-m}}{u+x_m},\frac{a_{l,2-m}}{u}) & = & -\log(p_1(x_m)) \\ \sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1}\max(\frac{a_{l,1-m}}{u},\frac{a_{l,2-m}}{u+x_m}) & = & -\log(p_2(x_m)) \end{cases} \tag{3.44}$$

or equivalently

$$\begin{cases} \sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1}\max(a_{l,1-m},\frac{a_{l,2-m}}{u/(u+x_1)}) & = & -\frac{\log(p_1(x_1))}{u+x_1} \\ \sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1}\max(a_{l,1-m},\frac{a_{l,2-m}}{(u+x_1)/u}) & = & -\frac{\log(p_2(x_1))}{u} \\ & \vdots & \\ \sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1}\max(a_{l,1-m},\frac{a_{l,2-m}}{u/(u+x_m)}) & = & -\frac{\log(p_1(x_m))}{u+x_m} \\ \sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1}\max(a_{l,1-m},\frac{a_{l,2-m}}{(u+x_m)/u}) & = & \frac{\log(p_2(x_m))}{u} \end{cases} \tag{3.45}$$

The left hand side is similar to (3.17) and hence according to Theorem 3.6 and assuming unit Fréchet margin for those observed values at extreme level, $P(Y_{i+1} < u+x,\ Y_i < u)$ and $P(Y_{i+1} < u,\ Y_i < u + x)$ uniquely determine all the parameters since we can take logarithm transformation to get (3.11), so the model is identified from (3.44) or (3.45).

Lemma 3.18 *Let*

$$\zeta_n = (\bar{X}_{A_1(x_1)},\ \bar{X}_{A_2(x_1)},\ \ldots, \bar{X}_{A_1(x_m)},\ \bar{X}_{A_2(x_m)})$$

and

$$\mu = (p_1(x_1),\ p_2(x_1), \ldots, p_1(x_m),\ p_2(x_m))$$

then

$$\sqrt{n}(\zeta_n - \mu) \xrightarrow{d} N(0, \Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\})$$

where the elements of Σ *have the forms*

$$\begin{aligned} &E[(I_{A_i(x_s)}(Y_1,Y_2) - p_i(x_s))(I_{A_j(x_t)}(Y_1,Y_2) - p_j(x_t))] \\ &\quad = EI_{A_i(x_s)\cap A_j(x_t)}(Y_1,Y_2) - p_i(x_s)p_j(x_t) \\ &\quad = p_{ij}(s,t) - p_i(x_s)p_j(x_t) \end{aligned}$$

with $i,j = 1,2;\ s,t = 1,2,\ldots,m,$ *and the elements of* W_k *have the forms*

$$\begin{aligned} &E[(I_{A_i(x_s)}(Y_1,Y_2) - p_i(x_s))(I_{A_j(x_t)}(Y_{1+k},Y_{2+k}) - p_j(x_t))] \\ &\quad = EI_{A_i(x_s)}(Y_1,Y_2)I_{A_j(x_t)}(Y_{1+k},Y_{2+k}) - p_i(x_s)p_j(x_t) \\ &\quad = p_{ij}^k(s,t) - p_i(x_s)p_j(x_t) \end{aligned}$$

with $i,j = 1,2;\ s,t = 1,2,\ldots,m.$

By setting

$$\left\{ \begin{array}{lcl} \sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1} \max(\frac{\widehat{a}_{l,1-m}}{u+x_1}, \frac{\widehat{a}_{l,2-m}}{u}) & = & -\log(\bar{X}_{A_1(x_1)}) \\ \sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1} \max(\frac{\widehat{a}_{l,1-m}}{u}, \frac{\widehat{a}_{l,2-m}}{u+x_1}) & = & -\log(\bar{X}_{A_2(x_1)}) \\ & \vdots & \\ \sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1} \max(\frac{\widehat{a}_{l,1-m}}{u+x_m}, \frac{\widehat{a}_{l,2-m}}{u}) & = & -\log(\bar{X}_{A_1(x_m)}) \\ \sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1} \max(\frac{\widehat{a}_{l,1-m}}{u}, \frac{\widehat{a}_{l,2-m}}{u+x_m}) & = & -\log(\bar{X}_{A_2(x_m)}) \end{array} \right. \tag{3.46}$$

we get

Theorem 3.19 *Suppose the model is identified from* $x_1, x_2, \ldots, x_m$*, where these values satisfy that* $\frac{u}{u+x_i}$ *or* $\frac{u+x_i}{u}$ *are different from the ratios of true parameter, then*

$$\sqrt{n}(\widehat{\mathbf{a}} - \mathbf{a}) \xrightarrow{d} N(0, B\Theta(\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\}))\Theta' B')$$

for $B = (C'C)^{-1}C, \Sigma$ *and* W_k *are defined the same as in Lemma 3.18 and* Θ *is defined by*

$$\Theta = diag(\frac{1}{p_1(x_1)}, \frac{1}{p_2(x_1)}, \ldots, \frac{1}{p_1(x_m)}, \frac{1}{p_2(x_m)})$$

3.6 Multivariate domains of attraction and non unit Fréchet margins

So far the models studied assumed the data follow the $M4$ process exactly. Some natural relaxations of the assumption would be that the distribution of underlying random variables belongs to the domain of attraction of a multivariate extreme value distribution and the marginal distribution of Z's is non unit Fréchet. In the univariate case, from these assumptions, efficient estimating methods based on threshold exceedances and generalized Pareto distribution have been developed. We now develop a similar estimating procedure in the multivariate context. First, we rephrase Theorem 5.4.3 of Galambos (1987) into Lemma 3.20 under the assumption the marginal distribution of bivariate extreme value distribution is unit Fréchet.

Lemma 3.20 *Let* $\mathbf{x} = (x_1,\ x_2)$*,* F *be the population distribution and*

$$F^n(\mathbf{a}_n\mathbf{x} + \mathbf{b}_n) \to H(\mathbf{x}) \tag{3.47}$$

where $H(\mathbf{x})$ *has Fréchet marginals* $H_{2,\xi_j}(x_j),\ j = 1, 2$*. Let* F *has the same univariate marginals* F_1 *and* F_2 *which are eventually strictly increasing, then* F *belongs to the bivariate domains of attraction of* H *if and only if*

$$\frac{1 - F(ux_1,\ ux_2)}{1 - F_1(u)} \to -\log H(x) \tag{3.48}$$

as $u \to \infty$*.*

We now use this lemma to construct estimators for all parameters $a_{l,k,d}$. Assume that F belongs to the bivariate domains of attraction of H which has the distribution

$$H(\mathbf{x}) = \exp[-\sum_{l=1}^{L}\sum_{m=1-K_2}^{2+K_1}\max(\frac{a_{l,1-m}}{x_1}, \frac{a_{l,2-m}}{x_2})] \tag{3.49}$$

Substitute $F(ux_1,\ ux_2)$ and $F_1(u)$ by $\frac{1}{n}\sum_{i=1}^{n} I_{(Y_i \le ux_1,\ Y_{i+1} \le ux_2)}$ and $\frac{1}{n}\sum_{i=1}^{n} I_{(Y_i \le u)}$ respectively, then

$$\Pr\Big\{\lim_{n\to\infty}\frac{1 - \frac{1}{n}\sum_{i=1}^{n} I_{(Y_i \le ux_1,\ Y_{i+1} \le ux_2)}}{1 - \frac{1}{n}\sum_{i=1}^{n} I_{(Y_i \le u)}} = -\log H(x)\Big\} = 1 \tag{3.50}$$

for all u.

From (3.50) we can construct estimation methods for parameters a_{lk}'s. Let $x_1 = 1$, then $-\log H(\mathbf{x}) = b(x_2)$ which was defined in (3.11), i.e.

$$\Pr\Big\{ \lim_{n\to\infty} \frac{1 - \frac{1}{n}\sum_{i=1}^{n} I_{(Y_i \le u,\ Y_{i+1} \le ux)}}{1 - \frac{1}{n}\sum_{i=1}^{n} I_{(Y_i \le u)}} = b(x) \Big\} = 1. \tag{3.51}$$

for all u.

For any fixed u, let $A_u(x) = (0,\ u) \times (0,\ ux)$ and

$$\bar{X}_{A_u(x)} = \frac{1}{n}\sum_{i=1}^{n} I_{A_u(x)}(Y_i,\ Y_{i+1})$$

then

$$\bar{X}_{A_u(x)} \xrightarrow{a.s.} P(Y_1 \le u,\ Y_2 \le ux) = p_u(x)$$

Lemma 3.21 *Let*

$$\zeta_{un} = (\bar{X}_{A_u(\infty)},\ \bar{X}_{A_u(x_1)}, \ldots, \bar{X}_{A_u(x_m)})$$

and

$$\mu_u = (p_u(\infty),\ p_u(x_1), \ldots, (p_u(x_m))$$

then

$$\sqrt{n}(\zeta_{un} - \mu_u) \xrightarrow{d} N(0, \Sigma_u + \sum_{k=1}^{K_1+K_2+1} \{W_{uk} + W'_{uk}\})$$

where the elements of Σ_u *have the forms*

$$\begin{aligned} &E[(I_{A_u(x_s)}(Y_1, Y_2) - p_u(x_s))(I_{A_u(x_t)}(Y_1, Y_2) - p_u(x_t))] \\ &\quad = EI_{A_u(x_s)\cap A_u(x_t)}(Y_1, Y_2) - p_u(x_s)p_u(x_t) \\ &\quad = p_u(x_s \wedge x_t) - p_u(x_s)p_u(x_t) \end{aligned}$$

with $s, t = 1, 2, \ldots, m$*, and the elements of* W_{uk} *have the forms*

$$\begin{aligned} &E[(I_{A_u(x_s)}(Y_1, Y_2) - p_u(x_s))(I_{A_u(x_t)}(Y_{1+k}, Y_{2+k}) - p_u(x_t))] \\ &\quad = EI_{A_u(x_s)}(Y_1, Y_2)I_{A_u(x_t)}(Y_{1+k}, Y_{2+k}) - p_u(x_s)p_u(x_t) \\ &\quad = p_{uk}(s, t) - p_u(x_s)p_u(x_t) \end{aligned}$$

with $s, t = 1, 2, \ldots, m$.

Let

$$b_u(x_i) = \frac{1 - p_u(x_i)}{1 - p_u(\infty)},\ i = 1, \ldots, m$$

then the Jacobian matrix of transforming $p_u(x_i)$ into $b_u(x_i)$ has the form:

$$J = \begin{bmatrix} \frac{1-p_u(x_1)}{(1-p_u(\infty))^2} & -\frac{1}{1-p_u(\infty)} & 0 & \cdots & 0 \\ \frac{1-p_u(x_2)}{(1-p_u(\infty))^2} & 0 & -\frac{1}{1-p_u(\infty)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1-p_u(x_m)}{(1-p_u(\infty))^2} & 0 & 0 & \cdots & -\frac{1}{1-p_u(\infty)} \end{bmatrix} \tag{3.52}$$

so we have

Lemma 3.22 *Let*

$$b_{un} = \left(\frac{1-\frac{1}{n}\sum_{i=1}^{n} I_{(Y_i\le u,\ Y_{i+1}\le ux_1)}}{1-\frac{1}{n}\sum_{i=1}^{n} I_{(Y_i\le u)}}, \dots, \frac{1-\frac{1}{n}\sum_{i=1}^{n} I_{(Y_i\le u,\ Y_{i+1}\le ux_m)}}{1-\frac{1}{n}\sum_{i=1}^{n} I_{(Y_i\le u)}}\right)'$$

$$b_u = (b_u(x_1), \dots, b_u(x_m))'$$

then

$$\sqrt{n}(b_{un} - b_u) \xrightarrow{d} N(0, J(\Sigma_u + \sum_{k=1}^{K_1+K_2+1} \{W_{uk} + W'_{uk}\})J')$$

Let

$$\widehat{b}_u(x) = \frac{1-\frac{1}{n}\sum_{i=1}^{n} I_{(Y_i\le u,\ Y_{i+1}\le ux)}}{1-\frac{1}{n}\sum_{i=1}^{n} I_{(Y_i\le u)}}.$$

Suppose the solutions of

$$\begin{cases} \sum_{l=1}^{L}[\widehat{a}_{l,K_2} & + \max(\widehat{a}_{l,K_2-1}, \frac{\widehat{a}_{l,K_2}}{x_1}) + \max(\widehat{a}_{l,K_2-2}, \frac{\widehat{a}_{l,K_2-1}}{x_1}) \\ & + \max(\widehat{a}_{l,K_2-3}, \frac{\widehat{a}_{l,K_2-2}}{x_1}) + \cdots + \max(\widehat{a}_{l,-K_1}, \frac{\widehat{a}_{l,-K_1+1}}{x_1}) + \frac{\widehat{a}_{l,-K_1}}{x_1}] = \widehat{b}_u(x_1) \\ \cdots & \cdots \\ \sum_{l=1}^{L}[\widehat{a}_{l,K_2} & + \max(\widehat{a}_{l,K_2-1}, \frac{\widehat{a}_{l,K_2}}{x_m}) + \max(\widehat{a}_{l,K_2-2}, \frac{\widehat{a}_{l,K_2-1}}{x_m}) \\ & + \max(\widehat{a}_{l,K_2-3}, \frac{\widehat{a}_{l,K_2-2}}{x_m}) + \cdots + \max(\widehat{a}_{l,-K_1}, \frac{\widehat{a}_{l,-K_1+1}}{x_m}) + \frac{\widehat{a}_{l,-K_1}}{x_m}] = \widehat{b}_u(x_m) \end{cases} \tag{3.53}$$

are $\widehat{\mathbf{a}}_u$. This is equivalent to $\widehat{C}_{un}\widehat{\mathbf{a}}_u = \widehat{\mathbf{b}}_u$ in matrix notations where C_{un} is uniquely determined by $\widehat{\mathbf{a}}_u$. And the solutions of

$$\begin{cases} \sum_{l=1}^{L}[a^u_{l,K_2} & + \max(a^u_{l,K_2-1}, \frac{a^u_{l,K_2}}{x_1}) + \max(a^u_{l,K_2-2}, \frac{a^u_{l,K_2-1}}{x_1}) \\ & + \max(a^u_{l,K_2-3}, \frac{a^u_{l,K_2-2}}{x_1}) + \cdots + \max(a^u_{l,-K_1}, \frac{a^u_{l,-K_1+1}}{x_1}) + \frac{a^u_{l,-K_1}}{x_1}] = b_u(x_1) \\ \cdots & \cdots \\ \sum_{l=1}^{L}[a^u_{l,K_2} & + \max(a^u_{l,K_2-1}, \frac{a^u_{l,K_2}}{x_m}) + \max(a^u_{l,K_2-2}, \frac{a^u_{l,K_2-1}}{x_m}) \\ & + \max(a^u_{l,K_2-3}, \frac{a^u_{l,K_2-2}}{x_m}) + \cdots + \max(a^u_{l,-K_1}, \frac{a^u_{l,-K_1+1}}{x_m}) + \frac{a^u_{l,-K_1}}{x_m}] = b_u(x_m) \end{cases} \tag{3.54}$$

are $\mathbf{a_u}$. And similarly this is equivalent to $C_u\mathbf{a_u} = \mathbf{b_u}$.

And the solutions of

$$\begin{cases} \sum_{l=1}^{L}[a_{l,K_2} & + & \max(a_{l,K_2-1}, \frac{a_{l,K_2}}{x_1}) + \max(a_{l,K_2-2}, \frac{a_{l,K_2-1}}{x_1}) \\ & + & \max(a_{l,K_2-3}, \frac{a_{l,K_2-2}}{x_1}) + \cdots + \max(a_{l,-K_1}, \frac{a_{l,-K_1+1}}{x_1}) + \frac{a_{l,-K_1}}{x_1}] = b(x_1) \\ \quad \cdots & & \cdots \\ \sum_{l=1}^{L}[a_{l,K_2} & + & \max(a_{l,K_2-1}, \frac{a_{l,K_2}}{x_m}) + \max(a_{l,K_2-2}, \frac{a_{l,K_2-1}}{x_m}) \\ & + & \max(a_{l,K_2-3}, \frac{a_{l,K_2-2}}{x_m}) + \cdots + \max(a_{l,-K_1}, \frac{a_{l,-K_1+1}}{x_m}) + \frac{a_{l,-K_1}}{x_m}] = b(x_m) \end{cases} \tag{3.55}$$

are $\mathbf{a}$. And similarly this is equivalent to $C\mathbf{a} = \mathbf{b}$.

The following theorem can be obtained.

Theorem 3.23 *Suppose the model is identified from $x_1, x_2, \ldots, x_m$, where these values are different from the ratios of true parameter, then*

$$\sqrt{n}(\widehat{\mathbf{a}}_u - \mathbf{a}_u) \xrightarrow{d} N(0, B_u J(\Sigma_u + \sum_{k=1}^{K_1+K_2+1} \{W_{uk} + W'_{uk}\}) J' B'_u)$$

for $B_u = (C'_u C_u)^{-1} C'_u, J, \Sigma_u$ and W_{uk} are defined the same as in Lemma 3.21, 3.22, (3.52).

We now study the limiting distribution of $\sqrt{n}(\widehat{\mathbf{a}}_u - \mathbf{a})$.

By Lemma 3.22, for each u and any vector α, we have

$$X_{un} = \sqrt{n}\frac{\alpha'(\widehat{b}_{un} - b_u)}{\sigma_{\alpha u}} \xrightarrow{d} N(0,1) \text{ as } n \to \infty$$

where $\sigma_{\alpha u} = \sqrt{\alpha' J(\Sigma_u + \sum_{k=1}^{K_1+K_2+1}\{W_{uk} + W'_{uk}\})J'\alpha}$. Denote the distribution function of X_{un} by $F_{un}(y)$, and standard normal distribution function $\Phi(y)$, then

$$F_{un}(y) \to \Phi(y), \ -\infty < y < \infty, \ as\ n \to \infty$$

Since $\Phi(y)$ is continuous, so

$$\lim_n \sup_y |F_{un}(y) - \Phi(y)| = 0,$$

for each u.

Now suppose $u(n)$ is a sequence of numbers chosen to satisfy the condition

$$\sqrt{n} \max_{1 \le i \le m} |b_{u(n)}(x_i) - b(x_i)| \to 0, \ as \ \ n \to \infty.$$

These $u(n)$ satisfy

$$\lim_n \sup_y |F_{u(n),n}(y) - \Phi(y)| = 0$$

which implies

$$X_{u(n),n} = \sqrt{n}\frac{\alpha'(\widehat{b}_{u(n),n} - b_{u(n)})}{\sigma_{\alpha u(n)}} \xrightarrow{d} N(0,1),$$

so

$$\sqrt{n}(b_{u(n),n} - b_{u(n)}) \xrightarrow{d} N(0,\Sigma)$$

where $\Sigma = \lim_u J(\Sigma_u + \sum_{k=1}^{K_1+K_2+1}\{W_{uk} + W'_{uk}\})J'$.

From the condition $\sqrt{n} \max_{1\le i\le m} |b_{u(n)}(x_i) - b(x_i)| \to 0,\ as\ \ n \to \infty$, we know the ratios of parameters from $b_{u(n)}(x_i)$ converge to the ratios of parameters from $b(x_i)$ since ratios are jump points. So the matrix $C_{u(n)}$ formed from (3.54) and the matrix formed from (3.55) are identical for sufficiently large n because all elements of $C_{u(n)}$ are also either 1, $\frac{1}{x_i}$ or $1 + \frac{1}{x_i}$, and this gives $\sqrt{n}(a_{u(n)} - a) \to 0$.

Let $B = (C'C)^{-1}C'$, then

$$\sqrt{n}(\widehat{a}_{u(n)} - a_{u(n)}) \xrightarrow{d} N(0, B\Sigma B'),$$

and

$$\sqrt{n}(\widehat{a}_{u(n)} - a) = \sqrt{n}(\widehat{a}_{u(n)} - a_{u(n)}) + \sqrt{n}(a_{u(n)} - a) \xrightarrow{d} N(0, B\Sigma B').$$

We form these results into the following corollary.

Corollary 3.24 *Suppose the model is identified from $x_1, x_2, \ldots, x_m$, where these values are different from the ratios of true parameter, and under the condition*

$$\sqrt{n} \max_{1\le i\le m} |b_{u(n)}(x_i) - b(x_i)| \to 0,\ as\ \ n \to \infty,$$

we have

$$\sqrt{n}(\widehat{a}_{u(n)} - a) \xrightarrow{d} N(0, B\Sigma B').$$

Chapter 4

Modeling Extreme Processes with Parametric Structures

In chapters 2 and 3, we studied probabilistic properties of M4 processes, proposed estimation methods, proved consistency and asymptotic normality of the estimators. All the methods are not restricted to the number of parameters. But in numerical computation aspect, the more parameters, the less precision of the estimates. It's not just a numerical stability issue, of course. The statistical precision of the estimates will be poor if the number of parameters is too large. So in this chapter, we will consider several parametric structures which could be imposed on the parameters. For instance we consider $a_{l,k,d}$ being symmetric about $k = 0$ for each l and d and some other specific possibilities. In section 4.1, we shall study a symmetric parameter structure model. In section 4.2, we shall study an asymmetric geometry parameter structure model. In section 4.3, a monotone parameter structure model will be discussed. The following fact will be used many times in this chapter. It should be no ambiguity when it is used without mentioning it.

Fact: If two lines

$$f(x) = ax + b \quad \text{and} \quad g(x) = cx + d$$

on the plane satisfy

$$f(x_1) = g(x_1) \quad \text{and} \quad f(x_2) = g(x_2)$$

for two different points x_1 and x_2, then

$$a = c, \quad \text{and } b = d.$$

4.1 Symmetric geometric parameter structure model

In this section we study a particular form of symmetric geometric parameter structure model, i.e., we assume

$$a_{l,k,d} = b_{ld}\lambda_{ld}^{|k|}, \quad k = -K_1, \dots, K_2,\ d = 1, \dots, D \tag{4.1}$$

for each l, where the unknown parameters are b_{ld} and λ_{ld}. We first consider the case $L = 1$ and then the case $L > 1$ in the following subsections.

4.1.1 Case $L = 1$

When $L = 1$, for simplicity we assume $K_1 = K_2$, then we have

$$a_{kd} = b_d\lambda_d^{|k|}, \quad k = -K, \dots, K,\ d = 1, \dots, D \tag{4.2}$$

Since $\sum_{k=-K}^{K} a_{kd} = b_d \sum_{k=-K}^{K} \lambda_d^{|k|} = 1$, $b_d = \frac{1}{\sum_{k=-K}^{K}\lambda_d^{|k|}}$. Let $\boldsymbol{\tau} = (0, 0, \dots, \tau_d, 0, \dots, 0)$, then by (1.25) we have

$$\theta(\boldsymbol{\tau}) = \frac{\max_k \max_d a_{kd}\tau_d}{\sum_k \max_d a_{kd}\tau_d} = \frac{\max_k b_d\lambda_d^{|k|}\tau_d}{\sum_k b_d\lambda_d^{|k|}\tau_d} = \frac{\max_k \lambda_d^{|k|}}{\sum_k \lambda_d^{|k|}} = \theta_d$$

which immediately implies

$$\theta_d = \max_k b_d\lambda_d^{|k|} = \max_k a_{kd}. \tag{4.3}$$

This tells that we can either from the estimation of $\max_k a_{kd}$ to get the estimation of θ_d or *vice versa*. Especially, if $\lambda_d < 1$, then $\theta_d = b_d$; if $\lambda_d > 1$, then $\theta_d = b_d\lambda_d^K$.

Since

$$\begin{aligned} \textstyle\sum_k \lambda_d^{|k|} &= 1 + 2\lambda_d(1 + \lambda_d + \cdots + \lambda_d^{K-1}) \\ &= 2(1 + \lambda_d + \lambda_d^2 + \cdots + \lambda_d^K) - 1 \end{aligned}$$

we have

$$1 + \lambda_d + \lambda_d^2 + \cdots + \lambda_d^K = \frac{1}{2}(\frac{1}{b_d} + 1). \tag{4.4}$$

Let $f(t) = 1 + t + \cdots + t^K$, then $f'(t) = 1 + 2t + \cdots + Kt^{K-1} > 0$, for $t > 0$. So $f(t)$ is strictly increasing and λ_d is uniquely determined by $\lambda_d = f^{-1}(\frac{1}{2}(\frac{1}{b_d} + 1))$. And so we have a theorem.

Proposition 4.1 *Under the parameter structure (4.2), we have*

$$\theta_d = \max_k a_{kd}, \quad \lambda_d = f^{-1}(\frac{1}{2}(\frac{1}{b_d} + 1)), \quad d = 1, \dots, D.$$

By the definition of $q(x)$ in (3.2), when $D = 1$ we have

$$\begin{aligned} q(x) \;\; = \;\; & b\lambda^K x + \max(b\lambda^{K-1}x, b\lambda^K) + \cdots + \max(b\lambda x, b\lambda^2) + \max(bx, b\lambda) \\ & + \max(b\lambda x, b) + \max(b\lambda^2 x, b\lambda) + \cdots + \max(b\lambda^K x, b\lambda^{K-1}) + b\lambda^K. \end{aligned} \tag{4.5}$$

When $x_2 > \max(\lambda, 1/\lambda)$,

$$q(x_2) = x_2 + b\lambda^K. \tag{4.6}$$

When $1/\lambda < x_1 < \lambda$,

$$q(x_1) = b(x_1 + 1)(2\lambda^K + \lambda^{K-1} + \cdots + \lambda). \tag{4.7}$$

When $\lambda < x_3 < 1/\lambda$

$$q(x_3) = b(x_3 + 1)(\lambda^K + \lambda^{K-1} + \cdots + 1). \tag{4.8}$$

When use $q(x_2)$ together $q(x_1)$ or $q(x_3)$, we can uniquely determine b and λ. Now let $x_1 < x_2$ be two points, the goal is to find a point x_2 such that $x_2 > \max(\lambda, 1/\lambda)$. We have the following two cases.

1. if $\frac{q(x_2)-q(x_1)}{x_2-x_1} = 1$, then $x_2 > x_1 > \max(\lambda, 1/\lambda)$,
2. if $\frac{q(x_2)-q(x_1)}{x_2-x_1} < 1$, calculate the intercept point $(x_3, q(x_3))$ of the line which goes through $(x_1, q(x_1))$ and $(x_2, q(x_2))$ to the line $y(x) = x$. Let $x_1 = x_2$, $x_2 = x_3$ and repeat this process until we have the case 1.

Note: if $x_2 > \max(\lambda, 1/\lambda)$, then $\frac{1}{x_2} < \min(\lambda, 1/\lambda)$.

Suppose now x and $q(x)$ are known, where $x > \max(\lambda, 1/\lambda)$, then we can get the value for b from

$$b^K f^{-1}(\frac{1}{2}(\frac{1}{b} + 1)) = (q(x) - x)^K \tag{4.9}$$

or equivalently

$$f^{-1}(\frac{1}{2}(\frac{1}{b} + 1)) = (\frac{q(x) - x}{b})^K$$

or

$$\frac{1}{2b} + \frac{1}{2} = f((\frac{q(x) - x}{b})^K).$$

When $q(x)$ is replaced by $\widehat{q}(x)$ in (4.9), we get the estimate of b, i.e. $\widehat{b}$. Since $\widehat{q}(x) \xrightarrow{a.s.} q(x)$, by continuous mapping theorem, $\widehat{b} \xrightarrow{a.s.} b$, and hence $\widehat{\lambda} \xrightarrow{a.s.} \lambda$. Since $q(x) = bf^{-1}(\frac{1}{2}(\frac{1}{b} + 1)) + x$, so we have

$$\begin{aligned} \frac{\partial q}{\partial b} \;\; &= \;\; f^{-K}(\tfrac{1}{2}(\tfrac{1}{b} + 1)) + \frac{Kbf^{-(K+1)}(\frac{1}{2}(\frac{1}{b}+1))}{f'(\frac{1}{2}(\frac{1}{b}+1))\frac{1}{2b^2}} \\ &= \;\; f^{-K}(\tfrac{1}{2}(\tfrac{1}{b} + 1)) + 2Kb^3 \frac{f^{-(K+1)}(\frac{1}{2}(\frac{1}{b}+1))}{f'(\frac{1}{2}(\frac{1}{b}+1))} \end{aligned}$$

so

$$\frac{\partial b}{\partial q} = \frac{f'(\frac{1}{2}(\frac{1}{b}+1))}{f'(\frac{1}{2}(\frac{1}{b}+1))f^{-K}(\frac{1}{2}(\frac{1}{b}+1)) + 2Kb^3 f^{-(K+1)}(\frac{1}{2}(\frac{1}{b}+1))} = \delta$$

This together with Corollary 3.4, We have proved the following theorem.

Theorem 4.2 *Suppose* $x > \max(\lambda, 1/\lambda)$*, then*

$$\sqrt{n}(\widehat{b} - b) \xrightarrow{d} N(0, x^2\sigma^2\delta^2)$$

where σ *is defined as in the Theorem 3.3.*

Since $\lambda = f^{-1}(\frac{1}{2}(\frac{1}{b}+1))$, $\frac{\partial \lambda}{\partial b} = \frac{2b^2}{f'(\frac{1}{2}(\frac{1}{b}+1))} = \lambda$, we then have the following corollary.

Corollary 4.3 *Suppose* $x > \max(\lambda, 1/\lambda)$*, then*

$$\sqrt{n}(\widehat{\lambda} - \lambda) \xrightarrow{d} N(0, x^2\sigma^2\delta^2\lambda^2)$$

In order to obtain asymptotic properties of $(\widehat{b}, \widehat{\lambda})$ we start from (4.6)-(4.8), the following is a corollary of Theorem 3.5.

Corollary 4.4 *Suppose* $\min(\lambda, 1/\lambda) < x_1 < \max(\lambda, 1/\lambda)$*,* $x_2 > \max(\lambda, 1/\lambda)$*, then*

$$\sqrt{n}(\widehat{\mathbf{q}} - \mathbf{q}) \xrightarrow{d} N(0, \Theta(\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\})\Theta'),$$

where

$$\widehat{\mathbf{q}} = \begin{bmatrix} \widehat{q}(x_1) \\ \widehat{q}(x_2) \end{bmatrix}, \mathbf{q} = \begin{bmatrix} q(x_1) \\ q(x_2) \end{bmatrix}, \Theta = \begin{bmatrix} \frac{x_1}{\mu_1} & 0 \\ 0 & \frac{x_2}{\mu_2} \end{bmatrix}.$$

and

$\mu_i = \Pr(Y_1 \le 1, Y_2 \le x_i)$, $\mu_{12} = \Pr(Y_1 \le 1, Y_2 \le x_1)$, $\sigma_{ij} = \mu_{ij} - \mu_i\mu_j$,
$w_k^{ij} = \Pr(Y_1 \le 1, Y_2 \le x_i, Y_{1+k} \le 1, Y_{2+k} \le x_j) - \mu_i\mu_j$, $\mu_{ii} = \mu_i$.

Theorem 4.5 *Suppose* $\min(\lambda, 1/\lambda) < x_1 < \max(\lambda, 1/\lambda)$*,* $x_2 > \max(\lambda, 1/\lambda)$*, then*

$$\sqrt{n}(\begin{bmatrix} \widehat{b} \\ \widehat{\lambda} \end{bmatrix} - \begin{bmatrix} b \\ \lambda \end{bmatrix}) \xrightarrow{d} N(0, J^{-1}\Theta(\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\})\Theta'(J^{-1})')$$

where

$$J = \begin{cases} J_1, & \text{if } 1/\lambda < x_1 < \lambda \\ J_2, & \text{if } \lambda < x_1 < 1/\lambda. \end{cases}$$

and

$$J_1 = \begin{bmatrix} \frac{(x_1+1)(2\lambda^K + \lambda^{K-1} + \cdots + \lambda)}{\lambda^K} & \frac{b(x_1+1)(2K\lambda^{K-1} + (K-1)\lambda^{K-2} + \cdots + 1)}{Kb\lambda^{K-1}} \end{bmatrix}$$

$$J_2 = \begin{bmatrix} \frac{(x_1+1)(\lambda^K + \lambda^{K-1} + \cdots + \lambda^2 + \lambda + 1)}{\lambda^K} & \frac{b(x_1+1)(K\lambda^{K-1} + (K-1)\lambda^{K-2} + \cdots + 2\lambda + 1)}{Kb\lambda^{K-1}} \end{bmatrix}$$

Proof. We prove the case when $\lambda < x_1 < 1/\lambda$. Since

$$\frac{\partial q(x_1)}{\partial b} = (x_1+1)(\lambda^K + \lambda^{K-1} + \cdots + \lambda^2 + \lambda + 1),$$

$$\frac{\partial q(x_1)}{\partial \lambda} = b(x_1+1)(K\lambda^{K-1} + (K-1)\lambda^{K-2} + \cdots + 2\lambda + 1),$$

$$\frac{\partial q(x_2)}{\partial b} = \lambda^K,\ \frac{\partial q(x_2)}{\partial \lambda} = Kb\lambda^{K-1}.$$

So the Jacorbian matrix is

$$J = \begin{bmatrix} \frac{\partial q(x_1)}{\partial b} & \frac{\partial q(x_1)}{\partial \lambda} \\ \frac{\partial q(x_2)}{\partial b} & \frac{\partial q(x_2)}{\partial \lambda} \end{bmatrix} = J_2.$$

The determinant of J is $|J_2| = b(x_1+1)(\lambda^{2K-2} + 2\lambda^{2K-3} + \cdots + (K-2)\lambda^{K+1} + (K-1)\lambda^K + K\lambda^{K-1}) > 0$, so J^{-1} exists. And then by the mean value theorem and Slutsky theorem the proof is completed. □

4.1.2 Case $L > 1$

We consider $D = 1$ in this subsection. Define

$$q_l(x) = b_l\lambda_l^K x + \max(b_l\lambda_l^{K-1}x, b_l\lambda_l^K) + \cdots + \max(b_l\lambda_l^K x, b_l\lambda_l^{K-1}) + b_l\lambda_l^K. \quad (4.10)$$

then

$$q(x) = q_1(x) + q_2(x) + \cdots + q_L(x). \quad (4.11)$$

Without loss of generality, we assume $\lambda_1 < \lambda_2 < \cdots < \lambda_L$. Since $q(x)$ is a piecewise linear function of x, the jumping points of $q'(x)$ are $1/\lambda_1,\ 1/\lambda_2, \ldots,\ 1/\lambda_L,\ \lambda_1,\ \lambda_2, \ldots, \lambda_L$. We assume all these points are different.

Suppose now

$$q^*(x) = q_1^*(x) + q_2^*(x) + \cdots + q_L^*(x) \quad (4.12)$$

where $q(x) = q^*(x)$ all x and

$$q_l^*(x) = b_l^*\lambda_l^{*K} x + \max(b_l^*\lambda_l^{*K-1}x, b_l^*\lambda_l^{*K}) + \cdots + \max(b_l^*\lambda_l^{*K}x, b_l^*\lambda_l^{*K-1}) + b_l^*\lambda_l^{*K}. \quad (4.13)$$

Lemma 4.6 *If $q(x) = q^*(x)$ all x, then*

$$(\lambda_1,\ \lambda_2, \ldots, \lambda_L) = (\lambda_1^*,\ \lambda_2^*, \ldots, \lambda_L^*)$$

$$(b_1, \ldots, b_L) = (b_1^*, \ldots, b_L^*).$$

Remark: Since both $q(x)$ and $q^*(x)$ are piecewise linear functions, we only need finite number of points such that the two are equal at those points. The following proof shows that.

Proof. $q^{*\prime}(x)$ has the same jumping points as $q'(x)$ has. Suppose

$$\begin{array}{l} \max(1/\lambda_1,\ 1/\lambda_2, \dots,\ 1/\lambda_L,\ \lambda_1,\ \lambda_2, \dots, \lambda_L) = \lambda_L, \\ \max(1/\lambda_1^*,\ 1/\lambda_2^*, \dots,\ 1/\lambda_L^*,\ \lambda_1^*,\ \lambda_2^*, \dots, \lambda_L^*) = 1/\lambda_1^*, \end{array} \tag{4.14}$$

Which imply $\lambda_L = 1/\lambda_1^*$. Let

$$f(\lambda) = \lambda + \lambda^2 + \cdots + \lambda^K.$$

Let $x > \lambda_L$ and y lies in the range between the second largest jumping point and λ_L. Then by the formulas (4.6)–(4.8), we have

$$q_l(x) = b_l x(2f(\lambda_l) + 1) + b_l \lambda_l^K$$

$$q(x) = \sum b_l x(2f(\lambda_l) + 1) + \sum b_l \lambda_l^K = x + \sum b_l \lambda_l^K \tag{4.15}$$

$$q_l(y) = b_l(y + 1)(\lambda_L^K + f(\lambda_l))$$

$$\begin{aligned} q(y) &= \sum_{l=1}^{L-1} b_l y(2f(\lambda_l) + 1) + \sum_{l=1}^{L-1} b_l \lambda_l^K + b_L(y + 1)(\lambda_L^K + f(\lambda_L)) \\ &= y + \sum_{l=1}^{L} b_l \lambda_l^K + b_L y(\lambda_L^K - f(\lambda_L) - 1) + b_L f(\lambda_L) \end{aligned} \tag{4.16}$$

(4.15)-(4.16) gives

$$q(x) - q(y) = x - y + b_L y(\lambda_L^K - f(\lambda_L) - 1) - b_L f(\lambda_L) \tag{4.17}$$

Now consider using λ^*'s and for $x > 1/\lambda_1^*$ we get

$$q^*(x) = x + \sum b_l^* \lambda_l^{*K} \tag{4.18}$$

For $\lambda_1^* < y < 1/\lambda_1^*$, we have

$$q_1^*(y) = b_1^*(y + 1)(f(\lambda_1^*) + 1)$$

$$\begin{aligned} q^*(y) &= b_1^*(y + 1)(f(\lambda_1^*) + 1) + \sum_{l=2}^{L} b_l^* y(2f(\lambda_l^*) + 1) + \sum_{l=2}^{L} b_l^* \lambda_l^{*K} \\ &= y - b_1^* y f(\lambda_1^*) + \sum_{l=2}^{L} b_l^* \lambda_l^{*K} + b_1^*(f(\lambda_1^*) + 1) \end{aligned} \tag{4.19}$$

(4.18)-(4.19) gives

$$q^*(x) - q^*(y) = x - y + b_1^* y f(\lambda_1^*) - b_1^*(f(\lambda_1^*) + 1 - \lambda_1^{*K}) \tag{4.20}$$

$q(x) - q(y) = q^*(x) - q^*(y)$ gives

$$b_L(f(\lambda_L) - \lambda_L^K + 1) = b_1^* f(\lambda_1^*) \tag{4.21}$$

$$b_L f(\lambda_L) = b_1^*(f(\lambda_1^*) + 1 - \lambda_1^{*K}) \tag{4.22}$$

Now suppose both x and y are between the maximum and the second maximum values of $\lambda,\ 1/\lambda$, then

$$q(x) - q(y) = x - y - b_L(y-x)(f(\lambda_L) - \lambda_L^K + 1)$$

$$q^*(x) - q^*(y) = x - y - b_1^*(x-y) f(\lambda_1^*)$$

which gives

$$b_L(f(\lambda_L) - \lambda_L^K + 1) = -b_1^* f(\lambda_1^*) \tag{4.23}$$

(4.21) and (4.23) can not be true simultaneously. So $\lambda_L = 1/\lambda_1^*$ can not be true, so $\lambda_L^* = \lambda_L$. And $b_L = b_L^*$ is obvious.

Now Suppose

$$\lambda_l = \lambda_l^*,\ b_l = b_l^*,\ l = k+1, \ldots, L$$

then

$$q_l(x) = q_l^*(x),\ l = k+1, \ldots, L$$

Suppose $1/\lambda_k < \lambda_k$ where λ_k is the $L-k$ largest among $1/\lambda_1,\ 1/\lambda_2, \ldots,\ 1/\lambda_L,\ \lambda_1,\ \lambda_2, \ldots,$ λ_L and $1/\lambda_k^* > \lambda_k^*$ where $1/\lambda_k^*$ is the $L-k$ largest among $1/\lambda_1^*,\ 1/\lambda_2^*, \ldots,\ 1/\lambda_L^*,\ \lambda_1^*,\ \lambda_2^*, \ldots,$ λ_L^*. So $\lambda_k = 1/\lambda_k^*$.

Suppose $x > \max(\lambda_l, 1/\lambda_l),\ l = 1, \ldots, k$ then

$$q(x) = \sum_{l=1}^{k} b_l x(2f(\lambda_l) + 1) + \sum_{l=1}^{k} b_l \lambda_l^K + \sum_{l=k+1}^{L} q_l(x)$$

For $1/\lambda_k < y < \lambda_k$,

$$\begin{aligned} q(y) &= \sum_{l=1}^{k-1} b_l y(2f(\lambda_l) + 1) + \sum_{l=1}^{k-1} b_l \lambda_l^K + b_k(y+1)(\lambda_k^K + f(\lambda_k)) + \sum_{l=k+1}^{L} q_l(y) \\ &= \sum_{l=1}^{k} b_l y(2f(\lambda_l) + 1) + \sum_{l=1}^{k} b_l \lambda_l^K + b_k(y+1)(\lambda_k^K + f(\lambda_k)) + \sum_{l=k+1}^{L} q_l(y) \\ &\quad -b_k y(2f(\lambda_k) + 1) - b_k \lambda_k^K \end{aligned}$$

$$\begin{aligned} q(x)-q(y) &= (x-y)\sum_{l=1}^{k} b_l y(2f(\lambda_l)+1) \\ &\quad + \sum_{l=k+1}^{L}(q_l(x)-q_l(y)) - b_k(y+1)(\lambda_k^K + f(\lambda_k)) \\ &\quad + b_k y(2f(\lambda_k)+1) + b_k\lambda_k^K \end{aligned}$$

Suppose $x > \max(\lambda_l^*, 1/\lambda_l^*),\ l = 1, \ldots, k$ then

$$q^*(x) = \sum_{l=1}^{k} b_l^* x(2f(\lambda_l^*)+1) + \sum_{l=1}^{k} b_l^*\lambda_l^{*K} + \sum_{l=k+1}^{L} q_l^*(x)$$

For $1/\lambda_k^* < y < \lambda_k^*$,

$$\begin{aligned} q^*(y) &= b_k^*(y+1)(f(\lambda_k^*)+1) + \sum_{l=1}^{k-1} b_l^* y(2f(\lambda_l^*)+1) + \sum_{l=1}^{k-1} b_l^*\lambda_l^{*K} + \sum_{l=k+1}^{L} q_l^*(y) \\ &= \sum_{l=1}^{k} b_l^* y(2f(\lambda_l^*)+1) + \sum_{l=1}^{k} b_l^*\lambda_l^{*K} - b_k^* y f(\lambda_k^*) + \sum_{l=k+1}^{L} q_l^*(y) \\ &\quad + b_k^*(f(\lambda_k^*)+1-\lambda_k^{*K}) \end{aligned}$$

$$\begin{aligned} q^*(x)-q^*(y) &= (x-y)\sum_{l=1}^{k} b_l^* y(2f(\lambda_l^*)+1) + \sum_{l=k+1}^{L}(q_l^*(x)-q_l^*(y)) + b_k^* y f(\lambda_k^*) \\ &\quad + b_k^*(f(\lambda_k^*)+1-\lambda_k^{*K}) \end{aligned}$$

From $q(x)-q(y) = q^*(x)-q^*(y)$ and $\sum_{l=1}^{k} b_l(2f(\lambda_l)+1) = \sum_{l=1}^{k} b_l^*(2f(\lambda_l^*)+1)$, we get

$$b_k(f(\lambda_k) - \lambda_k^K + 1) = b_k^* f(\lambda_k^*) \tag{4.24}$$

$$-b_k f(\lambda_k) = b_k^*(f(\lambda_k^*) - \lambda_k^{*K} + 1) \tag{4.25}$$

These two equations can not be true simultaneously. So $\lambda_k = \lambda_k^*,\ b_k = b_k^*$. So by induction, we have all $\lambda_l = \lambda_l^*,\ b_l = b_l^*$. Therefore $q(x)$ uniquely determines all parameters. □

Let $q(x)$ evaluate at $x_1, x_2, \ldots, x_m$ such that $q(x_1), q(x_2), \ldots, q(x_m)$ uniquely determine all parameters. This can be done as long as there are at least two points between every two adjacent jumping points.

Now suppose that $\widehat{q}(x_1), \widehat{q}(x_2), \ldots, \widehat{q}(x_m)$ are estimates of $q(x_1), q(x_2), \ldots, q(x_m)$, then

$$(\widehat{q}(x_1),\ \widehat{q}(x_2),\ \ldots,\ \widehat{q}(x_m)) \xrightarrow{a.s.} (q(x_1),\ q(x_2),\ \ldots,\ q(x_m)) \tag{4.26}$$

where $\widehat{q}(x)$ defined by (3.12). By (4.26), we have the following theorem.

Theorem 4.7 *The solution of*

$$\left\{\begin{array}{rcl}
\sum_{l=1}^{L}(\widehat{b}_l\widehat{\lambda}_l^K x_1+\max(\widehat{b}_l\widehat{\lambda}_l^{K-1}x_1,\widehat{b}_l\widehat{\lambda}_l^K)+\cdots+\max(\widehat{b}_l\widehat{\lambda}_l^K x_1,\widehat{b}_l\widehat{\lambda}_l^{K-1})+\widehat{b}_l\widehat{\lambda}_l^K) & = & \widehat{q}(x_1)\\
\sum_{l=1}^{L}(\widehat{b}_l\widehat{\lambda}_l^K x_2+\max(\widehat{b}_l\widehat{\lambda}_l^{K-1}x_2,\widehat{b}_l\widehat{\lambda}_l^K)+\cdots+\max(\widehat{b}_l\widehat{\lambda}_l^K x_2,\widehat{b}_l\widehat{\lambda}_l^{K-1})+\widehat{b}_l\widehat{\lambda}_l^K) & = & \widehat{q}(x_2)\\
\cdots & & \cdots\\
\sum_{l=1}^{L}(\widehat{b}_l\widehat{\lambda}_l^K x_m+\max(\widehat{b}_l\widehat{\lambda}_l^{K-1}x_m,\widehat{b}_l\widehat{\lambda}_l^K)+\cdots+\max(\widehat{b}_l\widehat{\lambda}_l^K x_m,\widehat{b}_l\widehat{\lambda}_l^{K-1})+\widehat{b}_l\widehat{\lambda}_l^K) & = & \widehat{q}(x_m)
\end{array}\right.$$

converges almost surely to the true parameter values.

The following is a corollary of Theorem 3.5.

Corollary 4.8

$$\sqrt{n}(\widehat{\mathbf{q}}-\mathbf{q}) \xrightarrow{d} N(0,\Theta(\Sigma+\sum_{k=1}^{K_1+K_2+1}\{W_k+W_k'\})\Theta'),$$

where

$$\widehat{\mathbf{q}}=\begin{bmatrix}\widehat{q}(x_1)\\ \widehat{q}(x_2)\\ \vdots\\ \widehat{q}(x_m)\end{bmatrix},\mathbf{q}=\begin{bmatrix}q(x_1)\\ q(x_2)\\ \vdots\\ q(x_m)\end{bmatrix},\Theta=\begin{bmatrix}\frac{x_1}{\mu_1} & 0 & \cdots & 0\\ 0 & \frac{x_2}{\mu_2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \frac{x_m}{\mu_m}\end{bmatrix}.$$

where
$\mu_i=\Pr(Y_1\le 1,Y_2\le x_i)$, $\mu_{ij}=\Pr(Y_1\le 1,Y_2\le \min(x_i,x_j))$, $\sigma_{ij}=\mu_{ij}-\mu_i\mu_j$,
$w_k^{ij}=\Pr(Y_1\le 1,Y_2\le x_i,Y_{1+k}\le 1,Y_{2+k}\le x_j)-\mu_i\mu_j$, $\mu_{ii}=\mu_i$.

As long as x is not a jumping point, and we view $q(x)$ as a function of all b_l and λ_l, then $q(x)$ has all continuous first order partial derivatives in a neighborhood of $b_1,\dots,b_L,\lambda_1,\dots,\lambda_L$. So we can construct the transformation Jacobian matrix J. And we have the following theorem.

Theorem 4.9

$$\sqrt{n}(\begin{bmatrix}\widehat{\mathbf{b}}\\ \widehat{\boldsymbol{\lambda}}\end{bmatrix}-\begin{bmatrix}\mathbf{b}\\ \boldsymbol{\lambda}\end{bmatrix}) \xrightarrow{d} N(0,J^{-1}\Theta(\Sigma+\sum_{k=1}^{K_1+K_2+1}\{W_k+W_k'\})\Theta'(J^{-1})')$$

where $\mathbf{b}=(b_1,\dots,b_L)'$, $\boldsymbol{\lambda}=(\lambda_1,\dots,\lambda_L)'$, $\widehat{\mathbf{b}}=(\widehat{b}_1,\dots,\widehat{b}_L)'$, $\widehat{\boldsymbol{\lambda}}=(\widehat{\lambda}_1,\dots,\widehat{\lambda}_L)'$.

4.2 Asymmetric geometric parameter structure model

In this section we study asymmetric geometric parameter structure model,

$$a_{l,k,d} = b_{ld}\lambda_{ld}^{(k)+}\phi_{ld}^{(k)-}, \quad k = -K_1, \ldots, K_2 \tag{4.27}$$

for each l. Here $a_{l0d} = b_{ld}$.

Note: the informal note by Smith and Weissman (1997) had a general form.

4.2.1 Case $L = 1$

When $L = 1$, the parameter structure becomes

$$a_{kd} = b_d\lambda_d^{(k)+}\phi_d^{(k)-}, \quad k = -K_1, \ldots, K_2, \ d = 1, \ldots, D. \tag{4.28}$$

By the definition of $q(x)$ in (3.2) when $D = 1$ we have

$$\begin{aligned} q(x) = b\lambda^{K_2}x + \max(b\lambda^{K_2-1}x, b\lambda^{K_2}) + \cdots + \max(b\lambda x, b\lambda^2) + \max(bx, b\lambda) \\ + \max(b\phi x, b) + \max(b\phi^2 x, b\phi) + \cdots + \max(b\phi^{K_1}x, b\phi^{K_1-1}) + b\phi^{K_1}. \end{aligned} \tag{4.29}$$

When $x_1 < 1/\phi, x_1 < \lambda$,

$$q(x_1) = b\lambda^{K_2}x_1 + 1. \tag{4.30}$$

When $1/\phi < x_2 < \lambda$,

$$q(x_2) = b[\lambda^{K_2}x_2 + (\lambda^{K_2} + \lambda^{K_2-1} + \cdots + \lambda) + x_2(\phi + \cdots + \phi^{K_1}) + \phi^{K_1}]. \tag{4.31}$$

When $\lambda < x_2 < 1/\phi$,

$$q(x_2) = b[x_2(\lambda^{K_2} + \lambda^{K_2-1} + \cdots + 1) + (1 + \phi + \cdots + \phi^{K_1})]. \tag{4.32}$$

When $x_3 > \lambda, \ x_3 > 1/\phi$,

$$q(x_3) = x_3 + b\phi^{K_1}. \tag{4.33}$$

Note: parameters either satisfy (4.30), (4.31), and (4.33) or (4.30), (4.32) and (4.33).

Consider now (4.30), (4.31), and (4.33). From (4.30) and (4.33), we have

$$\lambda^{K_2} = \frac{q(x_1) - 1}{x_1(q(x_3) - x_3)}\phi^{K_1}.$$

Subtract $1 = b(\lambda^{K_2} + \lambda^{K_2-1} + \cdots + 1 + \phi + \cdots + \phi^{K_1})$ from (4.31) and substitute λ^{K_2} with $\frac{q(x_1)-1}{x_1(q(x_3)-x_3)}\phi^{K_1}$, we have

$$\frac{q(x_2) - 1}{q(x_3) - x_3}\phi^{K_1} = \frac{x_2(q(x_1) - 1)}{x_1(q(x_3) - x_3)}\phi^{K_1} + (x_2 - 1)(\phi + \cdots + \phi^{K_1}) + \phi^{K_1} - 1$$

or equivalently

$$\frac{x_2(q(x_1)-1)-x_1(q(x_2)-1)+x_2x_1(q(x_3)-x_3)}{x_1(q(x_3)-x_3)}\phi^{K_1}+(x_2-1)(\phi+\cdots+\phi^{K_1})-1=0 \tag{4.34}$$

There are multiple solutions for this equation. To overcome this problem, we need to introduce additional points such that all parameters can be uniquely determined.

Suppose $x_1' < x_1 < \min(1/\phi,\lambda) < x_2' < x_2 < \max(1/\phi,\lambda) < x_3' < x_3$, and $q(x_1'),\ q(x_1),\ q(x_2'),\ q(x_2),\ q(x_3'),\ q(x_3)$ are known, then we can have the following three lines:

$$L_1:\ y=\frac{q(x_1)-q(x_1')}{x_1-x_1'}x+\frac{x_1q(x_1')-x_1'q(x_1)}{x_1-x_1'}$$
$$L_2:\ y=\frac{q(x_2)-q(x_2')}{x_2-x_2'}x+\frac{x_2q(x_2')-x_2'q(x_2)}{x_2-x_2'}$$
$$L_3:\ y=\frac{q(x_3)-q(x_3')}{x_3-x_3'}x+\frac{x_3q(x_3')-x_3'q(x_3)}{x_3-x_3'}$$

The intercept points of L_1 and L_2, L_2 and L_3 determine the values of $\frac{1}{\phi}$ and λ. What we need is to distinguish the values of ϕ and λ from the jumping points of $q'(x)$, i.e. the intercept points. By solving the intercept points, we then get $b,\ \lambda,\ \phi$ each is a function of $(q(x_1'),\ q(x_1),\ q(x_2'),\ q(x_2),\ q(x_3'),\ q(x_3))$ and hence can calculate the transformation Jacobian matrix.

Now suppose $q(x)$ satisfies (4.30), (4.31), and (4.33) while $q^*(x)$ satisfies (4.30), (4.32), and (4.33). Then $\phi^*=1/\lambda,\ \lambda^*=1/\phi$. And $b^*\phi^{*K_1}=b\phi^{K_1},\ b^*\lambda^{*K_2}=b\lambda^{K_2}$ imply $(\lambda\phi)^{K_2}=(\lambda\phi)^{K_1}$ which implies $K_2=K_1$. Thus if $K_2\neq K_1$, there are no such $q(x)$ and $q^*(x)$. Assume $K_1=K_2=K$, then we have

$$\begin{aligned}&b[\lambda^{K_2}x+(\lambda^{K_2}+\lambda^{K_2-1}+\cdots+\lambda)+x(\phi+\cdots+\phi^{K_1})+\phi^{K_1}]=\\&b^*[x(\lambda^{*K_2}+\lambda^{*K_2-1}+\cdots+\lambda^*+1)+(1+\phi^*+\cdots+\phi^{*K_1})]\end{aligned}$$

So

$$b\lambda^{K_2}+b(\phi+\cdots+\phi^{K_1})=b^*\lambda^{*K_2}+b^*(\lambda^{*K_2}+\lambda^{*K_2-1}+\cdots+\lambda^*+1)$$

and

$$b(\lambda^{K_2}+\lambda^{K_2-1}+\cdots+\lambda)+\phi^{K_1}=b^*(1+\phi^*+\cdots+\phi^{*K_1})$$

we have

$$b(\phi+\cdots+\phi^{K_1})=b^*(\lambda^{*K_2-1}+\lambda^{*K_2-1}+\cdots+\lambda^*+1)$$

and

$$b(\lambda^{K_2}+\lambda^{K_2-1}+\cdots+\lambda)=b^*(1+\phi^*+\cdots+\phi^{*K_1-1})$$

From these two equations we have

$$\frac{\frac{\phi-\phi^{K_1+1}}{1-\phi}}{\frac{\lambda-\lambda^{K_1+1}}{1-\lambda}} = \frac{\frac{1-\phi^{*K_2}}{1-\phi^*}}{\frac{1-\lambda^{*K_1}}{1-\lambda^*}}$$

Substitute $\phi^* = 1/\lambda$, $\lambda^* = 1/\phi$ in the above equation, we get

$$\frac{\phi - \phi^K}{1-\phi^K} = \frac{\lambda - \lambda^K}{1-\lambda^K}$$

Let

$$f(x) = \frac{x - x^t}{1 - x^t}$$

then

$$f'(x) = \frac{(t-1)x^t - tx^{t-1} + 1}{(1-x^t)^2}$$

By induction we can prove $f'(x) > 0$ for $t > 1$, so $f(x)$ is strictly increasing. Therefore $\lambda = \phi$. So if the model follows (4.30), (4.31) and (4.33), then the values of ϕ and λ will not satisfy (4.30), (4.32) and (4.33), or *vice versa.*

Summarize all the arguments we have a theorem in this subsection.

Theorem 4.10 *Under (4.28), when* $x_1' < x_1 < \min(1/\phi, \lambda) < x_2' < x_2 < \max(1/\phi, \lambda) < x_3' < x_3$, *and* $q(x_1'),\ q(x_1),\ q(x_2'),\ q(x_2),\ q(x_3'),\ q(x_3)$ *are known, then (4.30)-(4.33) uniquely determine all parameters.*

We have the following corollary.

Corollary 4.11 *When replace* $q(x_1'),\ q(x_1),\ q(x_2'),\ q(x_2),\ q(x_3'),\ q(x_3)$ *in (4.30)-(4.33) by* $\widehat{q}(x_1'),\ \widehat{q}(x_1),\ \widehat{q}(x_2'),\ \widehat{q}(x_2),\ \widehat{q}(x_3'),\ \widehat{q}(x_3)$ *and denote the solutions by* $\widehat{b},\ \widehat{\lambda},\ \widehat{\phi}$, *then*

$$(\widehat{b},\ \widehat{\lambda},\ \widehat{\phi}) \xrightarrow{a.s.} (b,\ \lambda,\ \phi) \tag{4.35}$$

as $n \to \infty$.

Proof. Since

$$(\widehat{q}(x_1'),\ \widehat{q}(x_1),\ \widehat{q}(x_2'),\ \widehat{q}(x_2),\ \widehat{q}(x_3'),\ \widehat{q}(x_3)) \xrightarrow{a.s.} (q(x_1'),\ q(x_1),\ q(x_2'),\ q(x_2),\ q(x_3'),\ q(x_3))$$

and the uniqueness of solutions of (4.30)-(4.33) by Theorem 4.10, (4.35) is true, and so the proof is completed. □

By following Corollary 4.8 and the arguments followed, the following theorem tells the limit joint distribution of $(\widehat{b},\ \widehat{\lambda},\ \widehat{\phi})$ after suitably normalized tends to a multivariate normal distribution.

Theorem 4.12 *Suppose* $x'_1 < x_1 < \min(1/\phi, \lambda) < x'_2 < x_2 < \max(1/\phi, \lambda) < x'_3 < x_3$, *then*

$$\sqrt{n}\left(\begin{bmatrix} \widehat{b} \\ \widehat{\lambda} \\ \widehat{\phi} \end{bmatrix} - \begin{bmatrix} b \\ \lambda \\ \phi \end{bmatrix} \right) \xrightarrow{d} N(0, J\Theta(\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W'_k\})\Theta' J')$$

where W_k, Σ, Θ *are defined in Corollary 4.8 with* $m = 6$ *and the elements of* J *are* $J_{1j} = \frac{\partial b}{\partial q(x_j)}$, $J_{2j} = \frac{\partial \lambda}{\partial q(x_j)}$, $J_{1j} = \frac{\partial \phi}{\partial q(x_j)}$, $j = 1, \dots, 6$.

Proof. By the same arguments in Theorem 2.6 and 2.7. $\square$

4.2.2 Case $L > 1$

We consider $D = 1$ in this subsection. Define

$$\begin{aligned} q_l(x) = {} & b_l\lambda_l^{K_2}x + \max(b_l\lambda_l^{K_2-1}x, b_l\lambda_l^{K_2}) + \cdots + \max(b_l\lambda_l x, b_l\lambda_l^2) + \max(b_l x, b_l\lambda_l) \\ & + \max(b_l\phi_l x, b_l) + \max(b_l\phi_l^2 x, b_l\phi_l) + \cdots + \max(b_l\phi_l^{K_1}x, b_l\phi_l^{K_1-1}) + b_l\phi_l^{K_1}. \end{aligned} \tag{4.36}$$

then

$$q(x) = q_1(x) + q_2(x) + \cdots + q_L(x). \tag{4.37}$$

Without loss of generality, we assume $\lambda_1 < \lambda_2 < \cdots < \lambda_L$. Since $q(x)$ is a piecewise linear function of x, the jumping points of $q'(x)$ are $1/\phi_1,\ 1/\phi_2, \dots,\ 1/\phi_L,\ \lambda_1,\ \lambda_2, \dots, \lambda_L$. We assume all these points are different.

Suppose now

$$q^*(x) = q_1^*(x) + q_2^*(x) + \cdots + q_L^*(x) \tag{4.38}$$

where $q(x) = q^*(x)$ all x and

$$\begin{aligned} q_l^*(x) = {} & b_l^*\lambda_l^{*K_2}x + \max(b_l^*\lambda_l^{*K_2-1}x, b_l^*\lambda_l^{*K_2}) + \cdots + \max(b_l^*\lambda_l^* x, b_l^*\lambda_l^{*2}) + \max(b_l^* x, b_l^*\lambda_l^*) \\ & + \max(b_l^*\phi_l^* x, b_l^*) + \max(b_l^*\phi_l^{*2}x, b_l^*\phi_l^*) + \cdots + \max(b_l^*\phi_l^{*K_1}x, b_l^*\phi_l^{*K_1-1}) + b_l^*\phi_l^{*K_1}. \end{aligned} \tag{4.39}$$

Then $q^{*\prime}(x)$ has the same jumping points as $q'(x)$ has. With out loss of generality we can assume

$$\max(1/\phi_1,\ 1/\phi_2, \dots,\ 1/\phi_L,\ \lambda_1,\ \lambda_2, \dots, \lambda_L) = \lambda_L, \tag{4.40}$$

or

$$\max(1/\phi_1,\ 1/\phi_2, \dots,\ 1/\phi_L,\ \lambda_1,\ \lambda_2, \dots, \lambda_L) = 1/\phi_1, \tag{4.41}$$

since the order in index l does not matter in the $M4$ process. We need to show that (4.37) and (4.38) agree with each other. We state this as the following lemma.

Lemma 4.13 *If* $q(x) = q^*(x)$ *all* x, *then*

$$(1/\phi_1,\ 1/\phi_2, \dots,\ 1/\phi_L,\ \lambda_1,\ \lambda_2, \dots, \lambda_L) = (1/\phi_1^*,\ 1/\phi_2^*, \dots,\ 1/\phi_L^*,\ \lambda_1^*,\ \lambda_2^*, \dots, \lambda_L^*) \tag{4.42}$$

is true for $L = 2$.

We leave the proof in the proofs section of this chapter. There are possibilities that this lemma can be generalized to case $L > 2$. But we will restrict our discussion in case $L = 2$ in this subsection.

Now let $x_1' < x_1 < x_2' < x_2 < x_3' < x_3 < x_4' < x_4 < x_5' < x_5$ and there is one jumping point that belongs to $(x_i',\ x_{i+1})$, $i = 1, 2, 3, 4$. Let $\widehat{q}(x_i)$ be the estimation of $q(x_i)$, and $\widehat{\mathbf{b}} = (\widehat{b}_1,\ \widehat{b}_2)'$, $\widehat{\boldsymbol{\lambda}} = (\widehat{\lambda}_1,\ \widehat{\lambda}_2)'$, $\widehat{\boldsymbol{\phi}} = (\widehat{\phi}_1,\ \widehat{\phi}_2)'$ be estimations of $\mathbf{b} = (b_1,\ b_2)'$, $\boldsymbol{\lambda} = (\lambda_1,\ \lambda_2)'$, $\boldsymbol{\phi} = (\phi_1,\ \phi_2)'$ respectively. Then the following theorem follows immediately.

Theorem 4.14 $(\widehat{\mathbf{b}}',\ \widehat{\boldsymbol{\lambda}}', \widehat{\boldsymbol{\phi}}') \xrightarrow{a.s.} (\mathbf{b}',\ \boldsymbol{\lambda}',\ \boldsymbol{\phi}')$ *as* $n \to \infty$.

Suppose J is the transformation Jacobian matrix, then the following theorem follows by the arguments before Theorem 4.9 and Theorem 4.14.

Theorem 4.15

$$\sqrt{n}\Big(\begin{bmatrix} \widehat{\mathbf{b}} \\ \widehat{\boldsymbol{\lambda}} \\ \widehat{\boldsymbol{\phi}} \end{bmatrix} - \begin{bmatrix} \mathbf{b} \\ \boldsymbol{\lambda} \\ \boldsymbol{\phi} \end{bmatrix} \Big) \xrightarrow{d} N(0, J\Theta(\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\})\Theta' J')$$

where $W_k,\ \Sigma,\ \Theta$ *are defined in Corollary 4.8 with* $m = 10$.

4.3 Monotone parameter structure model

In this section we study monotone parameter structure model, i.e.

$$a_{lk} = b_{-K_1} \lambda_l^{k+K_1}, \quad k = -K_1, \dots, K_2 \tag{4.43}$$

for each l.

Note: this is a special asymmetric form.

4.3.1 Case $L = 1$

We have

$$a_{kd} = b_d \lambda_d^{k+K_1}, \quad k = -K_1, \ldots, K_2,\ d = 1, \ldots, D \tag{4.44}$$

Since $\sum_{k=-K_1}^{K_2} a_{kd} = b_d \sum_{k=-K_1}^{K_2} \lambda_d^{k+K_1} = 1$, $b_d = \frac{1}{\sum_{k=-K_1}^{K_2} \lambda_d^{k+K_1}}$.
Let $\boldsymbol{\tau} = (0, 0, \ldots, \tau_d, 0, \ldots, 0)$, then by (1.25) we have

$$\theta(\boldsymbol{\tau}) = \frac{\max_k \max_d a_{kd}\tau_d}{\sum_k \max_d a_{kd}\tau_d} = \frac{\max_k b_d \lambda_d^{k+K_1}\tau_d}{\sum_k b_d \lambda_d^{k+K_1}\tau_d} = \frac{\max_k \lambda_d^{k+K_1}}{\sum_k \lambda_d^{k+K_1}} = \theta_d$$

which immediately implies

$$\theta_d = \max_k b_d \lambda_d^{k+K_1} = \max_k a_{kd}. \tag{4.45}$$

This tells that we can either from the estimation of $\max_k a_{kd}$ to get the estimation of θ_d or *vice versa*. Especially, if $\lambda_d < 1$, then $\theta_d = b_d$; if $\lambda_d > 1$, then $\theta_d = b_d \lambda_d^K$, where $K = K_1 + K_2$ and hereafter.

we have

$$1 + \lambda_d + \lambda_d^2 + \cdots + \lambda_d^K = \frac{1}{b_d}. \tag{4.46}$$

Let $f(t) = 1 + t + \cdots + t^K$, then $f'(t) = 1 + 2t + \cdots + Kt^{K-1} > 0$, for $t > 0$. So $f(t)$ is strictly increasing and λ_d is uniquely determined by $\lambda_d = f^{-1}(\frac{1}{b_d})$. And so we have a corollary.

Corollary 4.16 *Under the parameter structure (4.44), we have*

$$\theta_d = \max_k a_{kd}, \quad \lambda_d = f^{-1}(\frac{1}{b_d}), \quad d = 1, \ldots, D.$$

By the definition of $q(x)$ in (3.2) when $D = 1$ we have

$$q(x) = b + \max(bx, b\lambda) + \cdots + \max(b\lambda^{K-1}x, b\lambda^K) + b\lambda^K x. \tag{4.47}$$

When $x > \lambda$,

$$q(x) = b + bx + b\lambda x + \cdots + b\lambda^{K-1}x + b\lambda^K x = b + x. \tag{4.48}$$

So $b = q(x) - x$ and $\lambda = f^{-1}(\frac{1}{q(x)-x})$. When $q(x)$ is replaced by $\widehat{q}(x)$, we then have

$$\widehat{b} = \widehat{q}(x) - x,\ \widehat{\lambda} = f^{-1}(\frac{1}{\widehat{q}(x) - x}). \tag{4.49}$$

Since $\widehat{q}(x) \xrightarrow{a.s.} q(x)$ and mapping $f^{-1} \circ h$ is continuous, where $h(q) = \frac{1}{q-x}$, so $\widehat{b} \xrightarrow{a.s.} b$, $\widehat{\lambda} \xrightarrow{a.s.} \lambda$. We have the following corollary which immediately follows Corollary 3.4.

Corollary 4.17 *When $x > \lambda$, we have* $\sqrt{n}(\widehat{b} - b) \xrightarrow{d} N(0, x^2\sigma^2)$.

Since $b = \frac{1-\lambda}{1-\lambda^{K+1}}$, so $\frac{\partial b}{\partial \lambda} = \frac{-1+(K+1)\lambda^K - K\lambda^{K+1}}{(1-\lambda^{K+1})^2} = \delta$. Then a corollary immediately follows.

Corollary 4.18 *When $x > \lambda$, we have* $\sqrt{n}(\widehat{\lambda} - \lambda) \xrightarrow{d} N(0, x^2\sigma^2\delta^2)$.

In order to obtain asymptotic properties of $(\widehat{b},\ \widehat{\lambda})$, we consider $x_1 < \lambda$ and $x_2 > \lambda$. We have

$$q(x_1) = 1 + b\lambda^K x_1,\ q(x_2) = x_2 + b. \tag{4.50}$$

So

$$\frac{\partial q(x_1)}{\partial b} = \lambda^K x_1,\ \frac{\partial q(x_1)}{\partial \lambda} = bK\lambda^{K-1}x_1,\ \frac{\partial q(x_2)}{\partial b} = 1,\ \frac{\partial q(x_2)}{\partial \lambda} = 0. \tag{4.51}$$

So the transformation Jacobian matrix is

$$J = \begin{pmatrix} \lambda^K x_1 & bK\lambda^{K-1}x_1 \\ 1 & 0 \end{pmatrix} \tag{4.52}$$

We then have the following theorem.

Theorem 4.19 *Suppose $x_1 < \lambda < x_2$, then*

$$\sqrt{n}\left(\begin{bmatrix} \widehat{b} \\ \widehat{\lambda} \end{bmatrix} - \begin{bmatrix} b \\ \lambda \end{bmatrix}\right) \xrightarrow{d} N(0, J^{-1}\Theta(\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k'\})\Theta'(J^{-1})')$$

where Θ, Σ, W_k are defined as in Corollary 4.4.

4.3.2 Case $L > 1$

As usual we consider $D = 1$ only. Define

$$q_l(x) = b_l + \max(b_l x, b_l\lambda_l) + \cdots + \max(b_l\lambda_l^{K-1}x, b_l\lambda_l^K) + b_l\lambda_l^K x. \tag{4.53}$$

then

$$q(x) = q_1(x) + q_2(x) + \cdots + q_L(x).$$

Without loss of generality, we assume $\lambda_1 < \lambda_2 < \cdots < \lambda_L$ as change points. Let

$$f(b, \lambda) = b + b\lambda + \cdots + b\lambda^K$$

where $\sum_{l=1}^{L} f(b_l, \lambda_l) = 1$.

When $x > \lambda_l$,

$$q_l(x) = b_l + f(b_l, \lambda_l)x$$

When $x < \lambda_l$,

$$q_l(x) = f(b_l, \lambda_l) + b_l \lambda_l^K x$$

Lemma 4.20 *If $q(x) = q^*(x)$ all x, then*

$$(\lambda_1,\ \lambda_2, \ldots, \lambda_L) = (\lambda_1^*,\ \lambda_2^*, \ldots, \lambda_L^*)$$

$$(b_1, \ldots, b_L) = (b_1^*, \ldots, b_L^*)$$

Proof. Suppose now $x_1,\ x_1',\ x_2,\ x_2',\ \ldots, x_L,\ x_L',\ x_{L+1}$ satisfy

$$x_1' < x_1 < \lambda_1 < x_2' < x_2 < \lambda_2 < \cdots < x_L' < x_L < \lambda_L < x_{L+1}$$

then

$$\begin{aligned}
q(x_1') &= 1 + x_1'(b_1\lambda_1^K + b_2\lambda_2^K + \cdots + b_L\lambda_L^K) \\
q(x_1) &= 1 + x_1(b_1\lambda_1^K + b_2\lambda_2^K + \cdots + b_L\lambda_L^K) \\
q(x_2) &= \sum_{l=2}^{L} f(b_l, \lambda_l) + x_2(b_2\lambda_2^K + \cdots + b_L\lambda_L^K) + b_1 + f(b_1, \lambda_1)x_2 \\
&= 1 - f(b_1, \lambda_1) + x_2(b_1\lambda_1^K + b_2\lambda_2^K + \cdots + b_L\lambda_L^K) - x_2 b_1\lambda_1^K + b_1 + f(b_1, \lambda_1)x_2
\end{aligned}$$

Similarly

$$\begin{aligned}
q^*(x_1') &= 1 + x_1'(b_1^*\lambda_1^K + b_2^*\lambda_2^K + \cdots + b_L^*\lambda_L^K) \\
q^*(x_1) &= 1 + x_1(b_1^*\lambda_1^K + b_2^*\lambda_2^K + \cdots + b_L^*\lambda_L^K) \\
q^*(x_2) &= \sum_{l=2}^{L} f(b_l^*, \lambda_l) + x_2(b_2^*\lambda_2^K + \cdots + b_L^*\lambda_L^K) + b_1^* + f(b_1^*, \lambda_1)x_2 \\
&= 1 - f(b_1^*, \lambda_1) + x_2(b_1^*\lambda_1^K + b_2^*\lambda_2^K + \cdots + b_L^*\lambda_L^K) - x_2 b_1^*\lambda_1^K + b_1^* + f(b_1^*, \lambda_1)x_2
\end{aligned}$$

From $q(x_1) - q(x_1') = q^*(x_1) - q^*(x_1')$ we have

$$b_1\lambda_1^K + \cdots + b_L\lambda_L^K = b_1^*\lambda_1^K + \cdots + b_L^*\lambda_L^K$$

From $q(x_2) - q(x_1) = q^*(x_2) - q^*(x_1)$ we have

$$b_1 + f(b_1, \lambda_1) = b_1^* + f(b_1^*, \lambda_1)$$

$$b_1\lambda_1^K - f(b_1, \lambda_1) = b_1^*\lambda_1^K - f(b_1^*, \lambda_1)$$

Which give $b_1 = b_1^*$.

Suppose when $l = k$ we have

$$b_1 = b_1^*, \ \ldots, b_k = b_k^*.$$

Let $\lambda_k < x < \lambda_{k+1}$

$$q(x) = \sum_{l=1}^{k}(b_l + f(b_l, \lambda_l)x) + \sum_{l=k+1}^{L}(f(b_l, \lambda_l) + b_l\lambda_l^K x)$$

$$q^*(x) = \sum_{l=1}^{k}(b_l^* + f(b_l^*, \lambda_l)x) + \sum_{l=k+1}^{L}(f(b_l^*, \lambda_l) + b_l^*\lambda_l^K x)$$

Which imply

$$b_{k+1}\lambda_{k+1}^K + \cdots + b_L\lambda_L^K$$

When $\lambda_{k+1} < x < \lambda_{k+2}$

$$q(x) = \sum_{l=1}^{k}(b_l + f(b_l, \lambda_l)x) + b_{k+1} + f(b_{k+1}, \lambda_{k+1})x + \sum_{l=k+2}^{L}(f(b_l, \lambda_l) + b_l\lambda_l^K x)$$

$$q^*(x) = \sum_{l=1}^{k}(b_l^* + f(b_l^*, \lambda_l)x) + b_{k+1}^* + f(b_{k+1}^*, \lambda_{k+1})x + \sum_{l=k+2}^{L}(f(b_l^*, \lambda_l) + b_l^*\lambda_l^K x)$$

Which imply

$$b_{k+1} + \sum_{l=k+2}^{L} f(b_l, \lambda_l) = b_{k+1}^* + \sum_{l=k+2}^{L} f(b_l^*, \lambda_l)$$

$$b_{k+1} - f(b_{k+1}, \lambda_{k+1}) = b_{k+1}^* - f(b_{k+1}^*, \lambda_{k+1})$$

which imply

$$b_{k+1} = b_{k+1}^*$$

so by induction all $b_l = b_l^*$. □

Let $\widehat{q}(x_i)$ be the estimation of $q(x_i)$, and $\widehat{\mathbf{b}} = (\widehat{b}_1, \ldots \widehat{b}_L)'$, $\widehat{\boldsymbol{\lambda}} = (\widehat{\lambda}_1, \ldots \widehat{\lambda}_L)'$, $\widehat{\boldsymbol{\phi}} = (\widehat{\phi}_1, \ldots \widehat{\phi}_L)'$ be estimations of $\mathbf{b} = (b_1, \ldots b_L)'$, $\boldsymbol{\lambda} = (\lambda_1, \ldots \lambda_L)'$, $\boldsymbol{\phi} = (\phi_1, \ldots \phi_L)'$ respectively. Then the following theorem follows immediately.

Theorem 4.21 $(\widehat{\mathbf{b}}', \ \widehat{\boldsymbol{\lambda}}', \widehat{\boldsymbol{\phi}}') \xrightarrow{a.s.} (\mathbf{b}', \ \boldsymbol{\lambda}', \ \boldsymbol{\phi}')$ *as* $n \to \infty$.

The transformation Jacobian matrix J can be easily calculated. And a theorem then follows.

Theorem 4.22

$$\sqrt{n}\Big(\begin{bmatrix}\widehat{\mathbf{b}}\\ \widehat{\boldsymbol{\lambda}}\\ \widehat{\boldsymbol{\phi}}\end{bmatrix} - \begin{bmatrix}\mathbf{b}\\ \boldsymbol{\lambda}\\ \boldsymbol{\phi}\end{bmatrix}\Big) \xrightarrow{d} N(0, J\Theta(\Sigma + \sum_{k=1}^{K_1+K_2+1}\{W_k + W_k'\})\Theta' J')$$

where $W_k,\ \Sigma,\ \Theta$ is defined in Corollary 4.8 with $m = L+1$.

4.4 Proof of Lemma 4.13

Suppose we have $x_1,\ x_2,\ x_3,\ x_4,\ x_5$ satisfy

$$x_1 < j_1 < x_2 < j_2 < x_3 < j_3 < x_4 < j_4 < x_5 \tag{4.54}$$

where $j_1,\ j_2,\ j_3,\ j_4$ are jumping points.

Since the order in index l does not matter, we assume that $\lambda_1 < \lambda_2$ if one of λ_1 and λ_2 is the biggest number among all jumping points and that $1/\phi_2 < 1/\phi_1$ if one of $1/\phi_1$ and $1/\phi_2$ is the biggest number among all jumping points.
When λ_2=max$(\lambda_1,\ \lambda_2,\ 1/\phi_1,\ 1/\phi_2)$ we have the following six possible combinations:

$$\begin{array}{ll}\Lambda_1: & \frac{1}{\phi_1} < \lambda_1 < \frac{1}{\phi_2} < \lambda_2\\ \Lambda_2: & \frac{1}{\phi_2} < \lambda_1 < \frac{1}{\phi_1} < \lambda_2\\ \Lambda_3: & \frac{1}{\phi_1} < \frac{1}{\phi_2} < \lambda_1 < \lambda_2\\ \Lambda_4: & \frac{1}{\phi_2} < \frac{1}{\phi_1} < \lambda_1 < \lambda_2\\ \Lambda_5: & \lambda_1 < \frac{1}{\phi_1} < \frac{1}{\phi_2} < \lambda_2\\ \Lambda_6: & \lambda_1 < \frac{1}{\phi_2} < \frac{1}{\phi_1} < \lambda_2\end{array}$$

When $1/\phi_1$=max$(\lambda_1,\ \lambda_2,\ 1/\phi_1,\ 1/\phi_2)$ we have the following six possible combinations:

$$\begin{array}{ll}\Phi_1: & \lambda_1 < \lambda_2 < \frac{1}{\phi_2} < \frac{1}{\phi_1}\\ \Phi_2: & \lambda_2 < \lambda_1 < \frac{1}{\phi_2} < \frac{1}{\phi_1}\\ \Phi_3: & \lambda_1 < \frac{1}{\phi_2} < \lambda_2 < \frac{1}{\phi_1}\\ \Phi_4: & \lambda_2 < \frac{1}{\phi_2} < \lambda_1 < \frac{1}{\phi_1}\\ \Phi_5: & \frac{1}{\phi_2} < \lambda_1 < \lambda_2 < \frac{1}{\phi_1}\\ \Phi_6: & \frac{1}{\phi_2} < \lambda_2 < \lambda_1 < \frac{1}{\phi_1}\end{array}$$

The notation Λ_1^* in this section means

$$\Lambda_1^*: \ \frac{1}{\phi_1^*} < \lambda_1^* < \frac{1}{\phi_2^*} < \lambda_2^*$$

and similarly for other notations if used.

When Λ_1 is satisfied, we have

$$x_1 < 1/\phi_1 < x_2 < \lambda_1 < x_3 < 1/\phi_2 < x_4 < \lambda_2 < x_5$$

which gives

$$\begin{aligned}
&q_1(x_1) = b_1\lambda_1^{K_2}x_1 + 1\\
&q_1(x_2) = b_1[\lambda_1^{K_2}x_2 + (\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + x_2(\phi_1 + \cdots + \phi_1^{K_1}) + \phi_1^{K_1}]\\
&q_1(x_3) = x_3 + b_1\phi_1^{K_1}\\
&q_1(x_4) = x_4 + b_1\phi_1^{K_1}\\
&q_1(x_5) = x_5 + b_1\phi_1^{K_1}\\
&q_2(x_1) = b_2\lambda_2^{K_2}x_1 + 1\\
&q_2(x_2) = b_2\lambda_2^{K_2}x_2 + 1\\
&q_2(x_3) = b_2\lambda_2^{K_2}x_3 + 1\\
&q_2(x_4) = b_2[\lambda_2^{K_2}x_4 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_4(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
&q_2(x_5) = x_5 + b_2\phi_2^{K_1}
\end{aligned}$$

When Λ_2 is satisfied, we have

$$x_1 < 1/\phi_2 < x_2 < \lambda_1 < x_3 < 1/\phi_1 < x_4 < \lambda_2 < x_5$$

which gives

$$\begin{aligned}
&q_1(x_1) = b_1\lambda_1^{K_2}x_1 + 1\\
&q_1(x_2) = b_1\lambda_1^{K_2}x_2 + 1\\
&q_1(x_3) = b_1[x_3(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
&q_1(x_4) = x_4 + b_1\phi_1^{K_1}\\
&q_1(x_5) = x_5 + b_1\phi_1^{K_1}\\
&q_2(x_1) = b_2\lambda_2^{K_2}x_1 + 1\\
&q_2(x_2) = b_2[\lambda_2^{K_2}x_2 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_2(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
&q_2(x_3) = b_2[\lambda_2^{K_2}x_3 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_3(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
&q_2(x_4) = b_2[\lambda_2^{K_2}x_4 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_4(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
&q_2(x_5) = x_5 + b_2\phi_2^{K_1}
\end{aligned}$$

When Λ_3 is satisfied, we have

$$x_1 < 1/\phi_1 < x_2 < 1/\phi_2 < x_3 < \lambda_1 < x_4 < \lambda_2 < x_5$$

which gives

$$
\begin{aligned}
q_1(x_1) &= b_1\lambda_1^{K_2}x_1 + 1\\
q_1(x_2) &= b_1[\lambda_1^{K_2}x_2 + (\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + x_2(\phi_1 + \cdots + \phi_1^{K_1}) + \phi_1^{K_1}]\\
q_1(x_3) &= b_1[\lambda_1^{K_2}x_3 + (\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + x_3(\phi_1 + \cdots + \phi_1^{K_1}) + \phi_1^{K_1}]\\
q_1(x_4) &= x_4 + b_1\phi_1^{K_1}\\
q_1(x_5) &= x_5 + b_1\phi_1^{K_1}\\
q_2(x_1) &= b_2\lambda_2^{K_2}x_1 + 1\\
q_2(x_2) &= b_2\lambda_2^{K_2}x_2 + 1\\
q_2(x_3) &= b_2[\lambda_2^{K_2}x_3 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_3(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
q_2(x_4) &= b_2[\lambda_2^{K_2}x_4 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_4(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
q_2(x_5) &= x_5 + b_2\phi_2^{K_1}
\end{aligned}
$$

When Λ_4 is satisfied, we have

$$x_1 < 1/\phi_2 < x_2 < 1/\phi_1 < x_3 < \lambda_1 < x_4 < \lambda_2 < x_5$$

which gives

$$
\begin{aligned}
q_1(x_1) &= b_1\lambda_1^{K_2}x_1 + 1\\
q_1(x_2) &= b_1\lambda_1^{K_2}x_2 + 1\\
q_1(x_3) &= b_1[\lambda_1^{K_2}x_3 + (\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + x_3(\phi_1 + \cdots + \phi_1^{K_1}) + \phi_1^{K_1}]\\
q_1(x_4) &= x_4 + b_1\phi_1^{K_1}\\
q_1(x_5) &= x_5 + b_1\phi_1^{K_1}\\
q_2(x_1) &= b_2\lambda_2^{K_2}x_1 + 1\\
q_2(x_2) &= b_2[\lambda_2^{K_2}x_2 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_2(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
q_2(x_3) &= b_2[\lambda_2^{K_2}x_3 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_3(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
q_2(x_4) &= b_2[\lambda_2^{K_2}x_4 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_4(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
q_2(x_5) &= x_5 + b_2\phi_2^{K_1}
\end{aligned}
$$

When Λ_5 is satisfied, we have

$$x_1 < \lambda_1 < x_2 < 1/\phi_1 < x_3 < 1/\phi_2 < x_4 < \lambda_2 < x_5$$

which gives

$$
\begin{aligned}
q_1(x_1) &= b_1\lambda_1^{K_2}x_1 + 1\\
q_1(x_2) &= b_1[x_2(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
q_1(x_3) &= x_3 + b_1\phi_1^{K_1}\\
q_1(x_4) &= x_4 + b_1\phi_1^{K_1}\\
q_1(x_5) &= x_5 + b_1\phi_1^{K_1}\\
q_2(x_1) &= b_2\lambda_2^{K_2}x_1 + 1\\
q_2(x_2) &= b_2\lambda_2^{K_2}x_2 + 1\\
q_2(x_3) &= b_2\lambda_2^{K_2}x_3 + 1\\
q_2(x_4) &= b_2[\lambda_2^{K_2}x_4 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_4(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
q_2(x_5) &= x_5 + b_2\phi_2^{K_1}
\end{aligned}
$$

When Λ_6 is satisfied, we have

$$x_1 < \lambda_1 < x_2 < 1/\phi_2 < x_3 < 1/\phi_1 < x_4 < \lambda_2 < x_5$$

which gives

$$\begin{aligned}
q_1(x_1) &= b_1\lambda_1^{K_2}x_1 + 1\\
q_1(x_2) &= b_1[x_2(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
q_1(x_3) &= b_1[x_3(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
q_1(x_4) &= x_4 + b_1\phi_1^{K_1}\\
q_1(x_5) &= x_5 + b_1\phi_1^{K_1}\\
q_2(x_1) &= b_2\lambda_2^{K_2}x_1 + 1\\
q_2(x_2) &= b_2\lambda_2^{K_2}x_2 + 1\\
q_2(x_3) &= b_2[\lambda_2^{K_2}x_3 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_3(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
q_2(x_4) &= b_2[\lambda_2^{K_2}x_4 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_4(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
q_2(x_5) &= x_5 + b_2\phi_2^{K_1}
\end{aligned}$$

When Φ_1 is satisfied, we have

$$x_1 < \lambda_1 < x_2 < \lambda_2 < x_3 < 1/\phi_2 < x_4 < 1/\phi_1 < x_5$$

which gives

$$\begin{aligned}
q_1(x_1) &= b_1\lambda_1^{K_2}x_1 + 1\\
q_1(x_2) &= b_1[x_2(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
q_1(x_3) &= b_1[x_3(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
q_1(x_4) &= b_1[x_4(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
q_1(x_5) &= x_5 + b_1\phi_1^{K_1}\\
q_2(x_1) &= b_2\lambda_2^{K_2}x_1 + 1\\
q_2(x_2) &= b_2\lambda_2^{K_2}x_2 + 1\\
q_2(x_3) &= b_2[x_3(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + 1) + (1 + \phi_2 + \cdots + \phi_2^{K_1})]\\
q_2(x_4) &= x_4 + b_2\phi_2^{K_1}\\
q_2(x_5) &= x_5 + b_2\phi_2^{K_1}
\end{aligned}$$

When Φ_2 is satisfied, we have

$$x_1 < \lambda_2 < x_2 < \lambda_1 < x_3 < 1/\phi_2 < x_4 < 1/\phi_1 < x_5$$

which gives

$$
\begin{aligned}
q_1(x_1) &= b_1\lambda_1^{K_2}x_1 + 1\\
q_1(x_2) &= b_1\lambda_1^{K_2}x_2 + 1\\
q_1(x_3) &= b_1[x_3(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
q_1(x_4) &= b_1[x_4(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
q_1(x_5) &= x_5 + b_1\phi_1^{K_1}\\
q_2(x_1) &= b_2\lambda_2^{K_2}x_1 + 1\\
q_2(x_2) &= b_2[x_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + 1) + (1 + \phi_2 + \cdots + \phi_2^{K_1})]\\
q_2(x_3) &= b_2[x_3(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + 1) + (1 + \phi_2 + \cdots + \phi_2^{K_1})]\\
q_2(x_4) &= x_4 + b_2\phi_2^{K_1}\\
q_2(x_5) &= x_5 + b_2\phi_2^{K_1}
\end{aligned}
$$

When Φ_3 is satisfied, we have

$$x_1 < \lambda_1 < x_2 < 1/\phi_2 < x_3 < \lambda_2 < x_4 < 1/\phi_1 < x_5$$

which gives

$$
\begin{aligned}
q_1(x_1) &= b_1\lambda_1^{K_2}x_1 + 1\\
q_1(x_2) &= b_1[x_2(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
q_1(x_3) &= b_1[x_3(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
q_1(x_4) &= b_1[x_4(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
q_1(x_5) &= x_5 + b_1\phi_1^{K_1}\\
q_2(x_1) &= b_2\lambda_2^{K_2}x_1 + 1\\
q_2(x_2) &= b_2\lambda_2^{K_2}x_2 + 1\\
q_2(x_3) &= b_2[\lambda_2^{K_2}x_3 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_3(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
q_2(x_4) &= x_4 + b_2\phi_2^{K_1}\\
q_2(x_5) &= x_5 + b_2\phi_2^{K_1}
\end{aligned}
$$

When Φ_4 is satisfied, we have

$$x_1 < \lambda_2 < x_2 < 1/\phi_2 < x_3 < \lambda_1 < x_4 < 1/\phi_1 < x_5$$

which gives

$$
\begin{aligned}
q_1(x_1) &= b_1\lambda_1^{K_2}x_1 + 1\\
q_1(x_2) &= b_1\lambda_1^{K_2}x_2 + 1\\
q_1(x_3) &= b_1\lambda_1^{K_2}x_3 + 1\\
q_1(x_4) &= b_1[x_4(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
q_1(x_5) &= x_5 + b_1\phi_1^{K_1}\\
q_2(x_1) &= b_2\lambda_2^{K_2}x_1 + 1\\
q_2(x_2) &= b_2[x_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + 1) + (1 + \phi_2 + \cdots + \phi_2^{K_1})]\\
q_2(x_3) &= x_3 + b_2\phi_2^{K_1}\\
q_2(x_4) &= x_4 + b_2\phi_2^{K_1}\\
q_2(x_5) &= x_5 + b_2\phi_2^{K_1}
\end{aligned}
$$

When Φ_5 is satisfied, we have

$$x_1 < 1/\phi_2 < x_2 < \lambda_1 < x_3 < \lambda_2 < x_4 < 1/\phi_1 < x_5$$

which gives

$$\begin{aligned}
q_1(x_1) &= b_1\lambda_1^{K_2}x_1 + 1\\
q_1(x_2) &= b_1\lambda_1^{K_2}x_2 + 1\\
q_1(x_3) &= b_1[x_3(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
q_1(x_4) &= b_1[x_4(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
q_1(x_5) &= x_5 + b_1\phi_1^{K_1}\\
q_2(x_1) &= b_2\lambda_2^{K_2}x_1 + 1\\
q_2(x_2) &= b_2[\lambda_2^{K_2}x_2 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_2(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
q_2(x_3) &= b_2[\lambda_2^{K_2}x_3 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_3(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
q_2(x_4) &= x_4 + b_2\phi_2^{K_1}\\
q_2(x_5) &= x_5 + b_2\phi_2^{K_1}
\end{aligned}$$

When Φ_6 is satisfied, we have

$$x_1 < 1/\phi_2 < x_2 < \lambda_2 < x_3 < \lambda_1 < x_4 < 1/\phi_1 < x_5$$

which gives

$$\begin{aligned}
q_1(x_1) &= b_1\lambda_1^{K_2}x_1 + 1\\
q_1(x_2) &= b_1\lambda_1^{K_2}x_2 + 1\\
q_1(x_3) &= b_1\lambda_1^{K_2}x_3 + 1\\
q_1(x_4) &= b_1[x_4(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + (1 + \phi_1 + \cdots + \phi_1^{K_1})]\\
q_1(x_5) &= x_5 + b_1\phi_1^{K_1}\\
q_2(x_1) &= b_2\lambda_2^{K_2}x_1 + 1\\
q_2(x_2) &= b_2[\lambda_2^{K_2}x_2 + (\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + x_2(\phi_2 + \cdots + \phi_2^{K_1}) + \phi_2^{K_1}]\\
q_2(x_3) &= x_3 + b_2\phi_2^{K_1}\\
q_2(x_4) &= x_4 + b_2\phi_2^{K_1}\\
q_2(x_5) &= x_5 + b_2\phi_2^{K_1}
\end{aligned}$$

We need to show that if

$$q(x) = q^*(x)$$

for all x then

$$(1/\phi_1,\ 1/\phi_2,\ \lambda_1,\ \lambda_2,) = (1/\phi_1^*,\ 1/\phi_2^*,\ \lambda_1^*,\ \lambda_2^*) \tag{4.55}$$

where $q(x)$ corresponds to $(1/\phi_1,\ 1/\phi_2,\ \lambda_1,\ \lambda_2,)$, while $q^*(x)$ corresponds to $(1/\phi_1^*,\ 1/\phi_2^*,\ \lambda_1^*,\ \lambda_2^*)$. In other words we need to show every combination from all combinations Λ_i and Λ_j^*, Λ_i and Φ_j^*, Φ_i and Φ_j^* results in a contradiction or an impossible result. The total combinations are $C_6^2 + 6 \times 6 + C_6^2 = 15 + 36 + 15 = 66$. We start from Λ_i and Λ_j^*, then Φ_i and Φ_j^*, and finally Λ_i and Φ_j^*.

For all combinations we have

$$q(x_1) = q^*(x_1) \;\Rightarrow\; b_1\lambda_1^{K_2} + b_2\lambda_2^{K_2} = b_1^*\lambda_1^{*K_2} + b_2^*\lambda_2^{*K_2}, \tag{4.56}$$

$$q(x_5) = q^*(x_5) \;\Rightarrow\; b_1\phi_1^{K_1} + b_2\phi_2^{K_1} = b_1^*\phi_1^{*K_1} + b_2^*\phi_2^{*K_1}. \tag{4.57}$$

Case Λ_1 **and** Λ_2^*: From $q(x_3) = q^*(x_3)$, we get

$$\sum_{l=1}^{2} b_l^*[\lambda_l^{*K_2} + (1 + \phi_l^* + \cdots + \phi_l^{*K_1})] = b_2\lambda_2^{K_2} + 1,$$

$$\sum_{l=1}^{2} b_l^*[\phi_l^{*K_1} + (\lambda_l^{*K_2} + \cdots + \lambda_l^* + 1)] = b_1\phi_1^{K_1} + 1.$$

But this is not possible since the LHS is less than 1, while the RHS is greater than 1.
Case Λ_1 **and** Λ_3^*: Use the same arguments as in Case Λ_1 and Λ_2^*.
Case Λ_1 **and** Λ_4^*: Use the same arguments as in Case Λ_1 and Λ_2^*.
Case Λ_1 **and** Λ_5^*: From $q(x_2) = q^*(x_2)$, we get

$$\sum_{l=1}^{2} b_l[(\lambda_l^{*K_2} + \cdots + \lambda_l^*) + \phi_l^{*K_1}] = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1}) + 1$$

which is not possible for the same reason as in the Case Λ_1 and Λ_2^*.
Case Λ_1 **and** Λ_6^*: Use the same arguments as in Case Λ_1 and Λ_5^*.
Case Λ_2 **and** Λ_3^*: We have

$$\phi_2 = \phi_1^*,\; \lambda_1 = \frac{1}{\phi_2^*},\; \phi_1 = \frac{1}{\lambda_1^*},\; \lambda_2 = \lambda_2^*. \tag{4.58}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1^*(\phi_1^* + \cdots + \phi_1^{*K_1}) \tag{4.59}$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + b_2\phi_2^{K_1} = b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + \lambda_1^*) + b_1^*\phi_1^{*K_1} \tag{4.60}$$

From (4.58) and (4.59) we get $b_2 = b_1^*$. And from (4.60) we get $\lambda_2 = \lambda_1^*$, which is a contradiction to $\lambda_1^* < \lambda_2^*$.
Case Λ_2 **and** Λ_4^*: We have

$$\phi_2 = \phi_2^*,\; \lambda_1 = \frac{1}{\phi_1^*},\; \phi_1 = \frac{1}{\lambda_1^*},\; \lambda_2 = \lambda_2^*.$$

From $q(x_2) = q^*(x_2)$, we get

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + b_2\phi_2^{K_1} = b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) + b_2^*\phi_2^{*K_1}$$

which implies $b_2 = b_2^*$, and $q_2(x) = q_2^*(x)$. So $q_1(x) = q_1^*(x)$ which can be viewed as a case of $L = 1$ which has been shown to have uniqueness and it required $\lambda_1 = \phi_1$.
Case Λ_2 and Λ_5^*: We have

$$\phi_1 = \phi_2^*,\ \lambda_1 = \frac{1}{\phi_1^*},\ \phi_2 = \frac{1}{\lambda_1^*},\ \lambda_2 = \lambda_2^*.$$

From $q(x_2) = q^*(x_2)$, we get

$$b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1})$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + b_2\phi_2^{K_1} = b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) + b_2^*\phi_2^{*K_2}$$

From $q(x_4) = q^*(x_4)$, we get

$$b_2\lambda_2^{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_2^*\lambda_2^{*K_2} + b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1})$$

So $b_2 = b_2^*$ and $\phi_2 = \phi_2^* = \phi_1$, which is a contradiction to $\phi_1 \neq \phi_2$.
Case Λ_2 and Λ_6^*: From $q(x_2) = q^*(x_2)$, we get

$$\begin{aligned} &b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + b_2\phi_2^{K_1} \\ &\quad = b_1^*(1 + \cdots + \phi_1^{*K_1}) + b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) + b_2^*\phi_2^{*K_1} \end{aligned} \tag{4.61}$$

From $q(x_3) = q^*(x_3)$, we get

$$\begin{aligned} &b_1(1 + \phi_1 + \cdots + \phi_1^{K_1-1}) + b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) \\ &\quad = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) + b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) \end{aligned} \tag{4.62}$$

By subtracting (4.62) from (4.61) we get

$$b_1^*\phi_1^{*K_1} + b_2^*\phi_2^{*K_1} = b_2\phi_2^{K_1} - b_1(1 + \phi_1 + \cdots + \phi_1^{K_1-1})$$

which is not possible because of (4.57).
Case Λ_3 and Λ_4^*: We have

$$\phi_1 = \phi_2^*,\ \phi_2 = \phi_1^*,\ \lambda_1 = \lambda_1^*,\ \lambda_2 = \lambda_2^*.$$

From $q(x_2) = q^*(x_2)$, we get

$$b_1(\phi_1 + \cdots + \phi_1^{K_1}) = b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1})$$

$$b_1(\lambda_1^{K_2}+\lambda_1^{K_2-1}+\cdots+\lambda_1)+b_1\phi_1^{K_1}=b_2^*(\lambda_2^{*K_2}+\lambda_2^{*K_2-1}+\cdots+\lambda_2^*)+b_2\phi_2^{*K_1}$$

which imply $\lambda_1=\lambda_2^*=\lambda_2$. But this is a contradiction to $\lambda_1<\lambda_2$.

Case Λ_3 **and** Λ_5^*: From $q(x_3)=q^*(x_3)$, then follow the same arguments in Λ_1 and Λ_2^*.

Case Λ_3 **and** Λ_6^*: We have

$$\phi_1=\frac{1}{\lambda_1^*},\ \phi_2=\phi_2^*,\ \lambda_1=\frac{1}{\phi_1^*},\ \lambda_2=\lambda_2^*. \tag{4.63}$$

From $q(x_2)=q^*(x_2)$, we get

$$b_1(\phi_1+\cdots+\phi_1^{K_1})=b_1^*(\lambda_1^{*K_2-1}+\cdots+1) \tag{4.64}$$

$$b_1(\lambda_1^{K_2}+\lambda_1^{K_2-1}+\cdots+\lambda_1)+b_1\phi_1^{K_1}=b_1^*(1+\phi_1^*+\cdots+\phi_1^{*K_1}) \tag{4.65}$$

From $q(x_3)=q^*(x_3)$, we get

$$\begin{aligned}b_1(\phi_1+\cdots+\phi_1^{K_1})+b_2(\phi_2+\cdots+\phi_2^{K_1})\\ =b_1^*(\lambda_1^{*K_2-1}+\cdots+1)+b_2^*(\phi_2^*+\cdots+\phi_2^{*K_1})\end{aligned} \tag{4.66}$$

$$\begin{aligned}b_1(\lambda_1^{K_2}+\lambda_1^{K_2-1}+\cdots+\lambda_1)+b_2(\lambda_2^{K_2}+\lambda_2^{K_2-1}+\cdots+\lambda_2)\\ =b_1^*(1+\phi_1^*+\cdots+\phi_1^{*K_1-1})+b_2^*(\lambda_2^{*K_2}+\lambda_2^{*K_2-1}+\cdots+\lambda_2^*)\end{aligned} \tag{4.67}$$

(4.63), (4.64), (4.66) imply $b_2=b_2^*$, so $q_2(x)=q_2^*(x)$. Thus $q_1(x)=q_1^*(x)$ which can be viewed as Case $L=1$.

Case Λ_4 **and** Λ_5^*: From $q(x_3)=q^*(x_3)$, then follow the same arguments in Λ_1 and Λ_2^*.

Case Λ_4 **and** Λ_6^*: We have

$$\phi_1=\phi_2^*,\ \phi_2=\frac{1}{\lambda_1^*},\ \lambda_1=\frac{1}{\phi_1^*},\ \lambda_2=\lambda_2^*.$$

From $q(x_4)=q^*(x_4)$, we get

$$b_2(\lambda_2^{K_2}+\lambda_2^{K_2-1}+\cdots+\lambda_2)=b_2^*(\lambda_2^{*K_2}+\lambda_2^{*K_2-1}+\cdots+\lambda_2^*)$$

$$b_2(\phi_2^{K_2}+\phi_2^{K_2-1}+\cdots+\phi_2)=b_2^*(\phi_2^{*K_2}+\phi_2^{*K_2-1}+\cdots+\phi_2^*)$$

which imply $b_2=b_2^*$ and $\phi_2=\phi_2^*=\phi_1$, which is a contradiction to $\phi_1\neq\phi_2$.

Case Λ_5 **and** Λ_6^*: From $q(x_3)=q^*(x_3)$, then follow the same arguments in Λ_1 and Λ_2^*.

Case Φ_1 **and** Φ_2^*: We have

$$\lambda_1=\lambda_2^*,\ \lambda_2=\lambda_1^*,\ \phi_1=\phi_1^*,\ \phi_2=\phi_2^*.$$

From $q(x_2) = q^*(x_2)$, we get

$$b_1(\lambda_1^{K_2-1} + \lambda_1^{K_2-2} + \cdots + 1) = b_2^*(\lambda_2^{*K_2-1} + \lambda_2^{*K_2-2} + \cdots + 1)$$

$$b_1(1 + \phi_1 + \cdots + \phi_1^{K_1}) = b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1})$$

which imply $b_1 = b_2^*$, $\phi_1 = \phi_2^* = \phi_2$. But this is a contradiction to $\phi_1 < \phi_2$.

Case Φ_1 and Φ_3^*: We have

$$\lambda_1 = \lambda_1^*,\ \lambda_2 = 1/\phi_2^*,\ \phi_1 = \phi_1^*,\ \phi_2 = 1/\lambda_2^*.$$

From $q(x_2) = q^*(x_2)$, we get

$$b_1(\lambda_1^{K_2-1} + \lambda_1^{K_2-2} + \cdots + 1) = b_1^*(\lambda_1^{*K_2-1} + \lambda_1^{*K_2-2} + \cdots + 1)$$

$$b_1(1 + \phi_1 + \cdots + \phi_1^{K_1}) = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1})$$

which imply $b_1 = b_1^*$, $\lambda_1 = \lambda_1^* = \lambda_2$. But this is a contradiction to $\lambda_1 \neq \lambda_2$.

Case Φ_1 and Φ_4^*: From $q(x_3) = q^*(x_3)$, we get

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + 1) = b_1^*\lambda_1^{K_2} + 1$$

which is an impossible result.

Case Φ_1 and Φ_5^*: We have

$$\lambda_1 = 1/\phi_2^*,\ \lambda_2 = \lambda_1^*,\ \phi_1 = \phi_1^*,\ \phi_2 = 1/\lambda_2^*. \tag{4.68}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_1(\lambda_1^{K_2-1} + \lambda_1^{K_2-2} + \cdots + 1) = b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1}) \tag{4.69}$$

$$b_1(1 + \phi_1 + \cdots + \phi_1^{K_1}) = b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) + b_2^*\phi_2^{*K_1} \tag{4.70}$$

From $q(x_3) = q^*(x_3)$, we get

$$\begin{aligned} b_1(\lambda_1^{K_2-1} + \lambda_1^{K_2-2} + \cdots + 1) + b_2(\lambda_2^{K_2-1} + \lambda_2^{K_2-2} + \cdots + 1) \\ = b_1^*(\lambda_1^{*K_2-1} + \cdots + \lambda_1^*) + b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1}) \end{aligned} \tag{4.71}$$

$$\begin{aligned} b_1(1 + \phi_1 + \cdots + \phi_1^{K_1}) + b_2(1 + \phi_2 + \cdots + \phi_2^{K_1}) \\ = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1}) + b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) + b_2^*\phi_2^{*K-1} \end{aligned} \tag{4.72}$$

From (4.69)-(4.72) we get $b_1 = b_1^* = b_2$. From $q(x_4) = q^*(x_4)$, we get

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) = b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + 1) \tag{4.73}$$

which gives $\lambda_1 = \lambda_1^* = \lambda_2$. But this is a contradiction to $\lambda_1 \neq \lambda_2$.

Case Φ_1 and Φ_6^*: From $q(x_3) = q^*(x_3)$, we get

$$b_1(\lambda_1^{K_2-1} + \cdots + 1) + b_2(\lambda_2^{K_2-1} + \cdots + 1) = b_1^* \lambda_1^{K_2} + 1$$

which is not impossible for the same reason as in the Case Λ_1 and Λ_2^*.

Case Φ_2 and Φ_3^*: We have

$$\lambda_1 = 1/\phi_2^*,\ \lambda_2 = \lambda_1^*,\ \phi_1 = \phi_1^*,\ \phi_2 = 1/\lambda_2^*.$$

From $q(x_2) = q^*(x_2)$, we get

$$b_2(\lambda_2^{K_2-1} + \lambda_2^{K_2-2} + \cdots + 1) = b_1^*(\lambda_1^{*K_2-1} + \lambda_1^{*K_2-2} + \cdots + 1)$$

$$b_2(1 + \phi_2 + \cdots + \phi_2^{K_1}) = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1})$$

which give $b_2 = b_1^*$ and $\phi_2 = \phi_1^* = \phi_1$, which is a contradiction to $\phi_1 < \phi_2$.

Case Φ_2 and Φ_4^*: From $q(x_3) = q^*(x_3)$, then follow the same arguments in **Case Φ_1 and Φ_4^***.

Case Φ_2 and Φ_5^*: We have

$$\lambda_1 = \lambda_1^*,\ \lambda_2 = 1/\phi_2^*,\ \phi_1 = \phi_1^*,\ \phi_2 = 1/\lambda_2^*.$$

From $q(x_4) = q^*(x_4)$, we get

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) = b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + 1)$$

which gives $b_1 = b_1^*$, so $q_1(x_4) = q_1^*(x_4)$ and $q_2(x_4) = q_2^*(x_4)$. This is a case of $L = 1$.

Case Φ_2 and Φ_6^*: From $q(x_3) = q^*(x_3)$, then follow the same arguments in Case Φ_1 and Φ_6^*.

Case Φ_3 and Φ_4^*: From $q(x_3) = q^*(x_3)$, then follow the same arguments in Case Φ_1 and Φ_4^*.

Case Φ_3 and Φ_5^*: We have

$$\lambda_1 = 1/\phi_2^*,\ \lambda_2 = \lambda_2^*,\ \phi_1 = \phi_1^*,\ \phi_2 = 1/\lambda_1^*.$$

From $q(x_4) = q^*(x_4)$, we get

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) = b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + 1)$$

$$b_1(1 + \phi_1 + \cdots + \phi_1^{K_1-1}) = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1})$$

which give $b_1 = b_1^*$, $\lambda_1 = \lambda_1^*$, $\phi_2 = \phi_2^*$.

Case Φ_3 and Φ_6^*: From $q(x_3) = q^*(x_3)$, we get

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + b_2\lambda_2^{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1\lambda_1^{*K_2} + 1$$

which is not possible for the same reason as in the Case Λ_1 and Λ_2^*.

Case Φ_4 and Φ_5^*: From $q(x_3) = q^*(x_3)$, we get

$$b_1\lambda_1^{K_2} + 1 = b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + 1) + b_2^*\lambda_2^{*K_2} + b_2(\phi_2^* + \cdots + \phi_2^{*K_1})$$

which is not possible because the LHS is greater than 1, but the RHS is less than 1.

Case Φ_4 and Φ_6^*: We have

$$\lambda_1 = \lambda_1^*,\ \lambda_2 = 1/\phi_2^*,\ \phi_1 = \phi_1^*,\ \phi_2 = 1/\lambda_2^*.$$

From $q(x_4) = q^*(x_4)$, we get

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) = b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + 1)$$

which gives $b_1 = b_1^*$, so $q_1(x_4) = q_1^*(x_4)$ and hence $q_2(x_4) = q_2^*(x_4)$. This is the Case $L = 1$ and hence $\lambda_2 = \lambda_2^*$, $\phi_2 = \phi_2^*$.

Case Φ_5 and Φ_6^*: From $q(x_3) = q^*(x_3)$, we get

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + b_2\lambda_2^{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1\lambda_1^{*K_2} + 1$$

which is not possible for the same reason as in the Case Λ_1 and Λ_2^*.

Case Λ_1 and Φ_1^*: From $q(x_3) = q^*(x_3)$, we get

$$\sum_{l=1}^{2} b_l^*(\lambda_l^{*K_2} + \lambda_l^{*K_2-1} + \cdots + 1) = b_2\lambda_2^{K_2} + 1$$

which is not possible for the same reason as in the Case Λ_1 and Λ_2^*.

Case Λ_1 and Φ_2^*: Just as the case Λ_1 and Φ_1^*.

Case Λ_1 and Φ_3^*: From $q(x_3) = q^*(x_3)$, we get

$$b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + 1) + b_2^*\lambda_2^{*K_2} + b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1}) = b_2\lambda_2^{K_2} + 1$$

which is not possible for the same reason as in the Case Λ_1 and Λ_2^*.

Case Λ_1 and Φ_4^*: we have

$$\phi_1 = 1/\lambda_2^*,\ \lambda_1 = 1/\phi_2^*,\ \phi_2 = 1/\lambda_1^*,\ \lambda_2 = 1/\phi_1^* \tag{4.74}$$

From $q(x_3) = q^*(x_3)$, we get

$$b_1\phi_1^{K_1} = b_2^*\phi_2^{*K_1} \;\Rightarrow\; b_2\phi_2^{K_1} = b_1^*\phi_1^{*K_1} \tag{4.75}$$

$$b_2\lambda_2^{K_2} = b_1^*\lambda_1^{*K_2} \;\Rightarrow\; b_1\lambda_1^{K_2} = b_2^*\lambda_2^{*K_2} \tag{4.76}$$

(4.75) and (4.76) give

$$\frac{\phi_1^{K_1}}{\lambda_1^{K_2}} = \frac{\phi_2^{*K_1}}{\lambda_2^{*K_2}} = \frac{\phi_1^{K_2}}{\lambda_1^{K_1}}$$

which results in $K_1 = K_2$.

Suppose now $K_1 = K_2 = K$, From $q(x_2) = q^*(x_2)$, we get

$$b_1(\phi_1 + \cdots + \phi_1^{K_1}) = b_2^*(\lambda_1^{*K_2-1} + \cdots + 1) \tag{4.77}$$

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + b_1\phi_1^{K_1} = b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1}) \tag{4.78}$$

(4.78) is equivalent to

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) = b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1-1}) \tag{4.79}$$

(4.77) and (4.79) give

$$\frac{\phi_1 + \cdots + \phi_1^{K_1}}{\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1} = \frac{\lambda_1^{*K_2-1} + \lambda_1^{*K_2-1} + \cdots + 1}{1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}} = \frac{\lambda_1^{K-1}}{\phi_1^{K-1}}\frac{1 + \phi_1 + \cdots + \phi_1^{K-1}}{\lambda_1^{K-1} + \lambda_1^{K-2} + \cdots + 1}$$

which implies $\lambda_1 = \phi_1$, $\lambda_2^* = \phi_2^*$.

Similarly from $q(x_4) = q^*(x_4)$, we can get $\lambda_2 = \phi_2$, $\lambda_1^* = \phi_1^*$.

Summarize the above arguments, we get down to a symmetric case which we have proved the uniqueness of the parameters.

Case Λ_1 and Φ_5^*: From $q(x_3) = q^*(x_3)$, we get

$$b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + 1) + b_2^*\lambda_2^{*K_2} + b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1}) = b_2\lambda_2^{K_2} + 1$$

which is not possible for the same reason as in the Case Λ_1 and Λ_2^*.

Case Λ_1 and Φ_6^*: we have

$$\phi_1 = \phi_2^*,\ \lambda_1 = \lambda_2^*,\ \phi_2 = 1/\lambda_1^*,\ \lambda_2 = 1/\phi_1^*$$

From $q(x_2) = q^*(x_2)$, we get

$$b_1(\phi_1 + \cdots + \phi_1^{K_1}) = b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1})$$

which implies $b_1 = b_2^*$. And so $q_1(x_2) = q_2^*(x_2)$, $q_2(x_2) = q_1^*(x_2)$. And this is the case $L = 1$.

Case Λ_2 **and** Φ_1^*: we have

$$\phi_1 = \phi_2^*,\ \lambda_1 = \lambda_2^*,\ \phi_2 = 1/\lambda_1^*,\ \lambda_2 = 1/\phi_1^* \tag{4.80}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_2^*(\lambda_1^{*K_2-1} + \cdots + 1) \tag{4.81}$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + b_2\phi_2^{K_1} = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1}) \tag{4.82}$$

From $q(x_3) = q^*(x_3)$, we get

$$\begin{aligned} &b_1(\lambda_1^{K_2-1} + \cdots + 1) + b_2(\phi_2 + \cdots + \phi_2^{K_1}) \\ &\quad = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) + b_2^*(\lambda_2^{*K_2-1} + \cdots + 1) \end{aligned} \tag{4.83}$$

$$\begin{aligned} &b_1(1 + \phi_1 + \cdots + \phi_1^{K_1-1}) + b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) \\ &\quad = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) + b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1-1}) \end{aligned} \tag{4.84}$$

From $q(x_4) = q^*(x_4)$, we get

$$b_2\lambda_2^{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + 1) \tag{4.85}$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) \tag{4.86}$$

(4.80), (4.81) and (4.83) imply $b_1 = b_2^*$, then follow the same arguments as in Λ_1 and Φ_6^*.

Case Λ_2 **and** Φ_2^*: we have

$$\phi_1 = \phi_2^*,\ \lambda_1 = \lambda_1^*,\ \phi_2 = 1/\lambda_2^*,\ \lambda_2 = 1/\phi_1^* \tag{4.87}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_2^*(\lambda_2^{*K_2-1} + \cdots + 1) \tag{4.88}$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + b_2\phi_2^{K_1} = b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1}) \tag{4.89}$$

From $q(x_3) = q^*(x_3)$, we get

$$\begin{aligned} &b_1(\lambda_1^{K_2-1} + \cdots + 1) + b_2(\phi_2 + \cdots + \phi_2^{K_1}) \\ &\quad = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) + b_2^*(\lambda_2^{*K_2-1} + \cdots + 1) \end{aligned} \tag{4.90}$$

$$\begin{aligned} & b_1(1+\phi_1+\cdots+\phi_1^{K_1-1})+b_2(\lambda_2^{K_2}+\lambda_2^{K_2-1}+\cdots+\lambda_2) \\ & \quad = b_1^*(1+\phi_1^*+\cdots+\phi_1^{*K_1-1})+b_2^*(1+\phi_2^*+\cdots+\phi_2^{*K_1-1}) \end{aligned} \tag{4.91}$$

From $q(x_4)=q^*(x_4)$, we get

$$b_2\lambda_2^{K_2}+b_2(\phi_2+\cdots+\phi_2^{K_1})=b_1^*(\lambda_1^{*K_2}+\lambda_1^{*K_2-1}+\cdots+1) \tag{4.92}$$

$$b_2(\lambda_2^{K_2}+\lambda_2^{K_2-1}+\cdots+\lambda_2)=b_1^*(1+\phi_1^*+\cdots+\phi_1^{*K_1-1}) \tag{4.93}$$

(4.87), (4.91) and (4.93) imply $b_1=b_2^*$. (4.87), (4.88) and (4.90) imply $b_1=b_1^*$. Then (4.88) and (4.91) imply $\phi_1=\phi_1^*=\phi_2^*$, which is a contradiction to $\phi_1^*<\phi_2^*$.

Case Λ_2 and Φ_3^*: we have

$$\phi_1=1/\lambda_2^*,\ \lambda_1=1/\phi_2^*,\ \phi_2=1/\lambda_1^*,\ \lambda_2=1/\phi_1^* \tag{4.94}$$

From $q(x_2)=q^*(x_2)$, we get

$$b_2(\phi_2+\cdots+\phi_2^{K_1})=b_1^*(\lambda_1^{*K_2-1}+\cdots+1) \tag{4.95}$$

$$b_2(\lambda_2^{K_2}+\lambda_2^{K_2-1}+\cdots+\lambda_2)+b_2\phi_2^{K_1}=b_1^*(1+\phi_1^*+\cdots+\phi_1^{*K_1}) \tag{4.96}$$

From $q(x_3)=q^*(x_3)$, we get

$$\begin{aligned} & b_1(\lambda_1^{K_2-1}+\cdots+1)+b_2(\phi_2+\cdots+\phi_2^{K_1}) \\ & \quad = b_1^*(\lambda_1^{*K_2-1}+\cdots+1)+b_2^*(\phi_2^*+\cdots+\phi_2^{*K_1}) \end{aligned} \tag{4.97}$$

$$\begin{aligned} & b_1(1+\phi_1+\cdots+\phi_1^{K_1-1})+b_2(\lambda_2^{K_2}+\lambda_2^{K_2-1}+\cdots+\lambda_2) \\ & \quad = b_1^*(1+\phi_1^*+\cdots+\phi_1^{*K_1-1})+b_2^*(\lambda_2^{*K_2}+\lambda_2^{*K_2-1}+\cdots+\lambda_2^*) \end{aligned} \tag{4.98}$$

From $q(x_4)=q^*(x_4)$, we get

$$b_2\lambda_2^{K_2}+b_2(\phi_2+\cdots+\phi_2^{K_1})=b_1^*(\lambda_1^{*K_2}+\lambda_1^{*K_2-1}+\cdots+1) \tag{4.99}$$

$$b_2(\lambda_2^{K_2}+\lambda_2^{K_2-1}+\cdots+\lambda_2)=b_1^*(1+\phi_1^*+\cdots+\phi_1^{*K_1-1}) \tag{4.100}$$

From (4.95) and (4.99), we get

$$b_2\lambda_2^{K_2}=b_2^*\lambda_2^{*K_2} \tag{4.101}$$

From (4.96) and (4.100), we get

$$b_2\phi_2^{K_1}=b_2^*\phi_2^{*K_1} \tag{4.102}$$

From (4.101) and (4.102), we get

$$\lambda_2^{K_2}\phi_1^{*K_1}=\lambda_1^{*K_2}\phi_2^{K_1}$$

Suppose $K_2 > K_1$, since $\lambda_2\phi_1^* = \lambda_1^*\phi_2 = 1$, we have $\lambda_2 = \lambda_1^*$. So $\lambda_2 = 1/\phi_2$, which violates the condition $1/\phi_2 < \lambda_2$. Similar situation for $K_2 < K_1$. So $K_2 = K_1$.

Suppose now $K_1 = K_2 = K$, (4.95) and (4.97) give

$$b_1(\lambda_1^{K-1} + \cdots + 1) = b_2^*(\phi_2^* + \cdots + \phi_2^{*K}) \tag{4.103}$$

(4.98) and (4.100) give

$$b_1(1 + \phi_1 + \cdots + \phi_1^{K-1}) = b_2^*(\lambda_2^{*K} + \lambda_2^{*K-1} + \cdots + \lambda_2^*) \tag{4.104}$$

(4.103) and (4.104) give

$$\frac{\lambda_1^{K-1} + \cdots + 1}{1 + \phi_1 + \cdots + \phi_1^{K-1}} = \frac{\phi_2^* + \cdots + \phi_2^{*K}}{\lambda_2^{*K} + \lambda_2^{*K-1} + \cdots + \lambda_2^*}$$

which implies $\phi_1 = \lambda_1$ and then $\phi_2^* = \lambda_2^*$. (4.96) and (4.99) give

$$b_2(\lambda_2^K + \lambda_2^{K-1} + \cdots + 1) = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K-1})$$

$$b_2(\phi_2 + \cdots + \phi_2^K) = b_1^*(\lambda_1^{*K-1} + \cdots + 1)$$

which imply $\phi_2 = \lambda_2$ and $\phi_1^* = \lambda_1^*$. But this is a symmetric case.

Case Λ_2 and Φ_4^*: we have

$$\phi_1 = 1/\lambda_1^*,\ \lambda_1 = 1/\phi_2^*,\ \phi_2 = 1/\lambda_2^*,\ \lambda_2 = 1/\phi_1^* \tag{4.105}$$

From $q(x_3) = q^*(x_3)$, we get

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + b_2\lambda_2^{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = 1 + b_1^*\lambda_1^{K_2}$$

which is not possible for the same reason as in the Case Λ_1 and Λ_2^*.

Case Λ_2 and Φ_5^*: we have

$$\phi_1 = 1/\lambda_2^*,\ \lambda_1 = \lambda_1^*,\ \phi_2 = \phi_2^*,\ \lambda_2 = 1/\phi_1^* \tag{4.106}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1}) \tag{4.107}$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + b_2\phi_2^{K_1} = b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) + b_2^*\phi_2^{*K_1} \tag{4.108}$$

(4.106) and (4.107) gives $b_2 = b_2^*$. Then (4.108) gives $\lambda_2 = \lambda_2^*$ which violates condition $\lambda_2^* < 1/\phi_1^*$.

Case Λ_2 **and** Φ_6^*: Just as Λ_2 and Φ_4^*, from $q(x_3) = q^*(x_3)$, we get

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + b_2\lambda_2^{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = 1 + b_1^*\lambda_1^{K_2}$$

which is not possible for the same reason as in the Case Λ_1 and Λ_2^*.
Case Λ_3 **and** Φ_1^*: we have

$$\phi_1 = 1/\lambda_1^*,\ \lambda_1 = 1/\phi_2^*,\ \phi_2 = 1/\lambda_2^*,\ \lambda_2 = 1/\phi_1^* \tag{4.109}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_1(\phi_1 + \cdots + \phi_1^{K_1}) = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) \tag{4.110}$$

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + b_1\phi_1^{K_1} = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1}) \tag{4.111}$$

From $q(x_3) = q^*(x_3)$, we get

$$\begin{aligned} b_1(\phi_1 + \cdots + \phi_1^{K_1}) + b_2(\phi_2 + \cdots + \phi_2^{K_1}) \\ = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) + b_2^*(\lambda_2^{*K_2-1} + \cdots + \lambda_2^* + 1) \end{aligned} \tag{4.112}$$

$$\begin{aligned} b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) \\ = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) + b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1-1}) \end{aligned} \tag{4.113}$$

From $q(x_4) = q^*(x_4)$, we get

$$b_2\lambda_2^{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + 1) \tag{4.114}$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) \tag{4.115}$$

From (4.110) and (4.112) we get

$$b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_2^*(\lambda_2^{*K_2-1} + \cdots + 1) \tag{4.116}$$

From (4.113) and (4.115) we get

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) = b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1-1}) \tag{4.117}$$

From (4.110), (4.111), (4.114) and (4.115) we get

$$b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + \lambda_1^*) + b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1}) = \frac{1}{2} \tag{4.118}$$

$$b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) + b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1}) = \frac{1}{2} \tag{4.119}$$

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + b_1(1 + \phi_1 + \cdots + \phi_1^{K_1}) = \frac{1}{2} \tag{4.120}$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + b_2(1 + \phi_2 + \cdots + \phi_2^{K_1}) = \frac{1}{2} \tag{4.121}$$

From (4.120), we know that when $\lambda_1,\ \phi_1$ are given, then b_1 can be determined. From (4.120) and (4.110), we determine b_1^*. From (4.115), if additionally λ_2 is also given, then b_2 can be determined. In other words, provided that $\lambda_1,\ \phi_1,\ \lambda_2$ are known, then b_1 and b_2 can be determined, even ϕ_2 can be determined also. But it is not possible to determine the true values of b_2 and ϕ_2 based on $\lambda_1,\ \phi_1,\ \lambda_2$ since we can use only equation (4.121) which has two unknown variables.

Case Λ_3 and Φ_2^*: we have

$$\phi_1 = 1/\lambda_2^*,\ \lambda_1 = 1/\phi_2^*,\ \phi_2 = 1/\lambda_1^*,\ \lambda_2 = 1/\phi_1^* \tag{4.122}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_1(\phi_1 + \cdots + \phi_1^{K_1}) = b_2^*(\lambda_1^{*K_2-1} + \cdots + 1) \tag{4.123}$$

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + b_1\phi_1^{K_1} = b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1}) \tag{4.124}$$

From $q(x_3) = q^*(x_3)$, we get

$$\begin{aligned} &b_1(\phi_1 + \cdots + \phi_1^{K_1}) + b_2(\phi_2 + \cdots + \phi_2^{K_1}) \\ &\quad = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) + b_2^*(\lambda_2^{*K_2-1} + \cdots + \lambda_2^* + 1) \end{aligned} \tag{4.125}$$

$$\begin{aligned} &b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) \\ &\quad = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) + b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1-1}) \end{aligned} \tag{4.126}$$

From $q(x_4) = q^*(x_4)$, we get

$$b_2\lambda_2^{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + 1) \tag{4.127}$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) \tag{4.128}$$

From (4.126) and (4.128), we get

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) = b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1-1}) \tag{4.129}$$

From (4.124) and (4.129), we get

$$b_1\phi_1^{K_1} = b_2^*\phi_2^{*K_1} \ \Rightarrow\ b_2\phi_2^{K_1} = b_1^*\phi_1^{*K_1} \tag{4.130}$$

From (4.123) and (4.125), we get

$$b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) \tag{4.131}$$

From (4.127) and (4.131), we get

$$b_2\lambda_2^{K_2} = b_1^*\lambda_1^{*K_2} \;\Rightarrow\; b_1\lambda_1^{K_2} = b_2^*\lambda_2^{*K_2} \tag{4.132}$$

From (4.130) and (4.132), we get

$$\frac{\phi_1^{K_1}}{\lambda_1^{K_2}} = \frac{\phi_2^{*K_1}}{\lambda_2^{*K_2}} \;\Rightarrow\; K_1 = K_2 = K$$

Suppose now $K_1 = K_2 = K$, (4.128) and (4.131) give

$$\frac{\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2}{\phi_2 + \cdots + \phi_2^{K_1}} = \frac{1 + \phi_2^* + \cdots + \phi_2^{*K_1-1}}{\lambda_1^{*K_2-1} + \cdots + 1}$$

which implies $\lambda_2 = \phi_2$ and $\lambda_1^* = \phi_1^*$.

Similarly (4.123) and (4.129) imply $\lambda_1 = \phi_1$ and $\lambda_2^* = \phi_2^*$. So the cases here correspond to symmetric cases.

Case Λ_3 and Φ_3^*: we have

$$\phi_1 = 1/\lambda_1^*,\; \lambda_1 = \lambda_2^*,\; \phi_2 = \phi_2^*,\; \lambda_2 = 1/\phi_1^* \tag{4.133}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_1(\phi_1 + \cdots + \phi_1^{K_1}) = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) \tag{4.134}$$

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + b_1\phi_1^{K_1} = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1}) \tag{4.135}$$

From $q(x_3) = q^*(x_3)$, we get

$$\begin{aligned} b_1(\phi_1 + \cdots + \phi_1^{K_1}) + b_2(\phi_2 + \cdots + \phi_2^{K_1}) \\ = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) + b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1}) \end{aligned} \tag{4.136}$$

$$\begin{aligned} b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + 1) \\ = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) + b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) \end{aligned} \tag{4.137}$$

From $q(x_4) = q^*(x_4)$, we get

$$b_2\lambda_2^{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + 1) \tag{4.138}$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) \tag{4.139}$$

From (4.137) and (4.139), we get

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) = b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) \tag{4.140}$$

From (4.134) and (4.136), we get

$$b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1}) \tag{4.141}$$

From (4.140) and (4.141), we get

$$b_2 = b_2^* = b_1$$

So $q_1(x) = q_2^*(x)$, $q_2(x) = q_1^*(x)$. So we get $L = 1$ case.
Case Λ_3 and Φ_4^*: From $q(x_3) = q^*(x_3)$ we have

$$b_1\lambda_1^{K_2} + b_2\lambda_2^{K_2} + \sum_{l=1}^{2}(\phi_l + \cdots + \phi_l^{K_1}) = b_1^*\lambda_1^{*K_2} + 1$$

which is not possible for the same reason as in the Case Λ_1 and Λ_2^*.
Case Λ_3 and Φ_5^*: we have

$$\phi_1 = 1/\phi_2^*,\ \lambda_1 = \lambda_2^*,\ \phi_2 = 1/\lambda_1^*,\ \lambda_2 = 1/\phi_1^* \tag{4.142}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_1(\phi_1 + \cdots + \phi_1^{K_1}) = b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1}) \tag{4.143}$$

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + b_1\phi_1^{K_1} = b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) + b_2^*\phi_2^{*K_1} \tag{4.144}$$

From $q(x_3) = q^*(x_3)$, we get

$$\begin{aligned} &b_1(\phi_1 + \cdots + \phi_1^{K_1}) + b_2(\phi_2 + \cdots + \phi_2^{K_1}) \\ &\quad = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) + b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1}) \end{aligned} \tag{4.145}$$

$$\begin{aligned} &b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + 1) \\ &\quad = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) + b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) \end{aligned} \tag{4.146}$$

From $q(x_4) = q^*(x_4)$, we get

$$b_2\lambda_2^{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + 1) \tag{4.147}$$

$$b_2(\lambda_2^{K_2}+\lambda_2^{K_2-1}+\cdots+\lambda_2)=b_1^*(1+\phi_1^*+\cdots+\phi_1^{*K_1-1}) \tag{4.148}$$

From (4.143), (4.145), (4.146) and (4.148), we get $b_1=b_2^*$ which implies $q_1(x)=q_2^*(x)$. And then we have the same situation as in the case Λ_3 and Φ_3^*

Case Λ_3 **and** Φ_6^*: From $q(x_3)=q^*(x_3)$ we get the same equation as in the Case Λ_3 and Φ_4^* which is not possible.

Case Λ_4 **and** Φ_1^*: we have

$$\phi_1=1/\lambda_2^*,\ \lambda_1=1/\phi_2^*,\ \phi_2=1/\lambda_1^*,\ \lambda_2=1/\phi_1^* \tag{4.149}$$

From $q(x_2)=q^*(x_2)$, we get

$$b_2(\phi_2+\cdots+\phi_2^{K_1})=b_1^*(\lambda_1^{*K_2-1}+\cdots+1) \tag{4.150}$$

$$b_2(\lambda_2^{K_2}+\lambda_2^{K_2-1}+\cdots+\lambda_2)+b_2\phi_2^{K_1}=b_1^*(1+\phi_1^*+\cdots+\phi_1^{*K_1}) \tag{4.151}$$

From $q(x_3)=q^*(x_3)$, we get

$$\begin{aligned}b_1(\phi_1+\cdots+\phi_1^{K_1})+b_2(\phi_2+\cdots+\phi_2^{K_1})\\=b_1^*(\lambda_1^{*K_2-1}+\cdots+1)+b_2^*(\lambda_2^{*K_2-1}+\cdots+\lambda_2^*+1)\end{aligned} \tag{4.152}$$

$$\begin{aligned}b_1(\lambda_1^{K_2}+\lambda_1^{K_2-1}+\cdots+\lambda_1)+b_2(\lambda_2^{K_2}+\lambda_2^{K_2-1}+\cdots+\lambda_2)\\=b_1^*(1+\phi_1^*+\cdots+\phi_1^{*K_1-1})+b_2^*(1+\phi_2^*+\cdots+\phi_2^{*K_1-1})\end{aligned} \tag{4.153}$$

From $q(x_4)=q^*(x_4)$, we get

$$b_2\lambda_2^{K_2}+b_2(\phi_2+\cdots+\phi_2^{K_1})=b_1^*(\lambda_1^{*K_2}+\cdots+\lambda_1^*+1) \tag{4.154}$$

$$b_2(\lambda_2^{K_2}+\lambda_2^{K_2-1}+\cdots+\lambda_2)=b_1^*(1+\phi_1^*+\cdots+\phi_1^{*K_1-1}) \tag{4.155}$$

From (4.150) and (4.152), we get

$$b_1(\phi_1+\cdots+\phi_1^{K_1})=b_2^*(\lambda_2^{*K_2-1}+\cdots+\lambda_2^*+1) \tag{4.156}$$

From (4.153) and (4.155), we get

$$b_1(\lambda_1^{K_2}+\lambda_1^{K_2-1}+\cdots+\lambda_1)=b_2^*(1+\phi_2^*+\cdots+\phi_2^{*K_1-1}) \tag{4.157}$$

From (4.150), (4.154), (4.151)and (4.155) we get

$$b_2\phi_2^{K_1}=b_1^*\phi_1^{*K_1}\ \Rightarrow\ b_1\phi_1^{K_1}=b_2^*\phi_2^{*K_1} \tag{4.158}$$

$$b_2\lambda_2^{K_2} = b_1^*\lambda_1^{*K_2} \;\Rightarrow\; b_1\lambda_1^{K_2} = b_2^*\lambda_2^{*K_2} \tag{4.159}$$

(4.158) and (4.159) give

$$\frac{\phi_1^{K_1}}{\lambda_1^{K_2}} = \frac{\phi_2^{*K_1}}{\lambda_2^{*K_2}} = \frac{\phi_1^{K_2}}{\lambda_1^{K_1}}$$

which results in $K_1 = K_2$.

Suppose now $K_1 = K_2 = K$, From (4.156) and (4.157), we get

$$\frac{\phi_1 + \cdots + \phi_1^{K_1}}{\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1} = \frac{\lambda_1^{*K_2-1} + \lambda_1^{*K_2-1} + \cdots + 1}{1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}} = \frac{\lambda_1^{K-1}}{\phi_1^{K-1}}\frac{1 + \phi_1 + \cdots + \phi_1^{K-1}}{\lambda_1^{K-1} + \lambda_1^{K-2} + \cdots + 1}$$

which implies $\lambda_1 = \phi_1$, $\lambda_2^* = \phi_2^*$.

Similarly from (4.154) and (4.155), we can get $\lambda_2 = \phi_2$, $\lambda_1^* = \phi_1^*$.

And then we get symmetric cases.

Case Λ_4 and Φ_2^*: we have

$$\phi_1 = 1/\lambda_1^*,\;\; \lambda_1 = 1/\phi_2^*,\;\; \phi_2 = 1/\lambda_2^*,\;\; \lambda_2 = 1/\phi_1^* \tag{4.160}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_2^*(\lambda_2^{*K_2-1} + \cdots + 1) \tag{4.161}$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + b_2\phi_2^{K_1} = b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1}) \tag{4.162}$$

From $q(x_3) = q^*(x_3)$, we get

$$\begin{aligned} &b_1(\phi_1 + \cdots + \phi_1^{K_1}) + b_2(\phi_2 + \cdots + \phi_2^{K_1}) \\ &\quad = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) + b_2^*(\lambda_2^{*K_2-1} + \cdots + \lambda_2^* + 1) \end{aligned} \tag{4.163}$$

$$\begin{aligned} &b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) \\ &\quad = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) + b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1-1}) \end{aligned} \tag{4.164}$$

From $q(x_4) = q^*(x_4)$, we get

$$b_2\lambda_2^{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1^*(\lambda_1^{*K_2} + \cdots + \lambda_1^* + 1) \tag{4.165}$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) \tag{4.166}$$

From (4.161) and (4.163), we get

$$b_1(\phi_1 + \cdots + \phi_1^{K_1}) = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) \tag{4.167}$$

From (4.164) and (4.166), we get

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) = b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1-1}) \tag{4.168}$$

From (4.161), (4.162), (4.165) and (4.166) we can get equations (4.120), (4.121), (4.118), (4.119). Then the arguments follow similarly in Case Λ_3 and Φ_1^*. And we say (4.161)-(4.166) can not be satisfied.

Case Λ_4 **and** Φ_3^*: we have

$$\phi_1 = \phi_2^*,\ \lambda_1 = \lambda_2^*,\ \phi_2 = 1/\lambda_1^*,\ \lambda_2 = 1/\phi_1^* \tag{4.169}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) \tag{4.170}$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + b_2\phi_2^{K_1} = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1}) \tag{4.171}$$

From $q(x_3) = q^*(x_3)$, we get

$$\begin{aligned} &b_1(\phi_1 + \cdots + \phi_1^{K_1}) + b_2(\phi_2 + \cdots + \phi_2^{K_1}) \\ &\quad = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) + b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1}) \end{aligned} \tag{4.172}$$

$$\begin{aligned} &b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) + b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) \\ &\quad = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) + b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) \end{aligned} \tag{4.173}$$

From $q(x_4) = q^*(x_4)$, we get

$$b_2\lambda_2^{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1^*(\lambda_1^{*K_2} + \cdots + \lambda_1^* + 1) \tag{4.174}$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) \tag{4.175}$$

From (4.170) and (4.172), we get

$$b_1(\phi_1 + \cdots + \phi_1^{K_1}) = b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1}) \tag{4.176}$$

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + \lambda_1) = b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) \tag{4.177}$$

From (4.176) and (4.177), we get $b_1 = b_2^*$ which implies $q_1(x) = q_2^*(x)$ and then $q_2(x) = q_1^*(x)$ which becomes a case of $L = 1$.

Case Λ_4 **and** Φ_4^*: just as Case Λ_3 and Φ_4^*.

Case Λ_4 **and** Φ_5^*: we have

$$\phi_1 = 1/\lambda_1^*,\ \lambda_1 = \lambda_2^*,\ \phi_2 = \phi_2^*,\ \lambda_2 = 1/\phi_1^* \tag{4.178}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1}) \quad (4.179)$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) + b_2\phi_2^{K_1} = b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) + b_2^*\phi_2^{*K_1} \quad (4.180)$$

(4.179) implies $b_2 = b_2^*$. (4.180) implies $\lambda_2 = \lambda_2^* = \lambda_1$ which violates $\lambda_1 < \lambda_2$.
Case Λ_4 **and** Φ_6^*: just as Case Λ_3 and Φ_4^*.
Case Λ_5 **and** Φ_1^*: From $q(x_3) = q^*(x_3)$, we get

$$1 + b_2\lambda_2^{K_2} = \sum_{l=1}^{2} b_l^*(\lambda_l^{*K_2} + \lambda_l^{*K_2-1} + \cdots + 1)$$

which is not possible for the same reason as in the Case Φ_4 and Φ_5^*.
Case Λ_5 **and** Φ_2^*: just as Case Λ_5 and Φ_1^*.
Case Λ_5 **and** Φ_3^*: From $q(x_3) = q^*(x_3)$, we get

$$1 + b_2\lambda_2^{K_2} = b_1^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + 1) + b_2\lambda_2^{*K_2} + b_2(\phi_2^* + \cdots + \phi_2^{*K_1})$$

which is not possible for the same reason as in the Case Φ_4 and Φ_5^*.
Case Λ_5 **and** Φ_4^*: we have

$$\phi_1 = \phi_2^*, \ \lambda_1 = \lambda_2^*, \ \phi_2 = 1/\lambda_1^*, \ \lambda_2 = 1/\phi_1^* \quad (4.181)$$

From $q(x_2) = q^*(x_2)$, we get

$$b_1(\lambda_1^{K_2-1} + \cdots + 1) = b_2^*(\lambda_1^{*K_2-1} + \cdots + 1) \quad (4.182)$$

$$b_1(1 + \phi_2 + \cdots + \phi_2^{K_1}) = b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1}) \quad (4.183)$$

(4.182) and (4.183) imply $b_1 = b_2^*$ and $q_1(x) = q_2^*(x)$. So this is a case of $L = 1$.
Case Λ_5 **and** Φ_5^*: just as Case Λ_5 and Φ_3^*.
Case Λ_5 **and** Φ_6^*: we have

$$\phi_1 = 1/\lambda_2^*, \ \lambda_1 = 1/\phi_2^*, \ \phi_2 = 1/\lambda_1^*, \ \lambda_2 = 1/\phi_1^* \quad (4.184)$$

From $q(x_2) = q^*(x_2)$, we get

$$b_1(\lambda_1^{K_2-1} + \cdots + 1) = b_2^*(\lambda_1^{*K_2-1} + \cdots + 1) \quad (4.185)$$

$$b_1(1 + \phi_1 + \cdots + \phi_1^{K_1}) = b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1}) \quad (4.186)$$

From $q(x_3) = q^*(x_3)$, we get

$$b_1\phi_1^{K_1} = b_2^*\phi_2^{*K_1} \;\Rightarrow\; b_2\phi_2^{K_1} = b_1^*\phi_1^{*K_1} \tag{4.187}$$

$$b_2\lambda_2^{K_2} = b_1^*\lambda_1^{*K_2} \;\Rightarrow\; b_1\lambda_1^{K_2} = b_2^*\lambda_2^{*K_2} \tag{4.188}$$

(4.187) and (4.188) imply $K_1 = K_2$. From $q(x_4) = q^*(x_4)$, we get

$$b_2\lambda_{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + 1) \tag{4.189}$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) \tag{4.190}$$

(4.185) and (4.186) imply $\lambda_1 = \phi_1$, $\lambda_2^* = \phi_2^*$. (4.187), (4.188), (4.189) and (4.190) imply $\lambda_2 = \phi_2$, $\lambda_1^* = \phi_1^*$. So we get the symmetric cases.

Case Λ_6 and Φ_1^*: we have

$$\phi_1 = \phi_2^*,\; \lambda_1 = \lambda_1^*,\; \phi_2 = 1/\lambda_2^*,\; \lambda_2 = 1/\phi_1^* \tag{4.191}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_1(\lambda_1^{K_2-1} + \cdots + 1) = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) \tag{4.192}$$

$$b_1(1 + \phi_1 + \cdots + \phi_1^{K_1}) = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1}) \tag{4.193}$$

(4.191) and (4.192) imply $b_1 = b_1^*$. (4.193) implies $\phi_1 = \phi_1^* = \phi_2^*$ which violates the condition $\phi_1^* < \phi_2^*$.

Case Λ_6 and Φ_2^*: we have

$$\phi_1 = \phi_2^*,\; \lambda_1 = \lambda_1^*,\; \phi_2 = 1/\lambda_2^*,\; \lambda_2 = 1/\phi_1^* \tag{4.194}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_1(\lambda_1^{K_2-1} + \cdots + 1) = b_2^*(\lambda_2^{*K_2-1} + \cdots + 1) \tag{4.195}$$

$$b_1(1 + \phi_1 + \cdots + \phi_1^{K_1}) = b_2^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1}) \tag{4.196}$$

(4.194) and (4.195) imply $b_1 = b_2^*$. (4.196) implies $\phi_1 = \phi_1^* = \phi_2^*$ which violates the condition $\phi_1^* < \phi_2^*$.

Case Λ_6 and Φ_3^*: we have

$$\phi_1 = 1/\lambda_2^*,\; \lambda_1 = \lambda_1^*,\; \phi_2 = \phi_2^*,\; \lambda_2 = 1/\phi_1^* \tag{4.197}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_1(\lambda_1^{K_2-1} + \cdots + 1) = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) \tag{4.198}$$

$$b_1(1 + \phi_1 + \cdots + \phi_1^{K_1}) = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1}) \tag{4.199}$$

We get $b_1 = b_1^*$, $\phi_1 = \phi_1^*$. So $q_1(x) = q_1^*(x)$, $q_2(x) = q_2^*(x)$ and the case becomes a case of $L = 1$.

Case Λ_6 **and** Φ_4^*: From $q(x_3) = q^*(x_3)$, we get

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + b_2\lambda_2^{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1^*\lambda_1^{*K_2} + 1$$

which is not possible for the same reason as in the Case Λ_1 and Λ_2^*.

Case Λ_6 **and** Φ_5^*: we have

$$\phi_1 = 1/\lambda_2^*,\ \lambda_1 = 1/\phi_2^*,\ \phi_2 = 1/\lambda_1^*,\ \lambda_2 = 1/\phi_1^* \tag{4.200}$$

From $q(x_2) = q^*(x_2)$, we get

$$b_1(\lambda_1^{K_2-1} + \cdots + 1) = b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1}) \tag{4.201}$$

$$b_1(1 + \phi_1 + \cdots + \phi_1^{K_1}) = b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) + b_2^*\phi_2^{*K_1} \tag{4.202}$$

From $q(x_3) = q^*(x_3)$, we get

$$\begin{aligned} &b_1(\lambda_1^{K_2-1} + \cdots + 1) + b_2(\phi_2 + \cdots + \phi_2^{K_1}) \\ &\quad = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) + b_2^*(\phi_2^* + \cdots + \phi_2^{*K_1}) \end{aligned} \tag{4.203}$$

$$\begin{aligned} &b_1(1 + \phi_1 + \cdots + \phi_1^{K_1-1}) + b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) \\ &\quad = b_1^*(1 + \phi_1^* + \cdots + \phi_1^{*K_1-1}) + b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) \end{aligned} \tag{4.204}$$

From $q(x_4) = q^*(x_4)$, we get

$$b_2\lambda_2^{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1^*(\lambda_1^{*K_2} + \lambda_1^{*K_2-1} + \cdots + 1) \tag{4.205}$$

$$b_2(\lambda_2^{K_2} + \lambda_2^{K_2-1} + \cdots + \lambda_2) = b_1^*(1 + \phi_2^* + \cdots + \phi_2^{*K_1-1}) \tag{4.206}$$

From (4.201) and (4.201), we get

$$b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1^*(\lambda_1^{*K_2-1} + \cdots + 1) \tag{4.207}$$

$$b_1(1 + \phi_2 + \cdots + \phi_2^{K_1-1}) = b_2^*(\lambda_2^{*K_2} + \lambda_2^{*K_2-1} + \cdots + \lambda_2^*) \tag{4.208}$$

From (4.205) and (4.205), we get

$$b_2\lambda_2^{K_2} = b_1^*\lambda_1^{*K_2} \;\Rightarrow\; b_1\lambda_1^{K_2} = b_2^*\lambda_2^{*K_2} \tag{4.209}$$

From (4.202) and (4.208), we get

$$b_1\phi_1^{K_1} = b_2^*\phi_2^{*K_1} \;\Rightarrow\; b_2\phi_2^{K_1} = b_1^*\phi_1^{*K_1} \tag{4.210}$$

(4.209) and (4.210) imply $K_1 = K_2$. But (4.201) and (4.208) imply $\lambda_1 = \phi_1$ and (4.206) and (4.206) imply $\lambda_2 = \phi_2$. This gets down to a symmetric case.
Case Λ_6 **and** Φ_6^*: From $q(x_3) = q^*(x_3)$, we get

$$b_1(\lambda_1^{K_2} + \lambda_1^{K_2-1} + \cdots + 1) + b_2\lambda_2^{K_2} + b_2(\phi_2 + \cdots + \phi_2^{K_1}) = b_1^*\lambda_1^{*K_2} + 1$$

which is not possible for the same reason as in the Case Λ_1 and Λ_2^*.

Chapter 5

Multivariate Extreme Value Analysis to Value at Risk

5.1 Introduction

How to manage a portfolio efficiently, with the highest expected return for a given level of risk, or equivalently, the least risk for a given level of expected return, is the key to the success or failure of a financial system.

J.P. Morgan's RiskMetrics^{TM} (1996) defines risk as the degree of uncertainty of future net return. Financial systems face many risks which may result in financial collapse if not appropriately managed. According to RiskMetrics, a common classification of risks gives four main categories of risk which are Credit Risk (the potential loss because of the inability of a counterparts to meet its obligation), Operational Risk (errors that can be made in instructing payments or setting transactions), Liquidity Risk (inability of a firm to fund its illicit assets), and Market Risk (loss resulting from a change in the value of traceable assets). This work focuses on ways of modeling market risk via multivariate extreme value theories and methods.

To be aware of and to understand risks which a manager will face are very important to the modern financial management. Especially, as financial trading systems have been extended broadly and become more sophisticated, there has been increased awareness of the dangers of very large losses. For example, see Smith (2000), large price movements in security markets may cause the failure of financial systems. Examples include the bankruptcy of Baring Bank, Daiwa Bank and Orange County in California. The most spectacular example to date was the near-collapse of the hedge fund Long Term Capital Management in September 1998. LTCM was trading a complex mixture of derivatives which, according to some estimates, gave it an exposure to market risk as high as \$200 billion. Things started to go wrong after the collapse of the Russian economy in the summer of 1998, and to avoid a total collapse of the company, 15 major banks contributed to a \$3.75 billion rescue package.

From the insurance industry, very large claims can cause insurance companies to go bankrupt. Embrechts et al. (1997) lists 30 most costly insurance losses 1970-1995 in table 1 (the largest one is 16,000 million dollars) and 30 worst catastrophes in terms of fatalities 1970-1995 in table 2 (the largest fatality is 300,000), both taken from Sigma (1996).

These events are relatively rare, but important. These and other examples have increased awareness of the need to quantify probabilities of large losses, and for risk management systems to control such events. A tool called Value at Risk (henceforth, VaR) has been increasingly employed by many banks. It gained a higher profile in 1994 when J.P. Morgan published its RiskMetrics system. The Basle Committee on Banking Supervision has proposed in 1996 that internal VaR models may be used in the determination of the capital requirements that banks must fulfill to back their trading activities (cf. Dave and Stahl 1999). Books, like Jorion (1996), Dowd (1998), aimed at financial academics and traders and explained the statistics basis behind VaR. Best (1998) is aimed at the risk management practitioners. Dave and Stahl (1999)'s working paper studied 5 different VaR models with real data performance analysis.

Among many applications and models, portfolio returns are assumed normally distributed, or tail normally distributed. Such assumption makes the estimation easy. However this may underestimate the risk of the system which actually has a fat-tailed distribution. Most financial data are actually distributed with fat tail. LTCM and banks have been criticized for not "stress-testing" risk models against extreme market movements (Embrechts, 1999). Also back to November 1995, the Director of the Federal Reserve, Mr. A. Greenspan stated "work that characterizes the distribution of extreme events would be useful as well" (Embrechts, 1999). The excellent recent book by Embrechts *et al.* (1997) surveys the mathematical theory of EVT and discusses its applications to both financial and insurance risk management.

Although the use of EVT in finance and insurance industries has a considerable literature on the subject, especially there is a much longer history of its use in the insurance industry, most applications are restricted in univariate stochastic process data. Again Embrechts et al. (1997) is an excellent literature. Smith (2000) presented a demonstration of the merits of combining established models for extreme values with modern statistical techniques including Bayesian inference, hierarchical models and Monte Carlo sampling for real insurance data. Tsay (1999) applied extreme value theory to investigate the occurrence times and excesses over some high thresholds of financial time series (S&P index). Both two works used similar methodology though they had different kinds of data. As stated in Embrechts et al. (1997), in 1900 Louis Bachelier showed that Brownian motion lies at the heart of any model for asset returns; around the same time, Filip Lundberg showed that the homogeneous Poisson process, after a suitable time transformation, is the key model for insurance liability data. As to multivariate aspects, not much work has been done. However, it is very natural to consider multivariate extremes in which a portfolio of asset returns is under high risk due to a combination of various processes at extreme levels. For instance, daily exchange rates for the value of 1 US dollar against foreign currency, or insurance models in which there are several types of claims each day. There exist dependence structures among the various assets in a portfolio. If the composition of the portfolio is held fixed, then it may be enough to only assess the composition risk of the portfolio, which can be done by applying univariate EVT. However, to manage a portfolio efficiently, or equivalently to optimize the portfolio, the real rationale for considering multivariate aspects is often to help design the portfolio. The famous mean-variance approach first introduced by Markowitz is broadly used in financial management (Markowitz 1952, 1987, Korn 1997, Michaud 1998). The approach is based on an assumption of multivariate normality for the joint distribution of assets or securities. One of its formulae is to maximize the expected return subject to given risk (which is the variance of a linear combination of assets). An alternative option is to use VaR as the constraints, which we will investigate further and has drawn much more attention in financial management. Conventional VaR theory is highly questionable due to the joint multivariate normal distribution assumption of log returns which may not be appropriate to the fat-tailed data and may result in an underestimate of the risk.

One approach to the problem is through multivariate EVT. Resnick (1987) is an excellent source of information on possible approaches. Due to no standard notion of order in high dimensional Euclidean space, most approaches to date have focused on the one dimensional case. The good news is that there is a considerable progress. Coles and Tawn (1991, 1994) have done an impressive progress on modeling extreme multivariate events and made multivariate extreme value theory into a very practical method of data analysis. Embrechts, de Haan and Huang (1999) presented an approach for modeling tail events and showed results in 2-dimensional case theoretically and numerically. Smith and Weissman (1996) have proposed some alternative representations of extreme value processes to characterize the joint distribution of extremes in multivariate time series. They showed under fairly general conditions, extremal properties of a wide class of multivariate time series may be calculated by approximating the process in the tails by one of M4 form.

5.2 VaR methods

In risk management, one of its functions is to measure risk the financial system is exposed to. Questions like: how much could a bank lose on a normal trading day? Or what kind of risk the bank exposes to the market? These and other questions are very often asked to the risk manager by the CEO of the bank. To answer these questions, we seek some risk measurements to quantify the risk of all trading positions of the bank. However, some traditional risk measurements have difficulties to answer questions like those we have asked. Traditional methods usually only calculate each individual risk of market variables invested in the market by a financial institution. The overall market risk cannot be efficiently measured because of the number of market variables (hundreds or thousands) and very long computing time needed. These methods may be of benefit to traders who manage the trading activities for each financial instrument, but are not very useful to senior risk managers or regulators. For example, the variance of the portfolio return tells how variable the return is, but does not tell us how likely and what amount of money the bank will lose. VaR methods can overcome those difficulties the traditional methods suffered. VaR is defined as *the value at risk is the maximum possible loss on a portfolio over a given time interval, with a given level of confidence.* Statistically, if we let X_t be the loss over time t within a time horizon, say $[0, T]$, and the confidence level is $1 - \alpha$, then the VaR is just the upper α percentile x_α of the random variable X_T, which is

$$\Pr(X_T < x_\alpha) > 1 - \alpha. \tag{5.1}$$

In the literature, there are three typical methods to calculate VaR, i.e. variance-covariance approach, historical simulation approach, and Monte-Carlo simulation approach. We will briefly describe these methods further in the following subsections.

5.2.1 Variance-Covariance approach

The variance-covariance approach is first widely used VaR calculation method. The references, among others, include Chapter 11 of Jorion (1997), Chapter 3 of Dowd (1998), Chapter 2 of Best (1998). This approach assumes the return has a normal distribution for a single asset or the returns have jointly multivariate normal distribution for a portfolio with multiple assets. How to implement VaR calculation is presented as in the following.

Consider now a portfolio of d assets and at time t the returns are $R_{1t}, R_{2t}, \ldots, R_{dt}$. We write $\mathbf{R}_t = (R_{1t}, R_{2t}, \ldots, R_{dt})'$. $\mathbf{R}_t$ is assumed to be jointly normally distributed, i.e. $\mathbf{R}_t \sim N(\mu, \Sigma)$. Let $\mathbf{w} = (w_{1t}, \ldots, w_{dt})'$ be the weights of each individual position and sum to unity. The return of the portfolio at time t is defined as

$$R_t^p = \sum_{i=1}^{d} w_{it} R_{it} = \mathbf{w}'\mathbf{R}_t. \tag{5.2}$$

So the VaR for the portfolio is calculated from

$$\Pr(R_t^p > c) = \alpha. \tag{5.3}$$

Since R_t^p is distributed as $N(\mathbf{w}'\mu, \mathbf{w}'\Sigma\mathbf{w})$ and hence

$$c = \mathbf{w}'\mu + z_\alpha(\mathbf{w}'\Sigma\mathbf{w})^{\frac{1}{2}} \tag{5.4}$$

where z_α is the standard normal upper α percentile. And finally the VaR for the portfolio is

$$VaR_p = V * c \tag{5.5}$$

where V is the original position value or the investment. In practice, however, it is customary to assume that the expected price change is zero with given time period. The VaR for a portfolio is calculated simply as

$$VaR_p = V * z_\alpha(\mathbf{w}'\Sigma\mathbf{w})^{\frac{1}{2}} = (\mathbf{v}'\Sigma\mathbf{v})^{\frac{1}{2}} \tag{5.6}$$

where $\mathbf{v}$ is a vector of VaRs for each individual position.

5.2.2 Historical simulation approach

The historical simulation approach is to use the empirical distribution based on the past data. Theoretically, the empirical distribution converges to the true distribution of the interest random variable, here it is the portfolio return. The idea is to use the empirical distribution to approximate the true distribution and to read off VaR from the empirical distribution curve, usually a histogram. The advantages of using historical simulation include no model assumption and no distribution required; VaR can be read directly from a spread sheet. It also can capture the fat tailed behavior of the data with enough observations and very large observed values. But there are some disadvantages. For example, it may not be a good way to represent very extreme events because there is no possibility of extrapolation beyond the observed range of the data. It also assumes stationarity in time.

5.2.3 Monte-Carlo simulation approach

While the historical simulation approach using historical data to construct empirical distribution, the Monte-Carlo simulation approach draws the data from a random process and uses the drawn values to construct empirical distributions and further read off the VaR from the constructed empirical distribution. Here we need a random process to represent the price change. To achieve this, a continuous stochastic process which can be written into a stochastic model is often used in the literature. A simple model of price change may have the form

$$dX_t = \sigma dW_t$$

where W_t is a Brownian motion, and σ is known or at least estimable. This simple model has a solution

$$X_t = X_0 + \sigma W_t.$$

From this representation, we can draw values of X_t from Monte-Carlo simulation scheme which draws a random number from a random number generator, and then transforms this number into a normal random variate, and finally gets a simulated return value. This procedure is repeated over and over, until we think we have enough values which can be used to construct an empirical distribution.

Note: since the example model here is rather simple, X_t itself is normally distributed. We can calculate the VaR from Variance-covariance approach directly. The Monte-Carlo simulation approach is usually taking advantage when the portfolio contains multiple assets and relatively complicated structure which may not be easily solved by simpler approaches.

5.3 Extreme value approaches

The three methods described in the previous section have their own disadvantages. In general VaR is regarded as an extreme quantile of the loss distribution because we are interested in a 95%, 99% or even higher confidence level. For example, Bankers Trust uses 99%; Chemical and Chase uses 97.5%; Citibank, a 95.4% level; Bank America and J. P. Morgan uses 95% (Jorion, p20, 1997). Financial returns are fat-tailed and exhibit the form of clustering which usually caused by extreme price movements. The variance-covariance approach may not be suitable because its normality assumption. Historical simulation may not work due to lack of extreme observations. Also it is hard to use historical simulation to characterize the dependence structure among assets of a portfolio. The Monte-Carlo simulation also has a normality assumption for the underlying stochastic process.

The extreme value approach has drawn a major attention in VaR study recently. It has advantages in analyzing fat tailed data, which financial data are, and extrapolating beyond the range of observed data. Just as the variance-covariance approach has a distributional assumption, extreme value approaches assume the underlying limiting (for the extreme values) distributions are extreme type distributions. There are three types of extreme value distributions which we stated in Chapter 1 and they can be written into a generalized form

$$H(x;\mu,\sigma,\xi) = \exp\{-[1+\frac{\xi(x-\mu)}{\sigma}]^{-1/\xi}\} \tag{5.7}$$

which is (1.6) in Chapter 1. Now to calculate VaR of the underlying loss distribution is equivalent to computing the extreme quantile of (5.1) in case a portfolio only has one position or is statistically univariate. It is known that the normal distribution is stable and the extreme value distributions are max-stable. As a result, if R_{it} in (5.2) follows one of the extreme type distributions, R_t^p won't follow any extreme value distribution. But the limiting form of $\mathbf{R}_t$ will follow multivariate extreme distribution which we are interested in. So the VaR calculation is transformed into the calculation of critical value c in (5.3) from a multivariate extreme value distribution.

The multivariate extreme value distribution has no explicit forms, unlike the univariate extreme value distributions. We need to adopt one of the existing multivariate extreme value distributions to model the data. But the observed data are not independent, so we then need to estimate the extreme index for the multivariate time series under stationarity conditions. Since financial data exhibits clustering, it may be a better choice to model the extreme value of the data by using one of the $M4$ process since we have seen under mild general conditions a stationary stochastic process can be approximated in the tails by a max-stable process and very closely a max-stable process can be approximated by an $M4$ process. We present applications of $M4$ process to the financial data in section 5.5.

5.4 Optimal portfolio theories

Not putting all one's eggs in one basket has been a basic concept for a long history if one suspects the basket is not completely secured. In finance, portfolio diversification has been thought as an essential component of modern risk management. To minimize the risk will result in investing money on those market variables with smaller risk. On the other hand, an investor expects to gain the maximum possible returns with his investment. In general, the higher the risk, the higher the return. These two investment strategies, low risk and high return, are opposite to each other. Naturally an investor would seek an optimal investment plan with which either to maximize the portfolio mean return such that the estimated risk is not higher than an upper risk limit or to minimize the risk such that the mean return is not lower than a lower mean return limit within a given time period. This is referred as the portfolio problem in the literature.

The mean-variance approach pioneered by Markowitz (Markowitz 1952, 1987, Korn 1997, Michaud 1998) is the earliest approach to solve a portfolio investment problem. Although it is a one time period model approach it is still highly valued. In 1990, Harry M. Markowitz, Merton M. Miller and William E. Sharpe gained the Nobel Prize in economic sciences for their pioneering work in the theory of financial economics.

Let's now assume a portfolio consists of d assets. At time $t = 0$, an investor has to decide how many shares of each asset to hold until time $t = T$. Suppose the proportion of total money invested on asset i is π_i and the price of asset i at time t is $P_i(t)$, then the return of asset i is $R_i = \frac{P_i(T)}{P_i(0)}$. Finally the portfolio return is

$$R = \sum_{i=1}^{d} \pi_i R_i$$

and the mean return is defined by

$$\mu(\pi) = E(R) = \sum_{i=1}^{d} \pi_i E(R_i) = \sum_{i=1}^{d} \pi_i \mu_i$$

and the variation of the portfolio return is

$$\sigma^2(\pi) = \mathrm{Var}(R) = \sum_{i,j=1}^{d} \pi_i \mathrm{Cov}(R_i, R_j)\pi_j = \sum_{i,j=1}^{d} \pi_i \sigma_{ij} \pi_j.$$

Then mean-variance approach is to decide the optimal investment plan π by solving one of the following two optimization problems.

$$\begin{cases} \min\limits_{\pi \in \mathbb{R}^d} \sigma^2(\pi) \\ \text{s.t. } \ \mu(\pi) \geq \mu_{\text{low}}, \ \pi_i \geq 0, \ \sum\limits_{i=1}^{d} \pi_i = 1 \end{cases} \tag{5.8}$$

where μ_{low} is lowest mean return. Or

$$\begin{cases} \max\limits_{\pi \in \mathbb{R}^d} \mu(\pi) \\ \text{s.t. } \ \sigma^2(\pi) \leq \sigma_{\text{max}}, \ \pi_i \geq 0, \ \sum\limits_{i=1}^{d} \pi_i = 1 \end{cases} \tag{5.9}$$

where σ_{max} is regarded as maximum risk one can take. (5.8) is a quadratic optimization problem. There is a unique solution. Under some conditions (5.8) and (5.9) are equivalent. In practice, (5.9) seems more natural. We adopt (5.9) as a basic model in this present work and extend it to the VaR constraints calculated from a extreme value distribution.

Traditional risk measurement, for example, the variance-covariance approach, may under-estimate the risk a financial institution exposed to. And so we model the portfolio returns by using a multivariate extreme value distribution function, especially in this present case we adopt $M4$ processes modeling. And so our optimization model is

$$\begin{cases} \max\limits_{\pi \in \mathbb{R}^d} \mu(\pi) \\ \text{s.t. } \ \Pr(R^p > c) = \alpha, \ \pi_i \geq 0, \ \sum\limits_{i=1}^{d} \pi_i = 1 \end{cases} \tag{5.10}$$

where $R^p = \sum\limits_{i=1}^{d} \pi_i R_i = \pi' \mathbf{R}$, and the limiting distribution of exceedances over a high threshold of $\mathbf{R}$ follows a multivariate extreme value distribution.

5.5 Dynamic financial data modeling

In this section we will model financial time series data as $M4$ processes. Stock prices of GE, CITIBANK and Pfizer will be studied. Parameter estimates of $M4$ processes are based on an multivariate time series of approximately 5000 days. We first look at those extreme values of the negative returns and check whether extreme value distribution fitting is appropriate or not by using mean excess plot, Z-plot and W-plot. The data are standardized using GARCH(1,1) model which gives estimated conditional standard deviation. Then we check whether extreme value distribution fitting is appropriate or not to the standardized time series. A generalized Pareto distribution is used to fit the data above certain threshold(.02 is used in this study) for each sequence. The data are then transformed into Fréchet scale from fitted GPD function. The transformed data are used in $M4$ processes modeling. We begin at introducing some concepts and basic backgrounds.

5.5.1 Mean excess plot

The mean excess plot is a plot of the mean of all excess values over a threshold u against u itself. It usually suggests whether a extreme value distribution fitting is appropriate or not. It is very useful for initial diagnostics and selecting the threshold. It is based on the following identity: if Y has a generalized Pareto distribution, provided $\xi < 1$, then for threshold $u > 0$, define the *mean excess function*

$$e(u) = E\{Y - u|Y > u\} = \frac{\alpha + \xi u}{1 - \xi}.$$

Thus, a sample plot of mean excess against threshold should be approximately a straight line if the model is correct.

5.5.2 Z-statistics and W-statistics

The underlying idea behind these analysis of Z-statistics and W statistics is the point-process approach to univariate extreme value modeling due to Smith (1989). According to this viewpoint, the exceedance times and excess values of a high threshold are viewed as a two-dimensional point process. If the process is stationary and satisfies a condition that there are asymptotically no clusters among the high-level exceedances, then its limiting form is non-homogeneous Poisson. Smith and Shively (1995) introduced a number of diagnostic devices to examine the fit of the generalized extreme value distributions. One idea is based on what they called Z-statistics

$$Z_k = \int_{T_{k-1}}^{T_k} \Lambda_s(u)ds$$

where T_k denotes the time of the k'th exceedance of u. $\Lambda_t(x)$ is given by

$$\Lambda_t(x) = (1 + \xi_t \frac{x - \mu_t}{\psi_t})_+^{-1/\xi_t}$$

the intensity of a nonhomogeneous Poisson process of exceedances of a level x. If the model is correct, then $Z_1, Z_2, \ldots,$ will be independent exponentially distributed random variables with mean 1. The Z-statistics are an indication of how closely the exceedances of a fixed level u are represented by a nonhomogeneous Poisson process, but they do not test the generalized Pareto distribution assumption for the distribution of excesses over the threshold. This can be done via W-statistics:

$$W_k = \frac{1}{\xi_{T_k}} \log[1 + \xi_{T_k} \frac{Y_k - u}{\psi_{T_k} + \xi_{T_k}\{u - \mu_{T_k}\}}].$$

Then $W_1, W_2, ...$ are also independent exponential random variables with mean 1, if the model is correct. These techniques have been broadly used in model diagnostics, for example, Tsay (1999), Smith and Goodman (2000).

5.5.3 GARCH(1,1) model

Traditional time series model AR(p) assumes constant variance cross the time which experience has shown not the case. GARCH, generalized autoregressive conditional heteroscedasticity, process model the residual of a time series regression. The model was proposed by Bollerslev (1986). It does not assume the constant variance. Research has shown that it has been quite successful to use GARCH model of fitting financial time series. We now introduce the GARCH model.

Suppose time series regression has the form

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + u_t,$$

where

$$u_t = \sqrt{h_t} v_t$$

$$h_t = \kappa + \delta_1 h_{t-1} + \cdots + \delta_r h_{t-r} + \theta_1 u_{t-1}^2 + \cdots + \theta_s u_{t-s}^2$$

$$v_t \sim N(0, 1).$$

These three formula together are called GARCH(r,s) model.

5.5.4 Initial diagnostics

Figure 5.1 are time series plots of three stock returns. We can see there are extreme observed values in each sequence and they are not stationary. A data transformation may be needed in order to have stationarity and apply M4 process modeling. Before we do that, we check whether extreme value distribution fittings are appropriate or not for the observations which are above certain threshold.

Figure 5.2 is initial diagnosis which suggests that extreme value distribution fitting for the Pfizer data but not for the other two data sets. Further diagnosis based on Z and W statistics are used.

All W-plots from Figures 5.3 -5.5 suggest a generalized extreme value distribution fitting is appropriate. Some caution should be taken since a few points, partly the result of Oct. 87 crash, are away from the straight line. But Z-plots do not suggest a generalized extreme value distribution fitting.

5.5.5 Data transformation

As we mentioned earlier, our goal is to model $M4$ process to the three time series data sets. We now use GARCH(1,1) to model the volatility. Figure 5.6 shows estimated conditional standard deviation. Figure 5.7 shows standardized time series. Visually they look stationary. Figure 5.8-5.10 suggest a generalized extreme value distribution fitting is appropriate. Notice that the earlier Z-plots were not consistent with the model, but now they are.

After fitting the generalized extreme value distribution, the data set are transformed into unit Fréchet scale and the transformed data are plotted in Figure 5.11.

Since an $M4$ process has double indexes, one for signature patterns and one for moving range, we need to determine the order of moving range and the number of signature patterns. We apply graphical diagnostic methods to determine the order and propose a criteria to determine the number of signature patterns.

Based on the properties that an $M4$ process appears clustered observations when an extreme observation occurs, we check those observed values which are larger than a certain threshold. Empirical counts can tell both the moving range order and the dependence range. We look at the counts of paired neg-daily returns on unit Fréchet scale in different ways. We count the days when two different stock products both had price drops over certain threshold. We count the days when a single stock product had

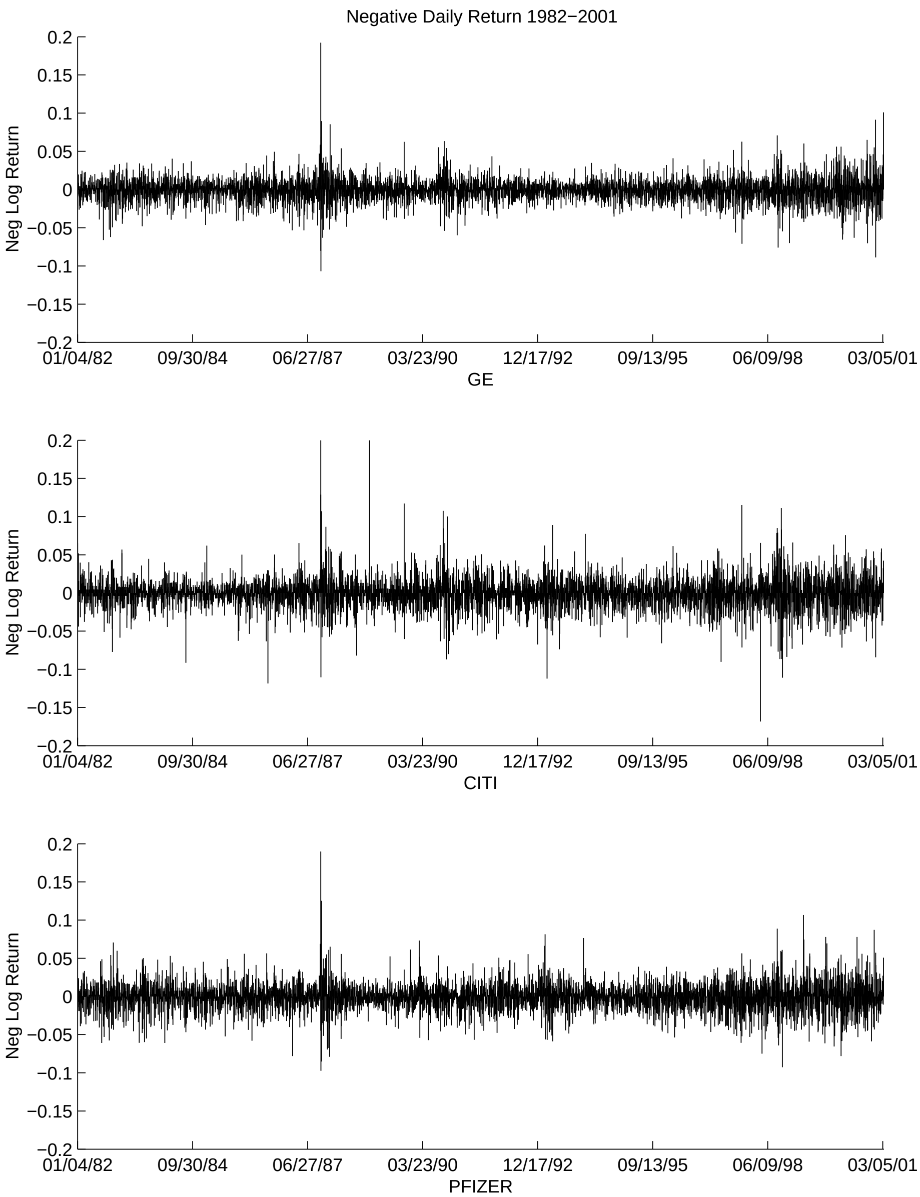

Figure 5.1: These figures show that there are extreme observations and the greatest drop happened in the same day in all three time series, i.e. October 19, 1987, the date of the Wall Street crash.

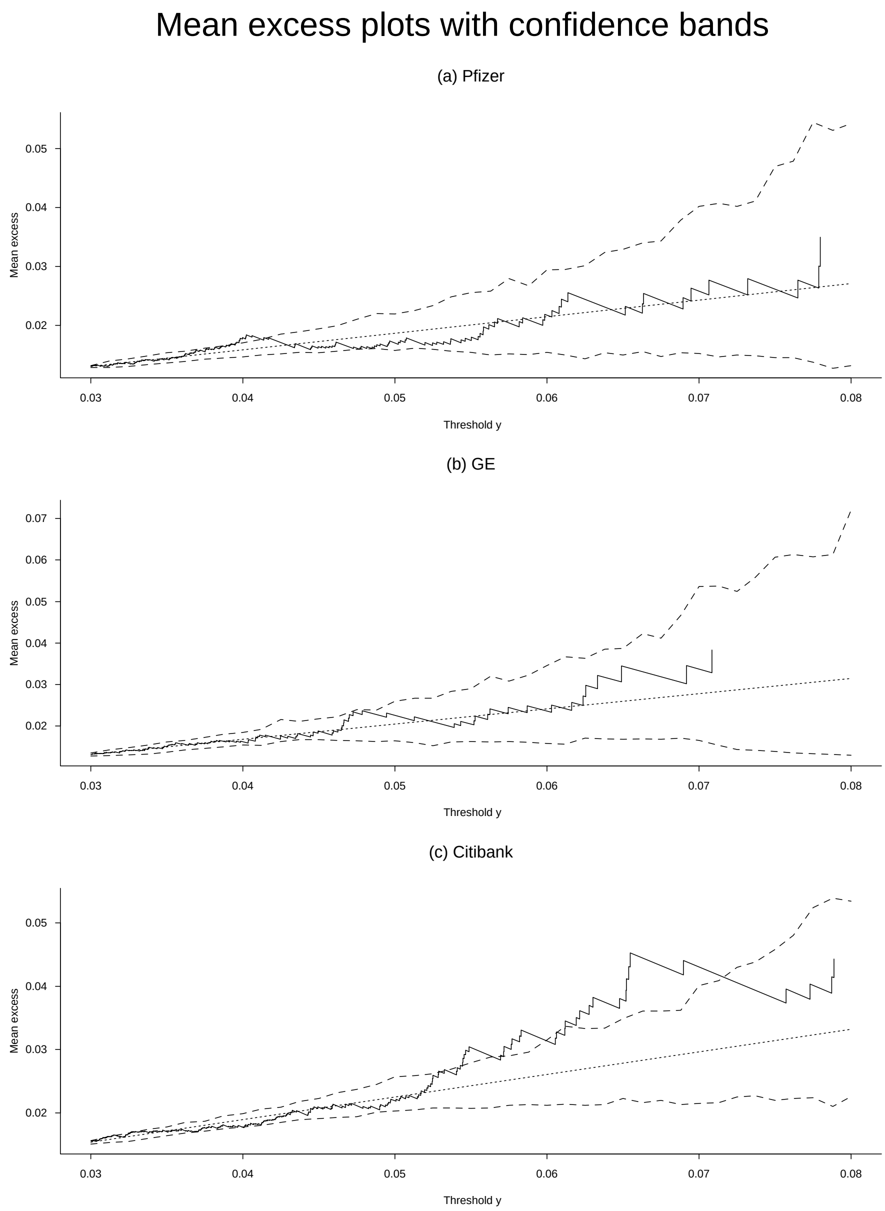

Figure 5.2: The mean excess plot usually suggests whether a extreme value distribution fitting is appropriate or not. The plot for the Pfizer data suggests extreme value distribution fitting since the plot is contained in its corresponding confidence interval. The other two are more doubtful since the plot goes outside the confidence bands, though further analysis shows that the extreme value approximation is reasonable in this case also.

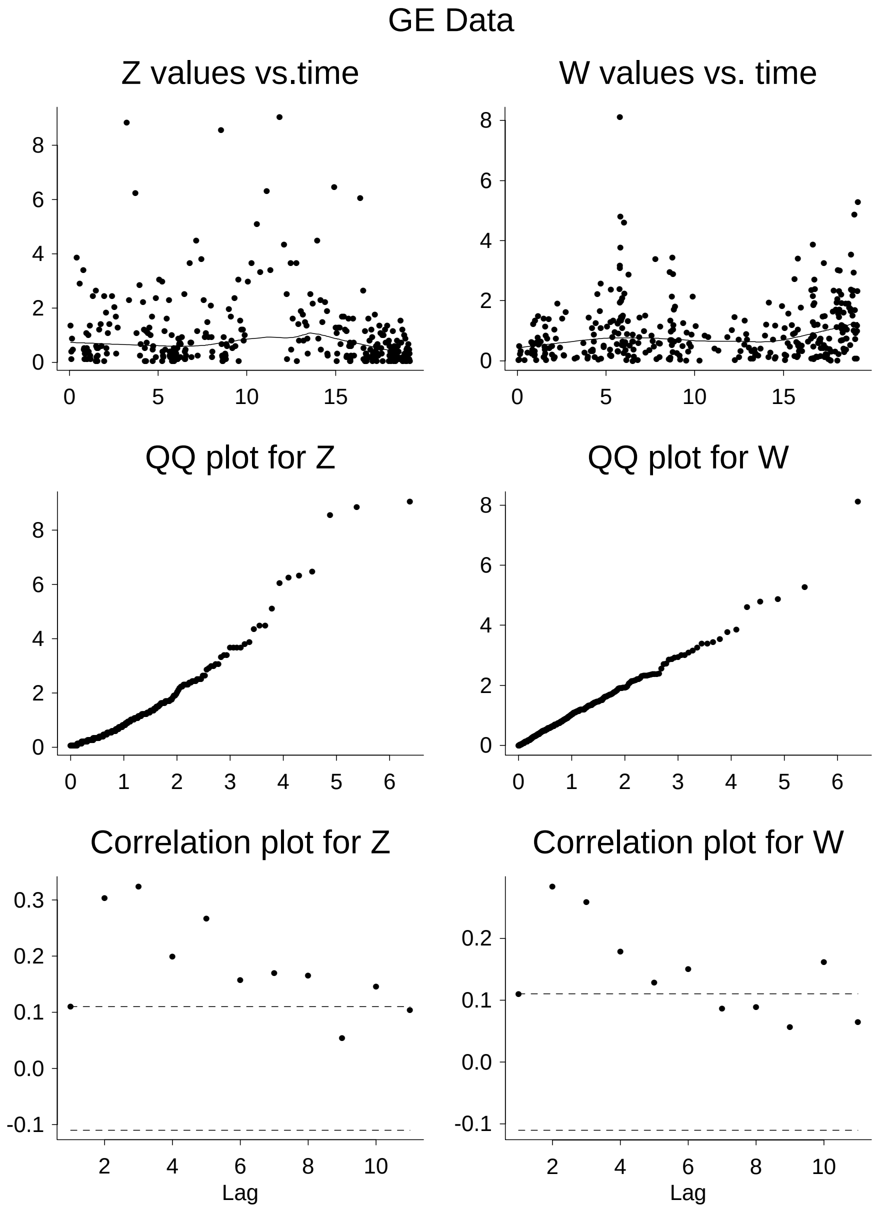

Figure 5.3: W-plots show a generalized extreme value distribution fitting is appropriate. Some caution should be given since a few points, partly the result of Oct87 crash, are away from the straight line.

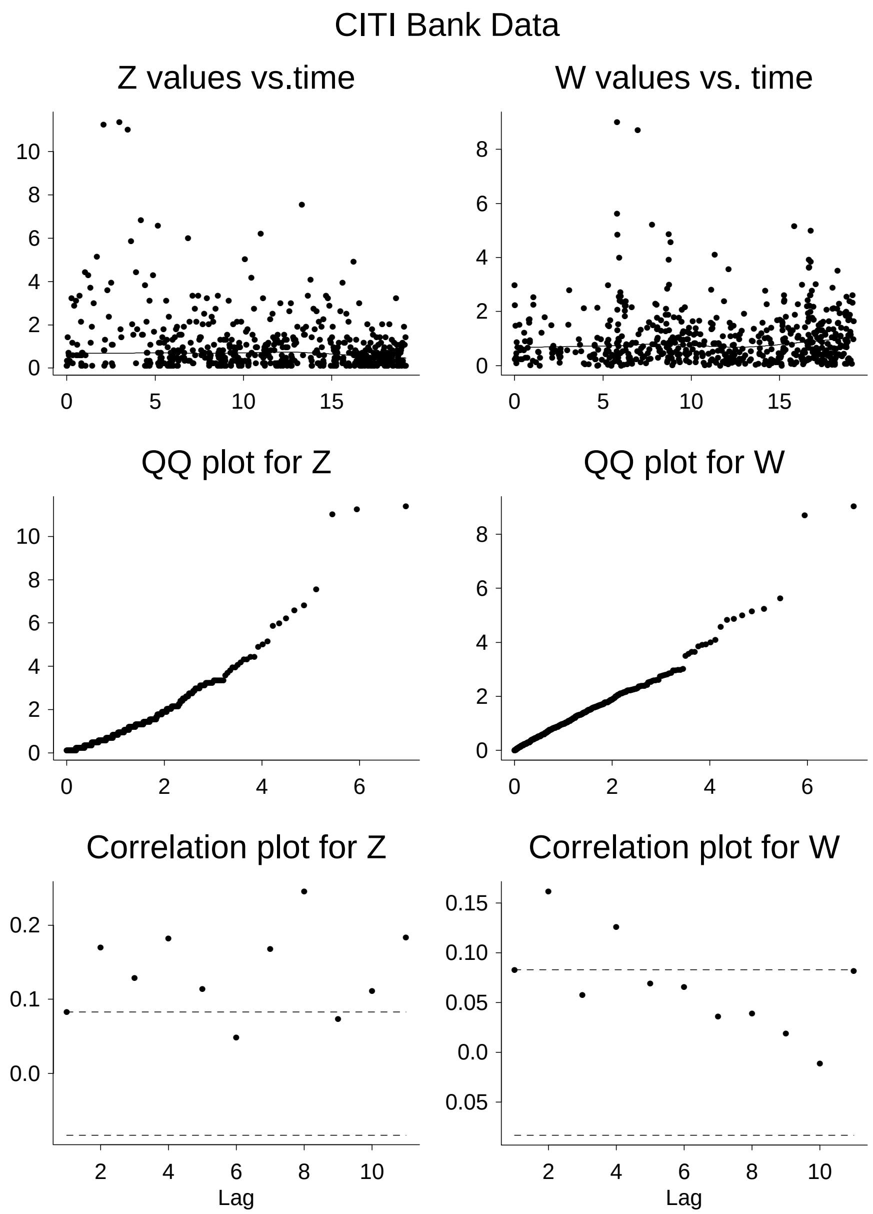

Figure 5.4: W-plots show a generalized extreme value distribution fitting is appropriate.

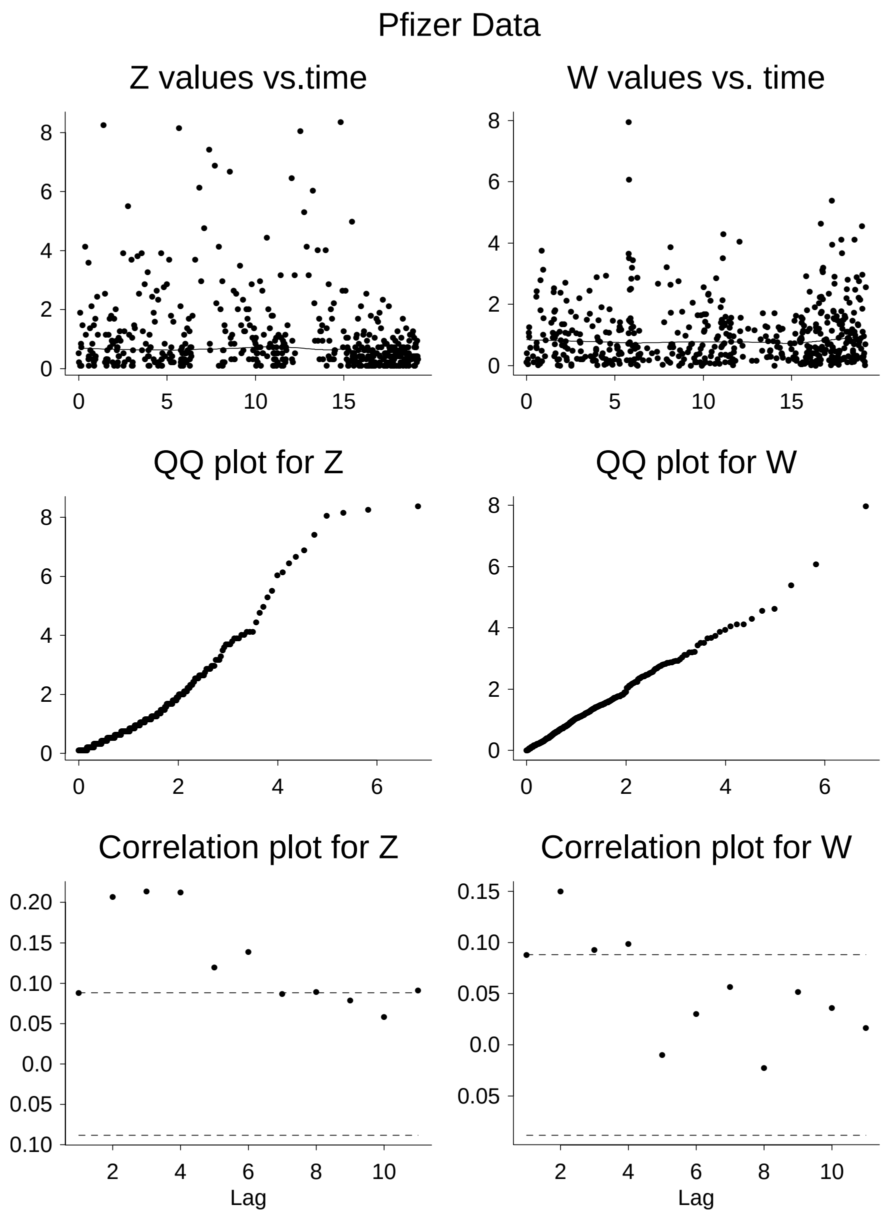

Figure 5.5: W-plots show a generalized extreme value distribution fitting is appropriate.

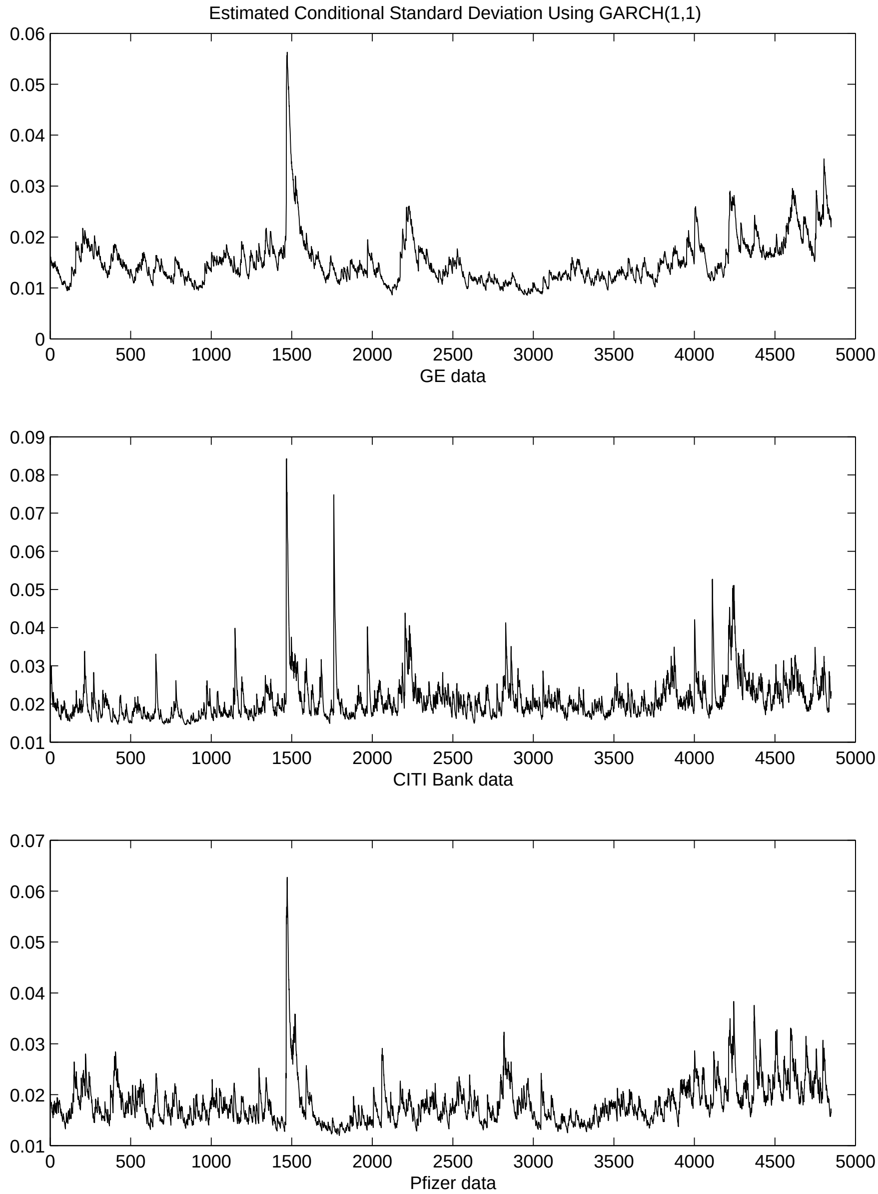

Figure 5.6: Estimated volatility

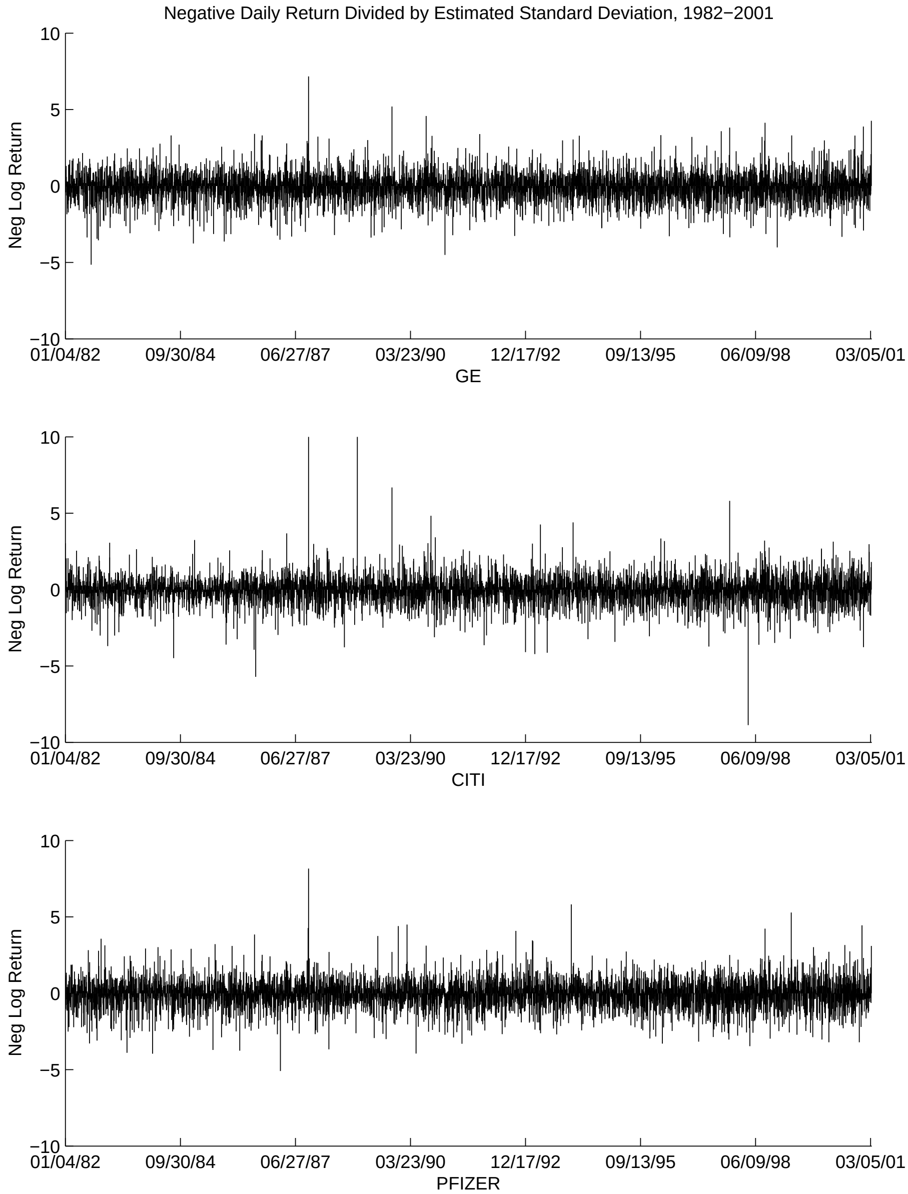

Figure 5.7: Standardized series

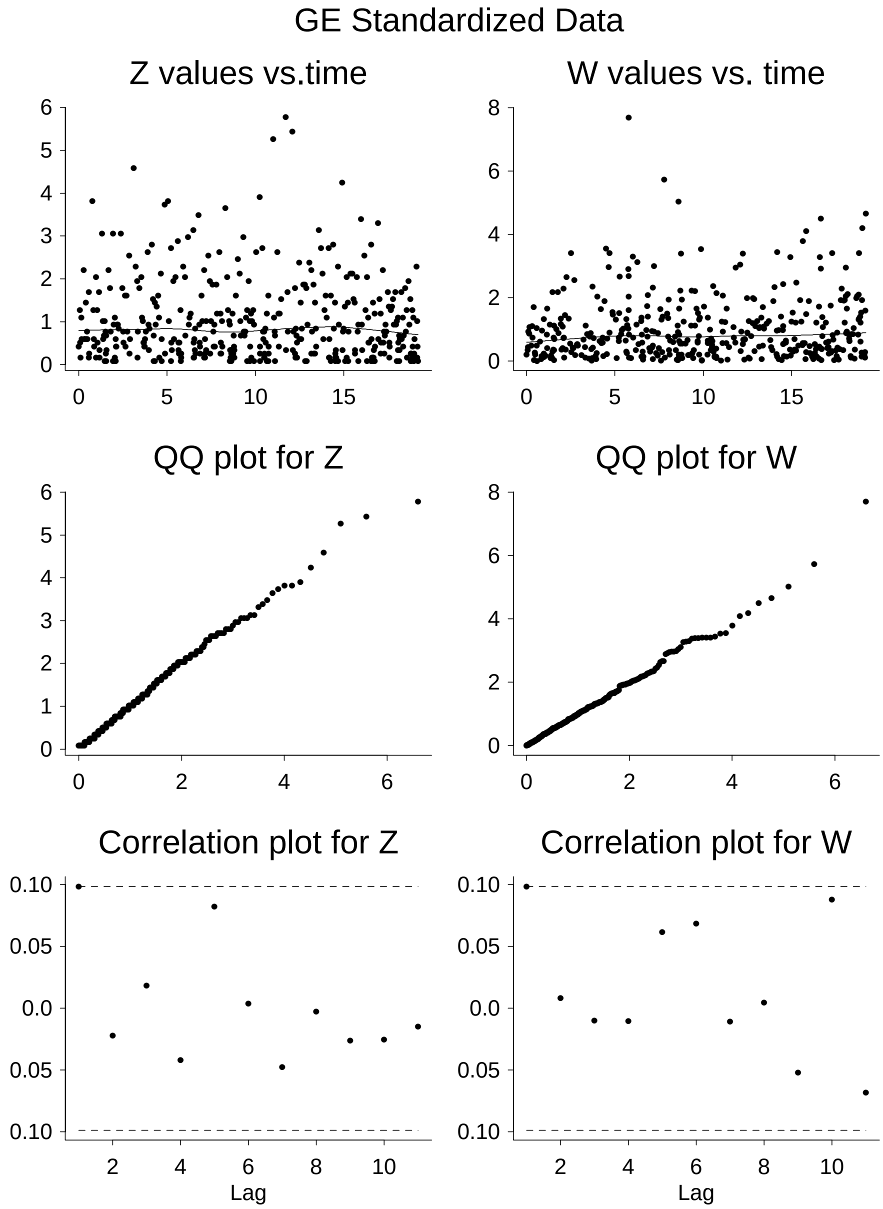

Figure 5.8: W-plots show a generalized extreme value distribution fitting is appropriate. Some caution should be given since a few points, partly the result of Oct87 crash, are away from the straight line.

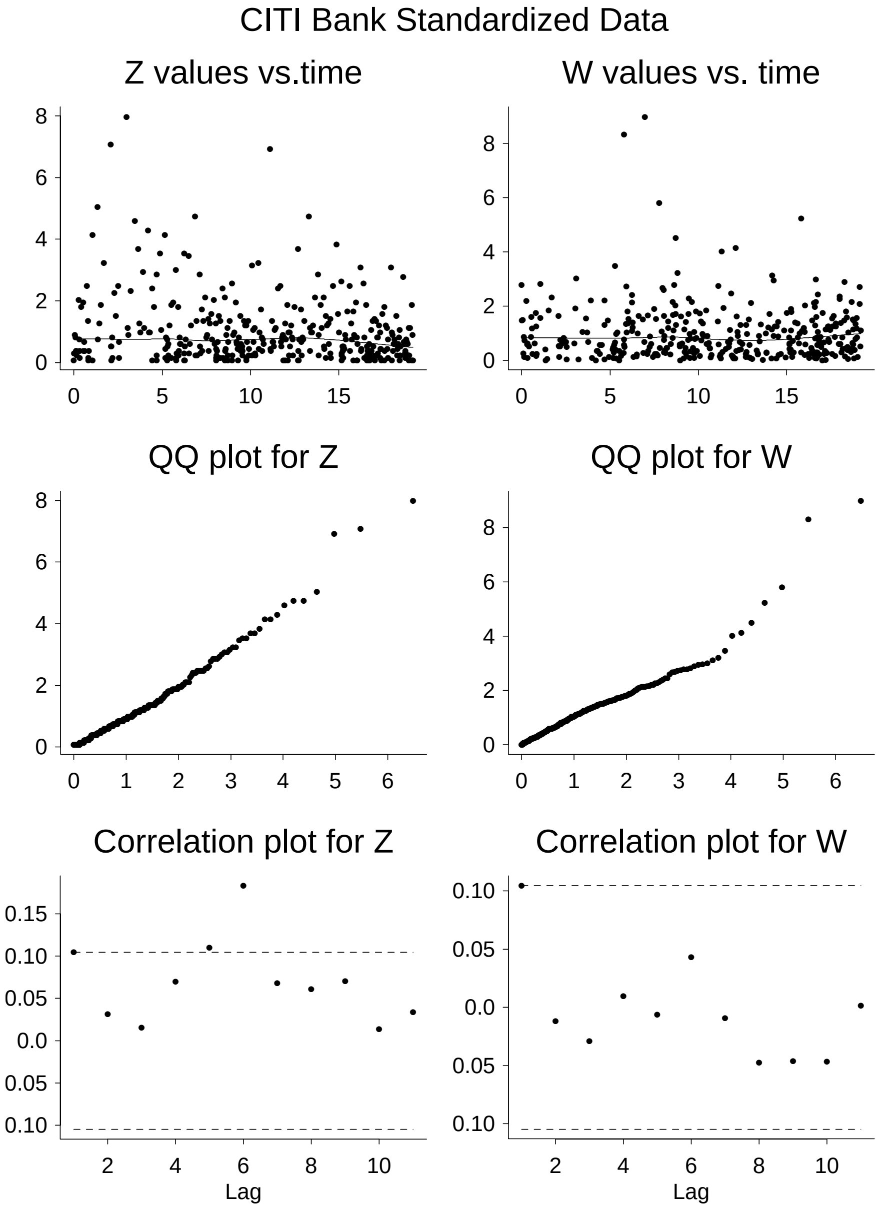

Figure 5.9: W-plots show a generalized extreme value distribution fitting is appropriate.

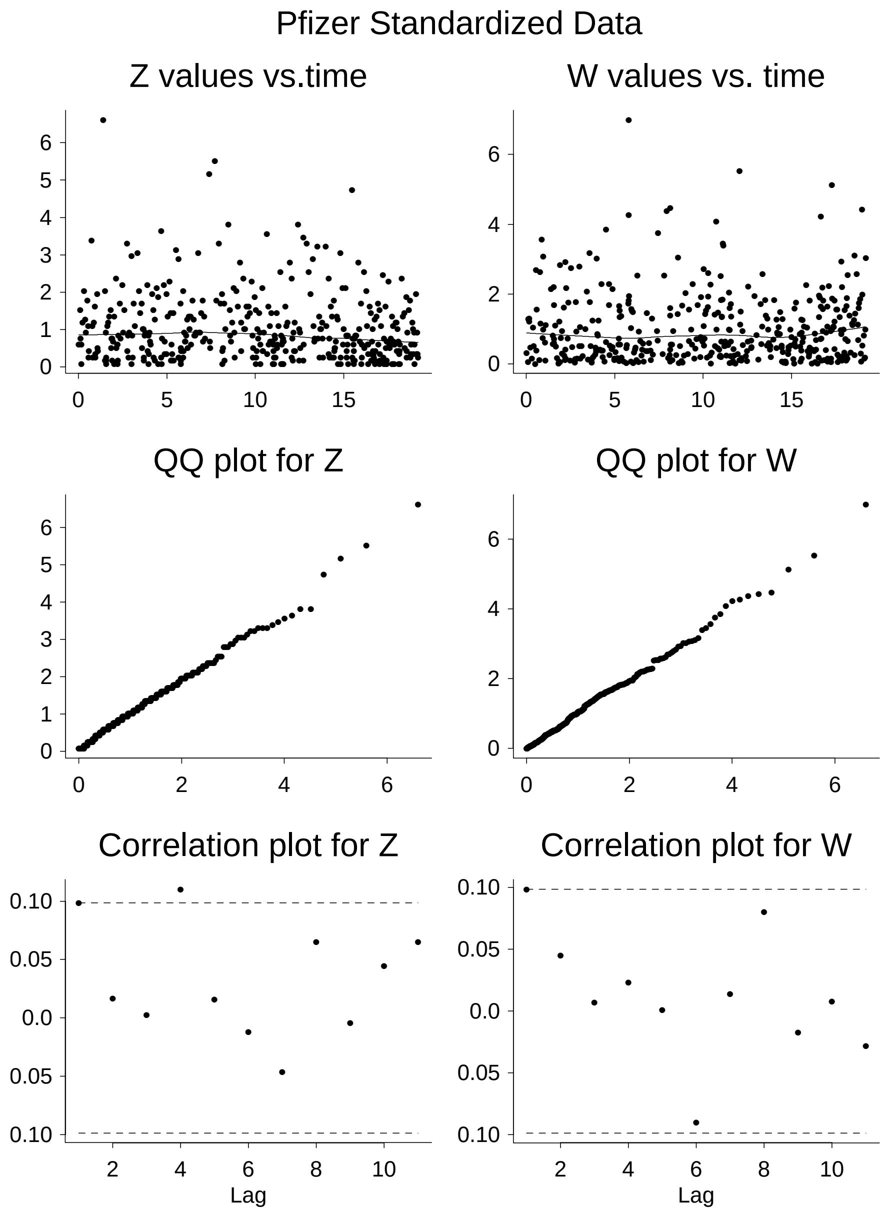

Figure 5.10: W-plots show a generalized extreme value distribution fitting is appropriate.

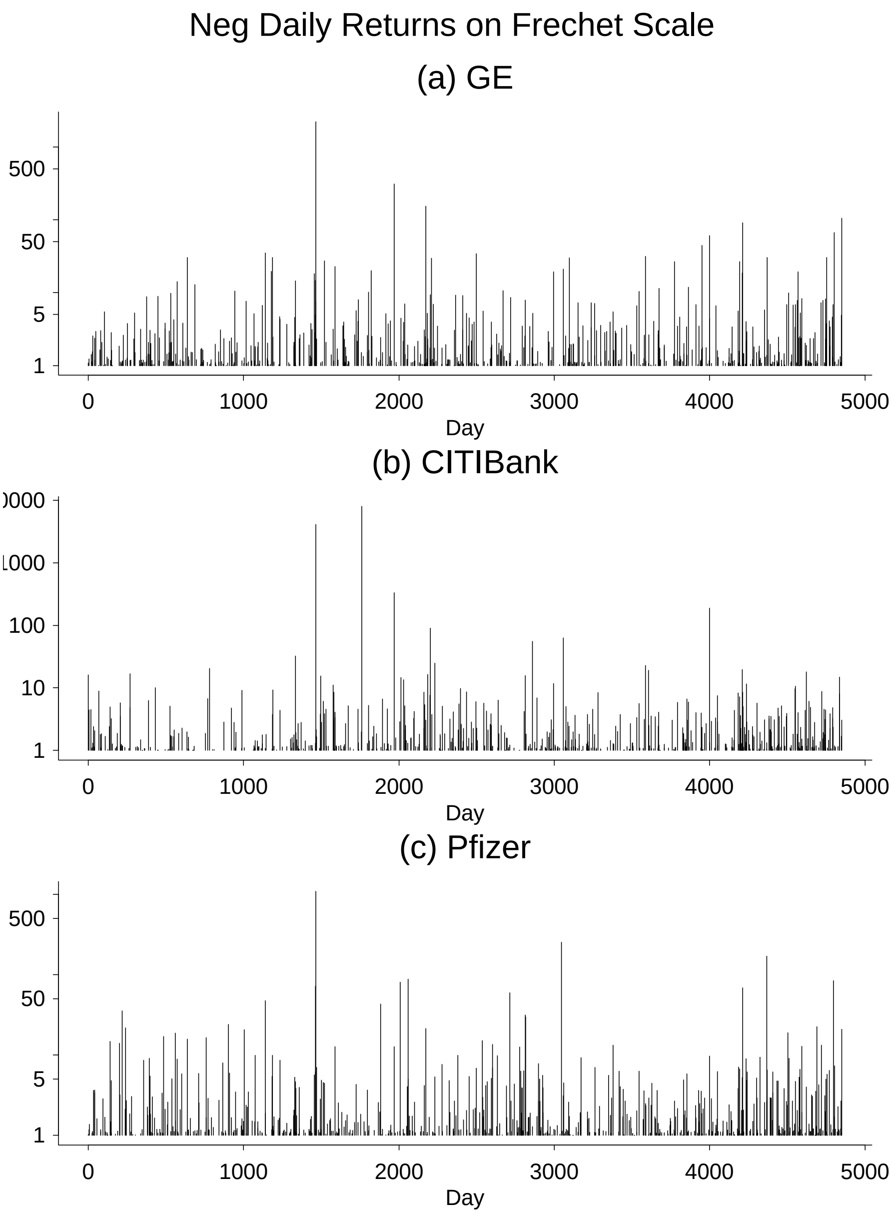

Figure 5.11: The negative returns after transformed into unit Fréchet scale.

two consecutive daily drops. We also count the days of two day range in which one stock price dropped in the first day, the other stock price dropped in the second day. We also calculate the expected counts under the assumption of independent.

We observe throughout that the "observed" values are larger than the "expected", and therefore we conclude that the variables are dependent. The choice of $M4$ processes to model the dependence may be appropriate.

5.5.6 Model selection and parameter estimations

All the figures suggest an $M4$ process fitting may be a good choice for financial time series data with multivariate temporal dependence.

Figure 5.11 and 5.12 suggest that a model of time dependance range of order 2 or 1 and at least 3 signature patterns. Some of these patterns have order of 2, corresponding to drops happened in two consecutive days, and one has order of 1, which corresponds to a single drop. Figure 5.12 shows that the strong dependence appears in the same day between series and in two consecutive days within each series.

We now use the following model to fit the data.

$$\begin{aligned} Y_{id} = \max(a_{1,-1,d}Z_{1,i-1}, &\quad a_{1,0,d}Z_{1,i}, \\ &\cdots \\ a_{L-1,-1,d}Z_{L-1,i-1}, &\quad a_{L-1,0,d}Z_{L-1,i}, \\ &\quad a_{L,0,d}Z_{L,i}) \end{aligned} \tag{5.11}$$

But we need to determine the number of signature patterns L. Define

$$Q_1(x) = \Pr(u + x > Y_{i+1} > u,\ Y_i > u)$$

$$Q_2(x) = \Pr(u + x > Y_{i+1} > u,\ Y_i < u)$$

$$A_1(x) = (u, +\infty) \times (u, u + x)$$

$$A_2(x) = (0, u) \times (u, u + x)$$

$$\bar{X}_{A_j(x)} = \frac{1}{n}\sum_{i=1}^{n} I_{A_j(x)}(Y_i, Y_{i+1}),\ j = 1, 2.$$

We extend Kolmogorov and Smirnov's distance into the following form

$$err_l = \sup_{x>0}[|Q_1(x) - \bar{X}_{A_1(x)}|, |Q_2(x) - \bar{X}_{A_2(x)}|]\}$$

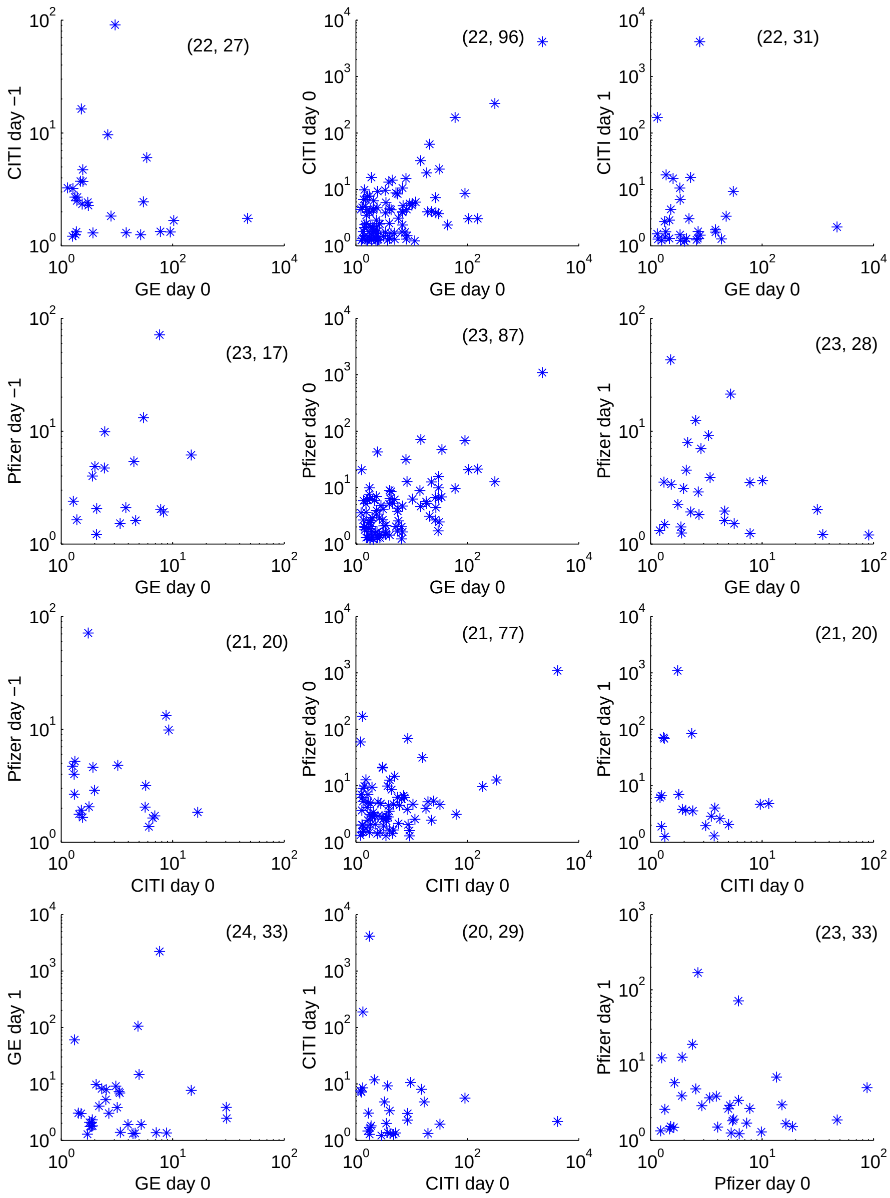

Figure 5.12: (i) All plots are based on Fréchet transformed exceedances of a high threshold based on negative log returns (so they represent price drops, not price rises); (ii) the purpose of the plots is to look for dependence among neighboring values; (iii) the numbers in parentheses show expected and observed numbers of simultaneous exceedances by the two variables, where “expected” is calculated on an independence assumption.

Where $Q_1(x)$ and $Q_2(x)$ are calculated under estimated parameter values.

The idea here is that we use the given information up to today and compute the probability of the event that the return falls into a certain range next day. Instead of computing conditional probability, we compute joint probability and joint empirical probability and compare the approximation errors.

For $l = 3, 4, 5, \ldots,$ we compute err_l at discrete points in interval $[u, 2u]$. Figure 5.13 shows the trends of err_l when l is increasing. We can see when $l = 5$ the curve reaches stability for CITI data. And similarly $l = 6$ for GE and Pfizer data.

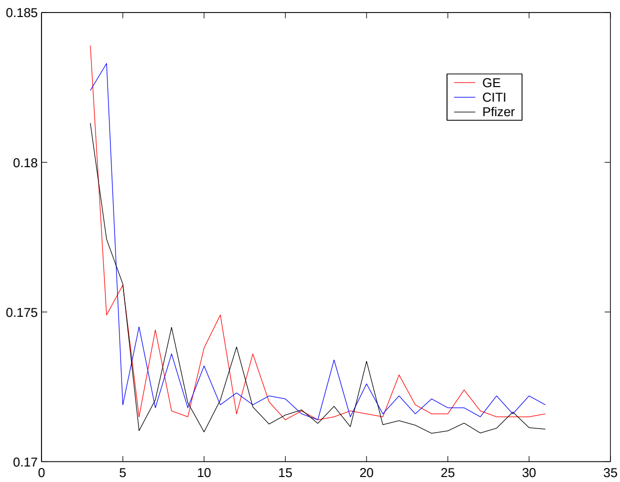

Figure 5.13: Number of signature patterns L vs Err plot.

We now fit three transformed time series data using the following model.

$$
\begin{aligned}
Y_{id} = \max(a_{1,-1,d}Z_{1,i-1}, \quad & a_{1,0,d}Z_{1,i}, \\
\cdots & \\
a_{5,-1,d}Z_{5,i-1}, \quad & a_{5,0,d}Z_{5,i}, \\
a_{6,-1,d}Z_{6,i-1}, \quad & a_{6,0,d}Z_{6,i}, \\
\cdots & \\
a_{9,-1,d}Z_{9,i-1}, \quad & a_{9,0,d}Z_{9,i}, \\
a_{10,-1,d}Z_{10,i-1}, \quad & a_{10,0,d}Z_{10,i}, \\
\cdots & \\
a_{14,-1,d}Z_{14,i-1}, \quad & a_{14,0,d}Z_{14,i}, \\
& a_{15,0,d}Z_{15,i})
\end{aligned}
\tag{5.12}
$$

where $a_{l,k,d} = 0, \ l = 1, \ldots, 5, \ d = 2, 3, \ a_{l,k,d} = 0, \ l = 6, \ldots, 9, \ d = 1, 3, \ a_{l,k,d} = 0, \ l = 10, \ldots, 14, \ d = 1, 2.$ We have considered here to treat those drops in two consecutive days as independent processes and single drops as dependent processes. We can apply more complex structure and use the arguments about how to model inter-series dependence discussed in section 3.3. But we adopt a relatively simple model to illustrate $M4$ process modeling to financial time series data here.

Table 5.5.6 is constructed based on the probabilistic properties of $M4$ processes. The returns above certain threshold in two consecutive days are clustered into 4 or 5 groups. Initial estimates are obtained by taking average of those points within each group. Then we can solve the system of nonlinear equations or minimize a weighted least squares functions. What we actually do here is to use the initial estimates as the base of selecting evaluation points $x_1, x_2, \cdots, x_m$ used in (3.11) and (3.12). Based on the initial estimates, we compute the adjacent parameter ratios. For each ratio value, say r_i, we let $x_j = .95$ and $x_{j+1} = 1.25r_i$. The value of m is equal to twice of the number of those ratios. The numbers .95 and 1.25 are arbitrary. Other numbers can be used as long as one is less than 1 and another one is larger than 1. Intuitively we choose the numbers as big as possible because the nature of function $\widehat{b}(x)$, which has less variability for large x, but we need to have two points between two adjacent ratios. We don't have a criterion to guide the choice of x_j's at this time. An optimization problem based on a's and x_j's may be useful, but we will not pursue this in the current work. After the x_j's have been chosen we solve (3.16) under a constraint that the matrix formed in the left hand side of (3.16) is the same as the one when the initial estimates are used. Using the estimated parameter values, we can compute asymptotic covariance matrix. Since we only have about 2200 observed drops but many parameters, the computation of the asymptotic covariance matrix of the joint distribution of the estimates is not efficient, i.e. the standard deviations of the estimators are not less than 1 as they are supposed to be. The results in Table 5.5.6 are simulation results based on 100 replications of sample size 1000.

Table 5.1: Estimations of Parameters

Param.	GE	CITI	Pfizer
$a_{1,-1}/a_{1,0}$ std	0.0614 / 0.0206 0.0169 / 0.0075		
$a_{2,-1}/a_{2,0}$ std	0.0778 / 0.0026 0.0381 / 0.0019		
$a_{3,-1}/a_{3,0}$ std	0.0220 / 0.0174 0.0089 / 0.0089		
$a_{4,-1}/a_{4,0}$ std	0.0070 / 0.0121 0.0043 / 0.0100		
$a_{5,-1}/a_{5,0}$ std	0.0020 / 0.0126 0.0011 / 0.0104		
$a_{6,-1}/a_{6,0}$ std		0.1937 / 0.0005 0.0482 / 0.0002	
$a_{7,-1}/a_{7,0}$ std		0.0086 / 0.0077 0.0079 / 0.0077	
$a_{8,-1}/a_{8,0}$ std		0.0012 / 0.0067 0.0014 / 0.0092	
$a_{9,-1}/a_{9,0}$ std		0.0003 / 0.0062 0.0003 / 0.0071	
$a_{10,-1}/a_{10,0}$ std			0.1594 / 0.0260 0.0162 / 0.0041
$a_{11,-1}/a_{11,0}$ std			0.0190 / 0.0105 0.0089 / 0.0074
$a_{12,-1}/a_{12,0}$ std			0.0062 / 0.0078 0.0027 /0.0060
$a_{13,-1}/a_{13,0}$ std			0.0029 / 0.0055 0.0014 / 0.0049
$a_{14,-1}/a_{14,0}$ std			0.0011 / 0.0062 0.0005/ 0.0052
$a_{15,0}$ std	0.7645 0.0128	0.7750 0.0274	0.7553 0.0042

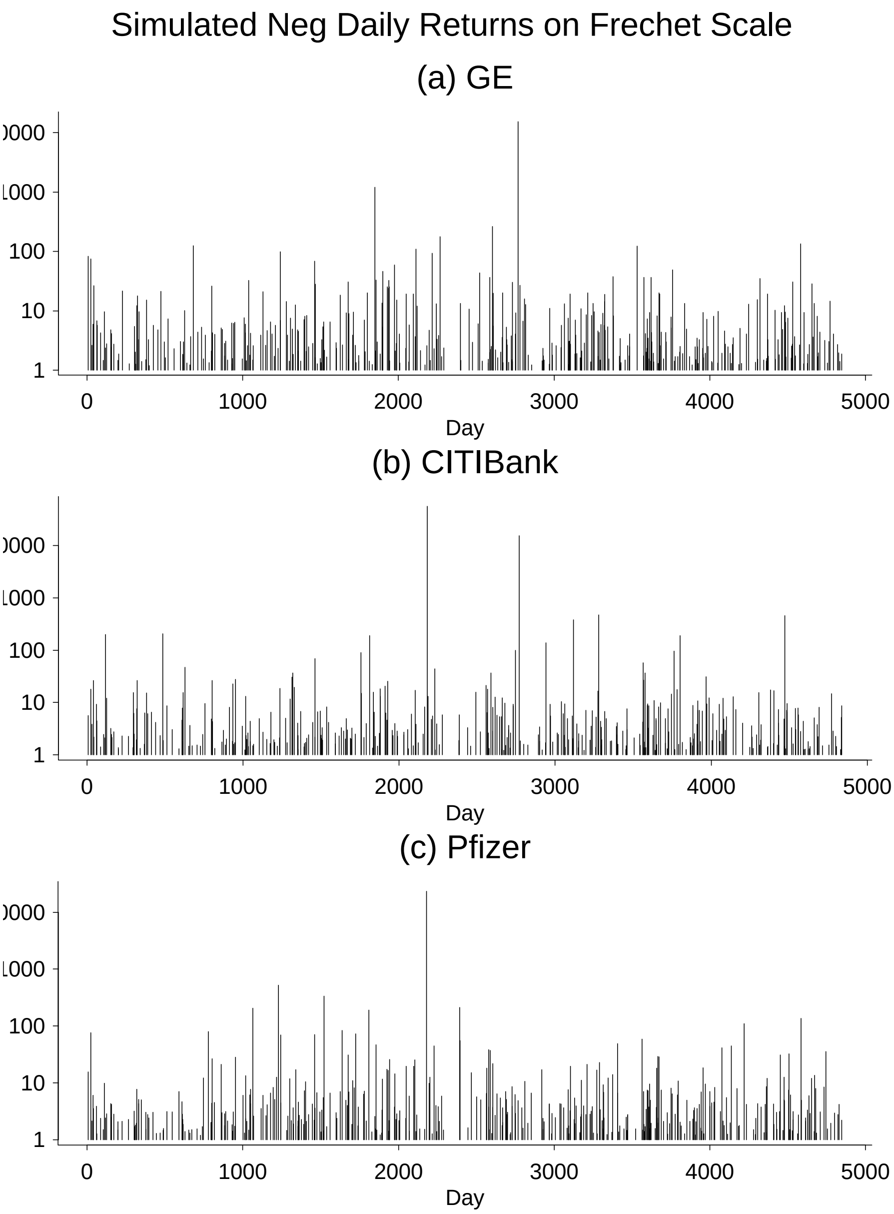

Figure 5.14: Simulated time series

5.6 VaR calculation and portfolio optimization

So far we have considered modeling the three time series to $M4$ processes. We now turn to do VaR calculation and portfolio optimization. For comparison, we will use variance-covariance approach and extreme value approach.

5.6.1 Using variance-covariance approach

We now compute the VaR of a portfolio containing three stock products, GE, CITI and Pfizer. At time t the returns are R_{1t}, R_{2t}, R_{3t}. We write $\mathbf{R}_t = (R_{1t}, R_{2t}, R_{3t})'$. $\mathbf{R}_t$ is assumed to be jointly normally distributed, i.e. $\mathbf{R}_t \sim N(\mu, \Sigma)$. We can estimate the covariance matrix Σ by the sample covariance matrix

$$\widehat{\Sigma} = \frac{1}{1000}\begin{pmatrix} 0.2539 & 0.0408 & 0.0180 \\ 0.0408 & 0.4591 & 0.1246 \\ 0.0180 & 0.1240 & 0.3557 \end{pmatrix}.$$

For the standardized time series we have the estimates

$$\widehat{\Sigma} = \frac{1}{1000}\begin{pmatrix} 0.2310 & 0.0298 & 0.0156 \\ 0.0298 & 0.4399 & 0.1136 \\ 0.0156 & 0.1136 & 0.3372 \end{pmatrix}.$$

The following figures plot calculated VaRs for different investment combinations, using original time series, standardized time series. We also plot the VaR versus Expected returns. In all plots, each point represents a portfolio investment plan or combination. For example, one point may represent (.25, .35, .4) which means 25% of money invested in GE, 35% of money invested in CITI and 40% of money invested in Pfizer.

Figures 5.15 and 5.16 are based on the original data. Figures 5.17 and 5.18 are based on the standardized data. Both figures 5.15 and 5.17 show risk diversification. In all figures, a green circle means investing all money on GE stock, a red cross means investing all money on CITI stock, and a blue diamond means investing all money on Pfizer stock. A green dot, a red dot or a blue dot means investing 50% or more of total money on stock GE, CITI or Pfizer respectively. A black dot means no individual stock received more than 50% of total money. From figures 5.16 and 5.18 we can optimize the portfolio with highest expected return for a given level of risk. It is the point of the upper curve when the risk is given.

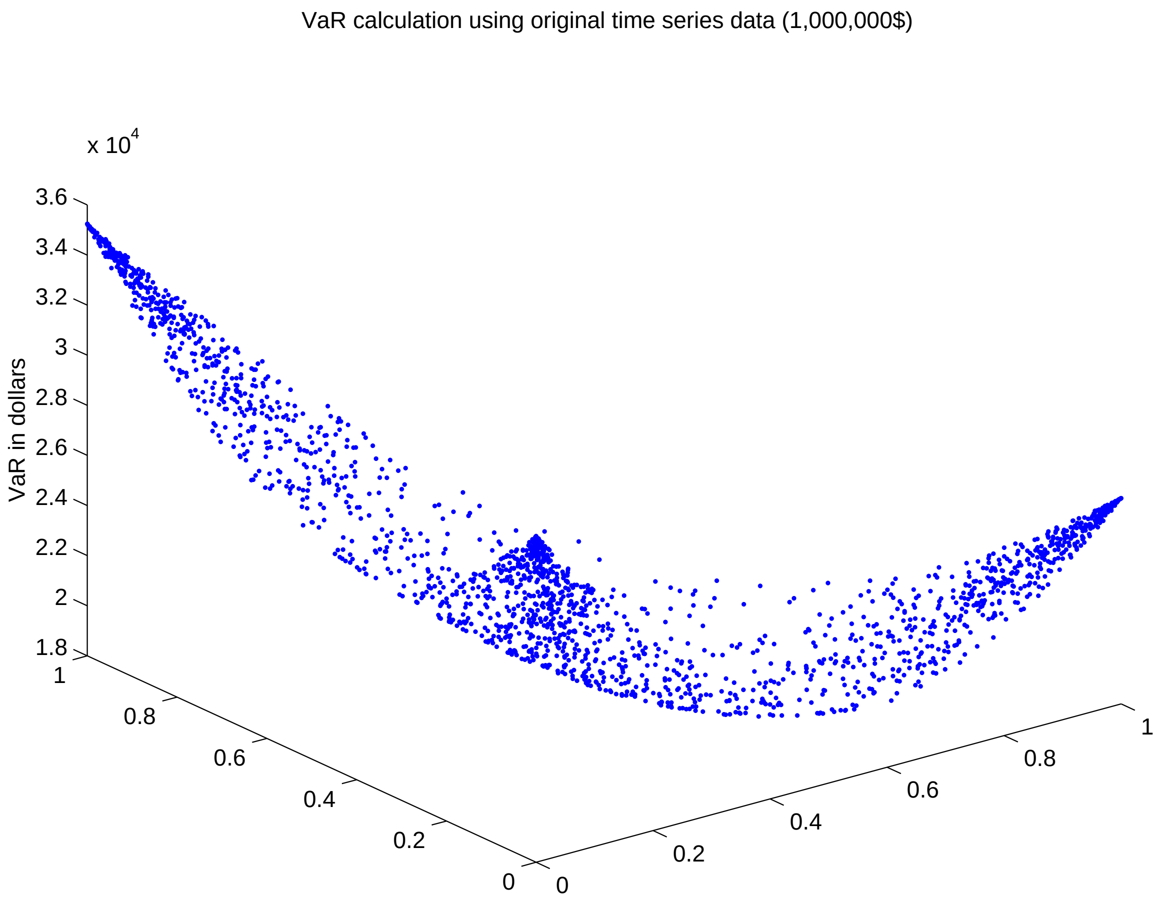

Figure 5.15: VaR for a portfolio of 3 stock products

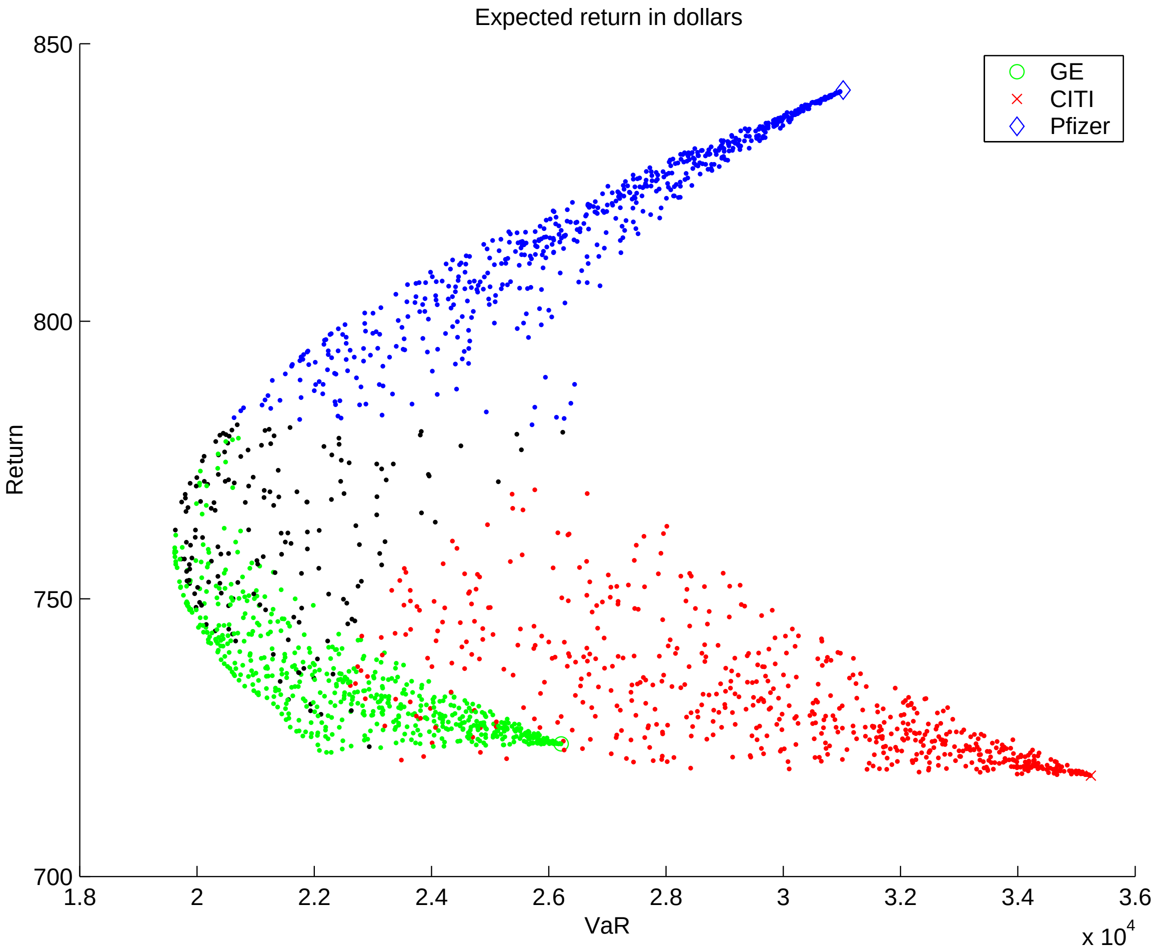

Figure 5.16: Expected returns for a portfolio of 3 stock products

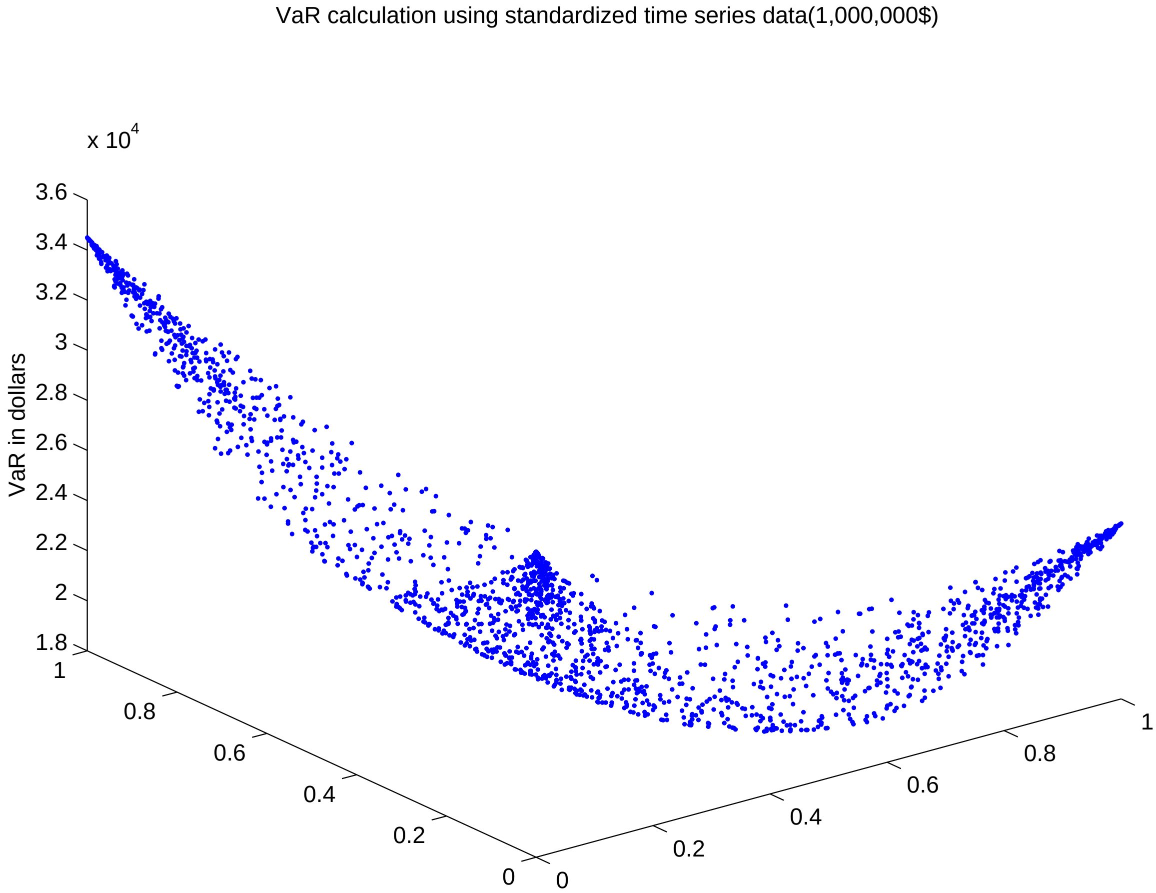

Figure 5.17: VaR for a portfolio of 3 stock products, standardized data

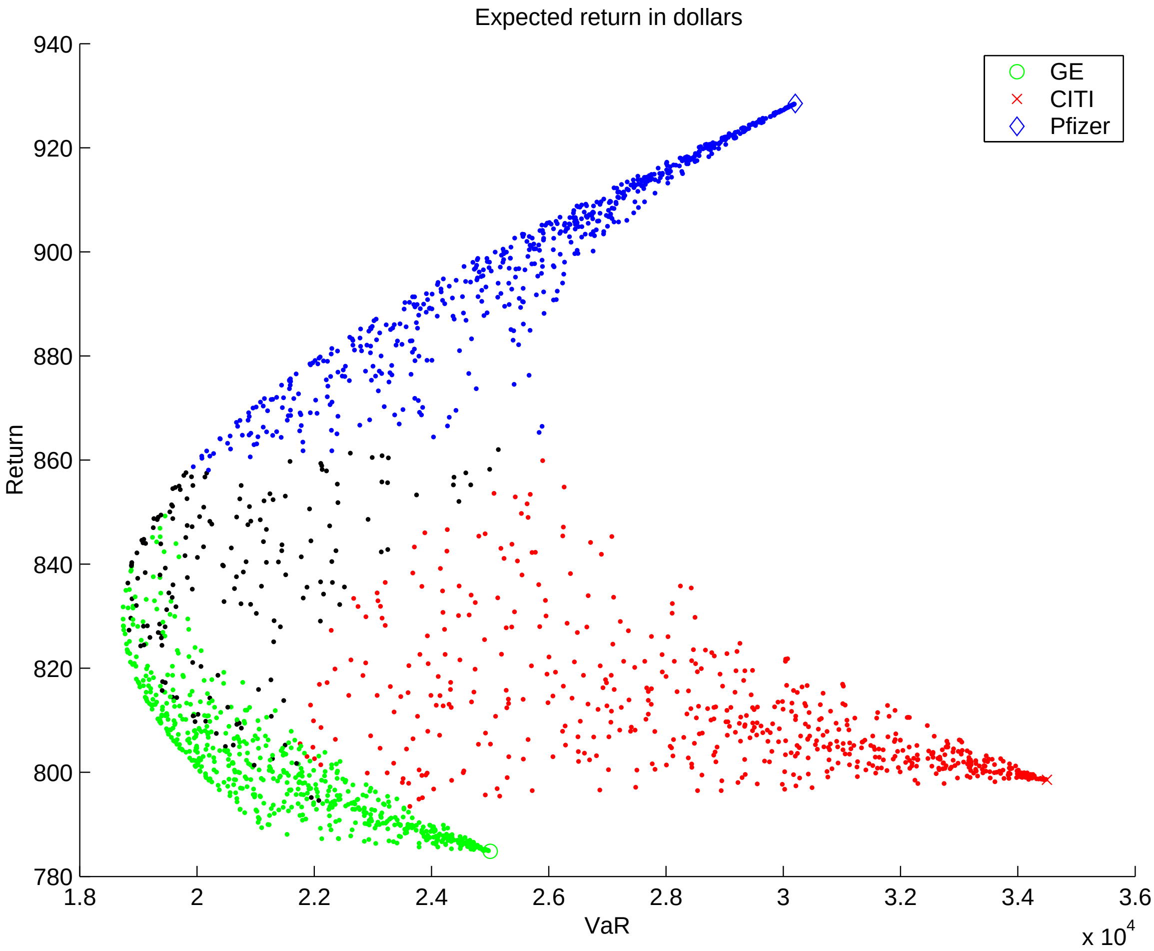

Figure 5.18: Expected returns for a portfolio of 3 stock products, standardized data

5.6.2 Using $M4$ process approach

The multivariate normal distribution assumption makes most statistical computations easier. But it may give inaccurate results if it doesn't fit the data and result in a wrong decision. In risk measurement it may under estimate the risk since Normal distribution is best fitting the centralized data. We now introduce a new method based on the $M4$ process modeling and use it to compute VaR and optimize a portfolio with the VaR constrains.

Suppose $c_1,\ c_2,\ c_3$ are proportions of stock products in a portfolio, VaR_p is the VaR of the portfolio return $c_1Y_1 + c_2Y_2 + c_3Y_3$ calculated from $P(c_1Y_1 + c_2Y_2 + c_3Y_3 > VaR_p) < \alpha$ for given level of confidence $1 - \alpha$. Due to the complex transformation to the original data, it's not possible to compute the maximal possible loss based on VaR_p since each individual stock behaves differently. One way is to model a univariate time series $c_1Y_{1i} + c_2Y_{2i} + c_3Y_{3i}$ in $M4$ process. But this is not practically applicable since the investments change all the time.

We propose the following procedure to determine the $VaR_p = d$ and individual risk factors $VaR_1 = d_1,\ VaR_2 = d_2$ and $VaR_3 = d_3$ simultaneously.

$$\begin{cases} \max\limits_{d_1>0,d_2>0,d_3>0} P\{c_1Y_1 \geq d_1,\ c_2Y_2 \geq d_2,\ c_3Y_3 \geq d_3\} \\ \text{s.t. } P(c_1Y_1 + c_2Y_2 + c_3Y_3 > d) < \alpha \\ \qquad d = d_1 + d_2 + d_3 \end{cases} \tag{5.13}$$

The constraints are very natural since they are the definitions of VaR of a portfolio. The objective is thought to have the highest probability for all individual risk factors beyond certain values when the portfolio is at the VaR.

Figure 5.19 draws the VaR versus the different combinations. Contrast to Figure 5.15 and 5.17, it has same trend as the other two have. But it doesn't show smoothed features. Figure 5.20 plots the VaR and expected returns. It is very different from Figures 5.16, 5.18. The maximal VaR of candidate portfolios under normal assumption is less than the minimal VaR of same candidate portfolios calculated by extreme value approach. In extreme value model, when the VaR is beyond certain value, the expected highest return is decreasing when the risk is increasing.

From Figure 5.18, one see that Pfizer gives highest return if invest all money to stock Pfizer. Stock CITI gives highest risk. Stock GE has lowest return. From Figure 5.20, the extreme portfolios (invest all, or almost all, money to one stock) behave similarly. But the overall structure is different.

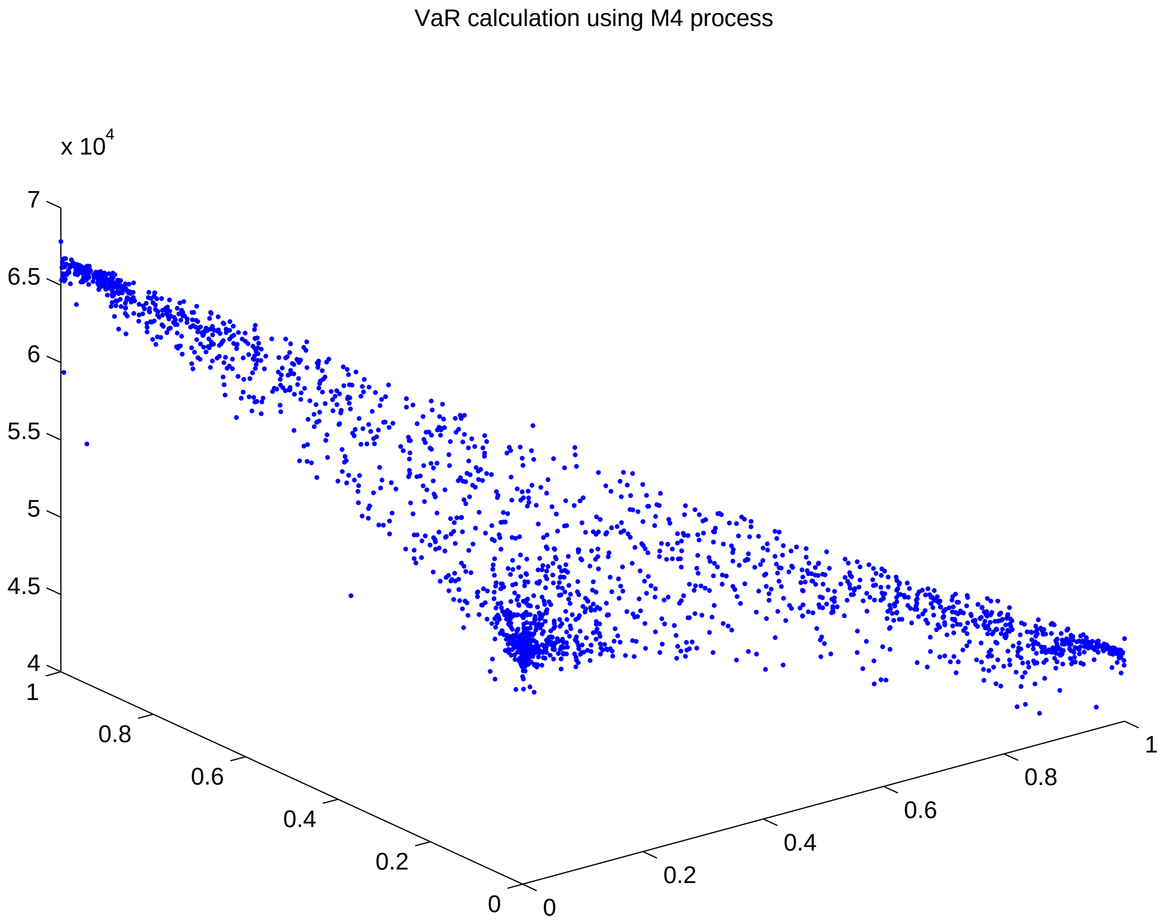

Figure 5.19: VaR for a portfolio of 3 stock products

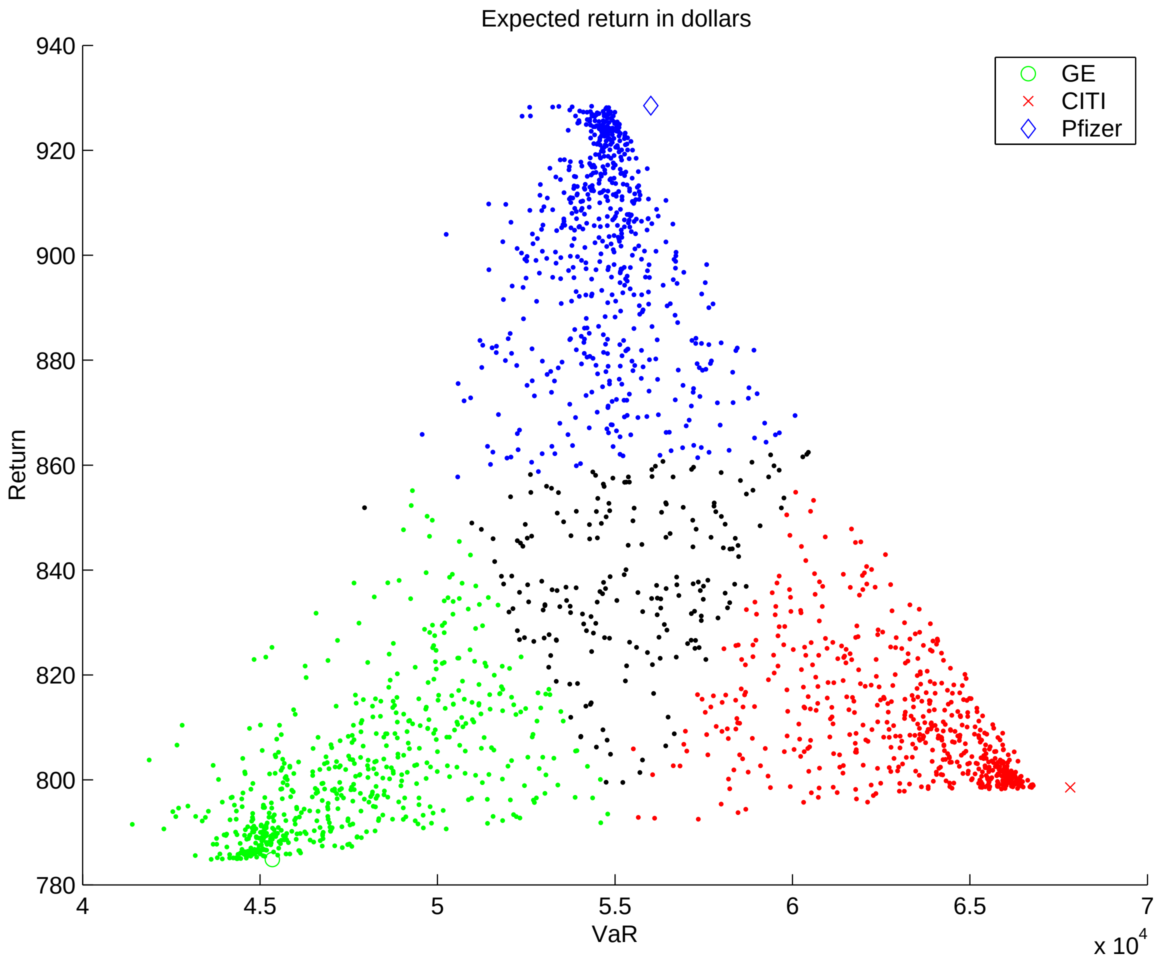

Figure 5.20: Portfolio Optimization Using Extreme Value Theory

5.6.3 Historical simulation approach

As we have seen that variance-covariance approach gives low VaR, while $M4$ process approach gives high VaR. We just can not simply tell which method is better. The goal of this section is to distinguish them more through the results from historical simulation approach.

The variance-covariance approach may be useful in routine risk management, while the extreme value approach should be the method used to extreme risk management. All VaR calculations in previous sections are unconditional, which means we didn't use the price move history. The variance-covariance approach can not be used to calculate the conditional VaR since they are under the independent assumption, but the extreme value approach can calculate the conditional VaR. Further comparisons between two approaches can be done using historical simulation approach which gives VaRs between VaRs obtained from the variance-covariance approach and the extreme value approach.

In Figure 5.21, we use those historical data when all three stocks had price drops simultaneously. As you can see the VaRs in Figure 5.21 are higher than those in Figure 5.16, but lower than those in Figure 5.20. If we use thresholds to historical simulation approach, we can see the VaRs move forward to the right, the VaRs in Figure 5.22 are very close to those, some are even higher than, in figure 5.20.

Since the historical simulation approach doesn't model dependence structure and is very difficult to calculate the conditional VaRs, the extreme value approach may be a better method, especially when considering extreme risk management.

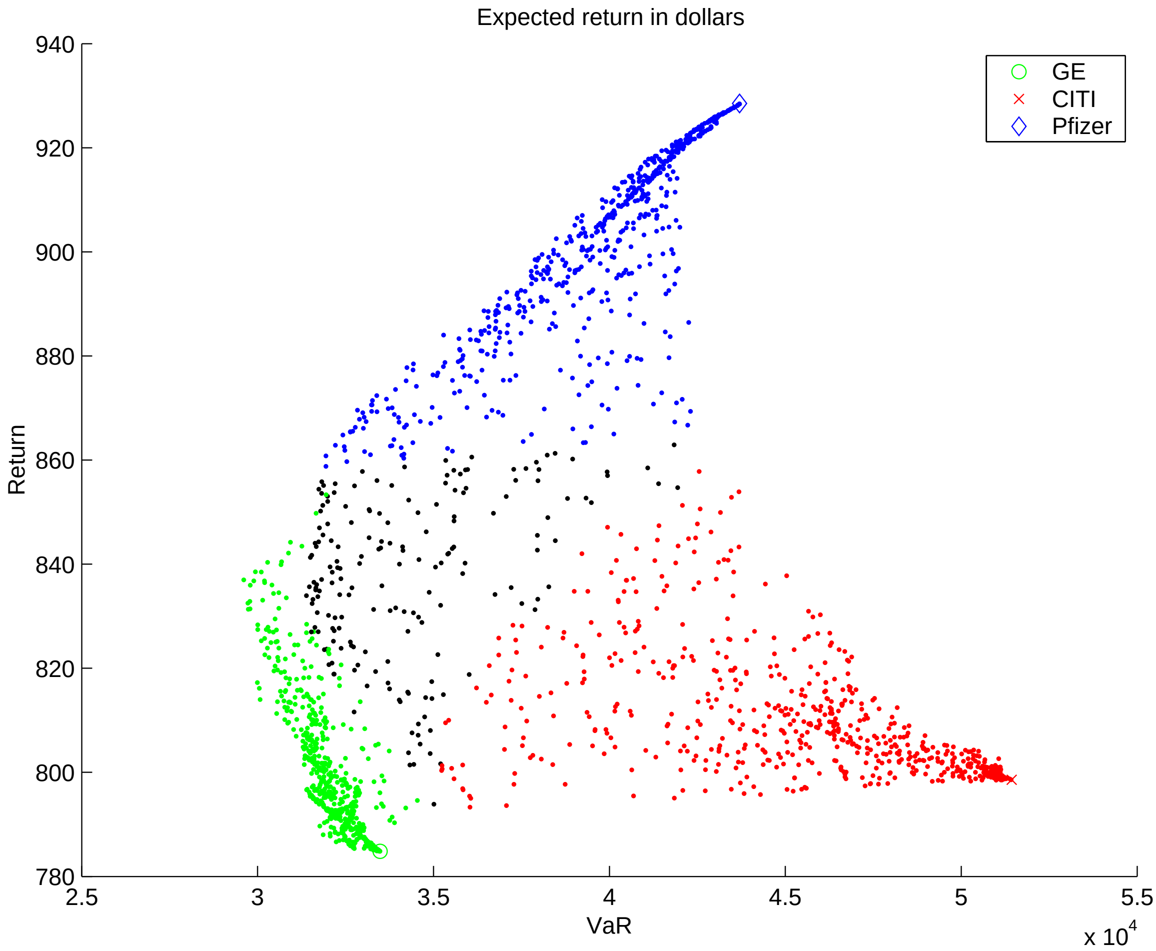

Figure 5.21: Portfolio Optimization Using historical simulation approach

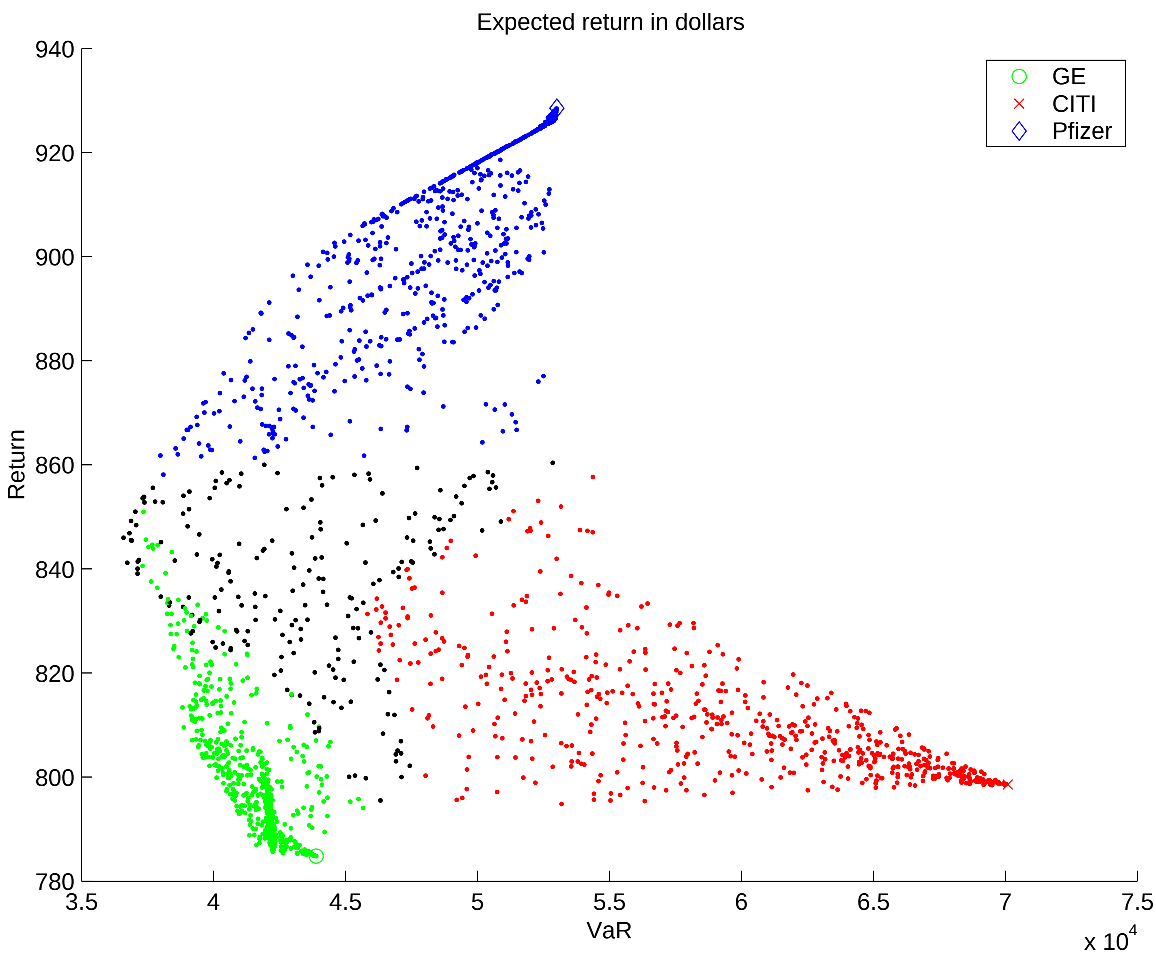

Figure 5.22: Portfolio Optimization Using historical simulation approach

Chapter 6

Summary

6.1 General discussion

The methods described here represent completely new approaches to the modeling of financial time series data. The main goal here is to propose an approach which can efficiently model multivariate time series which are both inter-serially and temporally dependent.

In order to achieve that goal, we have extended and proved some probabilistic properties of $M4$ processes. Then we have proposed estimating procedures when the ratios can be determined using probabilistic approach. We have also proposed a practically applicable method of $M4$ processes modeling. The consistency and asymptotic properties have been proved. We have also proposed a VaR calculation method based on $M4$ process modeling. The main theorems proved are Theorem 2.3, Theorem 2.4, Theorem 2.16, Theorem 3.7, Theorem 3.8, Proposition 3.12, Proposition 3.13, Theorem 3.15.

The results obtained can be used in many ways. For example, they can be used to compute VaR or to optimize the portfolio under VaR constraints and given information or historical data. Studies have shown financial data are fat tailed. They are not normally distributed. Compare with traditional assumption of normality of underlying distribution. These results provide more information to risk managers who may be most interested in the situation when an extreme price movement occurs what kind of risk the company is exposed to. The methods described can be used to other fields, such as modeling insurance data, environment data etc.

It may be possible to propose some variants of proposed estimators and to reduce the conditions imposed on the parameters. The choice of the points around the jump points and the selection of model need some further work.

6.2 Directions of future research

In this section, I list some research directions under extreme value settings.

- Some extreme theorist believe that multivariate extreme value study still has a long way to go. Besides $M4$ processes, it is worth to explore other dependence structures which have multivariate extreme value distribution representation.
- It may be worth to study modeling time series data through Markov process and extreme events, or Bernoulli jumps and extreme process, or Poisson jumps and extreme process.
- In a short term, it is worth to study the models:

$$Y_{id} = \max_{1\le l\le L} \max_{-K_1\le k\le K_2} a_{lkd} Z_{l,i-k} + N(0,\sigma^2), \quad d = 1, \cdots, D, \tag{6.1}$$

  where$\sum_{l=1}^{L}\sum_{k=-K_1}^{K_2} a_{lkd} = 1$ for $d = 1, \cdots, D$. Under the model we may be able to reduce the number of signature patterns and get more efficient estimates and their standard deviations.

$$Y_{id} = \max_{1\le l\le L} \max_{-K_1\le k\le K_2} a_{lkd} Z_{l,i-k} * N(0,\sigma^2), \quad d = 1, \cdots, D, \tag{6.2}$$

  where$\sum_{l=1}^{L}\sum_{k=-K_1}^{K_2} a_{lkd} = 1$ for $d = 1, \cdots, D$. Or other variants under which we can study both positive and negative returns and allow model to include short selling. It increases the flexibility and may be more practically useful.

- Estimation based on regression over threshold is worth to look into.
- It may be worth to look into multivariate max-stable $ARMA(p, q)$ processes for its natural link to $ARMA(p, q)$ process.
- It may be worth to study general volatility modeling in an extreme value settings.

Bibliography

[1] Arnold, S. F. (1990)*Mathematical Statistics.* Prentice Hall, Englewood Cliffs, New Jersey.

[2] Barnard, George A. (1965) The use of the likelihood function in statistical practice. (1967) *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (Berkeley, Calif., 1965/66) Vol. I: Statistics, pp. 27–40 Univ. California Press, Berkeley, Calif.

[3] Best, P. (1998) *Implementing value at risk.* John Wiley & Sons, England.

[4] Billingsley, P. (1995) *Probability and Measure*, 3rd edition, John & Wiley, New York.

[5] Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *J. Econometrics* **31**, 307-327

[6] Cambanis, S. and Leadbetter, M.R. (1994) *Measure and Probability*, Lecture Notes, Department of Statistics at the University Of North Carolina.

[7] Cartwright, D. E. (1958) On estimating the mean energy of sea waves from the highest waves in a record. *Proc. Roy. Soc. London. Ser. A* **247**, 22–28.

[8] Chen, X. R. (1981) *Introduction to Mathematical Statistics.* (in Chinese), Science Press, Beijing.

[9] Coles, S. G. and Tawn, J. A. (1991) Modelling Extreme Multivariate Events. *J. R. Statist. Soc. B,* **53**,pp. 377-392

[10] Coles, S. G. and Tawn, J. A. (1994) Statistical Methods for Multivariate Extrems: an Application to Structural Design(with discussion). *Appl. Statist.* **43**, No. 1, pp. 1-48

[11] Dave, R. D. and Stahl, G. (1999) On the Accuracy of VaR Estimates Based on the Variance-Covariance Approach.

[12] Davis, R. A. and Resnick, S. I. (1989) Basic Properties and Prediction of Max-ARMA Processes. *Adv. Appl. Prob.* , **21**, 781-803

[13] Davis, R. A. and Resnick, S. I. (1993) Prediction of Stationary Max-Stable Processes. *Ann. Appl. Prob.* , **3**, no 2, 497-525

[14] Davison, A. C. and Smith, R. L. (1990) Models for Exceedances over High Thresholds. *J. R. Statist. Soc. B,* **52**, No. 3, pp. 377-392

[15] Deheuvels, P. (1983) Point Processes and Multivariate Extreme Values. *Jour. of Multi. Anal.* **13**, pp. 257-272

[16] Dowd, K. (1998) *Beyond value at risk.* John Wiley & Sons, England.

[17] Embrechts, P. (1999) Extreme Value Theory in Finance and Insurance. http://math.ethz.ch

[18] Embrechts, P. (1999) Extreme Value Theory: Potential and Limitations as an Integrated Risk Management Tool. http://math.ethz.ch

[19] Embrechts, P., de Haan, L. and Huang, X. (2000) Modelling Multivariate Extrems. *Extremes and Integrated Risk Management*, pp. 59-70. (Edited by Paul Embrechts), UBS Warburg.

[20] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance,* Springer

[21] Fisher, R. A. and Tippett, L.H.C. (1928) Limiting forms of frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.*, **24**, 180-190

[22] Galambos, J. (1987) *Asymptotic theory of extreme order statistics*, 2nd edition. Krieger, Malabar, Florida.

[23] Giesbrecht, F. and Kempthorne, O. (1976) Maximum Likelihood Estimation in the Three-parameter Lognormal Distribution. *J. R. Statist. Soc.* **B 38**, 257-264

[24] Gnedenko, B. (1943) Sur la distribution limite du terme maximum d'une série aléatoire. (French) *Ann. of Math.* (2) **44**, 423–453

[25] de Haan, L. (1984) A Spectral Representation for Max-Stable Processes. *Ann. of Prob.* Vol 12, No. 4, 1194-1204

[26] de Haan, L. (1985) Extremes in higher dimensions: the model and some statistics. In *Proc. 45th Sess. Int. Statist. Inst.,* paper 26.3. The Hague: International Statistical Institute.

[27] de Haan, L. and Resnick, S. I. (1977) Limit Theory for Multivariate Sample Extremes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete,* **40**, pp.317-337

[28] Hall, P., Peng, L. and Yao. Q. (2001) Moving-maximum models for extrema of time series. Manuscript.

[29] Hsing, T. (1993) Extremal index estimation for a weakly dependent stationary sequence. *Ann. Statist.* **21**, no. 4, 2043–2071.

[30] Joe, H. (1989) Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statist. Probab. Lett.* **9**, 75-82.

[31] Jorion, P. (1996) *Value at Risk.*, Irwin

[32] Kempthorne, O. (1966) Some Aspects of Experimental Inferences. *J. Am. Statist. Ass.* **61**, 11-34

[33] Korn, R. (1997) Optimal Portfolio, World Scientific, Singapore.

[34] Leadbetter, M. R. (1983) Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete* **65**, no. 2, 291–306

[35] Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983) *Extremes and related properties of random sequences and processes.* Springer, Berlin.

[36] Leadbetter, M.R., Weissman, I., de Haan, L. and Rootzén, H. (1989) On clustering of high values in statistically stationary series. In: *Proc. 4th Int. Meet. Statistical Climatology* (ed: J. Sanson). Wellington: New Zealand Meteorological Service.

[37] Loynes, R. M. (1965) Extreme values in uniformly mixing staionary stochastic processes. *Ann. Math. Statist.* **36**, 993-999

[38] Mardia, K. V. (1970) *Families of bivariate distributions.* Griffin's Statistical Monographs and Courses, No. 27. Hafner Publishing Co., Darien, Conn.

[39] Markowitz, H. (1952) Portfolio selection. *J. of Finance*, **7**, 77-91.

[40] Markowitz, H. (1987) *Mean-varaince analysis in portfolio choice and capital markets.* Cambridge, MA: Balckwell.

[41] Marshall, A.W., Olkin, I. (1983) Domains of attraction of multivariate extreme value distributions. *Ann. Probab.* **11**, no. 1, 168–177.

[42] McFadden, D. (1978) Modelling the choice of residential location. In *Spatial Interaction Theory and Planning Models*(eds A. Karlqvist, L. Lundquist, F. Snickers and J. Weibull), pp. 75-96. Amsterdam: North-Holland.

[43] Michaud, R. O., (1998) *Efficient asset management.* Harvard Businees School Press, Boston.

[44] Nandagopalan, S. (1990) *Multivariate extremes and the estimation of the extremal index.* Ph.D. dissertation, Dept. Of Statistics, University of North Carolina, Chapel Hill.

[45] Nandagopalan, S. (1994) On the multivariate extremal index. *J. of Research, National Inst. of Standards and Technology.* **99**, 543-550.

[46] Newell, G. F. (1964) Asymptotic extremes for m-dependent random variables. *Ann. Math. Statist.* **35**, 1322-1325

[47] O'Brien, G. L. (1974) Limit theorems for the maximum term of a stationary process. *Ann. Probab.* **2**. 540-545

[48] O'Brien, G. L. (1987) Extreme Values for Stationary and Markov Sequences. *Ann. of Prob.* Vol 15, No. 1, 281-291

[49] Pickands, J. III (1975) Statistical Inference Using Extreme Order Statistics. *The Ann. of Statis.* **3**, No 1, 119-131

[50] Pickands, J. III (1981) Multivariate Extreme Value distributions. *Proc. 43rd Session I.S.I.* (Buenos Aires), 859-878.

[51] Resnick, S. I. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

[52] RiskMetrics (1996) Technical Document, the fourth edition.

[53] Robinson, M. E. and Tawn, J. A. (1995) Statistics for Exceptional Athletics Records. *Appl. Statist.* **44**, No. 4, pp. 499-511

[54] Sigma (1996) Natural catastophes and major losses in 1995: decrease compound to previous year but continually high level of losses since 1989. *Sigma publication No 2*, Swiss Re. Zürich.

[55] Smith, R. L. (1986) Extreme value theory based on the r largest annual events. *J. Hydrology*, **86**, 27-43.

[56] Smith, R. L. (1990) Extreme Value Theory. *Handbook of Applicable Mathematics (Supplement)*, W. Ledermann (ed.) John Wiley, Chichester.

[57] Smith, R. L. (1997) Statistics for Exceptional Athletics Records. *Appl. Statist.* **46**, No. 1, pp. 000-000

[58] Smith, R. L. (2000) Measuring Risk with Extreme Value Theory. *Extremes and Integrated Risk Management*, pp. 19-36. (Edited by Paul Embrechts), UBS Warburg.

[59] Smith, R. L. and Goodman, D. J. (2000) Bayesian Risk Ananlysis. *Extremes and Integrated Risk Management*, pp. 235-251. (Edited by Paul Embrechts), UBS Warburg.

[60] Smith, R. L., and Shively, T. S. (1995) Point Process Approach to Modeling Trends in Troposphereic Ozone Based on Exceedances of A High Threshold.*Atmospheric Environment,* **29**, No. 23, pp.3489-3499

[61] Smith, R. L., and Weissman, I. (1994) Estimating the extremal index. *J. Royal Statist. Soc. B.* **56**, 515-528

[62] Smith, R. L., and Weissman, I. (1996) Characterization and Estimation of the Multivariate Extremal Index.

[63] Smith, R. L., and Weissman, I. (1997) On the Estimation of Max-Stable Processes.

[64] Tawn, J. A. (1988) An extreme-value theory model for dependent observations. *J. Hydrology*, **101**, 227-250.

[65] Tawn, J. A. (1990) Modelling multivariate extreme value distribution. *Biometrika,* **77**, 245-253.

[66] Tsay, R. S. (1999) Extreme Value Analysis of Financial Data

[67] Weissman, I. (1978) Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* **73**, no. 364, 812–815

[68] Weissman, I. (1994) On the extremal index of stationary sequences. Technical report, presented at the Conference on Multivariate Extreme Value Estimation with Applications to Economics and Finance, Erasmus University, Rotterdam.