Time series and state space model with generalized extreme value distributed marginals and α-stable distributed errors

by

© **Jinjun Tong**

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Department of Mathematics and Statistics

Memorial University

May 2018

St. John's, Newfoundland and Labrador, Canada

Abstract

This thesis is mainly focused on the estimation and filtering of extreme events time series and models with generalized extreme value distributed marginals and the multiplicative errors from α-stable distribution.

First a non-linear time series with Fréchet distributed marginals and α-stable distributed errors is considered. To estimate the stability parameter, three recursive procedures are proposed. The first is based on the Hill estimation, the second is a modified Fan's estimation that uses the property of the α-stable distribution, and the last is an application of Kantorovich-Wasserstain metric.

For the state space model with generalized extreme value distributed marginals and α-stable distributed errors, the estimation is more complex, especially when the stability parameters are small. In the model with Gumbel distributed marginals, if one of the stability parameter is known, a procedure that generates an ensemble from the known error distribution by Monte Carlo followed by estimation is proposed. For a model with generalized extreme value distributed marginals and unknown stability parameters, first a recursive regression estimation is applied to obtain the generalized extreme valued parameters, then the Yule-Walker estimation or generalized least square regression model is used to estimate the stability parameters.

Regarding filtering, the estimation of unobserved states and their empirical conditional densities are our interests. The estimation of states is obtained numerically via Monte Carlo, based on the model structure. This procedure outperforms Kalman filter. As to the empirical conditional density, sequential importance sampling with different importance functions, particle filter with discrete sample space, auxiliary particle filter and plain linearization are used and compared.

The asymptotic properties and rates of convergence of the proposed estimations are studied analytically and through simulation. The methods and procedures developed in this thesis have been applied to analyze the air pollution data in New York city.

Lay summary

There are many practical situations that we deal with extreme events. It is also quite common to make the assumption of linearity and Gaussian errors. However, there is conspicuous evidence that these models may not be good enough since the errors may come from heavy-tailed distributions. We are mainly interested in the time series of extremes associated with errors from a special type of stable distribution.

A random variable is said to be stable if a linear combination of the samples from this distribution still belongs to the same family. Gaussian distribution is the most famous member of stable family. We are interested in a positive stable distribution, the α-stable distribution, which has no finite expectation and whose density function does not have a simple expression.

First, we studied a non-linear time series with the extreme events, which is Fréchet distributed, and an α-stable distributed error. The difficulty here is that some commonly used methods do not produce good estimates. We used the properties of the α-stable distributed errors in estimation. Some recursive procedures, based on Hill estimation, Fan's estimation and Kantorovich-Wasserstain metric, are proposed. These methods need to be applied recursive since the errors are unobserved.

Our next goal is the estimation of a state space model with an observed sequence (generalized extreme value distributed) and an unobserved sequence (Gumbel distributed, called the states) and two α-stable distributed errors. We would like to estimate the parameters as well as the unobserved states. The parameter estimation is complex because of the mixed effect of errors and the limited information driven from the observation sequence. We proposed a numerical method when one stability parameter is known, guessing the errors using the model structure. When both stability parameters are unknown, Yule-Walker estimation and the regression estimation are applied.

After parameter estimation, I proposed a Monte Carlo procedure to estimate the states. Such a procedure outperforms Kalman filter. Some filtering methods are applied to obtain the conditional density of the states, by generating guesses of the states with reasonable weights.

Finally, the time series model and the state space model were applied to the weekly maxima pollution data (CO, SO_2) in New York city.

Acknowledgements

I would like to express my sincere appreciation to those who have contributed to this thesis and helped me in those years.

First and foremost, I have to thank my supervisor Dr. JC Loredo-Osti, who offered me the chance to live a different life. I have learned a lot from you and there are still more in you for me to learn. You have set me a good example of excellent researcher. Thank you for your guidance and supervise through these years. Your discussion, ideas and feedback are invaluable.

I am grateful to my supervisor committee members, Dr. Awell Oyet, Dr. Asokan Variyath, for your insightful comments and encouragement. I would also like to express my appreciation to Dr. Hong Wang, who not only taught me in class but also gives me wonderful tips of life after class; Dr. Zhaozhi Fan, who I can always receive encouragement from; Dr. Taraneh Abarin, who is my teaching mentor; Dr. Yanqing Yi, Dr. Yuri Bahturin for your wonderful teaching. Besides this, I also need to thank Dr. Xiaoqiang Zhao and Dr. Jie Xiao, thank Ros, Tara, Leonce, Jennifer, Kim and Lawrence, for your wonderful work in the department.

I acknowledge and thank the Canadian Institutes of Health Research for the economical support through a grant to my supervisor.

Special thanks to my parents, for supporting their only child to leave home to see a bigger world and live the life she wants. In my heart, you are not only my parents, but also my earliest teachers and my best friends. I know that no matter what, you always mean well to me and be there for me. Thank you for everything.

I would also like to thank all of my friends: Haifeng Song and Yuxiang Zhang, Rui Peng and Min Yuan, who helped me a lot when I began my life in St John's; Tao Yi, Yin Yin, Fuxing, Yi Zhang, Zhaoyi, Jia Li, Wei Song, Hua Que, Zhixian Yu, Oleg and Solomon. I will always remember the joyful time we spent together. Thanks to MUN badminton club, Chienmin Yeh and Huifen, Jingjing and Jinliang, Pheobe and Zhiyu, Wei Chen and Sujin, Jiabin. Thank you all for been so nice to me, with or without your name being listed here. I sincerely wish you all the best.

Statement of contribution

The main original contributions of this manuscript are:

1. General results about the α-stable distribution are presented, and in some cases, alternative proofs are provided in the first chapter.

2. For the non-linear time series with Fréchet distributed marginals and the α-stable distributed errors, I developed three recursive procedures to estimate the unknown parameters, providing a way of using the properties of the error in estimation.

3. Since the observation sequence has generalized extreme value distributed marginal and satisfies the strong mixing condition, I proposed a recursive regression model to estimate the generalized extreme value parameters which works well with small chain size for heavy tailed marginal distributions.

4. In the state space model with Gumbel distributed marginals and α-stable distributed errors, I developed a Monte Carlo procedure based on the generated error samples and model structure to estimate the unobserved states when only one stability parameter is known. This procedure outperforms Kalman filter.

5. To obtain better estimates for the empirical filtering density of the state space model with generalized extreme value distributed marginals and α-stable distributed errors, I applied and compared three numerical filtering methods through simulations.

   This thesis is the product of the collaboration with my supervisor, Dr. JC Loredo-Osti. Dr. Loredo-Osti suggested the topic of this research and I implemented it, carried out the simulations and wrote the first draft of the manuscript. This manuscript was discussed with my supervisor and jointly edited.

Table of contents

List of tables

List of figures

Chapter 1

Introduction

1.1 Generalized extreme value distribution

1.1.1 Definition

When the maximum or minimum of a sequence is under consideration, the Generalized Extreme Value (GEV) distribution would be of relevance, since the GEV distribution is the limiting distribution for extreme values.

The research on extreme events has a long history. As early as year 1927, Fréchet [32] studied the properties of the distribution of the maximum. Later, researchers such as Fisher and Tippett [30], Gumbel [39], Leadbetter [62], Coles [15], Resnick [83] and many others, made substantial contributions to the topic.

Suppose that the sample consists of $\{X_k, 1 \leq k \leq n\}$. Since

$$\max\{X_k; 1 \leq k \leq n\} = -\min\{-X_k; 1 \leq k \leq n\},$$

there is not loss of generality if one considers only the maximum.

Similar to the Central Limit Theorem, which refers to the limiting distribution of the sample average, there is a theoretical framework that studies the limiting distribution of the sample maximum.

Let $\{X_n, n \geq 1\}$ be independent identically distributed random variables with common function F and define M_n as

$$M_n = \max\{X_k, 1 \leq k \leq n\}.$$

If there exist sequences $a_n > 0,\ b_n \in \mathbb{R}$ and a non-degenerate distribution G such that

$$\lim_{n\to\infty} P(M_n \le a_n x + b_n) = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x)$$

for every continuity point of G (Resnick denoted it as $F \in D(G)$ in [83]), thus G belongs to one of the three classes below:

(i) Gumbel ($\mu, \sigma > 0$)

$$F_{\mu,\sigma}(x) = \exp\left(-e^{-\frac{x-\mu}{\sigma}}\right),$$
$$f_{\mu,\sigma}(x) = \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma} - e^{-\frac{x-\mu}{\sigma}}\right), x \in \mathbb{R};$$

(ii) Fréchet ($\mu, \sigma > 0, \gamma > 0$)

$$F_{\mu,\sigma,\gamma}(x) = \begin{cases} 0, & x < \mu, \\ \exp\left(-(\frac{x-\mu}{\sigma})^{-\gamma}\right), & x \ge \mu; \end{cases}$$

$$f_{\mu,\sigma,\gamma}(x) = \begin{cases} 0, & x < \mu, \\ \frac{\gamma}{\sigma}(\frac{x-\mu}{\sigma})^{-\gamma-1}\exp\left(-(\frac{x-\mu}{\sigma})^{-\gamma}\right), & x \ge \mu; \end{cases}$$

(iii) Weibull ($\mu, \sigma > 0, \gamma > 0,$)

$$F_{\mu,\sigma,\gamma}(x) = \begin{cases} \exp\left(-(-\frac{x-\mu}{\sigma})^{\gamma}\right), & x < \mu, \\ 1, & x \ge \mu; \end{cases}$$

$$f_{\mu,\sigma,\gamma}(x) = \begin{cases} \frac{\gamma}{\sigma}(-\frac{x-\mu}{\sigma})^{\gamma-1}\exp\left(-(-\frac{x-\mu}{\sigma})^{\gamma}\right), & x < \mu, \\ 0, & x \ge \mu; \end{cases}$$

Under the random sampling assumption, i.e., $X_1, X_2, \ldots, X_n$ being identically and independently distributed, the GEV distribution with parameter μ, σ, γ, denoted as GEV(μ, σ, γ), is defined as:

$$\lim_{n\to\infty} P(M_n \le x) = \exp\left(-\left(1 + \gamma\frac{x-\mu}{\sigma}\right)^{-1/\gamma}\right) = F_{\mu,\delta,\gamma}(x) \tag{1.1}$$

for $1+\frac{\gamma}{\sigma}(x-\mu)>0, \sigma>0$, and its density function is

$$f(x;\mu,\sigma,\gamma)=\begin{cases}\frac{1}{\sigma}\left(1+\gamma(\frac{x-\mu}{\sigma})\right)^{-\frac{1}{\gamma}-1}\exp\left(-\left(1+\gamma(\frac{x-\mu}{\sigma})\right)^{-\frac{1}{\gamma}}\right), & \gamma\neq 0,\\ \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}-\exp(-\frac{x-\mu}{\sigma})\right), & \gamma=0.\end{cases}$$

1.1.2 Parameters and basic properties

The parameter $\mu\in\mathbb{R}$ is known as the location parameter, $\sigma>0$ is called the scale parameter, while $\gamma\in\mathbb{R}$ is the shape or tail parameter.

The behaviour of these three types of the distributions differs according to their tail, which is characterized by the tail index. Because of this feature, the tail parameter would be considered the most important parameter.

The support set of a GEV random variable, $1+\gamma\frac{x-\mu}{\sigma}>0$, depends on those parameters.

Using the density function of the GEV(μ,σ,γ),

- the mean of a random variable from Gumbel distribution ($\gamma=0$) is $\mu+\sigma\gamma_e$, where γ_e represents the Euler's constant ($\gamma_e\approx 0.57721$).
- When $\gamma<1$ and $\gamma\neq 0$, the mean of a GEV(μ,σ,γ) variable is $\mu+\sigma\frac{\Gamma(1-\gamma)-1}{\gamma}$, where $\Gamma(\cdot)$ is the Gamma function.
- When $\gamma>1$ the mean of a GEV(μ,σ,γ) variable does not exist.

1.1.3 Relationship between the members of GEV family

Taking the limit $\gamma\to 0$ to the Equation (1.1), we have

$$\lim_{\gamma\to 0,n\to\infty}P(M_n\leq x)=\exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right),\quad x\in\mathbb{R}$$

which is in Gumbel family, denoted as Gumbel(μ,σ).

When $\gamma<0$, the distribution corresponds to the Weibull family, denoted as Weibull(μ,σ,γ).

When $\gamma>0$, the distribution corresponds to the Fréchet family, denoted as Fréchet (μ,σ,γ).

The different types of the extreme value distribution could be transformed into each other, like if $X \sim \text{GEV}(\mu, \sigma, \gamma)$, $\gamma \neq 0$, then

$$\frac{1}{\gamma} \log \left(1 + \frac{\gamma}{\sigma}(X - \mu)\right) \sim \text{Gumbel}(0, 1)$$

and if $X \sim \text{Gumbel}(\mu, \sigma)$, then $\frac{X-\mu}{\sigma} \sim \text{Gumbel}(0, 1)$, and for $\gamma \neq 0$,

$$\frac{1}{\gamma} \left(e^{\gamma \frac{X-\mu}{\sigma}} - 1\right) \sim \text{GEV}(0, 1, \gamma).$$

1.1.4 Applications

extreme events happen in the nature with some regularity. Examples of these are floods, tornadoes, earthquakes, stock market crashes and soaring and so on. extreme events have the potential of a great impact to human society. Because of that, the study of GEV distributions and their prediction is important.

There are models to deal with events of small probability related to the GEV distribution. These models are widely used in risk management, finance, insurance, economics, hydrology, material sciences, telecommunications, and many other areas of application.

1.1.5 Stability postulate

If a random variable has distribution $G(x)$,

$$G^n(x) = G(a_n x + b_n),$$

and the limiting distribution exists and non-degenerate for some constant sequences $\{a_n\}$,$\{b_n\}$ depend on n, not on x, we say that the distribution G satisfies the maximum stability postulate.

This result was obtained by R. Fréchet and also by R. Fisher. B. Gnedenko showed that there are no distribution satisfying this postulate outside of the extreme value distribution family.

For an i.i.d. sequence $\{X_i, 1 \leq i \leq n\}$ where $X_i \sim \text{GEV}(\mu, \sigma, \gamma)$, $\gamma \neq 0$,

$$\begin{aligned} P(\max_{1\leq i\leq n}\{X_i\} \leq x) &= \exp\left(-n\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right) \\ &= \exp\left(-\left(1+\gamma\frac{x-\mu+\frac{\sigma}{\gamma}(1-n^{\gamma})}{\sigma n^{\gamma}}\right)^{-\frac{1}{\gamma}}\right), \end{aligned}$$

so $\max\{X_i\} \sim \text{GEV}(\mu - \frac{\sigma}{\gamma}(1-n^{\gamma}), \sigma n^{\gamma}, \gamma)$.

When $\gamma = 0$, i.e., Gumbel case,

$$P(\max_{1\leq i\leq n}\{X_i\} \leq x) = \exp(-ne^{-\frac{x-\mu}{\sigma}}) = \exp(-e^{-\frac{x-\mu-\sigma\log n}{\sigma}})$$

so $\max\{X_i\} \sim \text{Gumbel}(\mu + \sigma\log n, \sigma)$.

1.2 Stable distribution

Stability postulate describes the limiting distribution of extremes of i.i.d. random variables following the same distribution as the random variables come from. What about the linear combination of i.i.d. random variables follow the same distribution as the random variables come from? The distributions having this property are said to be stable.

1.2.1 Stable random variables

Normal distribution has the property that the linear combination of normal random variables is still normally distributed. Whether there are other distributions having this same property?

In 1920's, Paul Lévy studied the sums of independent identically distributed terms. The normal, Cauchy and Lévy distributions are the special cases of distributions having such a property.

Definition. We call a random variable stable if it has the property that a linear combination of two independent identically distributed random variables has the same distribution, up to location and scale parameters.

When talking about stable distribution, we can distinguish between stable in a broad sense and strictly stable random variable.

A random variable X is stable or stable in the *broad sense* if for X_1 and X_2, which are two independent copies of X, and any positive constants a and b,

$$aX_1 + bX_2 \stackrel{d}{=} cX + d$$

holds for some positive c and some $d \in \mathbb{R}$.

The random variable is called *strictly* stable if the equation holds with $d = 0$ for all admissible choices of a and b. There are equivalent definitions of stable random variables, such as the following

Definition. A non-degenerate X is stable if and only if for all $n > 1$, there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that

$$X_1 + \cdots + X_n \stackrel{d}{=} c_n X + d_n,$$

where $X_1, \ldots, X_n$ are independent copies of X. The random variable X is strictly stable if and only if $d_n = 0$ for all n.

A random variable is symmetrically stable if it is stable and symmetrically distributed around 0, i.e. $X \stackrel{d}{=} -X$. For the present work, we are mostly interested in the positive stable random variables.

1.2.2 The α-stable random variable

For reasons that will be clear later, we are particularly interested in a positive stable random variable S, whose Laplace transform can be written as

$$\mathrm{E}(e^{-tS}) = e^{-t^\alpha} \quad \text{for } \mathrm{Re}(t) > 0, \quad \alpha \in (0, 1). \tag{1.2}$$

The distribution of this random variable is known as the α-stable distribution (also called the Lévy stable distribution). We use $S \sim \mathcal{S}(\psi)$ to denote that S follows the α-stable distribution with parameter ψ.

If $S \sim \mathrm{S}(\alpha)$, from its Laplace transform, the corresponding characteristic function can be found,

$$\begin{aligned}\phi(t) &= \mathrm{E}(e^{itS}) = e^{-(-it)^{\alpha}} \\ &= \exp\left(-|t|^{\alpha}\left(\cos\left(\frac{\alpha\pi}{2}\right) - i\,\mathrm{sign}(t)\sin\left(\frac{\alpha\pi}{2}\right)\right)\right),\end{aligned}$$

where $i = \sqrt{-1}$.

Once that either the Laplace transform or the characteristic function of S is given, the density can be obtained using the approximate inversion formula, e.g.,

$$\begin{aligned}f_{\alpha}(x) &= \frac{1}{2\pi}\int_R \mathrm{Re}\{e^{-itx}\phi(t)\}dt \\ &= \frac{1}{2\pi}\int_R \mathrm{Re}\{e^{-|t|^{\alpha}\cos\left(\frac{\alpha\pi}{2}\right) - i[tx - \mathrm{sign}(t)|t|^{\alpha}\sin\left(\frac{\alpha\pi}{2}\right)]}\}dt \\ &= \frac{1}{2\pi}\int_R \cos\left(tx - \mathrm{sign}(t)|t|^{\alpha}\sin\left(\frac{\alpha\pi}{2}\right)\right)e^{-|t|^{\alpha}\cos\left(\frac{\alpha\pi}{2}\right)}dt \\ &= \frac{1}{\pi}\int_0^{\infty} \cos\left(tx - t^{\alpha}\sin\left(\frac{\alpha\pi}{2}\right)\right)e^{-t^{\alpha}\cos\left(\frac{\alpha\pi}{2}\right)}dt \qquad (1.3)\end{aligned}$$

for positive x, and the Gil-Peláez formula (see [35]) can be used to find the distribution of S, i.e.,

$$\begin{aligned}F_{\alpha}(x) = \quad & P(S \le x) = \frac{1}{2} + \frac{1}{2\pi}\int_0^{\infty}\mathrm{Re}\{\frac{e^{itx}\phi(-t) - e^{itx}\phi(t)}{it}\}dt \\ = \quad & \frac{1}{2} + \frac{1}{2\pi}\int_0^{\infty}\mathrm{Re}\{\frac{e^{itx - t^{\alpha}\cos\left(\frac{\alpha\pi}{2}\right) - it^{\alpha}\sin\left(\frac{\alpha\pi}{2}\right)} - e^{-itx - t^{\alpha}\cos\left(\frac{\alpha\pi}{2}\right) + it^{\alpha}\sin\left(\frac{\alpha\pi}{2}\right)}}{it}\}dt \\ = \quad & \frac{1}{2} + \frac{1}{2\pi}\int_0^{\infty}\frac{2\sin(tx - t^{\alpha}\sin\left(\frac{\alpha\pi}{2}\right))}{t}e^{-t^{\alpha}\cos\left(\frac{\alpha\pi}{2}\right)}dt \\ = \quad & \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty}\frac{\sin(tx - t^{\alpha}\sin\left(\frac{\alpha\pi}{2}\right))}{t}e^{-t^{\alpha}\cos\left(\frac{\alpha\pi}{2}\right)}dt\end{aligned}$$

when $x > 0$.

In 1959, Mikusinski [69] gave an alternative formula for the density of the α-stable distribution. For $x > 0$, if $S \sim \mathcal{S}(\alpha)$,

$$f_{\alpha}(x) = \frac{\alpha}{1-\alpha}\frac{1}{\pi x}\int_0^{\pi} b_{\alpha,x}(\theta)\exp\left(-b_{\alpha,x}(\theta)\right)d\theta,$$

where

$$b_{\alpha,x}(\theta) = \frac{\sin((1-\alpha)\theta)}{\sin\theta}\left(\frac{\sin(\alpha\theta)}{x\sin\theta}\right)^{\alpha/(1-\alpha)}, \quad 0 < \theta < \pi$$

and the distribution formula can be derived from the density,

$$F_\alpha(x) = \frac{1}{\pi}\int_0^\pi \exp\left(-b_{\alpha,x}(\theta)\right) d\theta, \quad 0 < \alpha < 1. \tag{1.4}$$

This representation of the density of a $\mathcal{S}(\alpha)$ random variable unveils a link between the exponential, uniform and the α-stable distributions.

Theorem 1. *Let* $U \sim$ Exponential(1) *and* $\Theta \sim$ Uniform$(0,\pi)$ *be two independent random variables, define*

$$X = \left(\frac{\sin((1-\alpha)\Theta)}{U\sin\Theta}\right)^{(1-\alpha)/\alpha}\frac{\sin(\alpha\Theta)}{\sin(\Theta)},$$

where $\alpha \in (0,1)$, *then* $X \sim \mathcal{S}(\alpha)$. *(*Chambers, Mallows and Stuck [11]*)*

This result provides an efficient way to simulate the α-stable random variables.

We can think of $\mathcal{S}(\alpha)$ as the kernel distribution in a scale family. An α-stable random variable with scale parameter $\sigma \neq 1$, and $\sigma > 0$ will be denoted as $S \sim \mathcal{S}(\alpha,\sigma)$, i.e. $\frac{S}{\sigma} \sim \mathcal{S}(\alpha)$. For this scale family, the Laplace transform and characteristic function can be written as

$$\begin{aligned}
&\mathrm{E}(e^{-tS}) = e^{-\sigma^\alpha t^\alpha}, \ \sigma > 0, \\
&\mathrm{E}(e^{itS}) = \exp\left(-\sigma^\alpha|t|^\alpha\left(\cos\left(\frac{\alpha\pi}{2}\right) - i\,\mathrm{sign}(t)\sin(\frac{\alpha\pi}{2})\right)\right).
\end{aligned}$$

If $S_1 \sim \mathcal{S}(\alpha,\sigma_1)$ and $S_2 \sim \mathcal{S}(\alpha,\sigma_2)$ are independent, for constants $c > 0, d > 0$ we have

$$cS_1 + dS_2 \sim \mathcal{S}(\alpha, (c^\alpha\sigma_1^\alpha + d^\alpha\sigma_2^\alpha)^{1/\alpha}). \tag{1.5}$$

To see this, notice that the Laplace transform of $cS_1 + dS_2$ can be written as

$$\mathrm{E}\left(e^{-t(cS_1+dS_2)}\right) = e^{-\sigma_1^\alpha c^\alpha t^\alpha - \sigma_2^\alpha d^\alpha t^\alpha} = e^{-\sigma^\alpha t^\alpha},$$

where $\sigma = (c^\alpha\sigma_1^\alpha + d^\alpha\sigma_2^\alpha)^{1/\alpha}$. Consequently, $S \sim \mathcal{S}(\alpha,\sigma)$ is a strictly stable distribution.

Denote the density function and distribution of $\mathcal{S}(\alpha,\sigma)$ as $f_{\alpha,\sigma}(x)$, $F_{\alpha,\sigma}(x)$ respectively. They can be obtained by using the linear transformation of $\mathcal{S}(\alpha)$, i.e.

$$f_{\alpha,\sigma}(x)=\frac{1}{\sigma}f_\alpha\left(\frac{x}{\sigma}\right),\qquad F_{\alpha,\sigma}(x)=F_\alpha\left(\frac{x}{\sigma}\right).$$

Although, $S\sim\mathcal{S}(\alpha)$ is a positive random variable whose mean and variance do not exist, its logarithmic moments exist, in particular,

$$\begin{aligned}\mathrm{E}(\log(S)) &= \gamma_e\left(\frac{1}{\alpha}-1\right),\\ \mathrm{Var}(\log(S)) &= \frac{\pi^2}{6}\left(\frac{1}{\alpha^2}-1\right),\end{aligned}$$

where $\gamma_e\approx 0.57721566$ is the Euler's constant. This result can be deduced by the following theorem.

Theorem 2. *For the α-stable distributed random variable $S\sim\mathcal{S}(\alpha),\alpha\in(0,1)$, we have*

$$\mathrm{E}\left(e^{-t\log S}\right)=\frac{\Gamma(1+t/\alpha)}{\Gamma(1+t)}$$

when $\mathrm{Re}(t)>-\alpha$.

Proof of Theorem 2.

$$\begin{aligned}&\mathrm{E}\left(e^{-t\log S}\right)=\int_0^\infty s^{-t}f_\alpha(s)ds\\ =\;&\frac{1}{\pi}\int_0^\infty\int_0^\infty s^{-t}\cos\left(us-u^\alpha\sin\frac{\alpha\pi}{2}\right)ds\cdot e^{-u^\alpha\cos\frac{\alpha\pi}{2}}du\\ =\;&\frac{\Gamma(1-t)}{\pi}\int_0^\infty\frac{1}{u^{1-t}}\sin\left(\frac{\alpha\pi}{2}+u^\alpha\sin\frac{\alpha\pi}{2}\right)e^{-u^\alpha\cos\frac{\alpha\pi}{2}}du\\ =\;&\frac{\Gamma(1-t)}{\alpha\pi}\int_0^\infty u^{\frac{t}{\alpha}-1}\sin\left(\frac{\alpha\pi}{2}+u\sin\frac{\alpha\pi}{2}\right)e^{-u\cos\frac{\alpha\pi}{2}}du\\ =\;&\frac{\Gamma(1-t)}{\pi}\frac{\Gamma\left(\frac{t}{\alpha}\right)}{\alpha}\sin(\pi t)\\ =\;&\frac{\Gamma(1+t/\alpha)}{\Gamma(1+t)}.\end{aligned}$$

□

Using this theorem, we have the following corollary.

Corollary 1. *If $S \sim \mathcal{S}(\alpha), \alpha \in (0,1)$, we have*

$$\mathrm{E}(e^{it \log S}) = \frac{\Gamma(1 - it/\alpha)}{\Gamma(1 - it)},$$

*(*Zolotarev [98]*) and*

$$\mathrm{E}(S^{-t}) = \frac{\Gamma(1 + t/\alpha)}{\Gamma(1 + t)} \qquad \textit{for} \qquad t > -\alpha.$$

1.2.3 Tail behaviour

According to its tail behaviour, a distribution can be classified into different families, e.g., heavy tailed, long tailed, sub-exponential.

The tail behaviour of an α-stable distribution is described in the following theorem.

Theorem 3 (Feller [29], p. 448)**.** *Let $S \sim \mathcal{S}(\alpha)$. For a large enough value x, the asymptotic behaviour of the α-stable distribution is described by*

$$P(S > x) \approx \frac{1}{\Gamma(1 - \alpha)} x^{-\alpha}. \tag{1.6}$$

Based on this approximation, we have that

$$f_\alpha(x) \approx \frac{\alpha}{\Gamma(1 - \alpha)} x^{-\alpha - 1} \, \mathrm{I}_{(x_\alpha, \infty)}(x),$$

where $\mathrm{I}(\cdot)$ is the indicator function and $x_\alpha = \Gamma^{-1/\alpha}(1 - \alpha)$. This density was first presented in Mikusinski [69], Titchmarch [91]. Notice that x must be greater than x_α to use these approximations.

This means that the upper tail of an α-stable distribution behaves asymptotically as the Pareto law with parameters α and x_α.

Feller's proof of Theorem 3 is a generalization of the Hardy-Littlewood Tauberian theorem (Hardy [43]). The core of the argument is the fact that the Laplace transform

of $\mathrm{P}(S > x)$ can be developed as

$$\begin{aligned}\int_0^\infty e^{-tx}\,\mathrm{P}(S>x)dx &= \frac{1}{t}-\frac{1}{t}\int_0^\infty e^{-tx}f_\alpha(x)dx \\ &= \frac{1}{t}\left(t^\alpha-\frac{1}{2}t^{2\alpha}+O_p(t^{3\alpha})\right) \\ &= \frac{1}{t^{1-\alpha}}-\frac{1}{2}\frac{1}{t^{1-2\alpha}}+O_p\left(\frac{1}{t^{1-3\alpha}}\right)\end{aligned} \tag{1.7}$$

for $t > 0$. Thus when $0 < \alpha < 1$ and t is small enough,

$$\mathrm{P}(S>x)=\frac{x^{-\alpha}}{\Gamma(1-\alpha)}+O_p\left(x^{-2\alpha}\right).$$

In fact, from the Feller's Tauberian theorem we have that Equation (1.7) and $\mathrm{P}(S > x) = \frac{x^{-\alpha}}{\Gamma(1-\alpha)}L(x)$ imply each other, where

$$L(x)=1-\frac{1}{2}x^{-\alpha}+\frac{1}{6}x^{-2\alpha}+O_p(x^{-3\alpha}).$$

Furthermore, this Tauberian theorem also implies that $e^{x^{-\alpha}}\,\mathrm{P}(S \le x) \to 0$ when $x \to 0$ (Feller [29]). In general, the Tauberian theorems relate the asymptotic behaviour of $\mathrm{P}(S > x)$ when $x \to \infty$ with the asymptotics of its Laplace transform, $\mathrm{E}(e^{-tS})$, when $t \to 0$ and, when dealing with extreme values, this relationship can be used as the foundation for an estimation procedure.

Heavy tail

A random variable X is said to have a (right) heavy tail, if for any positive λ,

$$\lim_{x\to\infty} e^{\lambda x}\,\mathrm{P}(X>x)=\infty,$$

which means that heavy-tailed distributions are those whose tails decay to zero at a rate slower than the exponential.

Some commonly heavy-tailed distributions are Weibull distribution, t-distribution, Pareto distribution and Cauchy distribution. More information about heavy-tailed distributions would be found in Teugels [90], Crovella [18], Pickands [78]. There are many ways to test whether a distribution function is heavy tailed, such as the Kolmogorov test, Berk-Jones test, score test and their integrated version (see Koning

and Liang [59], Kolmogorov [56], Berk and Jones [5]). The most widely used way to explore the data is graphical methods, i.e. quantile-quantile plot (Q-Q plots, see Wilk and Gnanadesikan [97]), Hill plots and the distribution of mean excess.

From the Equation (1.6) we can see that the random variable $S \sim \mathcal{S}(\alpha)$ has a heavy tail.

Long tail

A random variable X is said to have long tail (Asmussen [3]) if for all $c > 0$,

$$\lim_{x\to\infty} \mathrm{P}(X > x + c|X > x) = 1.$$

The long-tailed distribution like Pareto distribution and Lévy distribution have been widely used in business and marketing area.

The linear transformation, product function, maximum and minimum of independent long-tailed random variables still have long tails.

The α-stable distribution $\mathcal{S}(\alpha)$ also has a long tail. To see this, consider

$$\frac{\mathrm{P}(S > x + c)}{\mathrm{P}(S > x)} \approx \left(\frac{x}{x + c}\right)^{\alpha} \to 1, \quad \text{as } x \to \infty.$$

Any distribution with long tail is in the heavy-tailed family, but a heavy-tailed distribution may not have a long tail.

Sub-exponential distribution

For a positive, independent, identically distributed random variable sequence $X_i, 1 \leq i \leq n$ with $n \geq 2$, if

$$\lim_{x\to\infty} \frac{\mathrm{P}(X_1 + \cdots + X_n > x)}{\mathrm{P}(X_1 > x)} = n, \quad \text{or} \quad \lim_{x\to\infty} \frac{\mathrm{P}(X_1 + \cdots + X_n > x)}{\mathrm{P}(\max(X_1, \ldots, X_n) > x)} = 1,$$

the sampling distribution is said to belong to sub-exponential family (Teugels [90]).

Using the Equations (1.5) and (1.6), the α-stable distribution $\mathcal{S}(\alpha)$ is also sub-exponential, since $\sum_{i=1}^{n} S_i \sim \mathcal{S}(\alpha, n^{1/\alpha})$ for an i.i.d. $\alpha-$stable sequence $\{S_i, 1 \leq i \leq n\}$,

$$\lim_{x\to\infty} \frac{\mathrm{P}(S_1 + \cdots + S_n > x)}{\mathrm{P}(S_1 > x)} = \lim_{x\to\infty} \frac{\mathrm{P}(S_1 > n^{-1/\alpha}x)}{\mathrm{P}(S_1 > x)} = n.$$

Sub-exponential distributions all have long tails, but long tailed distributions may not all be sub-exponential.

1.2.4 General stable distributions

The α-stable distribution also belongs to a more general family, called the stable distribution, defined by the characteristic function (Kanter [53], Zolotarev [98], Nolan [74])

$$\log \mathrm{E}(e^{itX}) = \begin{cases} i\mu t - \sigma^\alpha |t|^\alpha \left(1 - i\beta \tan\left(\frac{\pi\alpha}{2}\right) \mathrm{sign}(t)\right), & \alpha \neq 1, \\ i\mu t - \sigma|t| \left(1 + i\beta \frac{2}{\pi} \mathrm{sign}(t) \log|t|\right), & \alpha = 1. \end{cases} \tag{1.8}$$

The distribution of a stable random variable X is denoted as $X \sim \mathcal{S}(\alpha, \beta, \mu, \sigma)$, where $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\sigma > 0$, $\mu \in \mathbb{R}$.

The followings are special cases of stable random variables:

- when $\alpha = 2, \beta = 0$, the stable distribution is a normal distribution;
- when $\alpha < 2$, the variance of a stable random variable does not exist; when $2 > \alpha \geq 1$, its mean does not exist as well;
- when $\alpha = 1, \beta = 0$, it is the Cauchy distribution, a special case of both the stable distribution and the t-distribution;
- when $\alpha = 1, \beta = 1$, it is the Landau distribution;
- when $\alpha = \frac{3}{2}, \beta = 0$, it is the Holtsmark distribution;
- when $\alpha = \frac{1}{2}, \beta = 1$, it is the Lévy distribution;
- when $\alpha = 0$, it is the Dirac delta function.

Stable distributions have the following property.

Theorem 4 (Theorem in Feller [29])**.** *All (non-degenerate) stable distributions are continuous distributions with an infinitely differentiable density.*

The continuity of non-degenerate stable distributions can be deduced by its infinitely divisible property. For example, the density Equation (1.3) ensures its infinitely differentiable.

Although this theorem guarantees the existence of stable random variable's density, except in a few special cases, there is no closed expression form of the density for a stable distribution that can be used to evaluate the distribution function, even we know that the density function could be deduced from its characteristic function or equivalently, from its Laplace transform, for computational propose.

There are other parameterizations that differently expose the density and distribution of an α-stable random variable. For example Nolan [73] described the distribution of a random variable $\mathrm{S}(\alpha, \delta_\alpha)$ with $\delta_\alpha = \cos^{1/\alpha}\left(\frac{\alpha\pi}{2}\right)$ and called it the α-stable distribution, with density $f_\alpha^N(y)$ given by

$$f_\alpha^N(y) = \frac{\alpha}{\pi(1-\alpha)} y^{\frac{1}{\alpha-1}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} V(\eta) e^{-V(\eta) y^{\frac{\alpha}{\alpha-1}}} d\eta,$$

$$V(\eta) = \left(\cos\left(\frac{\pi\alpha}{2}\right)\right)^{\frac{1}{\alpha-1}} \left(\frac{\cos(\eta)}{\sin(\frac{\alpha\pi}{2} + \alpha\eta)}\right)^{\frac{\alpha}{\alpha-1}} \frac{\cos(\frac{\alpha\pi}{2} + (\alpha-1)\eta)}{\cos(\eta)}.$$

However letting $x = \delta_\alpha y$ and $\theta = \eta + \frac{\pi}{2}$ to the density of x is the same as the expression presented in Equation (1.4).

1.2.5 Computer programs concerned with stable distribution

Many R packages are built to deal with the stable family. John Nolan has a personal website about stable distribution, from where many packages and papers can be found.

In 1999, Nolan [75] developed a computer program to obtain the densities of stable distribution. This program splits the region of integration up into intervals where the cosine term changes sign and does the calculation. It works well when $|x|$ is not large and $\alpha > 0.9$. When $|x|$ is large, the number of oscillations is large and when α is small, the intervals where the cosine term changes sign grows, under which conditions the integral is difficult to evaluate precisely.

An improved program STABLE was given in Nolan [73] to deal with the condition when $\alpha > 0.1$. It changes the integration interval. This program is said to be improved to give more accurate density calculations on the tails, which is necessary for accurate likelihood calculations. The program gives the approximation of the densities when $\alpha > 0.4$ using splines, does maximum likelihood estimation and diagnostics for assessing the stability of a data set.

In 2001, Nolan [76] presents some details about the program, including the Fisher

information matrix of a sample. When the parameters are on the interior of the parameter space, the maximum likelihood estimator follows the standard theory, i.e., the estimators are consistent and asymptotically normal distributed. When the parameter β is on the boundary, like when $\beta = 0$, the stable densities are symmetric and all the correlation coefficients involving β are 0. If $\beta = 1$, the variance of estimator of β is 0 and all the correlation coefficients involving β are undefined.

R package "stabledist" can be used to generate random variables from the stable distribution and compute their densities, distribution functions.

1.3 Time series

Time series is a stochastic process where a data sequence is collected in time order. Time series are commonly used in many applications like econometric, financial field (stock market like Dow Jones index sequence), engineering area (signal processing, control engineering) and weather forecasting.

The main statistical problems regarding a time series are estimation, prediction and there is plenty of literature addressing this problems, e.g., Box et al. [7], Cowpertwait [16], Durbin [25], Gershenfeld [72], Hamilton [42]. Nevertheless, the body of work addressing time series of extreme events is limited. Here, I will address the modeling of discrete time series of extreme events with the generating mechanism for error terms coming from the α-stable distributions.

1.3.1 Stationary and non-stationary time series

A time series is said to be strictly stationary if the jointly density distribution of the random variables in the series is independent of time. Formally,

Definition. Let $\{X_t, t \geq 1\}$ be a time series, if for any $s, t_1, \ldots, t_k \in \mathbb{Z}$,

$$\mathrm{P}(X_{t_1} \leq x_1, \ldots, X_{t_k} \leq x_k) = \mathrm{P}(X_{t_1+s} \leq x_1, \ldots, X_{t_k+s} \leq x_k),$$

the series is said to be strictly stationary.

A related concept is weak stationary, or simply, stationary.

Definition. For any $s, t, k \in \mathbb{Z}$, a weak stationary time series $\{X_t\}$ is a sequence satisfies

$$\begin{aligned} &\mathrm{E}(X_s) = \mathrm{E}(X_t), \quad \mathrm{Var}(X_s) = \mathrm{Var}(X_t), \\ &\mathrm{Cov}(X_t, X_s) = \mathrm{Cov}(X_{t+k}, X_{s+k}). \end{aligned}$$

For a normal process, the strict stationary is equivalent to the weak stationary. The innovations in a Wiener process measured in regular intervals in time is a simple example of the stationary time series.

For the analysis and the prediction of the stationary time series, there are some commonly used models like regression model, auto-regressive (AR) model, moving average (MA) model, auto-regressive moving average (ARMA) model. More details would be found in Box et al. [7], Gouriéroux and Monfort[37], Brockwell [9], Benjamin [54]. All these models are linear. The non-linear and non-stationary processes are usually more challenging to analyze and predict.

1.3.2 Time series with Fréchet distributed marginals and α-stable distributed errors

We are interested in non-linear stationary time series of extreme events, particularly, values whose marginals have GEV distributions. For this purpose, we explore the relationship between the GEV and α-stable family of distributions.

Let $X \sim$ Fréchet$(0, 1, \gamma)$ and $S \sim \mathcal{S}(\alpha)$ be independent random variables and define

$$Y = XS^{\frac{1}{\gamma}},$$

then, $Y \sim$ Fréchet$(0, 1, \alpha\gamma)$, since

$$\begin{aligned} \mathrm{P}(Y \leq y) = &\quad \mathrm{E}\left(\mathrm{P}(X \leq yS^{-\frac{1}{\gamma}}|S)\right) \\ = &\quad \mathrm{E}\left(\exp\left(-\left(yS^{-\frac{1}{\gamma}}\right)^{-\gamma}\right)\right) \\ = &\quad \exp(-y^{-\alpha\gamma}). \end{aligned}$$

Use this property we have the following result,

Theorem 5. *Let $X_0 \sim$ Fréchet$(0, 1, \gamma)$ be independent with the sequence $\{S_t\}$, where $S_t \sim \mathcal{S}(\alpha), 1 \leq t \leq n$, is a set of independent random variables. Define the variable*

X_t *as*

$$X_t = X_{t-1}^{\alpha} S_t^{\frac{\alpha}{\gamma}}, \quad 1 \leq t \leq n, \tag{1.9}$$

then X_t *has a marginal* Fréchet$(0, 1, \gamma)$ *distribution for* $1 \leq t \leq n$.

We know that the Fréchet, Gumbel, Weibull distribution can transformed to each other, thus, based on Equation (1.9), we can generate time series with GEV distributed marginals using independent α-stable random variables. The alternatives are given in the following corollaries.

Corollary 2. *Let* $X_0 \sim$ Gumbel$(0, 1)$ *and* $\{S_t\}$, *where* $S_t \sim \mathcal{S}(\alpha), 1 \leq t \leq n$, *is a set of independent random variables. Then the series*

$$X_t = \alpha X_{t-1} + \alpha \log(S_t), 1 \leq t \leq n, \tag{1.10}$$

has marginal Gumbel$(0, 1)$ *distribution for* $1 \leq t \leq n$, *i.e.* $X_t \sim$ Gumbel$(0, 1)$.

Corollary 3. *Let* $X_0 \sim$ Gumbel$(0, 1)$, $\{X_t\}$ *and* $\{S_t\}$ *are as in Corollary 2.* $\{\xi_t\}$ *is an independent* $\mathcal{S}(\psi)$ *random variable sequence with* $\psi \in (0, 1)$, *then the series*

$$Y_t = \mu - \frac{\sigma}{\gamma} + \frac{\sigma}{\gamma} e^{\psi\gamma X_t} \xi_t^{\psi\gamma}, \quad 1 \leq t \leq n, \tag{1.11}$$

Y_t *has marginal* GEV(μ, σ, γ) *distribution for* $1 \leq t \leq n$, *i.e.* $Y_t \sim$ GEV(μ, σ, γ).

Corollary 4. *Let* $\{X_t\}$ *and* $\{S_t\}$ *are as in Theorem 5.* $\{\xi_t\}$ *is an independent* $\mathcal{S}(\psi)$ *random variable sequence with* $\psi \in (0, 1)$, *then the series*

$$Y_t = \mu - \frac{\sigma}{\gamma'} + \frac{\sigma}{\gamma'} X_t^{\psi\gamma\gamma'} \xi_t^{\psi\gamma'}, \quad 1 \leq t \leq n, \tag{1.12}$$

Y_t *has marginal* GEV(μ, σ, γ') *distribution for* $1 \leq t \leq n$, *i.e.* $Y_t \sim$ GEV(μ, σ, γ').

The time series in Equation (1.9), (1.10) and the state space series in (1.11) (1.12) are Markovian.

1.4 State space model

The origin of state space models could be traced to dynamic systems in engineering including automatic control, communications, robotics, and aerospace systems. Merwe et al. [94] defined the state space model as follows

Definition. A state space model is such that the measurement Y_t recorded at time t is described by two equations, the observation and the state ones. The observation equation depicts the link between Y_t and X_t, where X_t represents an unobserved state variable. The state equation, also called the system equation, models the temporal dynamical structure. Many state space models can be represented as

$$\begin{aligned} Y_t &= F_t(X_t, \xi_t), \qquad \text{(observation equation)} \\ X_t &= G_t(X_{t-1}, \eta_t), \qquad \text{(state equation)}. \end{aligned}$$

where F_t, G_t are functions and ξ_t, η_t are noises.

The state variables X_t, though unobserved, are important to the model because they provide a data generating mechanism. The conditional density $f(X_t|Y_1, Y_2, \ldots, Y_t)$ is one of our primary interests in the state space modeling, which is called the filtering density.

1.4.1 Literature review

Because of its simplicity, linear state space models with Gaussian errors are most common in the literature.

Linear state space model. Let $\{u_t, t \geq 1\}$, $\{v_t, t \geq 1\}$ be independent normal sequences with zero means and variances σ_u^2, σ_v^2 respectively. For some non-zero constants ϕ_1, ϕ_2, the state space model is

$$\begin{aligned} Y_t &= \phi_1 X_t + v_t, \\ X_t &= \phi_2 X_{t-1} + u_t. \end{aligned}$$

For this kind of model, maximum likelihood estimation is a plausible option because of the linearity and the Gaussian noise. Covariance structure of the observation sequence $\{Y_t\}$ is also commonly used in estimation. When the hidden state X_t is of interest, Kalman filter is the preferred framework to address this kind models (see Kalman [52]).

In many practical situations under a state space model, the assumption of Gaussian noise can be relaxed to the requirement of white noise, without further complications. Whenever the model is not linear, the analysis becomes more complex, which is our

case.

Now, we review the relation of the Gumbel and α-stable distributions.

Let $S \sim \mathcal{S}(\alpha, \sigma)$ be an α-stable random variable with the scale parameter σ,

(i) Gumbel random variable can be represented as the linear combination of a Gumbel random variable and the logarithm of an independent α-stable random variable (see Corollary 1 and also Crowder [19], Hougaard [45], Fougéres, Nolan and Rootzén [31]), i.e. if $G \sim \text{Gumbel}(\mu, \sigma)$ is independent of S, then $G + \log S \sim \text{Gumbel}(\mu, \sigma/\alpha)$.

(ii) Gumbel random variable can be expressed as the maximum over an α-stable distributed number of independent blocks (see Fougéres, Nolan and Rootzén [31]) , i.e. if the maximum over an unit block is Gumbel(μ, σ) distributed, then the maximum over S blocks is Gumbel distributed with the location parameter μ, the scale parameter σ/α.

(iii) Gumbel random variable can be thought as a conditionally Poisson point process. If X is the maximum y-coordinate of a point process in $(0, 1] \times R$, with intensity $Se^{-(x-\mu)/\sigma}$, conditionally on stable variable S, then the unconditional distribution X is Gumbel with location μ, scale σ/α (see Tawn [89], Fougéres, Nolan and Rootzén [31]).

Using property (i), Toulemonde et al. [93] proposed the following linear Gumbel time series and a Gumbel state space model related to the logarithm of α-stable variables.

Linear Gumbel model. For $t \in \mathbb{Z}$, let

$$X_t = \alpha X_{t-1} + \alpha\sigma \log S_t,$$

where $\{S_t\}$ is α-stable noise with parameter $\alpha \in (0, 1)$ and parameter $\sigma > 0$, then $\{X_t\}$ is a linear Gumbel AR series.

This model has been applied to the daily maximum of methane CH_4 and daily maximum of nitrous oxide (N_2O) measured in Gif-sur-Yvette, France (Toulemonde et al. [93]).

Later, Toulemonde et al. extended this AR series to a Gumbel state space model

in [92].

$$Y_t = v_t - H_t\sigma\gamma_e(\frac{1}{\alpha_2} - 1) + H_t X_t + H_t\sigma\log\eta_t,$$
$$X_t = \alpha_1 X_{t-1} + \alpha_1\sigma\log\xi_t - \sigma\gamma_e(1-\alpha_1), \quad t \in \mathbb{Z}$$

where $\{\xi_t\}$ and $\{\eta_t\}$ are i.i.d. α-stable random variables, $\xi_t \sim \mathcal{S}(\alpha_1), \eta_t \sim \mathcal{S}(\alpha_2)$, parameters $H_t > 0$, $\alpha_1,\ \alpha_2 \in (0,1)$ and γ_e is the Euler's constant.

In this model, Y_t and X_t are both Gumbel variables, with parameters $\left(v_t - \frac{H_t\gamma_e\sigma}{\alpha_2}, H_t\frac{\sigma}{\alpha_2}\right)$ and $(-\gamma_e\sigma, \sigma)$, respectively. The parameter estimation was not presented in this paper. As to the approximation of the filtering density, the auxiliary particle filter proposed by Pitt and Sherphard [79] was applied to this model. Their results show some advantages when comparing with Kalman filter and bootstrap filter.

Max-stable Model. For independent Fréchet noise $\{\xi_t\}$ and $\{\eta_t\}$, with parameters $F_t > 0, G_t > 0$, using the stable postulate of GEV distributions, Naveau and Poncet [71] proposed a max-stable state space model with GEV distributed marginals.

$$Y_t = \max(F_t X_t, \xi_t),$$
$$X_t = \max(G_t X_{t-1}, \eta_t), \quad t \in \mathbb{Z}.$$

To see this, for simplicity, we let $F_t = F$ and $G_t = G$. From the state equation, we have

$$X_t = \max\left(\ldots, G^t\eta_0, G^{t-1}\eta_1, \ldots, G\eta_{t-1}, \eta_t\right). \tag{1.13}$$

Without loss of generality, assume that the state variable follows Fréchet distribution with location parameter 0, scale parameter 1, since if $\eta_t \sim$ Fréchet(μ, δ, γ), then

$$X_t = \mu + \delta\max\left(\ldots, G^t\eta_0', G^{t-1}\eta_1', \ldots, G\eta_{t-1}', \eta_t'\right),$$

where $\eta_t' \sim$ Fréchet$(0, 1, \gamma)$.

Use Equation (1.13),

$$\mathrm{P}(X_t \le x) = \prod_{i=-\infty}^{t} \mathrm{P}\left(\eta_i \le \frac{x}{G^{t-i}}\right) = \prod_{i=-\infty}^{t} e^{-\left(\frac{x}{G^{t-i}}\right)^{-\gamma}} = \exp\left(-\frac{x^{-\gamma}}{1-G^{\gamma}}\right),$$

which indicates $X_t \sim$ Fréchet $\left(0, (1-G^\gamma)^{-1/\gamma}, \gamma\right)$, thus $Y_t \sim$ Fréchet $\left(0, (\frac{F^\gamma+1-G^\gamma}{1-G^\gamma})^{1/\gamma}, \gamma\right)$ if $\xi_t \sim$ Fréchet$(0,1,\gamma)$.

In their paper, this model is used to fit the daily maximum of precipitation amounts recorded every three hours in Nîmes, France with some given value of F_t, G_t.

Gumbel state space model. Naveau and Poncet [71] also proposed the following state space model to include the heavy tailed distributions in climate studies,

$$\begin{aligned} Y_t &= F_t \log X_t + \xi_t, \\ X_t &= G_t X_{t-1} + \eta_t, \quad t \in \mathbb{Z}, \end{aligned}$$

where $\{\xi_t\}$ is an i.i.d. Gumbel noise, $\{\eta_t\}$ is an i.i.d. α-stable noise, parameters $F_t, G_t > 0$.

X_t here is an α-stable variable by the definition of stable distribution while Y_t is a Gumbel variable using Corollary 1.

This model is based on the results given by Fougéres, Nolan and Rootzén [31], where models with GEV distributed marginals constructed by α-stable variables, using the properties (i)-(iii) between the Gumbel distribution and the α-stable distribution. Unfortunately, the estimation and the filtering density of hidden state are unsolved in Naveau and Poncet [71], Fougéres, Nolan and Rootzén [31].

GEV-M3 Model. In 2010, Kunihama et al. [60] studied the model

$$\begin{aligned} Y_t &= \mu + \psi \frac{X_t^\gamma - 1}{\gamma} + \eta_t, \\ X_t &= \max_{0 \le k \le K} \alpha_k Z_{t-k}, \quad t \in \mathbb{Z} \end{aligned}$$

where $\{Z_t\}$ is a sequence of independent unit Fréchet random variables and $\eta_t \sim N(0, \sigma^2)$. The α_k's are constants satisfy $\sum_k \alpha_k = 1$, and μ, ψ, γ are the main parameters of the model.

The state equation ensures the state is unit Fréchet distributed since

$$\mathrm{P}(X_t \le x) = \prod_{0 \le k \le K} \mathrm{P}\left(Z_{t-k} \le \frac{x}{\alpha_k}\right) = e^{-x^{-1}\sum_k \alpha_k} = e^{-x^{-1}}.$$

Thus the observation Y_t follows GEV(μ, ψ, γ) only when there is no observation noise, i.e. $\sigma^2 = 0$.

This model was applied to the daily minimum of in-trade stock returns. This state equation shows a max-stable process, called the maxima of moving maxima (M3) process. Kunihama et al. [60] transformed the Fréchet sequence $\{Z_t\}$ to Gumbel sequence, then used normal structure to approximate the Gumbel sequence. They tried Monte Carlo method, bootstrap filter, particle filter to obtain the filtering density.

A similar model was considered by Peter Hall, Liang Peng and Qiwei Yao [41], Chamú Morales [12], called the moving-maximum model for extremes of time series.

Notice that, if the observation equation is defined as

$$Y_t = \mu - \frac{\delta}{\gamma} + \frac{\delta}{\gamma} X_t^{\psi\gamma} \eta_t^{\psi\gamma},$$

where η_t is the α-stable distributed errors with parameter $\psi \in (0, 1)$, then the observation is GEV(μ, ψ, γ) distributed.

GEV-AR and GEV-MA Model. Fitted to a monthly series of minimum returns of Tokyo daily stock data, Nakajima et al. [70] considered the model

$$\begin{aligned} Y_t &= \mu + \psi \frac{\exp(\sigma X_t) - 1}{\sigma} + \xi_t, \\ X_t &= \alpha X_{t-1} + \eta_t, \quad t = 1, \ldots, n \end{aligned}$$

where ξ_t is normally distributed with variance σ^2, $|\alpha| < 1$.

In their paper, Nakajima et al. assumed that the hidden state X_t is given by a stationary AR process driven by the Gumbel distributed noise η_t, ignored the fact that Gumbel is not in the stable family. Obviously, should that were the case, the marginal sequence $\{Y_t\}$ would be GEV distributed whenever there is no observation noise, i.e. $\sigma^2 = 0$.

Instead of the AR process in the GEV-AR model, the MA process

$$X_t = \eta_t + \theta \eta_{t-1}, \quad t \in \mathbb{Z}$$

where $|\theta| < 1$ and η_t is Gumbel distributed, is used as the state equation in a GEV-MA model in Nakajima et al. [70]. Only when both θ and σ are zero, this model

produces Gumbel observations Y_t, with simple linear structure.

1.5 Objective of the thesis

Our interest is the state space model with GEV distributed marginals and α-stable or exponential α-stable distributed errors, improved from the GEV-AR model and the Gumbel state space model.

For the state space model

$$\begin{cases} Y_j = \mu - \frac{\sigma}{\gamma} + \frac{\sigma}{\gamma} e^{\psi\gamma X_j} \xi_j^{\psi\gamma}, \\ X_{j+1} = \alpha X_j + \alpha \log S_{j+1}, \end{cases} \tag{1.14}$$

let $\gamma \neq 0$, $\{S_j \sim \mathcal{S}(\alpha), 1 \leq j\}$ and $\{\xi_j \sim \mathcal{S}(\psi), 1 \leq j\}$ be two independent α-stable random variable sequences with $\alpha, \psi \in (0, 1)$.

If $X_0 \sim \text{Gumbel}(0, 1)$ then $X_j \sim \text{Gumbel}(0, 1)$ and $Y_j \sim \text{GEV}(\mu, \sigma, \gamma)$.

The unsolved estimation problem in Naveau and Poncet [71] and Fougéres, Nolan and Rootzén [31]) is a special case of this model.

This thesis mainly focuses on the estimation and model filter for the model (1.14).

Chapter 2 is concerned on the estimation of a time series with Fréchet distributed marginals and α-stable distributed errors. Yule-Walker estimation, AR model and recursive Hill estimation and recursive moment estimation are used.

Chapter 3 is about the state space model with GEV distributed marginals and α-stable distributed errors. ARMA model, Yule-Walker estimation and adjusted regression estimation are considered.

The model filter and prediction are discussed in the Chapter 4. Different methods like Kalman filter, particle filer and auxiliary particle filter, plain linearization are compared.

Chapter 2

Time series with Fréchet distributed marginals and α-stable distributed errors

Before describing the inferences for state space models, we will consider the time series with GEV distributed marginals and α-stable distributed errors.

In this chapter, we mainly discuss a time series $\{X_t\}$ with Fréchet distributed marginals, as stated in Equation (1.9), Theorem 5. As a corollary to Theorem 5, we have that if $S_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha)$ with $\alpha \in (0,1), \gamma > 0$,

$$\begin{aligned} X_t = & \quad X_{t-1}^{\alpha} S_t^{\frac{\alpha}{\gamma}} \\ = & \quad X_{t-k}^{\alpha^k} \prod_{i=1}^{k} S_{t+1-i}^{\alpha^i/\gamma} \\ = & \quad \prod_{i=-\infty}^{t} S_i^{\alpha^{t+1-i}/\gamma}, \end{aligned}$$

where each X_t has a marginal Fréchet$(0,1,\gamma)$ distribution for $t \in \mathbb{Z}$.

Our aim is to estimate the stability parameter α and the Fréchet tail parameter γ in Equation (2.1) using the observation sequence $\{X_t\}$.

The sequence $\{X_t\}$ is strictly stationary, since

$$\mathrm{P}(X_{t_1} \leq x_{t_1}, X_{t_2} \leq x_{t_2}, \ldots, X_{t_s} \leq x_{t_s}) = \mathrm{P}(X_{t_1+k} \leq x_{t_1}, X_{t_2+k} \leq x_{t_2}, \ldots, X_{t_s+k} \leq x_{t_s})$$

for any $t_1, t_2, \ldots, t_s$.

Knowing the density of α-stable variable S_t, we can write the likelihood function of $\{X_t, 0 \leq t \leq n\}$ as

$$f(x_0, \ldots, x_n; \alpha, \gamma) = f(x_0) \left(\frac{\gamma}{\alpha}\right)^n \prod_{i=1}^{n} x_{i-1}^{-\gamma} X_i^{\frac{\gamma}{\alpha}-1} f_\alpha \left(x_{i-1}^{-\gamma} x_i^{\frac{\gamma}{\alpha}}\right)$$

where f_α denotes the density of $\mathcal{S}(\alpha)$.

Usually the maximum likelihood function is used to do the estimation, however it cannot be used here, since when the stability parameter α is small, in a neighbourhood of zero, the density around point zero is too large. For any x_i which is less than 1, $f_\alpha \left(x_{i-1}^{-\gamma} x_i^{\frac{\gamma}{\alpha}}\right)$increases as α goes to 0.

As an alternative to maximum likelihood estimation, moment estimation and Yule-Walker estimation can be considered. The mean and variance of the stationary sequence $\{X_t\}$ are finite under some constrains.

$$\begin{aligned} \mathrm{E}(X_t) &= \Gamma(1 - \tfrac{1}{\gamma}), && \text{when} \quad \gamma > 1, \\ \mathrm{Var}(X_t) &= \Gamma(1 - \tfrac{2}{\gamma}) - \Gamma^2(1 - \tfrac{1}{\gamma}), && \text{when} \quad \gamma > 2. \end{aligned}$$

However, all the moments of the logarithm of Fréchet variables exist. It would be natural to consider the time series of logarithm of X_t. From Equation (2.1) we have

$$\log X_t = \alpha \log X_{t-1} + \frac{\alpha}{\gamma} \log S_t. \tag{2.1}$$

We already know that $\log X_t \sim \mathrm{Gumbel}(0, \frac{1}{\gamma})$ for all $t \in \mathbb{Z}$, thus $\{\log X_t\}$ is a linear stationary sequence and

$$\begin{aligned} &\mathrm{E}(\log X_t) = \frac{\gamma_e}{\gamma}, \quad \mathrm{Var}(\log X_t) = \frac{\pi^2}{6} \frac{1}{\gamma^2}, \\ &\mathrm{Cov}(\log X_t, \log X_{t-k}) = \alpha^k \, \mathrm{Var}(\log X_{t-k}) = \alpha^k \frac{\pi^2}{6} \frac{1}{\gamma^2}, \end{aligned}$$

where γ_e denotes the Euler's constant.

2.1 Yule-Walker estimation for α

The sequence $\{\log X_t\}$ in Equation (2.1) is stationary. The estimator of α could be obtained by using Yule-Walker equation, while the estimator of γ could be obtained by the moment estimation.

$$\hat{\alpha} = \frac{\sum_{i=2}^{n}(\log X_i - \overline{\log X})(\log X_{i-1} - \overline{\log X})}{\sum_{i=1}^{n}(\log X_i - \overline{\log X})^2},$$
$$\hat{\gamma} = \sqrt{\frac{\pi^2(n-1)}{6\sum_{i=1}^{n}(\log X_i - \overline{\log X})^2}},$$

where $\overline{\log X}$ represents the mean of the sequence $\{\log X_i, 1 \leq i \leq n\}$.

The asymptotic properties of the Yule-Walker estimator $\hat{\alpha}$ and the moment estimator $\hat{\gamma}$ can be summarized using the result in Toulemonde et al. [93].

Theorem 6. *(Toulemonde et al. [93]) The Yule-Walker estimator $\hat{\alpha}$ and the moment estimator $\hat{\gamma}$ are almost surely consistent and $\sqrt{n}(\hat{\gamma}-\gamma, \hat{\alpha}-\alpha)'$ converges in distribution to a Gaussian vector with zero mean and covariance matrix*

$$\begin{pmatrix} \frac{11}{10}\gamma^2\frac{1+\alpha^2}{1-\alpha^2} & -\alpha\gamma \\ -\alpha\gamma & 1-\alpha^2 \end{pmatrix}$$

as $n \to \infty$.

The Ergodic theorem is used to make sure that the estimates converge almost surely, while the finite fourth moments of the logarithm of α-stable random variable, the classical result of Bartlett in Brockwell and Davis [9] together with the delta method guarantee the asymptotic normality of the estimate.

In Figure 2.1, simulation results of $\hat{\alpha}$ and $\hat{\gamma}$ with different chain size ($n = 50, 100$) are displayed. The first column is the plots of α estimate using Yule-Walker estimation, the second column is the plots of the estimate of Fréchet tail parameter γ, which takes the value 1.2 (black lines). The red curves are plotted by means of estimators in 500 repeated simulations and the green curves give their respective 95% confidence intervals.

Yule-Walker estimation works satisfactorily and efficiently for many models, but in our case, it produces non-null probability of estimators locating outside the parameter space. In our simulations (Figure 2.1), the Yule-Walker estimates are negative in about

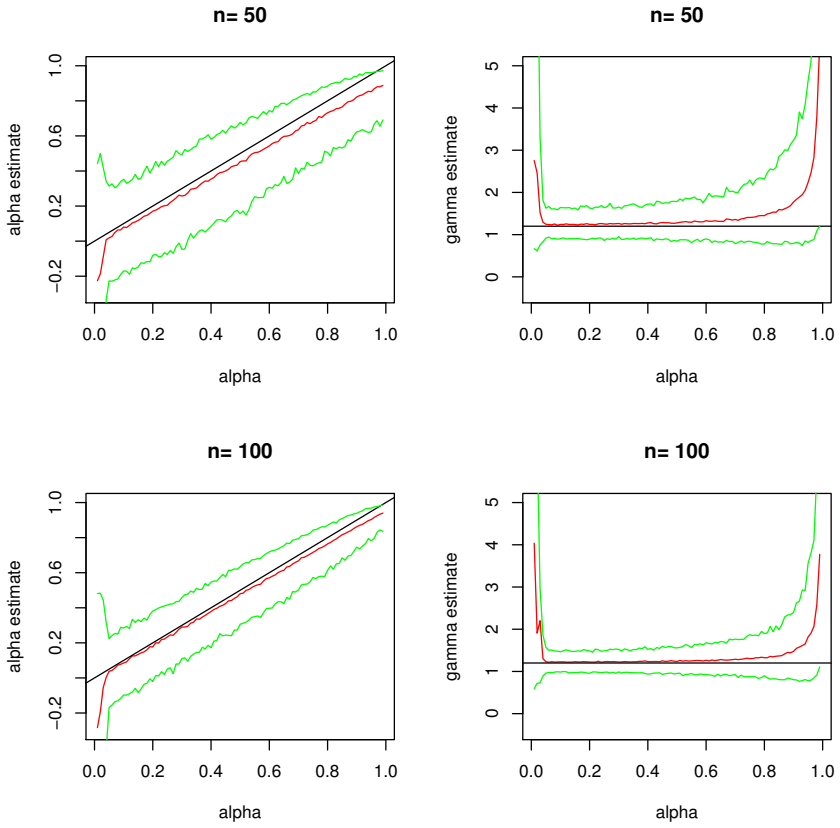

Figure 2.1: Yule-Walker estimate of α in the first column and moment estimate of γ in the second column, with means in red and 95% confidence interval in green.

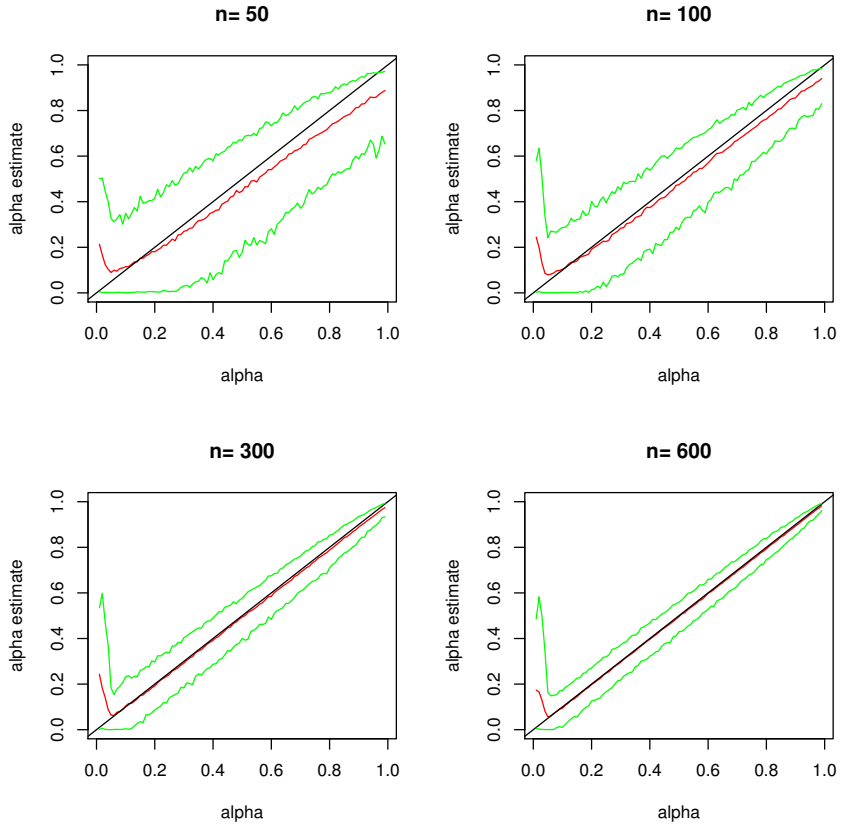

Figure 2.2: Adjusted Yule-Walker estimate of α with means in red and 95% confidence interval in green.

half the cases whenever α is close to 0 (e.g., with $\alpha = 0.06$, 42% of the estimates are negative). When α takes a value close to 1, some estimates exceed 1.

A way to avoid the estimates from being outside the parameter space is to use

$$\hat{\alpha}_k = \frac{n-1}{n-k+1} \frac{\sum_{i=k}^{n} (\log X_i - \overline{\log X})(\log X_{i-1} - \overline{\log X})}{\sum_{i=1}^{n} (\log X_i - \overline{\log X})^2}$$

for $k = 2, 3, \ldots,$ choose the estimate $\hat{\alpha} = \hat{\alpha}_k$, where $k = \arg\ \min\{k : \hat{\alpha}_k \in (0, 1)\}$.

The simulation result of this adjusted Yule-Walker estimation are shown in Figure 2.2. Compared with the original Yule-Walker estimators, the simulation result improves a little. However when the chain size is small ($n = 50$), this method overestimates α when α is close to 0.

Usually a larger chain size produces better estimates. In Figure 2.2, the estimates

around 0 show no improvement as chain size increases. One of the reasons is that it is hard to get a large chain sized non-zero observation $\{X_t\}$ when α is close to 0. Notice that the density of α-stable distributed errors around zero is large, which means at some time k, α-stable variable S_k would be nearly 0, which makes $X_t \approx 0$ for all $t \geq k$. Thus even the chain size n increases to a large number, the observation sequence provides little information about α, as long as $\alpha \to 0$.

2.2 Autoregressive model and conditional linear programming

Besides the Yule-Walker estimation, the AR model can be used for the estimation purposes. The sequence $\{\log X_t\}$ can be written as an AR(1) sequence, for $t = 1, 2, \ldots, n$,

$$\begin{aligned} \log X_t &= \alpha \log X_{t-1} + \frac{\alpha}{\gamma} \log S_t \\ &= \frac{\gamma_e}{\gamma}(1-\alpha) + \alpha \log X_{t-1} + u_t, \end{aligned} \tag{2.2}$$

where $u_t = \frac{\alpha}{\gamma} \log S_t - \frac{\gamma_e}{\gamma}(1-\alpha)$ is a zero mean noise.

The parameters α and γ could be estimated by obtaining the AR(1) coefficient and the constant value $\frac{\gamma_e}{\gamma}(1-\alpha)$.

Figure 2.3 shows the simulation result of $\alpha \in (0,1), \gamma = 1.2$(black horizontal line in the second column) using AR model with n=50 and 100 respectively, in 500 repetitions. Since the least square estimation is used in the AR(1) model, the simulation results are equivalent to the Yule-Walker estimates (see the first column in Figure 2.1). However, the estimates of γ are erratic, may be caused by the asymmetry and the skewness of u_t.

A method which can yield an $\hat{\alpha}$ within the parameter space (0,1) is needed. The constrained linear programming is considered.

The AR(1) model (2.2) can be thought as a simple linear regression equation, where $\{u_t\}$ is an i.i.d. sequence with zero mean.

To obtain estimator $\hat{\alpha} \in (0,1)$, we use constrained linear programming. Let

$$(\hat{\alpha}, \hat{\gamma})' = \underset{0<\alpha<1, \gamma>0}{\arg\min} \sum_{i=2}^{n} \left| \log X_i - \alpha \log X_{i-1} - \frac{\gamma_e(1-\alpha)}{\gamma} \right|.$$

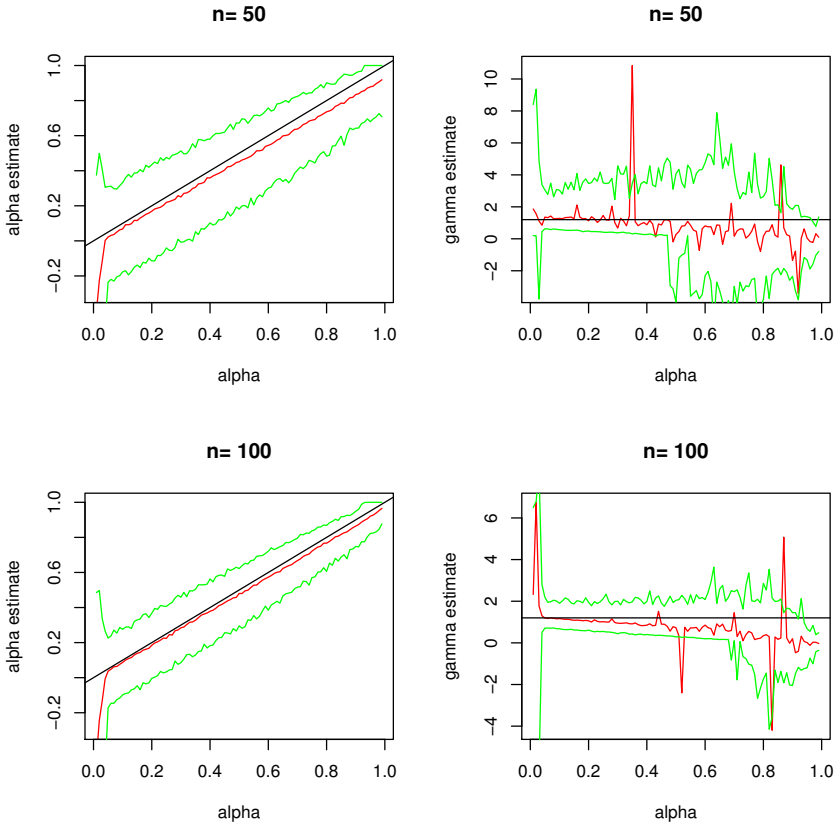

Figure 2.3: Estimation results obtained using the AR(1) model. Estimate of α in the first column and estimate of γ in the second column, with means in red and 95% confidence interval in green.

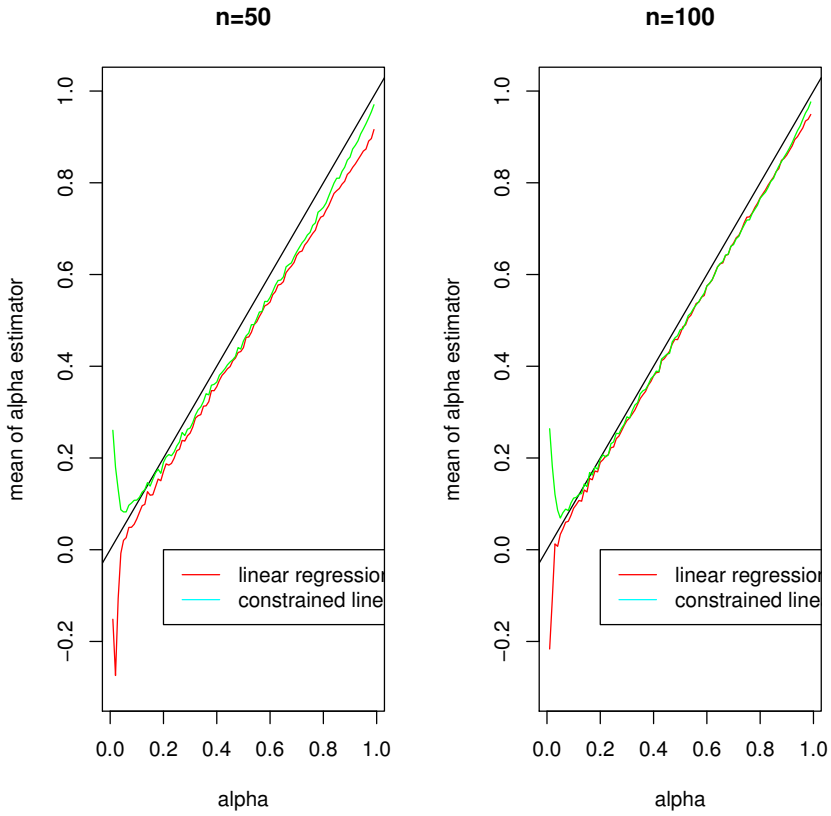

Figure 2.4: Comparison of α estimates obtained by simple linear regression and the constrained linear programming.

Figure 2.4 shows the comparison of α estimates obtained by the simple linear regression (also the original Yule-Walker estimation and the AR(1) model) and the constrained linear programming. Overall, the constrained linear programming performs better than the simple linear regression. Especially when $\alpha > 0.8$, where it can happen that some simple linear regression estimators exceed 1, while the constrained linear programming estimators are closer to the true value.

However, the constrained linear programming does not improve the estimation of γ. The constrained linear programming estimator of γ is still erratic. Furthermore, when $\hat{\alpha}$, using the constrained linear programming, hits the constrain bounds, we have trouble estimating the parameter γ.

2.3 Weak mixing condition

Lacking of the effective estimation for γ (the maximum likelihood estimation does not work for our observations and the AR model gives erratic estimates), the estimation ignoring the dependence structure of the observations seems to be unavoidable. Before doing this, the dependence structure of our observations should be studied.

In 1983, Leadbetter and Rootzén [62] proposed a weak mixing condition $D(u_n)$, which focused on the asymptotic independence of a stationary sequence. Later, Hüsler showed that the extreme values of non-stationary sequences satisfied the asymptotic independence condition in his papers [47], [46], [48] and the book of Falk, Hüsler and Reiss [27]. The weak mixing condition was extend to non-stationary random fields by Pereira and Ferreira in 2006 (see [77]).

Besides the weak mixing condition, there are other mixing conditions like the strong mixing assumption (Rootzén, Leadbetter and de Haan[86]), coordinate-wise mixing (Leadbetter and Rootzén [61]), ρ-mixing condition (Kolmogorov and Rozanov [57]) and so on. In Brandley [8], eight mixing conditions are connected and compared. More information about mixing conditions can be found in Brandley [8], Chen et al. [13].

Under certain mixing condition, the good properties, like the Central Limit Theorem, the weak invariance principles and rates of convergence hold for independent sequence still hold for stationary sequence.

The weak mixing condition $D(u_n)$ is defined as below.

Definition. Write

$$F_{i_1,\ldots,i_n}(x_1,\ldots,x_n) = \mathrm{P}(X_{i_1} \le x_1,\ldots,X_{i_n} \le x_n)$$

for the jointly distribution of $X_{i_1},\ldots,X_{i_n}$, and for brevity, $F_{i_1,\ldots,i_n}(u) = F_{i_1,\ldots,i_n}(u,\ldots,u)$ for each $n, i_1,\ldots,i_n, u$.

Let $\{u_n\}$ be a sequence of constants. Then the sequence $\{X_n\}$ is said to satisfy $D(u_n)$ if for each n, k_n and each choice of integers $i_1,\ldots,i_p, j_1,\ldots,j_{p'}$ such that

$$1 \le i_1 < i_2 < \cdots < i_p < j_1 < \cdots < j_{p'} \le n, j_1 - i_p \ge k_n, \tag{2.3}$$

we have

$$|F_{i_1,\ldots,i_p,j_1,\ldots,j_{p'}}(u_n) - F_{i_1,\ldots,i_p}(u_n)F_{j_1,\ldots,j_{p'}}(u_n)| < c_{n,k_n} \tag{2.4}$$

where $c_{n,k_n} \to 0$ as $n \to \infty$ for some sequence $k_n \to \infty,\ k_n = n$).

The sequence $\{u_n\}$ in the weak mixing condition $D(u_n)$ is a sequence of thresholds satisfying

$$\lim_{n\to\infty} n(1 - F(u_n)) = \tau$$

for some given $\tau > 0$.

If the sequence $\{X_t\}$ satisfies the weak mixing condition, under which the Central Limit Theorem holds, then the moment estimation of γ is unbiased and consistent, which is the same as the result in Theorem 6.

Theorem 7. *For the time series* $\{X_t, t \in \mathbb{Z}\}$, *if* $X_t = X_{t-1}^{\alpha} S_t^{\frac{\alpha}{\gamma}}$, $S_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha)$, *then* $X_t \sim$*Fréchet*$(0, 1, \gamma)$ *and the sequence* $\{X_t\}$ *satisfies the weak mixing condition* $D(u_n)$.

To show the inequality (2.4) holds, first we have

$$\begin{aligned} & |F_{i_1,\ldots,i_p,j_1,\ldots,j_{p'}}(u_n) - F_{i_1,\ldots,i_p}(u_n) F_{j_1,\ldots,j_{p'}}(u_n)| \\ = \ & F_{i_1}(u_n) F_{i_2|i_1}(u_n) \cdots F_{i_p|i_{p-1}}(u_n) \cdot \left|F_{j_1|i_p}(u_n) - F_{j_1}(u_n)\right| \cdot F_{j_{p'}|j_{p'-1}}(u_n) \cdots F_{j_2|j_1}(u_n) \\ \leq \ & \left|F_{j_1|i_p}(u_n) - F_{j_1}(u_n)\right|, \end{aligned}$$

where $F_{i_2|i_1}(\cdot)$ is the conditional probability of X_{i_2} given X_{i_1} and $i_1, \ldots, i_p, j_1, \ldots, j_{p'}$ are defined as in (2.3). For simplicity, we let $j_1 - i_p = k$ and denote $j_1 = t$ in the proof of Theorem 7.

Now the proof is simplified to find out the effect of X_{t-k} on X_t, i.e., compare the conditional probability $\mathrm{P}(X_t \leq u_n | X_{t-k} \leq u_n)$ with $\mathrm{P}(X_t \leq u_n)$. Define the threshold u_n as

$$u_n = \left(-\log\left(1 - \frac{\tau}{n}\right)\right)^{-\frac{1}{\gamma}}, \tag{2.5}$$

for some given $\tau > 0$.

The following lemma is needed in the proof of Theorem 7.

Lemma 8. *(See Feller [29] p.336) For independent Lévy stable distributions* $S_1 \sim \mathcal{S}(\alpha)$, $S_2 \sim \mathcal{S}(\psi)$,

$$S_1 S_2^{1/\alpha} \sim \mathcal{S}(\alpha\psi).$$

Proof. For $t > 0$,

$$\begin{aligned} \mathrm{E}\left(e^{-tS_1 S_2^{1/\alpha}}\right) &= \mathrm{E}\left(\mathrm{E}\left(e^{-tS_1 S_2^{1/\alpha}}|S_1\right)\right) \\ &= \mathrm{E}\left(e^{-t^\alpha S_2}\right) \\ &= e^{-t^{\alpha\psi}}, \end{aligned}$$

which is the Laplace transform of a $\mathcal{S}(\alpha\psi)$ random variable by the definition. □

Proof of Theorem 7. Use Lemma 8 we have $S_{t-k+1} S_{t-k+2}^{\frac{1}{\alpha}} \cdots S_t^{\frac{1}{\alpha^{k-1}}} \sim \mathcal{S}(\alpha^k)$.

Applying the stable tail behaviour in Equation (1.6), we have that

$$\mathrm{P}(S_{t-k+1} S_{t-k+2}^{\frac{1}{\alpha}} \cdots S_t^{\frac{1}{\alpha^{k-1}}} \leq y) \approx 1 - \frac{1}{\Gamma(1-\alpha^k)} y^{-\alpha^k}$$

for large y, so that the joint distribution of X_{t-k} and X_t can be approximated as

$$\begin{aligned} & \mathrm{P}(X_t \leq u_n, X_{t-k} \leq u_n) \\ = & \int_0^{u_n} \mathrm{P}(X_t \leq u_n | X_{t-k} = x) f_{X_{t-k}}(x) dx \\ = & \int_0^{u_n} \mathrm{P}\left(x^{\alpha^k} \left(S_{t-k+1} S_{t-k+2}^{\frac{1}{\alpha}} \cdots S_t^{\frac{1}{\alpha^{k-1}}} \right)^{\frac{\alpha^k}{\gamma}} \leq u_n | X_{t-k} = x \right) d \exp(-x^{-\gamma}) \\ = & \int_0^{u_n} F_{\alpha^k}\left(x^{-\gamma} u_n^{\gamma/\alpha^k} \right) d \exp(-x^{-\gamma}) \\ \approx & \int_0^{u_n} \left(1 - \frac{1}{\Gamma(1-\alpha^k)} u_n^{-\gamma} x^{\gamma\alpha^k} \right) d \exp(-x^{-\gamma}) \\ = & e^{-u_n^{-\gamma}} - \frac{1}{\Gamma(1-\alpha^k)} u_n^{-\gamma} \int_0^{u_n} x^{\gamma\alpha^k} d \exp(-x^{-\gamma}) \\ = & e^{-u_n^{-\gamma}} - \frac{1}{\Gamma(1-\alpha^k)} u_n^{-\gamma} \int_{u_n^{-\gamma}}^{\infty} u^{-\alpha^k} e^{-u} du \\ = & e^{-u_n^{-\gamma}} - u_n^{-\gamma} \mathrm{P}(U > u_n^{-\gamma}) \end{aligned}$$

where U is a Gamma$(1-\alpha^k, 1)$ random variable.

Combined with the Equation (2.5), we have

$$\begin{aligned} & \mathrm{P}(X_t \le u_n | X_{t-k} \le u_n) \\ = \; & 1 - \frac{u_n^{-\gamma}}{e^{-u_n^{-\gamma}}} \mathrm{P}(U > u_n^{-\gamma}) \\ = \; & 1 + \frac{\log(1-\frac{\tau}{n})}{1-\frac{\tau}{n}} \mathrm{P}(U > u_n^{-\gamma}) \\ = \; & 1 - \frac{\tau}{n} \mathrm{P}(U > u_n^{-\gamma}) + O_p\left(\frac{1}{n^2}\right), \end{aligned} \tag{2.6}$$

which means for $\alpha \in (0,1)$,

$$|F_{i_1,\ldots,i_p,j_1,\ldots,j_{p'}}(u_n) - F_{i_1,\ldots,i_p}(u_n) F_{j_1,\ldots,j_{p'}}(u_n)| \to 0.$$

Thus $\{X_t\}$ satisfies condition $D(u_n)$. □

Furthermore, notice that when $\alpha^{k_n} \to 0$ with $k_n = |j_1 - i_p| \to \infty$, U converges in distribution to Exponential(1), thus $\mathrm{P}(U > u_n^{-\gamma}) \to e^{-u_n^{-\gamma}}$,

$$|\,\mathrm{P}(X_t \le u_n | X_{t-k} \le u_n) - P(X_t \le u_n)| \;\; = O_p\left(\frac{1}{n^2}\right).$$

When $\alpha \to 1$, α^{k_n} either goes to 0 or 1 and if $\alpha^{k_n} \to 1$, U converges in distribution to Dirac(0). $\mathrm{P}(U \le u_n^{-\gamma}) \to 1$, which makes

$$|\,\mathrm{P}(X_t \le u_n | X_{t-k} \le u_n) - P(X_t \le u_n)| = \tau/n.$$

When $\alpha \to 1$, the dependence structure of $\{X_t\}$ is stronger than that of $\alpha \to 0$. This suggests that using the dependence structure to estimate the stability parameter when α is very small may be useless. On the other hand, when α is large, the estimation of Fréchet tail parameter obtained by the method of moments is okay since the observation sequence has stronger dependency, compared with the observation sequence generated by small α. However, in this case, be aware that the effective sample size would be much smaller that the chain size.

Figure 2.5 shows the estimators of γ using the maximum likelihood estimation ignoring the dependence structure. When $\alpha \to 0$, the MLE estimate is better than AR(1) estimate. When α is large, ignoring the dependence structure ends up with the increasing bias of the estimates due to the stronger dependency.

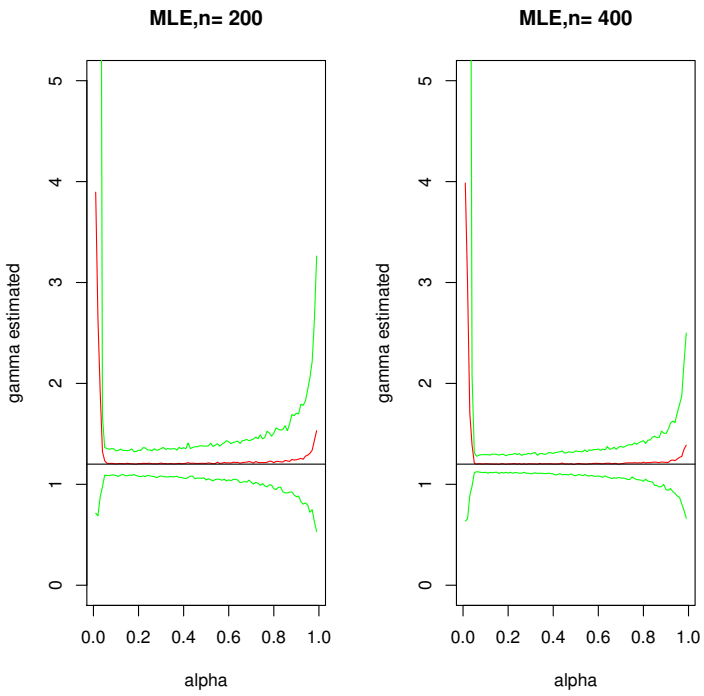

Figure 2.5: MLE of γ when ignoring the dependence structure (means in red and 95% in green).

2.4 Strong mixing condition

In 1956, Rosenblatt introduced the strong mixing condition, which is also called α-mixing condition in his paper [87].

Definition. (Strong mixing condition) Suppose $\{X_t\}$ is a sequence of random variables on a given probability space $(\Omega, \mathscr{F}, P)$. Let $\mathscr{F}_j^l$ denote the $\sigma-$field of events generated by the random variable $X_k, j \leq k \leq l$. For any two σ-fields $\mathscr{A}$ and $\mathscr{B} \subset \mathscr{F}$, define the measure of dependence as

$$\varrho(\mathscr{A}, \mathscr{B}) = \sup_{A \in \mathscr{A}, B \in \mathscr{B}} |\,\mathrm{P}(A \cap B) - \mathrm{P}(A)\,\mathrm{P}(B)|.$$

For the given random sequence $\{X_t\}$, for any positive integer n, define the dependence coefficient

$$\varrho(n) = \sup_j \varrho(\mathscr{F}_{-\infty}^j, \mathscr{F}_{j+n}^{\infty}).$$

The random sequence $\{X_t\}$ is said to be strongly mixing or α-mixing if $\varrho(n) \to 0$ as $n \to \infty$.

Rosenblatt [87] proved the Central Limit Theorem holds under the strong mixing condition. Besides Central Limit Theorem, weak invariance principles, laws of the iterated logarithm, almost sure invariance principles, and rates of convergence in the strong law of large numbers hold under the strong mixing condition (see McLeish [68], Reznik [84], Dehling and Philipp [21], Stoica [88], Brandley [8]). When these properties hold, they can be used to estimate the tail parameter in a weakly dependent process with Fréchet distributed marginals.

For the stationary sequence $\{\log X_t\}$,

$$\varrho(n) = \sup_j \varrho(\mathscr{F}_{-\infty}^j, \mathscr{F}_{j+n}^{\infty}) = \sup_{A \in \mathscr{F}_{-\infty}^j, B \in \mathscr{F}_{j+n}^{\infty}} |\,\mathrm{P}(A \cap B) - \mathrm{P}(A)\,\mathrm{P}(B)|.$$

Theorem 9. *The time series $\{X_t\}$ stated in Theorem 5 satisfies the strong mixing condition.*

The following lemma is needed to prove Theorem 9.

Lemma 10. *If $S \sim \mathcal{S}(\alpha)$, $\alpha \in (0,1)$, $\alpha \log(S)$ converges in distribution to* Gumbel$(0,1)$ *when $\alpha \to 0$ and*

$$\mathrm{P}(\alpha \log(S) \leq x) = \exp(-e^{-x}) + \exp(-x - e^{-x})\varepsilon_\alpha + O_p(\alpha),$$

where $\varepsilon_\alpha = \alpha(1 + x - \log \alpha)$.

This Lemma ensures that when $t \in \mathbb{Z}$, the observation X_t is Fréchet distributed.

Proof. Use the probability function of stable random variable S as shown in Equation (1.4), the distribution of $\alpha \log(S)$ is

$$\mathrm{P}(\alpha \log(S) \leq x) = \mathrm{P}(S \leq e^{x/\alpha}) = \frac{1}{\pi} \int_0^\pi \exp\left(-b_\alpha(\theta) e^{\frac{-x}{1-\alpha}}\right) d\theta, \tag{2.7}$$

with

$$b_\alpha(\theta) = \frac{\sin((1-\alpha)\theta)}{\sin\theta} \left(\frac{\sin(\alpha\theta)}{\sin\theta}\right)^{\alpha/(1-\alpha)}.$$

First, we show the general statement of the Lemma. For any $\theta \in (0, \pi)$

$$\begin{aligned}
\lim_{\alpha\to 0} \alpha \log \sin(\alpha\theta) &= \lim_{\alpha\to 0} \frac{\log \sin(\alpha\theta)}{1/\alpha} \\
&= \lim_{\alpha\to 0} \frac{\theta \cot(\alpha\theta)}{-\frac{1}{\alpha}^2} \\
&= -\theta \lim_{\alpha\to 0} \frac{2\alpha \cos(\alpha\theta) - \theta\alpha^2 \sin(\alpha\theta)}{\theta \cos(\alpha\theta)} \\
&= 0,
\end{aligned}$$

which means

$$\lim_{\alpha\to 0} [\sin(\alpha\theta)]^{\alpha/(1-\alpha)} = \lim_{\alpha\to 0} \exp\left(\frac{\alpha}{1-\alpha} \log \sin(\alpha\theta)\right) = 1.$$

Thus we have

$$\lim_{\alpha\to 0} b_\alpha(\theta) = \lim_{\alpha\to 0} \frac{\sin((1-\alpha)\theta)}{\sin\theta} \left(\frac{\sin(\alpha\theta)}{\sin\theta}\right)^{\alpha/(1-\alpha)} = 1$$

Because of Equation (2.7), this result shows that the distribution of $\alpha \log(S)$ goes to $e^{-e^{-x}}$ as $\alpha \to 0$.

To see the second part of the Lemma, for $0 \le t \le 1$,

$$\begin{aligned} & \exp\left(-b_\alpha(\theta) e^{-\frac{x}{1-\alpha}}\right) \\ = \; & \exp(-e^{-x}) + \alpha\,(1 + x - \log\alpha)\exp(-x - e^{-x}) \\ & -\frac{1}{2}\left(e^x\,(x - \log\alpha)^2 - (1 + x - \log\alpha)^2\right)\exp(-2x - e^{-x})\alpha^2 + O_p(\alpha\theta^2). \end{aligned}$$

From Equation (2.7),we have that

$$\begin{aligned} & \mathrm{P}(\alpha \log(S) \le x) \\ = \; & \exp(-e^{-x}) + e^{-x-e^{-x}}\,(1 + x - \log\alpha)\,\alpha \\ & -\frac{1}{2} e^{-2x-e^{-x}}\left(e^x\,(x - \log\alpha)^2 - (1 + x - \log\alpha)^2\right)\alpha^2 + O_p(\alpha) \\ = \; & \exp(-e^{-x}) + \exp(-x - e^{-x})\varepsilon_\alpha + O_p(\alpha) \end{aligned}$$

with $\varepsilon_\alpha = (1 + x - \log\alpha)\,\alpha$. □

Now we show the sequence $\{X_t\}$ satisfies the strong mixing condition.

Proof of Theorem 9. For any $x, y > 0$, integer j,

$$\begin{aligned} \varrho(n) \;&\le\; |\,\mathrm{P}(X_j \le x, X_{j+n} \le y) - \mathrm{P}(X_j \le x)\,\mathrm{P}(X_{j+n} \le y)| \\ &=\; |\,\mathrm{P}(X_j \le x, X_{j+n} \le y) - e^{-x^{-\gamma} - y^{-\gamma}}|. \end{aligned}$$

With Lemma 8 and Lemma 10, when $n \to \infty$,

$$\begin{aligned} & \mathrm{P}(X_j \le x, X_{j+n} \le y) \\ = \; & P\left(X_j \le x, X_j^{\alpha^n}(S_{j+1}\cdots S_{j+n}^{\alpha^{1-n}})^{\alpha^n/\gamma} \le y\right) \\ = \; & \int_0^x P\left(\alpha^n \log(S_{j+1}\cdots S_{j+n}^{\alpha^{1-n}}) \le \gamma(\log y - \alpha^n \log X_j)\right) de^{-X_j^{-\gamma}} \\ = \; & \int_0^x e^{-e^{-\gamma(\log y - \alpha^n \log X_j)}} de^{-X_j^{-\gamma}} + O_p(n\alpha^n) \\ = \; & \int_0^x e^{-y^{-\gamma}}\left(1 - \gamma\alpha^n y^{-\gamma}\log X_j + O_p(\alpha^{2n})\right) de^{-X_j^{-\gamma}} + O_p(n\alpha^n) \\ = \; & e^{-x^{-\gamma}-y^{-\gamma}} + O_p(n\alpha^n). \end{aligned}$$

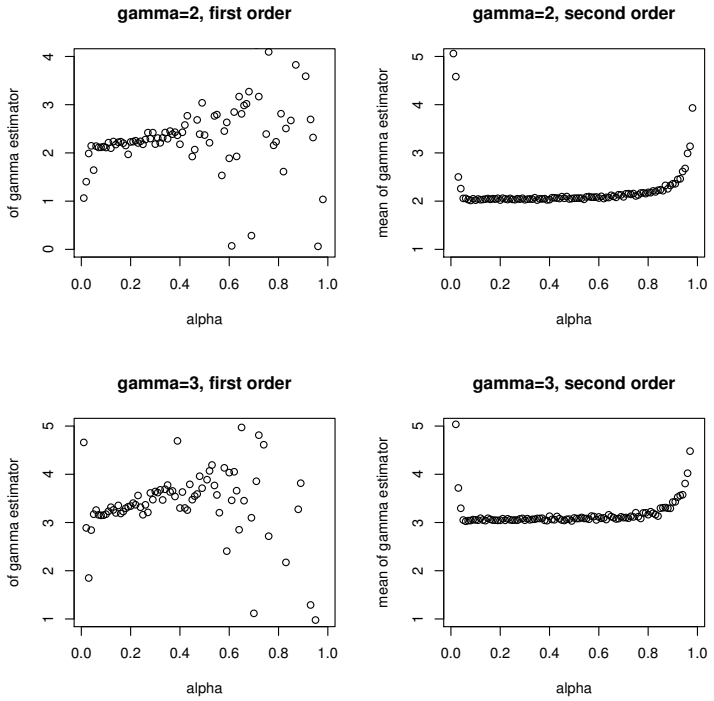

Figure 2.6: Moment estimation of γ, n=100.

The last step holds because

$$\int_0^x (\log X_j)^k de^{-X_j^{-\gamma}} = \gamma^{-k} \int_{-\infty}^{\gamma x} y^k de^{-e^{-y}},$$

which is less than the k-th moment of a Gumbel random variable, which is finite. Thus

$$\varrho(n) = O_p(n\alpha^n),$$

which means $\{X_t\}$ satisfies the strong mixing condition.

□

The observation sequence in the model (1.9) satisfies the weak mixing condition and the strong mixing condition. Thus the average of the observation $\overline{\log X}$ goes to a normal distribution with mean $\frac{\gamma_e}{\gamma}$ and variance $\frac{\pi^2}{6}\frac{1}{\gamma^2}\frac{n-2\alpha-n\alpha^2+2\alpha^{n+1}}{n^2(1-\alpha)^2}$, as our previous calculation for the covariance shows. Moment estimation can be used to obtain $\hat{\gamma}$.

Figure 2.6 is the simulation results of γ estimates by the method of moments from

the observation $\{\log X_t\}$ when $n = 100$. The first column is the estimates of $\hat{\gamma}_1$ using the first order moments, the second column pictures the plots of estimates $\hat{\gamma}_2$ using the second order moments,

$$\hat{\gamma}_1 = \frac{n\gamma_e}{\sum_{1\le i\le n} \log X_i}, \qquad \hat{\gamma}_2 = \sqrt{\frac{\pi^2(n-1)}{6\sum_{i=1}^{n}(\log X_i - \overline{\log X})^2}}.$$

$\hat{\gamma}_2$ is as same as the γ estimator stated in Section 2.1.

We can see that $\hat{\gamma}_2$ works much better than $\hat{\gamma}_1$ from Figure 2.6.

As discussed before, when α is small, takes α value close to 0, it is difficult to obtain an efficient estimate, because of the small non-zero chain size and the weak dependence structure.

When α goes to 1, α^n does not converge to 0. The condition of Lemma 10 is no longer satisfied, thus the strong mixing condition does not hold. In fact, with $\alpha \to 1$, the model (1.9) $X_t = X_{t-1}^{\alpha} S_t^{\alpha/\gamma}$ reduces to $X_t \approx X_{t-1}$ because $S_t \sim \mathcal{S}(\alpha)$ becomes degenerate (the Laplace transform goes to e^{-t}), which makes the effective sample size close to 1, regardless of n.

2.5 Extremal index

In 1988, Leadbetter and Rootzén [63] studied the asymptotic distribution of extreme values for a wide class of dependent stochastic sequences. The extreme values are considered as an over exceedance point process and the definition of extremal index, used to describe the dependence structure of a sequence, is defined as

Definition. For a stationary sequence $\{X_n\}$ with distribution F, and a constant sequence u_n satisfying $n(1 - F(u_n)) \to \tau$ with some $\tau > 0$, if

$$\mathrm{P}\left(\max_{1\le i\le n} X_i \le u_n\right) \to e^{-\theta\tau},$$

we say that the stationary sequence $\{X_n\}$ has extremal index $\theta, 0 \le \theta \le 1$.

Extremal index is not only used to define the asymptotic distribution of maximum, but also to find the asymptotic distribution of the other order statistics.

For an independent sequence, the extremal index, θ, is one.

If a dependent sequence has extremal index one, the asymptotic distributions of its order statistics are the same as the asymptotic distributions of the order statistics of an independent sequence from the same distribution.

Using the result in Equation (2.6), we have

$$\mathrm{P}\left(X_t \leq u_n | X_{t-1} \leq u_n\right) = 1 - u_n^{-\gamma}\theta_n,$$

where

$$\theta_n = \frac{\mathrm{P}(U > u_n^{-\gamma})}{e^{-u_n^{-\gamma}}},$$

with $U \sim \text{Gamma}(1-\alpha, 1)$ and the sequence $\{u_n\}$ is defined as in Equation (2.5). Thus

$$\begin{aligned}
\mathrm{P}(\max_{1\leq t\leq n}\{X_t\} \leq u_n) = \quad & \mathrm{P}(X_1 \leq u_n)\prod_{t=2}^{n} P(X_t \leq u_n | X_{t-1} \leq u_n) \\
\approx \quad & (1-\frac{\tau}{n})\left(1 - \frac{\tau}{n}\theta_n + O_p(\frac{1}{n^2})\right)^{n-1} \\
\approx \quad & e^{-\theta_n \tau},
\end{aligned}$$

when $0 < \alpha < 1$.

The extremal index of sequence $\{X_t\}$ is $\lim_{n\to\infty}\theta_n$. θ_n is a decreasing function of α. $\theta_n \to 1$ as $\alpha \to 0$ because U converges in distribution to Exponential(1) and $\theta_n \to 0$ as $\alpha \to 1$ since U converges in distribution to the Dirac delta distribution.

For any given $\alpha \in (0,1)$, $\theta_n \to 1$ when $n \to \infty$. A plot of θ_n as a function of α is shown in graph 2.7. Although, when α is large, the rate on which θ_n goes to 1 is much slower than that of when α is small.

Before considering the order statistics, notice that the tail of Fréchet$(0,1,\gamma)$ distribution has the property

$$\mathrm{P}(X_t > x) = 1 - e^{-x^{-\gamma}} = x^{-\gamma} + O_p(x^{-2\gamma})$$

as $x \to \infty$, which indicates that it is a heavy tail distribution. In fact,

$$\frac{\mathrm{P}(X_t > x+c)}{\mathrm{P}(X_t > x)} = \left(\frac{x}{x+c}\right)^{\gamma}\left(\frac{1 - \frac{1}{2}(x+c)^{-\gamma} + \frac{1}{6}(x+c)^{-2\gamma} + O_p((x+c)^{-3\gamma})}{1 - \frac{1}{2}x^{-\gamma} + \frac{1}{6}x^{-2\gamma} + O_p(x^{-3\gamma})}\right) \to 1$$

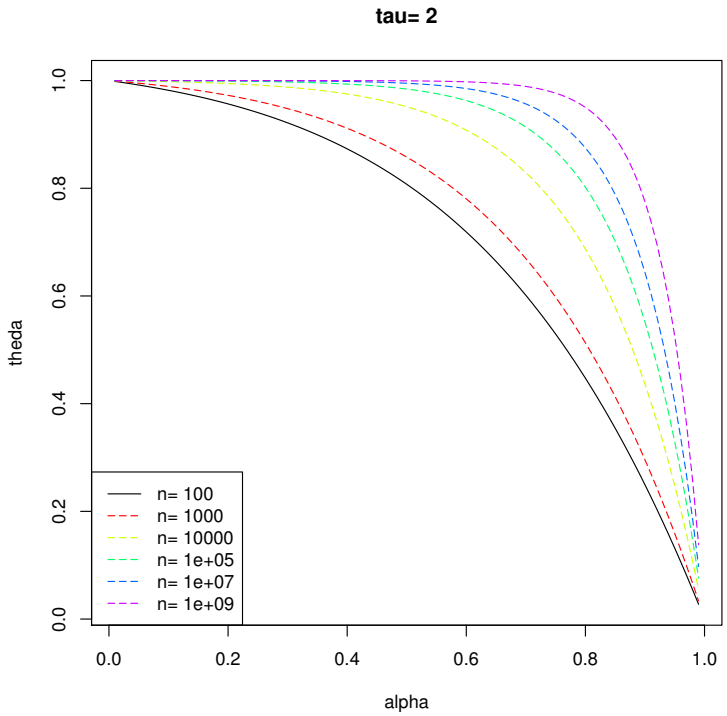

Figure 2.7: Extreme index of $\{X_t\}$ as α increases.

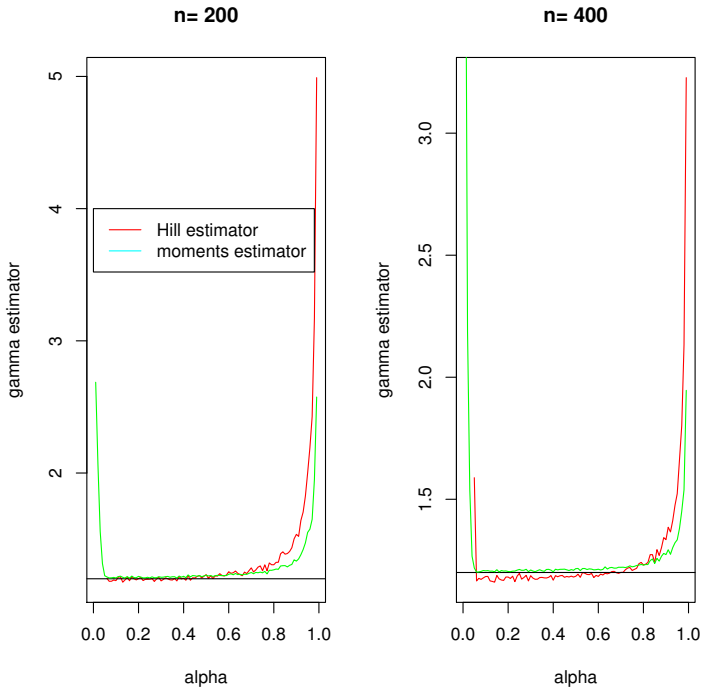

Figure 2.8: Comparison of γ estimates using Hill estimates and the moments estimates.

as $x \to \infty$, which means that Fréchet$(0, 1, \gamma)$ is long-tailed.

Hill estimation is commonly used for the heavy-tailed distributions. The asymptotic normality of Hill estimator is studied by researchers under the strong mixing condition, like in Ling and Peng [64]. Resnick and Stărică ([81], [82]) studied the behaviour of Hill estimator for the stationary sequence and in the AR model. Thus we can try to estimate γ using the Hill estimation.

Figure 2.8 are the comparison of Hill estimator and the moment estimator of γ. Hill estimator works, but behaves not as good as the moment estimator.

2.6 Recursive Hill estimation

For the independent error sequence, we already know that $S_t \sim \mathcal{S}(\alpha)$, $t = 1, 2, \ldots,$ is a sequence from a heavy tailed distribution.

Hill estimation, which depends on the order statistics, can be used here, at least when α is away from 1. It works efficiently for the heavy-tailed distributions. Since the stable distribution $\mathcal{S}(\alpha)$ has a heavy tail and in our model the errors are an independent sequence, it is natural to think of Hill estimation as a plausible option.

Before applying the Hill estimation, we compare the tail behaviour of stable distributions with a small parameter (α goes to zero) and a large parameter ($1-\alpha$ with α goes to 0). Denote $S' \sim \mathcal{S}(\alpha), S'' \sim \mathcal{S}(1-\alpha)$, then

$$\frac{\mathrm{P}(S' > x)}{\mathrm{P}(S'' > x)} \approx \frac{\Gamma(\alpha)}{\Gamma(1-\alpha)} x \quad \text{for large} \quad x. \tag{2.8}$$

This ratio goes to infinity as $\alpha \to 0$, which indicates that the tail of a stable distribution with the stability parameter goes to 1 is much thinner than the stable parameter goes to 0. If the Hill estimation is applied to the stable distribution, the results of observations generated by small stability parameter should be better than the results with large stability parameter. We reached a similar conclusion, when looking into θ_n, the extremal index as a function of α.

The problem here is that the Hill estimation is not applied to the error sequence, since the error sequence cannot be observed.

To obtain a satisfactory estimator of α and improve the estimator of γ, I proposed a recursive Hill estimation.

We write $S_t(\alpha, \gamma)$ as the unobserved sequence $X_t^{\frac{\gamma}{\alpha}} X_{t-1}^{-\gamma}$, i.e. $S_t(\alpha, \gamma) = X_t^{\frac{\gamma}{\alpha}} X_{t-1}^{-\gamma}$.

Denote $\hat{\gamma}_{[i]}$ as the estimate of γ at the i-th stage and by $\hat{\gamma}_{[1]}$, its initial value.

We begin with $i = 1$, apply the following steps recursively.

(a). For a given $\hat{\gamma}_{[i]}$, obtain $\alpha_{[i]}$ as

$$\hat{\alpha}_{[i]} = \arg\min_{0<\alpha<1} \left(\frac{1}{\alpha} - \frac{1}{k} \sum_{t=1}^{k} \log \hat{S}^{[i]}_{(n-t)}(\alpha) + \log \hat{S}^{[i]}_{(n-k-1)}(\alpha) \right)^2, \tag{2.9}$$

where $\{\hat{S}^{[i]}_{(t)}(\alpha)\}$ is denoted as the order statistics of the sequence $\{S_t(\alpha, \hat{\gamma}_{[i]})\}$ for all the time $t > 1$, and $\hat{S}^{[i]}_{(1)}(\alpha)$ is the minimum of $\{X_t^{\hat{\gamma}_{[i]}/\alpha} X_{t-1}^{-\hat{\gamma}_{[i]}}\}$ while $\hat{S}^{[i]}_{(n-1)}(\alpha)$ is the maximum.

The reason to do so is that by Hill estimation, the following approximate equation should hold for a reasonably large k,

$$\frac{1}{\alpha} \approx \frac{1}{k}\sum_{t=1}^{k} \log \hat{S}^{[i]}_{(n-t)}(\alpha) - \log \hat{S}^{[i]}_{(n-k-1)}(\alpha).$$

(b) Now we update the γ estimator using the $\hat{\alpha}_{[i]}$ obtained from Equation (2.9). For a reasonable number k', let

$$\hat{\gamma}_{[i+1]} = \hat{\alpha}^{-1}_{[i]} \left(\frac{1}{k'} \sum_{t=1}^{k'} \log \hat{U}^{[i]}_{(n-t)} - \log \hat{U}^{[i]}_{(n-k'-1)} \right)^{-1}, \tag{2.10}$$

where $\hat{U}^{[i]}_{(t)}$ is the i-th order statistic from the sequence $\{U_t(\hat{\alpha}_{[i]})\}$ and $U_t(\alpha) = X_t^{1/\alpha} X_{t-1}^{-1}$, which only depends on α, and has a heavy tail with index $\gamma\alpha$.

In this step, the recursive Hill estimation is applied again to update the tail parameter, since according to the Hill estimation, if $\hat{\alpha}_{[i]}$ is close to α, we have

$$\frac{1}{\alpha\gamma} \approx \frac{1}{k'}\sum_{t=1}^{k'} \log \hat{U}^{[i]}_{(n-t)} - \log \hat{U}^{[i]}_{(n-k'-1)}.$$

Repeat steps (a), (b) until little change for α, γ estimates are obtained.

Hill estimation works for sequence with heavy tail. When α is small, i.e. close to 0, the tail of a $\mathcal{S}(\alpha)$ random variable is long and easy to find. However, when α increases, the mode of the density of $\mathcal{S}(\alpha)$ moves to the right, the tail is thinner.

The simulation results of the recursive Hill estimation are shown in the Figure 2.9.

The estimation of γ when α is close to 1 seemed okay even when sample size is small (n=50). When α is small, recursive Hill estimation returns estimates within the parameter space, which is an improvement when comparing with the Yule-Walker estimation, which may yield negative estimators.

In step (a), the estimator of γ is used to update the α estimate. To see the effect of using $\hat{\gamma}$ instead of γ, we need to compare the distributions of the order statistics $\hat{S}_{(t)}(\alpha)$ generated by $\hat{\gamma}$.

Let $\hat{S}_t(\alpha) = X_t^{\frac{\hat{\gamma}}{\alpha}} X_{t-1}^{-\hat{\gamma}}$, where $\hat{\gamma}$ is the moment estimator of γ and $\hat{S}_{(t)}, 1 \le t \le n-1$ be the order statistics of $\hat{S}_t(\alpha), \hat{S}_{(1)} \le \hat{S}_{(2)} \le \cdots \le \hat{S}_{(n-1)}$ while $S_{(t)}, 1 \le t \le n-1$ are the order statistics of $S_t, 2 \le t \le n$.

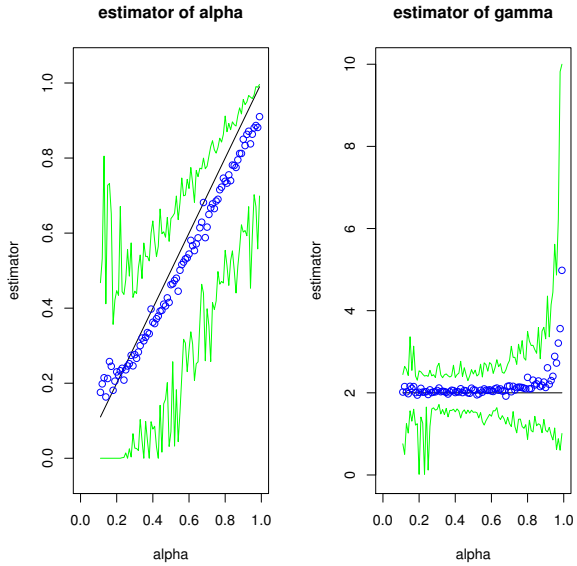

Figure 2.9: Recursive Hill estimates of α (left) and the estimates of γ (right) with n=50, means in blue dots, 95% confidence interval in green.

Denote the Hill estimator $\alpha_{[h]}$ and $\hat{\alpha}_{[h]}$ as

$$\frac{1}{\alpha_{[h]}} = \frac{1}{k}\sum_{i=1}^{k} \log \frac{S_{(n-i)}}{S_{(n-k-1)}}, \quad \frac{1}{\hat{\alpha}_{[h]}} = \frac{1}{k}\sum_{i=1}^{k} \log \frac{\hat{S}_{(n-i)}}{\hat{S}_{(n-k-1)}}.$$

We know that the Hill estimator has the asymptotic property of normality

$$\sqrt{k}(\alpha_{[h]} - \alpha) \to N(0, \alpha^2)$$

when $n \to \infty$. Now we would like to find the property of $\hat{\alpha} = \hat{\alpha}_1$ in Equation (2.9), using the moment estimator $\hat{\gamma}$.

Since $\hat{\alpha}$ depends on the order statistics of $\hat{S}_t(\alpha) = X_t^{\hat{\gamma}/\alpha} X_{t-1}^{-\hat{\gamma}}$ for $t > 1$, thus we have

$$\log \hat{S}_t(\alpha) - \log S_t = \frac{\hat{\gamma} - \gamma}{\alpha} \log X_t - (\hat{\gamma} - \gamma) \log X_{t-1} = \frac{\hat{\gamma} - \gamma}{\gamma} \log S_t,$$

which means $\log \hat{S}_t(\alpha) = \frac{\hat{\gamma}}{\gamma} \log S_t$ and we know that $\frac{\hat{\gamma}}{\gamma} \to N(1, \Omega_n)$ with $\Omega_n = \frac{11}{10n} \frac{1+\alpha^2}{1-\alpha^2}$ (Theorem 6).

We need to consider the distribution of $\log \hat{S}_t(\alpha)$ and the distribution of its order statistic. Since the Hill estimator uses the extremes beyond the threshold, when x is

large,

$$
\begin{aligned}
& \mathrm{P}(\log \hat{S}_t(\alpha) \le x) \\
= & \int_R \mathrm{P}\left(\left(1+\Omega_n^{1/2} w\right) \log S_t \le x\right) dF_w(w) \\
\approx & \int_R \left(1 - \frac{1}{\Gamma(1-\alpha)} \exp^{-\frac{\alpha x}{1+\Omega_n^{1/2} w}}\right) dF_w(w) \\
= & 1 - \frac{e^{-\alpha x}}{\Gamma(1-\alpha)} \int_R \left(1 + \alpha x \sqrt{\Omega_n} w + \left(\frac{\alpha^2 x^2}{2} - \alpha x\right) \Omega_n w^2 + O_p(\Omega_n^{3/2} w^3)\right) dF_w(w) \\
= & 1 - \frac{e^{-\alpha x}}{\Gamma(1-\alpha)} \left(1 + \alpha x \left(\frac{\alpha x}{2} - 1\right) \Omega_n + \alpha x \sqrt{\Omega_n} \left(2 - \frac{\alpha x}{2}\right) \frac{\phi(\frac{1}{\sqrt{\Omega_n}})}{\Phi(\frac{1}{\sqrt{\Omega_n}})} + O_p\left(\Omega_n^{3/2}\right)\right) \\
= & \left(1 - \frac{e^{-\alpha x}}{\Gamma(1-\alpha)}\right) \left(1 - \frac{\alpha x \left(\frac{\alpha x}{2} - 1\right) \Omega_n + \alpha x \sqrt{\Omega_n} \left(2 - \frac{\alpha x}{2}\right) \frac{\phi(\frac{1}{\sqrt{\Omega_n}})}{\Phi(\frac{1}{\sqrt{\Omega_n}})}}{\Gamma(1-\alpha) e^{\alpha x} - 1} + O_p\left(\Omega_n^{3/2}\right)\right) \\
\to & 1 - \frac{e^{-\alpha x}}{\Gamma(1-\alpha)},
\end{aligned}
$$

where $F_w(w) = \frac{\Phi(w) - \Phi(-1/\sqrt{\Omega_n})}{1 - \Phi(-1/\sqrt{\Omega_n})}$ is the truncated standard normal distribution, since the first integral in this calculation is convergent only in the region $w > -\frac{1}{\sqrt{\Omega_n}}$. Φ and ϕ is the distribution function and density function of a standard normal random variable.

This conclusion shows that the tail of $\hat{S}_t(\alpha)$ behaves the same as the tail of $\mathrm{S}(\alpha)$.

Lemma 11. *The tail of $\hat{S}_t(\alpha)$ converges in distribution to the tail of S_t for $1 \le t \le n$.*

To obtain the asymptotic density of the order statistic, we have

$$
\begin{aligned}
& \lim_{n\to\infty} \frac{\mathrm{P}^n(\log \hat{S}_t(\alpha) \le x)}{\mathrm{P}^n(\log S_t \le x)} \qquad (2.11) \\
\approx & \lim_{n\to\infty} \left(1 - \frac{\alpha x \left(\frac{\alpha x}{2} - 1\right) \Omega_n + \alpha x \sqrt{\Omega_n} \left(2 - \frac{\alpha x}{2}\right) \frac{\phi(\frac{1}{\sqrt{\Omega_n}})}{\Phi(\frac{1}{\sqrt{\Omega_n}})}}{\Gamma(1-\alpha) e^{\alpha x} - 1} + O_p\left(\Omega_n^{3/2}\right)\right)^n \\
= & \exp\left(-\frac{11}{20} \frac{1+\alpha^2}{1-\alpha^2} \frac{\alpha x(\alpha x - 2)}{e^{\alpha x}\Gamma(1-\alpha) - 1}\right),
\end{aligned}
$$

which is close to 1 when x is the 70% quantile of $\mathrm{S}(\alpha)$ (see Table 2.1) or larger (the

	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	$\alpha = 0.5$
Eq (2.11)	1.0000000	0.9999994	0.9415507	0.9472035	1.0693781
	$\alpha = 0.6$	$\alpha = 0.7$	$\alpha = 0.8$	$\alpha = 0.9$	
Eq (2.11)	1.1473225	1.1369867	1.0556348	1.0012682	

Table 2.1: The ratio in Equation (2.11) with x be the 70% quantile of $\mathrm{S}(\alpha)$.

ratio has maximum error 10^{-5} from 1 when x takes the value of 90% quantile of $\mathrm{S}(\alpha)$). We only use the extremes beyond the threshold to get α, thus if we define the tail reasonable, we could obtain good estimator of α.

Using $\mathrm{P}(\log \hat{S}_t(\alpha) \leq x)$, we can also obtain the ratio of the other order statistics of $\hat{S}_t(\alpha)$ and S_t, which is close to 1 when x is reasonably large.

2.7 Recursive Fan's estimation

So far, all the estimations are based on the logarithm of the observations, even the Hill estimation. After logarithm, some properties would be changed or lost. In this section, we do estimation with $\{X_i\}$ use the estimation proposed by Fan [28].

In Fan [28], the stability parameter is estimated using the property of strictly stable, i.e.

$$S_1 + S_2 \stackrel{d}{=} 2^{1/\alpha} S_1,$$

where the independent stable random variables $S_1, S_2 \sim S(\alpha)$. Thus

$$\hat{\alpha} = \frac{\log 2}{\mathrm{E}\left(\log(S_1 + S_2)\right) - \mathrm{E}(\log S_1)}.$$

The estimator obtained by Fan's method has been proved to be unbiased, consistent and approximate normal. Denote Fan's estimate as

$$\hat{\alpha}_F = \frac{n(n-1)}{2} \log 2 \left(\sum_{1 \leq i < j \leq n} \left(\log(S_i + S_j) - \frac{\log S_i + \log S_j}{2} \right) \right)^{-1}.$$

and $\sqrt{n}\left(\frac{1}{\hat{\alpha}_F} - \frac{1}{\alpha}\right)$ converges in distribution to a normal distribution with zero mean and variance, which is the sample variance of the sequence

$$\left\{ \frac{1}{n \log 2} \sum_{j \neq i} \left(\log(S_i + S_j) - \frac{\log S_i + \log S_j}{2} \right), 1 \leq i \leq n \right\}.$$

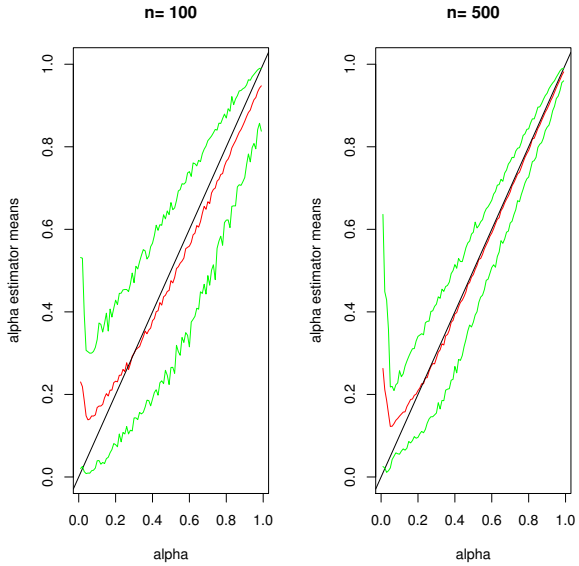

Figure 2.10: Estimators of α using Fan's estimation, with means in red, 95% confidence interval in green.

For our time series, the error sequence $\{S_t\}$ is unobserved. A modified recursive Fan's estimation can be applied to obtain the α estimate.

With initial estimators $\hat{\alpha}_s, \hat{\gamma}_s, s = 1$, obtain

$$\hat{S}_t = X_t^{\hat{\gamma}_s/\hat{\alpha}_s} X_{t-1}^{-\hat{\gamma}_s}$$

for $2 \leq t \leq n$. This dependent sequence can be used to update the stability parameter α, by applying the strictly stable property to $\hat{S}_i, \hat{S}_j$ which are not located next to each other. Let

$$\hat{\alpha}_{s+1} = \arg\min_{0<\alpha<1} \left| \frac{(n-2)(n-3)}{2} \frac{\log 2}{\alpha} - \sum_{2\leq i<j\leq n, j-i>1} \left(\log(\hat{S}_i + \hat{S}_j) - \frac{\log \hat{S}_i + \log \hat{S}_j}{2} \right) \right|.$$

And denote

$$\hat{S}_t^{1/\gamma} = X_t^{1/\hat{\alpha}_{s+1}} X_{t-1}^{-1},$$

update the estimator of γ using $\hat{\alpha}_{s+1}$.

The simulation results of chain size 100 and 500 are shown in Figure 2.10.

Use the moment estimator $\hat{\gamma}$ from $\{\log X_t\}$, as the initial estimator, in the recursive Fan's estimation, we have $\hat{S}_t = S_t^{\hat{\gamma}/\gamma}$. Denote $\hat{\gamma}/\gamma \to 1 + \sqrt{\Omega_n}Z$, where $Z \sim N(0,1)$, $\Omega_n = \frac{11}{10n}\frac{1+\alpha^2}{1-\alpha^2}$ as before (in Theorem 6).

Notice that

$$\hat{S}_t = S_t\left(1 + \sqrt{\Omega_n}Z\log S_t + O_p(\Omega_n)\right),$$

thus

$$\begin{aligned}
& \log\left(\hat{S}_i + \hat{S}_j\right) - \frac{\log\hat{S}_i + \log\hat{S}_j}{2} \\
= \quad & \log\left(\hat{S}_i + \hat{S}_j\right) - \log\left(S_i + S_j\right) + \log\left(S_i + S_j\right) - \frac{\log S_i + \log S_j}{2} \\
& + \frac{\log S_i + \log S_j}{2} - \frac{\log\hat{S}_i + \log\hat{S}_j}{2} \\
= \quad & \log\left(1 + \frac{S_i\log S_i + S_j\log S_j}{S_i + S_j}\sqrt{\Omega_n}Z + O_p(\sqrt{\Omega_n})\right) + \log\left(S_i + S_j\right) - \frac{\log S_i + \log S_j}{2} \\
& - \frac{1}{2}\log\left(1 + \sqrt{\Omega_n}Z\log S_i + O_p(\Omega_n)\right) - \frac{1}{2}\log\left(1 + \sqrt{\Omega_n}Z\log S_j + O_p(\Omega_n)\right) \\
= \quad & \frac{S_i\log S_i + S_j\log S_j}{S_i + S_j}\sqrt{\Omega_n}Z + \log\left(S_i + S_j\right) - \frac{\log S_i + \log S_j}{2} \\
& - \frac{\log S_i + \log S_j}{2}\sqrt{\Omega_n}Z + O_p(\Omega_n) \\
= \quad & \frac{(S_i - S_j)(\log S_i - \log S_j)}{2(S_i + S_j)}\sqrt{\Omega_n}Z + \log\left(S_i + S_j\right) - \frac{\log S_i + \log S_j}{2} + O_p(\Omega_n),
\end{aligned}$$

where $\Omega_n \to 0$ as $n \to \infty$.

$$\begin{aligned}
\mathrm{E}\left(\log(\hat{S}_i + \hat{S}_j) - \frac{\log\hat{S}_i + \log\hat{S}_j}{2}\right) &= \mathrm{E}\left(\log\left(S_i + S_j\right) - \frac{\log S_i + \log S_j}{2}\right) + O_p(\Omega_n) \\
&= \frac{\log 2}{\alpha} + O_p(\Omega_n).
\end{aligned}$$

Thus the convergence rate of recursive Fan's estimate is $O_p(1/n)$.

2.8 Estimation using Kantorovich-Wasserstein Metric

We know that the Laplace transform is an alternative way to characterize the distribution of a random variable. Also as mentioned before, by the Feller's Tauberian theorem we know of an important relationship between the behaviour of large values from α-stable distribution and its Laplace transform near the origin. With this in mind, here we propose the Laplace transform of $S_t^{\alpha/\gamma}$ to estimate the unknown parameters.

Since

$$X_t X_{t-1}^{-\alpha} = S_t^{\alpha/\gamma} \tag{2.12}$$

and $\mathrm{E}\left(e^{-uS_t^{\alpha/\gamma}}\right) \leq \infty$, we have

$$\omega(u) = \mathrm{E}\left(e^{-uX_t X_{t-1}^{-\alpha}}\right) = \mathrm{E}\left(e^{-uS_t^{\alpha/\gamma}}\right) \tag{2.13}$$

for all time t and $u > 0$. The Laplace transform of $S_t^{\alpha/\gamma}$, $\omega(u)$, depends on the integral

$$\int_0^\infty \cos(s) e^{-us^{\alpha/\gamma}} ds,$$

which needs to be calculated numerically.

The idea of the estimation is to minimize the distance between the empirical estimator of $\omega(u)$ and $\omega(u)$ itself, i.e.,

$$\hat{\alpha} = \underset{\alpha \in (0,1)}{\arg\min} \int |\hat{\omega}(u) - \omega(u)| du,$$

where $\hat{\omega}(u) = \frac{1}{n} \sum_{t=2}^{n} e^{-uX_t X_{t-1}^{-\alpha}}$.

To test how the Laplace transform works on estimating α, we simulated the time series with a known $\gamma = 2$ and a small chain size ($n = 50$), with the result shown in Figure 2.11. From the figure we can see this method has smaller bias when α is large (close to 1). As α increases, the dependence structure of observations is stronger and easier to be captured by the Equation (2.13).

To estimate α using the Laplace transform, we need a good estimator of γ first. We use the same idea, the Kantorovich-Wasserstein metric, to obtain the Fréchet shape parameter, knowing that because of the mixing conditions, the empirical distribution

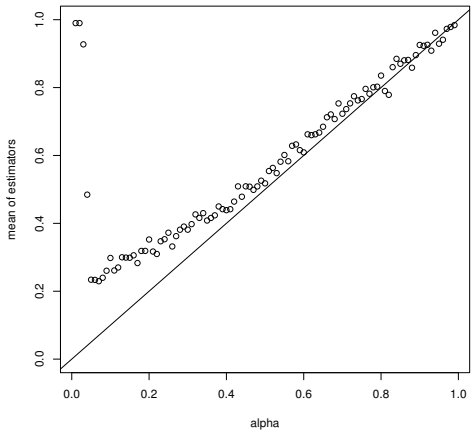

Figure 2.11: Averages of α estimates using Laplace transformation, $n = 50$.

of a sequence from the model (1.9) is asymptotically normal, i.e.,

$$\sqrt{n}\left(\frac{\sum_{i=1}^{n} \mathrm{I}(X_i \leq x)}{n} - \exp(-x^{-\gamma})\right) \to N\left(0, e^{-x^{-\gamma}}\left(1 - e^{-x^{-\gamma}}\right)\right),$$
$$\mathrm{P}\left(\sqrt{n}\left|\frac{\sum_{i=1}^{n} \mathrm{I}(X_i \leq x)}{n} - \exp(-x^{-\gamma})\right| > x\right) < 2e^{-2x^2},$$

where I() is the indicator function. Let

$$\hat{\gamma} \;=\; \arg\min_{\gamma>0} \int_0^\infty \left|\frac{\sum_{i=1}^{n} I(X_i \leq x)}{n} - \exp(-x^{-\gamma})\right| dx,$$

which ignores the dependence structure of the observation sequence, and use $\hat{\gamma}$ to obtain

$$\hat{\alpha} = \arg\min_{\alpha\in(0,1)} \int_0^\infty \left|\hat{\omega}(u) - \mathrm{E}(e^{-uS_t^{\alpha/\hat{\gamma}}})\right| du.$$

The simulation result of using Kantorovich-Wasserstein distance in estimation is shown in the Figure 2.12. The behaviour of γ estimates is better than what we had before. Surprisingly, even when α is large, which brings the stronger dependence between the observations, minimizing the distance between the empirical distribution and the marginal distribution gives the estimators with small bias compared with the other methods we tried. Besides this, when α is large, $\hat{\alpha}$ behaves good and it seems not to be affected by the differences between $\hat{\gamma}$ and γ.

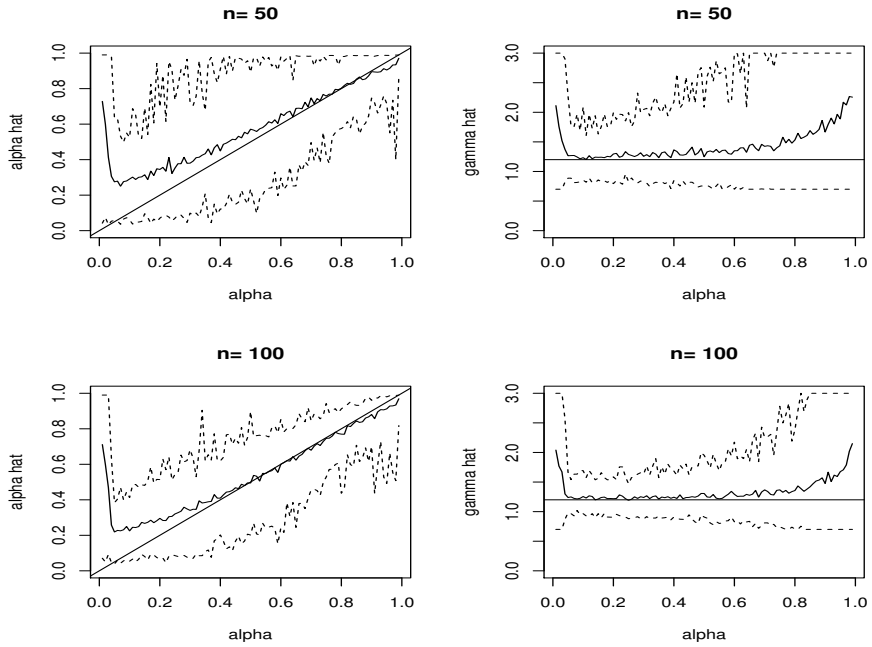

Figure 2.12: Means and 95% confidence interval of α estimates (first column) and the γ estimates using Kantorovich-Wasserstein distance (second column).

As the chain size increases, the behaviour of estimators obtained by Kantorovich-Wasserstein distance shows consistency (in Figure 2.13). With a large stability parameter (like $\alpha = 0.8$ in the third row), to obtain a good estimator of γ, larger sample size is needed when comparing with the estimation in the time series with a small stability parameter. When α is large, the dependence structure should be considered to estimated γ.

To get better estimates, expressed in another way, to update the estimates, a similar method is applied repeatedly with the new estimators, either to update the γ estimator by applying the Kantorovich-Wasserstein distance to both sides of the Equation (2.12), i.e.,

$$\hat{\gamma}_{u_1} = \arg\min_{\gamma>0} \int_0^\infty \left| \frac{\sum_{i=2}^{n} I(X_i X_{i-1}^{-\hat{\alpha}} \leq x)}{n-1} - F_{\hat{\alpha}}(x^{\gamma/\hat{\alpha}}) \right| dx,$$

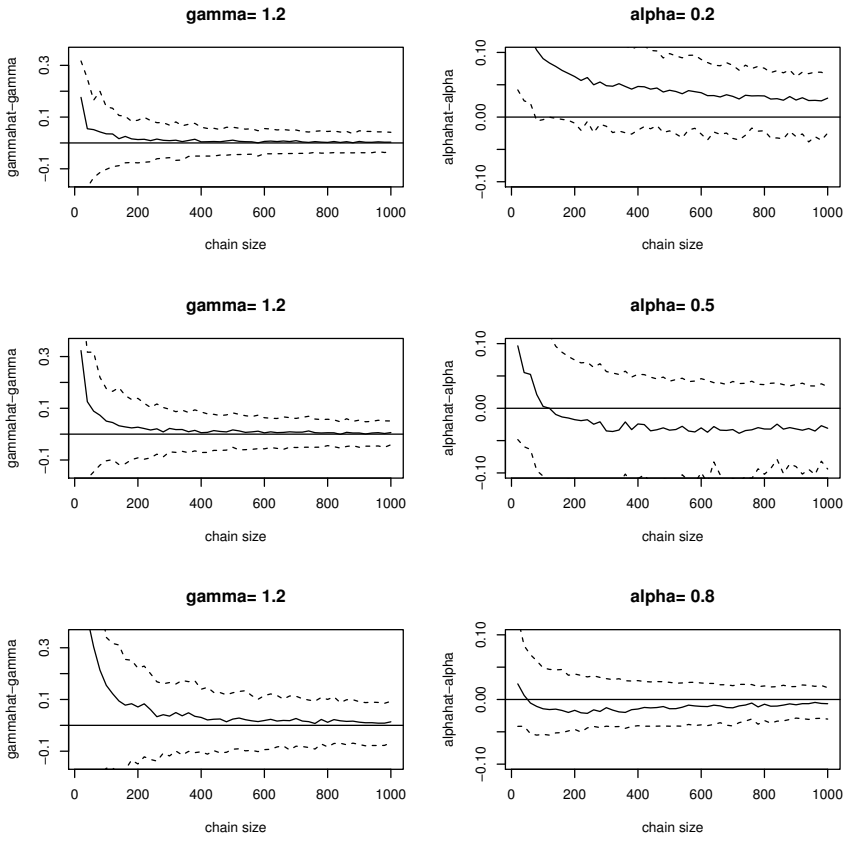

Figure 2.13: Means and 95% confidence interval of γ estimates (first column), α estimates (second column) using the Kantorovich-Wasserstein distance.

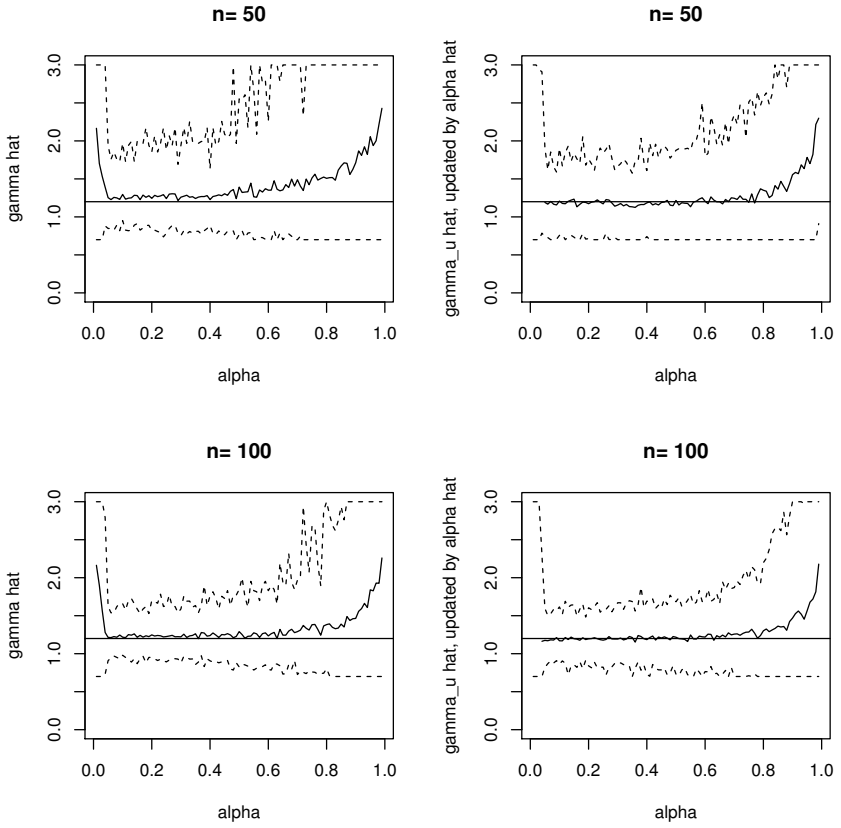

Figure 2.14: Means and 95% confidence interval of $\hat{\gamma}$(first column) and $\hat{\gamma}_{u_1}$(second column).

or use the maximum likelihood estimation with the new $\hat{\alpha}$, which is

$$\begin{aligned}\hat{\gamma}_{u_2} &= \arg\max_{\gamma>0} \prod_{i=2}^{n} f_{S_i^{1/\gamma}}\left(X_i^{1/\hat{\alpha}} X_{i-1}^{-1}\right) \\ &= \arg\max_{\gamma>0} \left((n-1)\log(\gamma) + (r-1)\sum_{i=2}^{n} \log(X_i X_{i-1}^{-\hat{\alpha}}) + \sum_{i=2}^{n} \log f_{\hat{\alpha}}(X_i^{\gamma} X_{i-1}^{-\hat{\alpha}\gamma}) \right).\end{aligned}$$

Figure 2.14 shows the comparison of $\hat{\gamma}$ and $\hat{\gamma}_{u_1}$ when $n = 50, 100$. As α increases, some improvement of $\hat{\gamma}_{u_1}$ can be seen, since $\hat{\gamma}_{u_1}$ depends on $\hat{\alpha}$, considers the dependence structure.

To obtain the convergence rate of $\frac{\hat{\gamma}}{\gamma} - 1$, the convergence rate of $\frac{\sum_{i=1}^{n} \mathrm{I}(X_i \le x)}{n} - \exp(-x^{-\gamma})$ is considered first. Use the Law of the iterated logarithm, Hartman and Wintner [44] showed that for an i.i.d. sequence, if $S_n - S_{n-1}$ has zero mean and finite

variance δ^2, S_n denote the partial sum of the sequence, then

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\delta^2 \log\log n}} = 1 \qquad \text{a.s.}$$

R. Davis [20] extended the result to the stationary sequence under some conditions. When $0 < \gamma < 2$ and for some $k = o_p(n), k \to \infty$, if

$$\limsup_{n\to\infty} n \sum_{t=1}^{[n/k]-1} \mathrm{P}\left(X_1 > u_n, X_{1+t} > u_n\right) = o_p(1),$$

then the convergence rate of the dependent sequence partial sum is the same as that of an dependent sequence. Since

$$\begin{aligned}
\mathrm{P}\left(X_1 > u_n, X_{1+t} > u_n\right) &\approx \frac{u_n^{-\gamma}}{\Gamma(1-\alpha^t)} \int_0^{u_n^{-\gamma}} x^{-\alpha^t} de^{-x}, \\
n \sum_{t=1}^{[n/k]-1} \mathrm{P}\left(X_1 > u_n, X_{1+t} > u_n\right) &\approx n u_n^{-\gamma} \sum_{t=1}^{[n/k]-1} \frac{1}{\Gamma(1-\alpha^t)} \int_0^{u_n^{-\gamma}} x^{-\alpha^t} de^{-x} \\
&\approx \tau \sum_{t=1}^{[n/k]-1} \frac{u_n^{-\gamma(1-\alpha^t)}}{\Gamma(2-\alpha^t)} \\
&< \frac{\tau^2}{n} \sum_{t=1}^{[n/k]-1} \left(\frac{\tau}{n}\right)^{-\alpha^t} \\
&\to 0.
\end{aligned}$$

Applying this result we have that

$$\begin{aligned}
\left| \frac{\sum_{i=1}^n \mathrm{I}(X_i \le x)}{n} - \exp(-x^{-\gamma}) \right| &= \sqrt{\frac{2e^{-x^{-\gamma}}(1-e^{-x^{-\gamma}}) \log\log n}{n}} \\
&= O_p\left(\sqrt{\frac{\log\log n}{n}}\right) \quad \text{a.s.}
\end{aligned}$$

Thus, we can expect that

$$\limsup_{n\to\infty} \left|\exp(-x^{-\hat{\gamma}}) - \exp(-x^{-\gamma})\right| = O_p\left(\sqrt{\frac{\log\log n}{n}}\right) \qquad \text{a.s.}$$

On the other hand, by Taylor expansion, we have

$$\begin{aligned}\exp(-x^{-\hat{\gamma}}) - \exp(-x^{-\gamma}) &= -x^{-\gamma}\exp(-x^{-\gamma})(\gamma-\hat{\gamma})\log x + O_p((\gamma-\hat{\gamma})^2) \\ &= O_p\left(|\gamma-\hat{\gamma}|\right),\end{aligned}$$

i.e., $|\gamma-\hat{\gamma}| = O_p\left(\sqrt{\frac{\log\log n}{n}}\right)$ a.s.

Thus we can write

$$\frac{1}{\hat{\gamma}} = \frac{1}{\gamma} - \frac{1}{\gamma^2}(\hat{\gamma}-\gamma) + O_p\left(\hat{\gamma}-\gamma\right)^2),$$

which implies

$$\begin{aligned}S_t^{\alpha/\hat{\gamma}} &= S_t^{\alpha/\gamma - \frac{\alpha}{\gamma^2}(\hat{\gamma}-\gamma)+O_p\left((\hat{\gamma}-\gamma)^2\right)} \\ &= S_t^{\alpha/\gamma}\left(1+(\hat{\gamma}-\gamma)\log S_t^{-\frac{\alpha}{\gamma^2}} + O_p\left((\hat{\gamma}-\gamma)^2\right)\right), \\ e^{-uS_t^{\alpha/\hat{\gamma}}} &= e^{-uS_t^{\alpha/\gamma}}\left(1-u(\hat{\gamma}-\gamma)S_t^{\alpha/\gamma}\log S_t^{-\frac{\alpha}{\gamma^2}} + O_p\left((\hat{\gamma}-\gamma)^2\right)\right).\end{aligned}$$

So

$$\mathrm{E}\left(e^{-uX_tX_{t-1}^{-\alpha}}\right) - \mathrm{E}\left(e^{-uS_t^{\alpha/\hat{\gamma}}}\right) = u\,\mathrm{E}\left((\hat{\gamma}-\gamma)S_t^{\alpha/\gamma}e^{-uS_t^{\alpha/\gamma}}\log S_t^{-\frac{\alpha}{\gamma^2}}\right) + O_p\left(\mathrm{E}(\hat{\gamma}-\gamma)^2\right).$$

Chapter 3

State space model with GEV distributed marginals and α-stable distributed errors

In this chapter, we consider the state space model with GEV distributed marginals and α-stable distributed errors

$$\begin{cases} Y_t = \mu - \frac{\sigma}{\gamma} + \frac{\sigma}{\gamma} e^{\psi\gamma X_t} \xi_t^{\psi\gamma}, & \text{(observation equation)} \\ X_{t+1} = \alpha X_t + \alpha \log S_{t+1}, & \text{(state equation)} \end{cases} \tag{3.1}$$

where $\{S_t\}$ and $\{\xi_t\}$ are independent error sequences following α-stable distributions, $S_t \sim \mathcal{S}(\alpha)$ and $\xi_t \sim \mathcal{S}(\psi)$ with $\alpha, \psi \in (0,1)$, for $t \geq 1, \gamma \neq 0$.

If X_0 follows standard Gumbel distribution, i.e. Gumbel(0,1), then the marginal distribution of X_t is the standard Gumbel distribution, and the marginal distribution of Y_t follows GEV(μ, σ, γ) for all $t \geq 1$.

The state space model with Gumbel distributed marginals and exponential α-stable distributed errors is

$$\begin{cases} Y_t' = \mu + \sigma\psi X_t + \sigma\psi \log \xi_t, & \text{(observation equation)} \\ X_{t+1} = \alpha X_t + \alpha \log S_{t+1}, & \text{(state equation)} \end{cases}$$

where $S_t \sim \mathcal{S}(\alpha)$, $\xi_t \sim \mathcal{S}(\psi)$, are independent α-stable distributed errors. The observation $Y_t' \sim$ Gumbel(μ, σ) for $t \in \mathbb{Z}$.

3.1 Dependence of transformed observation sequence

First, we have a look into the dependence of observation sequence $\{Y_t, t \geq 1\}$ in model (3.1). For simplicity, denote

$$\begin{cases} Z_t = \frac{1}{\gamma} \log \frac{Y_t - \mu + \sigma/\gamma}{\sigma/\gamma}, & \text{when } \gamma \neq 0 \\ Z_t = \frac{Y_t - \mu}{\sigma}, & \text{when } \gamma = 0 \end{cases}$$

so that we have $Z_t \sim$Gumbel(0,1).

Our main estimation interests are the GEV shape parameter γ, and the stability parameters α and ψ. In this chapter we will discuss the state space model with $\mu = 0$, $\sigma = 1$, i.e.

$$\mathrm{P}(Y_t \leq y) = e^{-(1+\gamma y)^{-1/\gamma}}, \ 1 + \gamma y > 0.$$

We check the dependent structure of $\{Z_t\}$ instead of $\{Y_t\}$. If $\{Z_t\}$ satisfies the strong mixing condition, then $\{Y_t\}$ satisfies the weak or strong mixing conditions.

Now we have the state space model with standard Gumbel distributed marginals

$$\begin{cases} Z_t = \psi X_t + \psi \log \xi_t, & \text{(observation equation)} \\ X_{t+1} = \alpha X_t + \alpha \log S_{t+1}. & \text{(state equation)} \end{cases} \tag{3.2}$$

This state space model can be rewritten as

$$\begin{aligned} Z_t &= \alpha\psi X_{t-1} + \alpha\psi \log S_t + \psi \log \xi_t \\ &= \alpha Z_{t-1} + \psi \log \xi_t - \alpha\psi \log \xi_{t-1} + \alpha\psi \log S_t \qquad (3.3) \\ &= \alpha^k \psi X_{t-k} + \psi \log \xi_t + \sum_{i=1}^{k} \alpha^i \psi \log S_{t+1-i}. \\ &= \alpha^k Z_{t-k} + \psi \log \xi_t - \alpha^k \psi \log \xi_{t-k} + \sum_{i=1}^{k} \alpha^i \psi \log S_{t+1-i}. \qquad (3.4) \end{aligned}$$

The sequence $\{Z_t, t \geq 1\}$ is a stationary sequence with covariance

$$\mathrm{Cov}(Z_t, Z_{t-k}) = \frac{\pi^2}{6} \alpha^k \psi^2, \quad 1 \leq k. \tag{3.5}$$

To check whether $\{Z_t\}$ satisfies the strong mixing condition, we calculate the difference between $\mathrm{P}(Z_t \leq x_0, Z_{t+n} \leq x_1)$ and $\mathrm{P}(Z_t \leq x_0)\,\mathrm{P}(Z_{t+n} \leq x_1)$.

Lemma 12. *For any $x_0, x_1 \in \mathbb{R}$, when $n \to \infty$,*

$$|\,\mathrm{P}(Z_t \le x_0, Z_{t+n} \le x_1) - \mathrm{P}(Z_t \le x_0)\,\mathrm{P}(Z_{t+n} \le x_1)| = O_p(n\alpha^n).$$

Proof. Use the Equation (3.4), we have

$$Z_{t+n} = \alpha^n Z_t + \psi \log \xi_{t+n} - \alpha^n \psi \log \xi_t + \sum_{i=1}^{n} \alpha^i \psi \log S_{t+n+1-i},$$

where $\psi \log \xi_{t+n} + \sum_{i=1}^{n} \alpha^i \psi \log S_{t+n+1-i} = \alpha^n \psi \log S_n^*$ with $S_n^* \sim \mathcal{S}(\alpha^n \psi)$. Thus

$$\begin{aligned}
& \mathrm{P}(Z_t \le x_0, Z_{t+n} \le x_1) \\
= \quad & \int_{-\infty}^{x_0} \mathrm{P}(Z_{t+n} \le x_1 | Z_t) d e^{-e^{-Z_t}} \\
= \quad & \int_{-\infty}^{x_0} \mathrm{P}(\alpha^n \psi \log S_n^* - \alpha^n \psi \log \xi_t \le x_1 - \alpha^n Z_t | Z_t) d e^{-e^{-Z_t}} \\
= \quad & \int_{-\infty}^{x_0} \int_{0}^{\infty} \mathrm{P}(\alpha^n \psi \log S_n^* \le x_1 - \alpha^n Z_t + \alpha^n \psi \log \xi_t | Z_t, \xi_t) d F_\psi(\xi_t) d e^{-e^{-Z_t}}.
\end{aligned}$$

Denote $C_{Z_t} = e^{-x_1 + \alpha^n Z_t}$, use Lemma 10 in Chapter 2.4, we have

$$\begin{aligned}
& \mathrm{P}(Z_t \le x_0, Z_{t+n} \le x_1) \\
= \quad & \int_{-\infty}^{x_0} \int_{0}^{\infty} \exp(-C_{Z_t} e^{-\alpha^n \psi \log \xi_t}) d F_\psi(\xi_t) d e^{-e^{-Z_t}} \\
& + \alpha^n \psi \int_{-\infty}^{x_0} \int_{0}^{\infty} C_{Z_t} \xi_t^{-\alpha^n \psi} e^{-C_{Z_t} \xi_t^{-\alpha^n \psi}} \\
& \qquad \times (1 + x_1 - \alpha^n Z_t + \alpha^n \psi \log \xi_t - \log(\alpha^n \psi))\, d F_\psi(\xi_t) d e^{-e^{-Z_t}} \\
& + O_p(n\alpha^n) \\
= \quad & A_n + \alpha^n \psi B_n + O_p(n\alpha^n).
\end{aligned}$$

Using Taylor's expansion,

$$\begin{aligned}
A_n &= \int_{-\infty}^{x_0}\int_0^{\infty} \exp\left(-C_{Z_t}e^{-\alpha^n\psi\log\xi_t}\right) dF_\psi(\xi_t)de^{-e^{-Z_t}} \\
&= \int_{-\infty}^{x_0}\int_0^{\infty} e^{-C_{Z_t}}\left(1+\alpha^n\psi C_{Z_t}\log\xi_t + O_p(\alpha^{2n}\psi^2\log^2\xi_t(C_{Z_t}+C_{Z_t}^2)\right) dF_\psi(\xi_t)de^{-e^{-Z_t}} \\
&= \int_{-\infty}^{x_0} e^{-C_{Z_t}}\left(1+\alpha^n\psi C_{Z_t}\,\mathrm{E}(\log\xi_t) + O_p\left(\alpha^{2n}\psi^2(C_{Z_t}^2+C_{Z_t})\,\mathrm{E}(\log^2\xi_t)\right)\right) de^{-e^{-Z_t}}.
\end{aligned}$$

Since

$$\begin{aligned}
&\int_{-\infty}^{x_0} e^{-C_{Z_t}}de^{-e^{-Z_t}} \\
=\;& e^{-e^{-x_1}}\int_{-\infty}^{x_0}\left(1-\alpha^n e^{-x_1}Z_t + O_p(\alpha^{2n}e^{-2x_1}Z_t^2)\right)de^{-e^{-Z_t}} \\
=\;& e^{-e^{-x_1}}\left(e^{-e^{-x_0}} - \alpha^n e^{-x_1}\int_{-\infty}^{x_0} Z_t de^{-e^{-Z_t}} + O_p\left(\alpha^{2n}\int_{-\infty}^{x_0} Z_t^2 de^{-e^{-Z_t}}\right)\right) \\
=\;& e^{-e^{-x_1}-e^{-x_0}} + O_p(\alpha^n),
\end{aligned}$$

for $0 < \int_{-\infty}^{x_0} Z_t de^{-e^{-Z_t}} < \infty$, and $\int_{-\infty}^{x_0} C_{Z_t}e^{-C_{Z_t}}de^{-e^{-Z_t}} = e^{-x_1-e^{-x_1}}[e^{-e^{-x_0}} + O_p(\alpha^n)]$ is also finite, together with the fact that the moments of log α-stable distribution are finite, we have

$$A_n = e^{-e^{-x_1}-e^{-x_0}} + O_p(\alpha^n).$$

Similarly, we can obtain that

$$B_n = (1+x_1)\int_{-\infty}^{x_0} C_{Z_t}e^{-C_{Z_t}}de^{-e^{Z_t}} + O_p(n),$$

thus

$$|\,\mathrm{P}(Z_t \le x, Z_{t+n} \le y) - \mathrm{P}(Z_t \le x)\,\mathrm{P}(Z_{t+n} \le y)| = O_p(n\alpha^n).$$

□

When $0 < \alpha < 1$, the sequence $\{Z_t\}$ satisfies the strong mixing condition since $n\alpha^n \to 0$ as $n \to \infty$. However this mixing condition no longer holds when $\alpha \to 1$. Furthermore, when α is large, say close to 1, the large value of n makes $O_p(n\alpha^n)$ non-negligible. This is something that needs to be considered when using the result.

3.2 Estimation of γ

In the estimation of the time series discussed in Chapter 2, γ can be estimated using moment estimation. While in the estimation of the state space model with GEV distributed marginals, moment estimation is not applicable since the mean and variance of GEV random variable are finite only under certain conditions ($\gamma < 0.5$).

Instead of moment estimation, maximum likelihood estimation, Hill estimation (when a heavy tail is detected) and the Kantorovich-Wasserstein distance can be used to estimate γ, since the strong mixing condition is satisfied.

Figure 3.1 compares the results of three estimations, maximum likelihood estimation, Hill estimation and Kantorovich-Wasserstein distance, to obtain γ estimates, ignoring the dependence structure of the observations which are from the state space model with GEV(0, 1, $\gamma = 1.2$) marginals. The first two rows show γ estimates when α takes the value of 0.2, 0.4, 0.6, 0.8, respectively. The last two rows show the results of γ estimates when ψ takes the value of 0.2, 0.4, 0.6, 0.8, respectively. The chain size here is 100. From Figure 3.1 we can see that when both of the stability parameters are large, the estimation ignoring the dependence is not good. If the stability parameters are not large (not close to 1) and not in the neighbourhood of 0, the γ estimates ignoring the dependence works because in such a case the dependence is very weak to be captured by a small sized observation sequence.

3.2.1 Regression estimation of GEV parameters

Here we will consider the estimation of the GEV parameters in the state space model with observations $\{Y_t, 1 \leq t \leq n\}$, which have marginal distribution GEV(μ, σ, γ).

Since

$$Z_t = \frac{1}{\gamma} \log \frac{Y_t - \mu + \sigma/\gamma}{\sigma/\gamma}$$

and $\{Z_t\}$ satisfies the strong mixing condition, thus $\{Y_t\}$ satisfies the strong mixing condition.

To find a good estimation method, the tail behaviour should be studied first. If $\{Y_t\}$ does not have a heavy tail, casual estimation technique can be used; otherwise alternative estimation procedures for heavy tailed distributions like the method using Kantorovich-Wasserstein metric and a regression method presented in the next section.

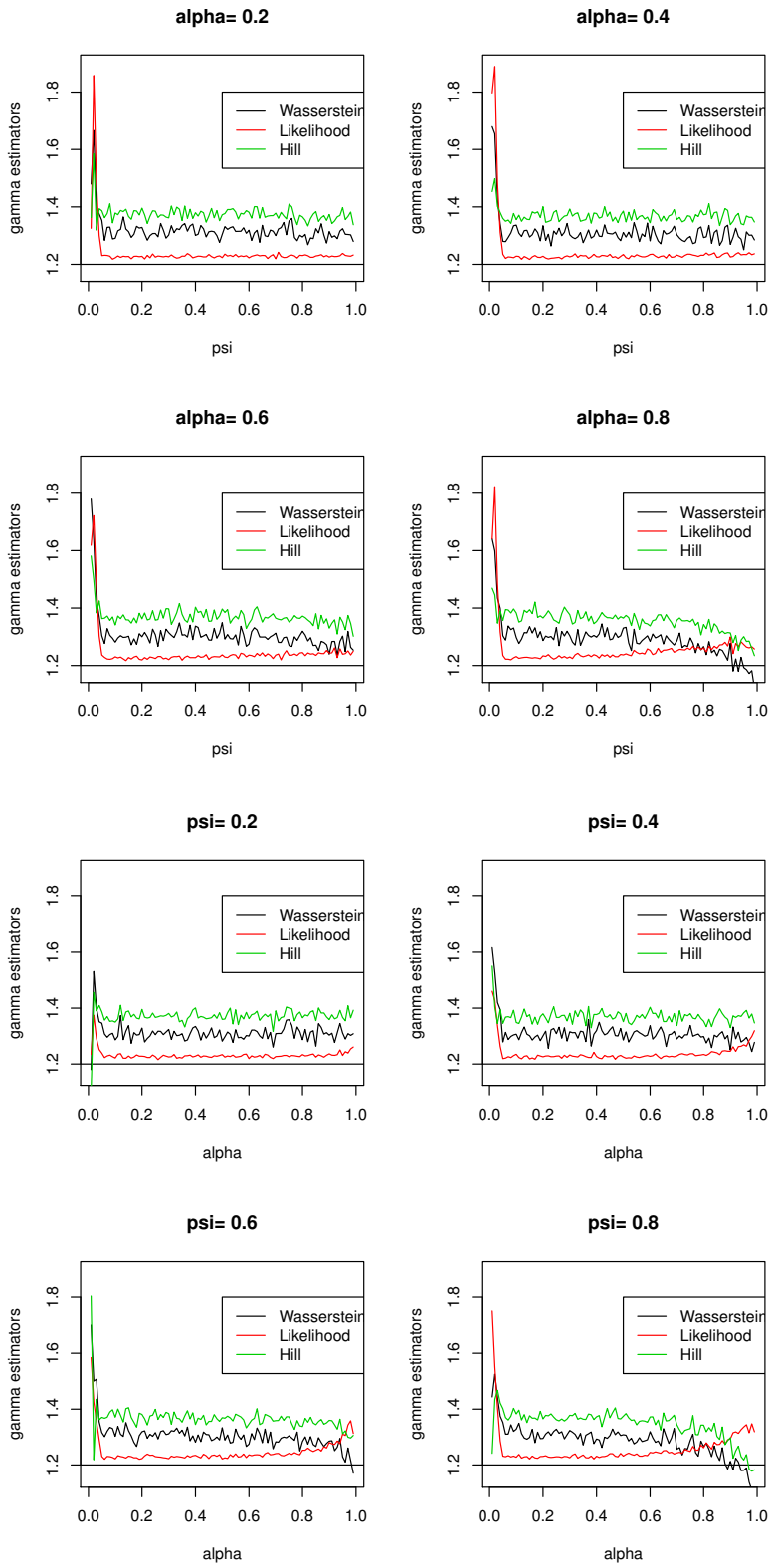

Figure 3.1: Estimates of γ ignoring the dependence, with the chain size n=100, $\gamma = 1.2$.

3.2.2 Estimation of γ for independent GEV distributed observations

As we mentioned before, the strong mixing condition assumes many properties for a stationary sequence, like the usual limit theories and principles like the Central Limit Theorem, weak invariance principle, laws of the iterated logarithm and almost sure invariance principles, as well as the rates of convergence in the strong law of large numbers. In this part, we consider the estimation of γ for independent GEV observations first.

For an independent sequence $\{Y_t\}$, with marginal GEV(μ, σ, γ) distribution and $\gamma < -1$ (heavy-tailed), we could use the order statistics in a regression model to estimate all the GEV parameters. This procedure goes as follows. Let $Y_{(1)} \leq Y_{(2)} \leq \ldots Y_{(n)}$ be the order statistics of this independent GEV(μ, σ, γ) sequence.

Theorem 13. *For an independent sequence $\{Y_t, 1 \leq t \leq n\}$, with marginal distribution GEV(μ, σ, γ) and a heavy tail, the points*

$$(x_i, y_i) = \left(\log\left(-\log\left(\frac{i}{n+1} \right) \right), \log\left(Y_{(n)} - Y_{(i)} \right) \right), 1 \leq i \leq n-1,$$

are located on the line

$$y_i = -\gamma x_i - \log\left(\frac{-\gamma}{\sigma} \right) + \eta_{n,i}, \quad 1 \leq i \leq n-1 \tag{3.6}$$

with

$$\begin{aligned} \mathrm{E}(\eta_{n,i}) = \quad & -\frac{E_n}{E_i} + \frac{\partial F^{-1}\left(\frac{i}{n+1}\right) \partial F^{-1}(\frac{n}{n+1})}{2E_i^2} \frac{i}{(n+1)^2(n+2)} \\ & -\frac{E_n(\partial F^{-1}(\frac{n}{n+1}))^2}{E_i^3} \frac{in+i-i^2}{(n+1)^2(n+2)} + o_p\left(\frac{1}{n+2} \right), \end{aligned}$$

$$
\begin{aligned}
\mathrm{Var}(\eta_{n,i}) = \quad & \frac{(\partial F^{-1}(\frac{n}{n+1}))^2}{E_i^2}\frac{n}{(n+1)^2(n+2)} \\
& + \left(\frac{\gamma}{\frac{i}{n+1}\log\frac{i}{n+1}} - \frac{E_n \partial F^{-1}\left(\frac{i}{n+1}\right)}{E_i}\right)^2 \frac{in+i-i^2}{(n+1)^2(n+2)} \\
& + 2\frac{\partial F^{-1}(\frac{n}{n+1})}{E_i}\left(\frac{\gamma}{\frac{i}{n+1}\log\frac{i}{n+1}} - \frac{E_n \partial F^{-1}\left(\frac{i}{n+1}\right)}{E_i}\right)\frac{i}{(n+1)^2(n+2)} \\
& + o_p\left(\frac{1}{n+2}\right),
\end{aligned}
$$

$$
\begin{aligned}
& \mathrm{Cov}(\eta_{n,i}, \eta_{n,j}) = \\
& \left(\frac{\gamma}{\frac{i}{n+1}\log\frac{i}{n+1}} - \frac{E_n \partial F^{-1}(\frac{j}{n+1})}{E_i}\right)\left(\frac{\gamma}{\frac{j}{n+1}\log\frac{j}{n+1}} - \frac{E_n \partial F^{-1}\left(\frac{i}{n+1}\right)}{E_j}\right)\frac{in+i-ij}{(n+1)^2(n+2)} \\
& + \frac{\partial F^{-1}(\frac{n}{n+1})}{E_j}\left(\frac{\gamma}{\frac{i}{n+1}\log\frac{i}{n+1}} - \frac{E_n \partial F^{-1}\left(\frac{i}{n+1}\right)}{E_i}\right)\frac{in+i-i^2}{(n+1)^2(n+2)} + \\
& \frac{\partial F^{-1}(\frac{n}{n+1})}{E_i}\left(\frac{\gamma}{\frac{j}{n+1}\log\frac{j}{n+1}} - \frac{E_n \partial F^{-1}(\frac{j}{n+1})}{E_j}\right)\frac{jn+j-j^2}{(n+1)^2(n+2)} \\
& + \frac{\partial F^{-1}(\frac{n}{n+1})}{E_i}\frac{\partial F^{-1}(\frac{n}{n+1})}{E_j}\frac{n}{(n+1)^2(n+2)} + o_p\left(\frac{1}{n+2}\right),
\end{aligned}
$$

for $i < j$ and $1 \le i, j \le n$, where F is the distribution function of GEV(μ, σ, γ) and

$$
\begin{aligned}
E_i = \quad & \mathrm{E}\left(\mu - \frac{\sigma}{\gamma} - Y_{(i)}\right) \\
= \quad & -\frac{\sigma}{\gamma}\left(\log\frac{n+1}{i}\right)^{-\gamma} - \frac{\sigma}{2}\frac{(r+1)(\log\frac{n+1}{i})^{-\gamma-2} - \log(\frac{n+1}{i})^{-\gamma-1}}{i}\frac{n+1-i}{n+2} \\
& + o_p\left(\frac{in+i-i^2}{(n+1)^2(n+2)}\right).
\end{aligned}
$$

.

For the regular linear regression, error term has constant mean and variance. Strictly speaking, $\eta_{n,i}$ here is not an error term. The mean and variance of $\eta_{n,i}$ depend on not only the GEV parameters, but also the order i. Under this circumstance, recursive regression, instead of the general linear regression, is used to obtain the estimators.

The derivation of Theorem 13 is presented in Appendix A.

n=100	mean of $\hat{\gamma}$	variance of $\hat{\gamma}$
$\gamma = -1.5$	-1.499622	0.02049879
$\gamma = -2$	-1.967157	0.03893676
$\gamma = -2.5$	-2.458674	0.05371618
$\gamma = -3$	-2.936059	0.08363773
n=30	mean of $\hat{\gamma}$	variance of $\hat{\gamma}$
$\gamma = -1.5$	-1.499851	0.06325039
$\gamma = -2$	-1.967599	0.111452
$\gamma = -2.5$	-2.416473	0.1700445
$\gamma = -3$	-2.864989	0.2330282
n=10	mean of $\hat{\gamma}$	variance of $\hat{\gamma}$
$\gamma = -1.5$	-1.572666	0.2224078
$\gamma = -2$	-1.991599	0.3382196
$\gamma = -2.5$	-2.413584	0.4942586
$\gamma = -3$	-2.812775	0.6781313

Table 3.1: Simulation result of γ with different sample sizes.

3.2.3 Simulation results for the state space model

Table 3.1 shows the means and variances of 1000 repeated estimates of γ with the chain size $n = 100$, 30, 10 separately. Recursive regression estimation works well even when n is small. Its γ estimates are close to the true value under the condition that the chain size downs to 10.

Remark 1. *When n is small, like $n \leq 30$, $\frac{i}{n}$ is a better estimator for empirical distribution function of the i-th order statistic than $\frac{i}{n+1}$.*

Remark 2. *For an independent* Fréchet(μ, σ, γ) *sequence, which is* GEV$(\mu + \sigma, \frac{\sigma}{\gamma}, \frac{1}{\gamma})$ *distributed, the points*

$$(x_i, y_i) = \left(\log\left(-\log\left(\frac{i}{n+1} \right) \right), \log\left(Y_{(i)} - Y_{(1)} \right) \right), \quad 2 \leq i \leq n,$$

are located on the line

$$y_i = -\frac{1}{\gamma} x_i + \log \sigma.$$

The location parameter μ can be estimated by the extreme $Y_{(1)}$.

In the estimation described in Theorem 13, μ and σ are estimated recursive. The estimator of μ is affected by $\hat{\gamma}$, since the tail of the distribution affects the efficiency

of the estimator of μ. Simulation result for estimates of σ, μ are presented in Table 3.2, 3.3 respectively.

$n = 100$	mean of $\hat{\sigma}$	variance of $\hat{\sigma}$
$\sigma = 1$	1.04072	0.05266952
$\sigma = 2$	2.085132	0.2130584
$\sigma = 3$	3.0128904	0.5176407
n=30	mean of $\hat{\sigma}$	variance of $\hat{\sigma}$
$\sigma = 1$	1.16218	0.1923123
$\sigma = 2$	2.307019	0.7524202
$\sigma = 3$	3.435808	1.889715

Table 3.2: Estimates of the scale parameter σ with different sample sizes.

n=100	mean of $\hat{\mu}$	variance of $\hat{\mu}$
GEV(0,1,-1.5)	-0.09072988	0.04996528
GEV(2,3,-2)	2.008903	0.2102406
GEV(5,4,-1)	4.949279	0.1819464
n=30	mean of $\hat{\mu}$	variance of $\hat{\mu}$
GEV(0,1,-1.5)	-0.02423031	0.01314015
GEV(2,3,-2)	1.6987	0.617337
GEV(5,4,-1)	4.760389	0.6830012

Table 3.3: Estimates of the location parameter μ with different sample sizes.

3.3 Estimation of stability parameters

Compared with the estimation of GEV parameters, the estimation of the stability parameter is more difficult. From Figure 3.1, we can see that the change of the stability parameters rarely affects the model, unless both stability parameters are large, close to 1. All the estimation methods related to the likelihood (maximum likelihood, iterated filtering, et al.) are not applicable here, due to the same argument discussed in Chapter 2. Besides, we only have the information of one observed sequence, but need to estimate two stability parameters. Here, we assume that one stability parameter is known and estimate the other unknown stability parameter first. For simplicity, we consider the state space model with Gumbel distributed marginals $\{Z_t\}$ in this section, since the GEV distributed observations can be transformed to Gumbel using the estimators of the GEV parameters.

3.3.1 One stability parameter is known

Estimation of ψ

Suppose that α is known. The covariance structure of the observations can be used to estimate ψ. Figure 3.2 contains the results of using the first order sample covariance of $\{Z_t\}$,

$$r(1) = \mathrm{Cov}(Z_t, Z_{t-1}) = \alpha\psi^2\frac{\pi^2}{6}, \quad (\text{“covariance1” in Figure 3.2})$$

and the first order covariance of sequence $\{Z_t - \alpha Z_{t-1}\}$,

$$\mathrm{Cov}(Z_t - \alpha Z_{t-1}, Z_{t-1} - \alpha Z_{t-2}) = -\alpha(1-\psi^2)\frac{\pi^2}{6}, \quad (\text{“covariance2” in Figure 3.2})$$

to estimate ψ.

From these plots, we can see that the estimations are not good when α is small. The plausible explanation is that when α is small, the information of $\{Z_t\}$ regarding the dependence structure is weak.

A numerical method, performing like filtering, is applied here to estimate ψ and the hidden states when α is known.

First, particles at time t are generated by the equation

$$\hat{X}_{t,i} = \alpha\hat{X}_{t-1,i} + \alpha\log\hat{S}_{t,i}, \quad 1 \le i \le m,$$

with the corresponding weight $f_{\psi_0}(Z_t|\hat{X}_{t,i})$, where ψ_0 is the Yule-Walker estimator of ψ. If the moment estimator of ψ is less than 0 or greater than 1, let $\psi_0 = 0.01$ or 0.99 respectively. The mode of the particles at time t is used to estimate the hidden state X_t. The estimator of ψ can be obtained by applying the linear regression between the observation Z_t and the estimator of X_t. Of Course, this two stage estimation procedure may have reliability issues that we do not discuss in this document.

The simulation results of this numerical method is shown in Figure 3.3.

Estimation of α

On the other hand, if ψ is known, there are other ways, besides using the moments, to estimate α. Here, we applied regression and ARMA model to the sequence $\left\{\frac{Z_t}{\psi}\right\}$

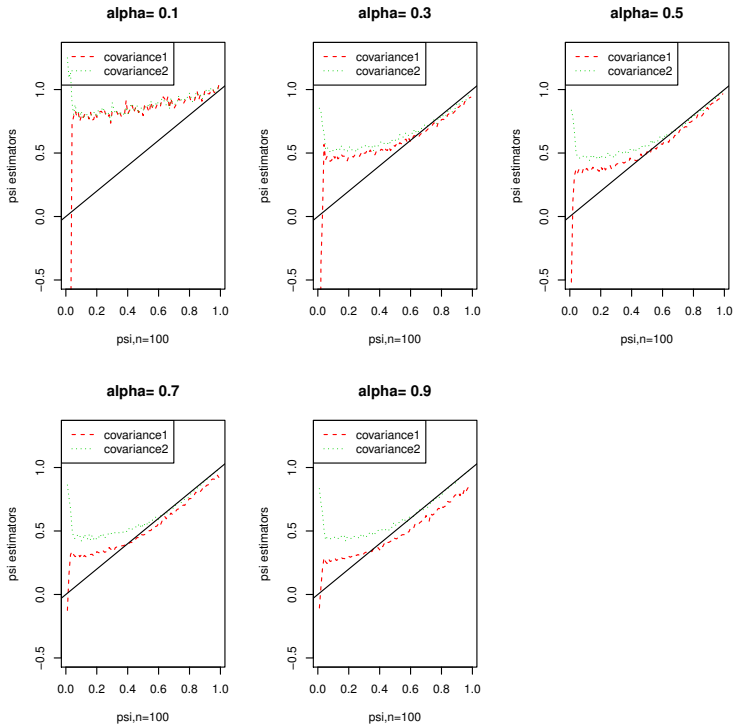

Figure 3.2: The ψ estimates obtained using the first order covariance when α is known.

(see Figure 3.4) since

$$\frac{Z_t}{\psi} - \alpha\frac{Z_{t-1}}{\psi} = \alpha \log S_t + \log \xi_t - \alpha \log \xi_{t-1}.$$

When ψ is large, α estimates obtained by using the moments (denoted as the "covariance" in the plots) behave better than using ARMA and regression model. When ψ is small, none of those methods works. An interesting observation is that when α is close to 1, all these estimation methods underestimate α.

The estimation of α is not straightforward, even when ψ is known. Notice that the covariance of the observations in equation (3.5) and $\mathrm{E}(Z_t - \alpha Z_{t-1}) = \gamma_e(1-\alpha)$ only depend on α, thus the estimation of α is important to the study of our state space model.

Usually, likelihood estimation and the generalized linear regression can be used to estimate the unknown parameter. Here the maximum likelihood estimation does not work for our model, due to the same reason we discussed in Chapter 2. Numerical results are used to approximate the likelihood to avoid the propensity. However, the

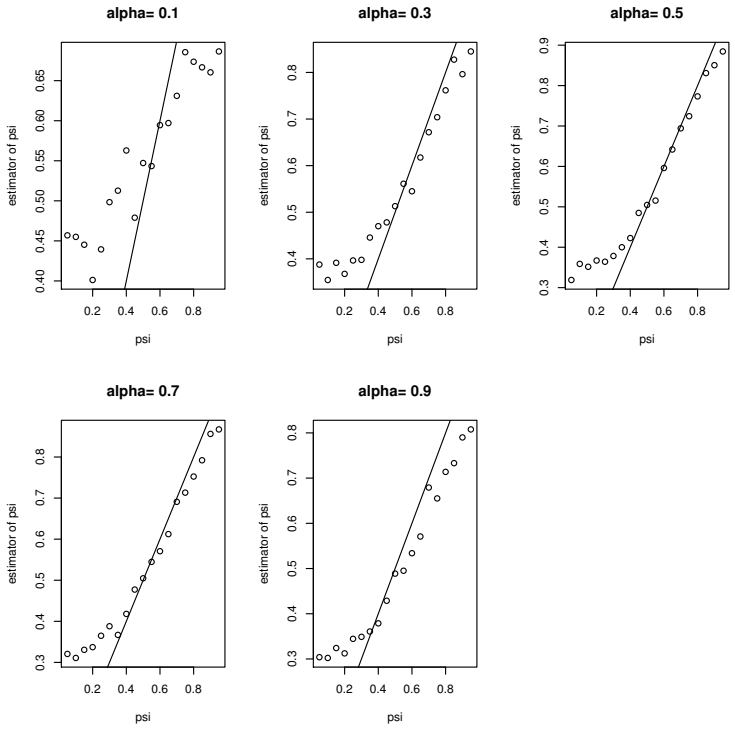

Figure 3.3: The ψ estimates obtained by filtering when α is known.

simulation result of the approximated likelihood estimation is erratic. Thus we tried to estimate α in a numerical way.

When ψ is known, we can generate a random sample from $\mathcal{S}(\psi)$, assuming it is the error sequence used in the model. Now the challenge is to decide which one in the generated $\mathcal{S}(\psi)$ sequence, denoted as $\{\tilde{\xi}_s, 1 \leq s \leq n\}$, is the closest to ξ_t for all the time points t, $1 \leq t \leq n$.

This idea is similar to the particle filter. For a given time t, assign a reasonable weight w_s to this generated sequence, let $\hat{\xi}_t = \tilde{\xi}_k$, where w_k is the maximum of the weight sequence.

The differences between this method and the particle filter is that once a generated random variable $\tilde{\xi}_k$ is picked to be the "estimator" of ξ_t, it should be removed from the generated error sequence. The rest random variables $\tilde{\xi}_s$ are used to estimate the other states to make sure the sequence of $\{\xi_t\}$ estimates does fairly represent a sample generated from $\mathcal{S}(\psi)$.

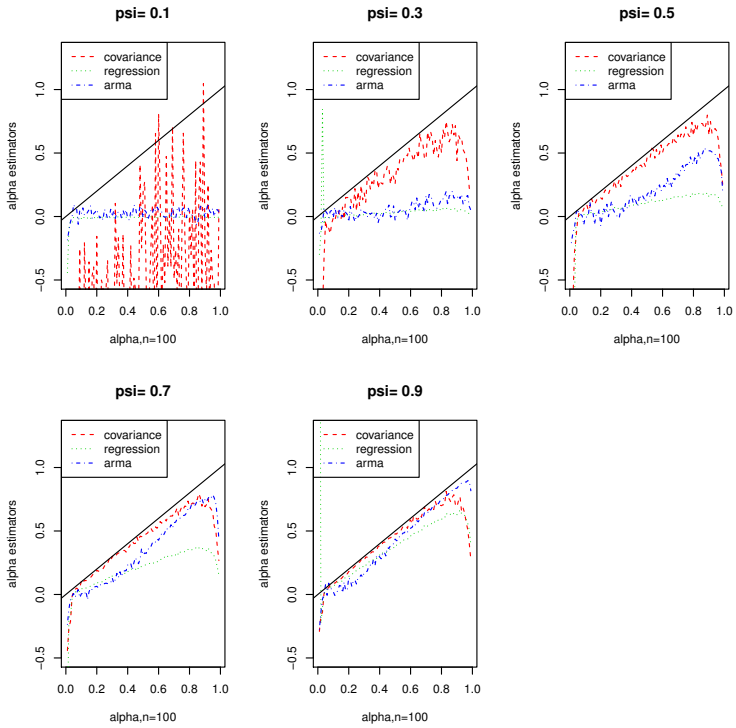

Figure 3.4: The α estimates when ψ is known.

The weight function w_s is important here. The conditional density

$$f(\xi_t|Z_t) \propto f_\psi(\xi_t) f(Z_t|\xi_t) \propto \xi_t f_\psi(\xi_t) e^{-\xi_t e^{-Z_t/\psi}}$$

is a reasonable choice for the weight function.

In addition to this conditional density, we also consider ordering stable sample due to the order of $\{Z_t\}$, since the value of the conditional probability $\mathrm{P}(\varepsilon_t > \varepsilon_{t-1}|Z_t > Z_{t-1})$, which is

$$\frac{\int_0^\infty \int_{\varepsilon_{t-1}}^\infty \left(1 - F_\alpha\left(\varepsilon_{t-1}\varepsilon_t^{-1/\alpha} e^{\frac{1-\alpha}{\alpha\psi} Z_{t-1}}\right) f_\psi(\varepsilon_t) f_\psi(\varepsilon_{t-1})\right) d\varepsilon_t d\varepsilon_{t-1}}{\int_0^\infty \int_0^\infty \left(1 - F_\alpha\left(\varepsilon_{t-1}\varepsilon_t^{-1/\alpha} e^{\frac{1-\alpha}{\alpha\psi} Z_{t-1}}\right) f_\psi(\varepsilon_t) f_\psi(\varepsilon_{t-1})\right) d\varepsilon_t d\varepsilon_{t-1}},$$

is almost 1 when α, ψ are greater than 0.5, and close to 1 for small $\alpha, \psi \in (0, 1)$ in numerical calculation.

Thus the ordered stable sample is assigned to the time t according to the order of $\{Z_t\}$, which is, if for a given time t, $Z_t = Z_{(s)}$, then $\hat{\xi}_t = \tilde{\xi}_{(s)}$ where $\tilde{\xi}_{(s)}$ is the order statistics of the generated $\mathcal{S}(\psi)$ random sample.

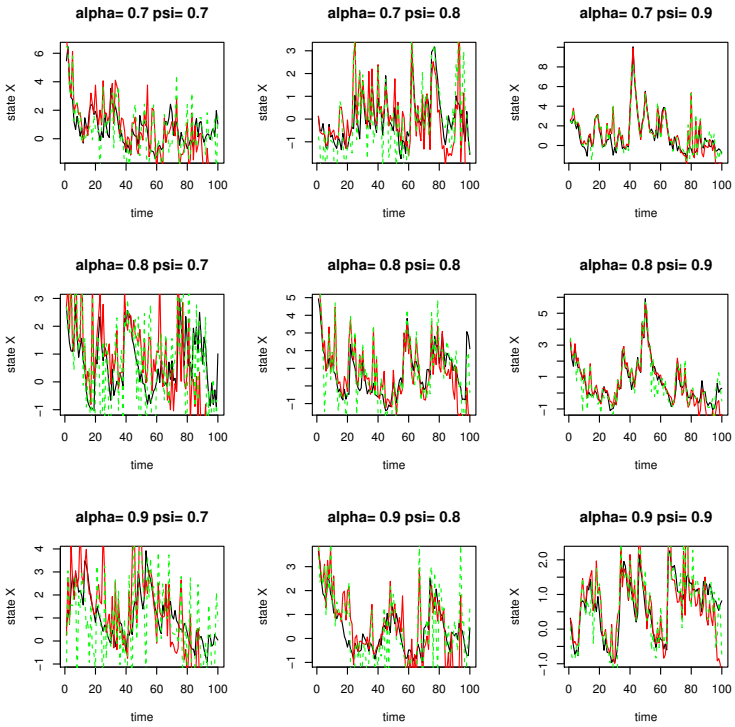

Figure 3.5: Estimates of $\{X_t\}$ when ψ is known using different weight functions (red curve is the estimators using $f(\xi_t|Z_t)$ while the green is using the order of observations).

The simulation results of the hidden state estimates are shown in Figure 3.5. The true values of the state variables are lined in black, while their estimators, obtained by using the weight $f(\xi_t|Z_t)$, are in red and the estimators, obtained by using the order of observations, are in green.

With the estimators of the hidden states, estimator of α can be obtained by the methods discussed in Chapter 2.

This numerical method is based on the supposition that the set of generated α-stable random variables is close enough to the set composed by the errors in the model. This is a reasonable presumption except when ψ is small, in which case $\{\tilde{\xi}_t\}$ may differ from $\{\xi_t\}$ due to the heavy tail of $\mathrm{S}(\psi)$. So in such a case, this procedure could return the state estimates with large, unreasonable errors.

When the stability parameter is large, this numerical method works well (in Figure 3.5), for the tail is thinner than the tail of a stable distribution with a small stability parameter.

The stability estimates here are not as good as the results showed in Chapter 2,

since the model depends on more unknown parameters but contains the same amount of information (only the observations).

If both of the stability parameters are unknown, commonly used estimations like the moment estimation, the regression model are applied. Besides these methods, some numerical methods like the approximate Bayesian computation, forward backward algorithm were used. However these methods take long computation time and produce unstable results, thus their simulation results are not presented here.

3.3.2 Moment estimation

Here we focus on the state space model with the standardized Gumbel distributed marginals $Z_t, 1 \leq t \leq n$.

$$\begin{cases} Z_t = \psi X_t + \psi \log \xi_t, \\ X_{t+1} = \alpha X_t + \alpha \log S_{t+1}. \end{cases}$$

Equation (3.5) shows the relation between covariance and the stability parameters. To distinguish the variance, covariance and the sample variance, sample covariance, we use $\mathrm{var}(\cdot), \mathrm{cov}(\cdot,\cdot)$ to denote the sample variance and covariance.

Estimators of α, ψ can be obtained by

$$\hat{\alpha} = \frac{\mathrm{cov}(Z_t, Z_{t-2})}{\mathrm{cov}(Z_t, Z_{t-1})}, \qquad \hat{\psi}^2 = \frac{\mathrm{cov}(Z_t, Z_{t-1})}{\frac{\pi^2}{6}\alpha} = \frac{6}{\pi^2}\frac{\mathrm{cov}^2(Z_t, Z_{t-1})}{\mathrm{cov}(Z_t, Z_{t-2})}.$$

The asymptotic normality holds for the covariance (shown in Section 3.3.4), but no longer holds for $\hat{\alpha}, \hat{\psi}$. The distribution of $\hat{\alpha}, \hat{\psi}$ can be derived from the asymptotic distribution of these covariance.

The problem here is that the value of covariance $\mathrm{Cov}(Z_t, Z_{t-k})$ depends on the power of stability parameters $\alpha^k\psi$, which is smaller than 1. However the sample covariance $\mathrm{cov}(Z_t, Z_{t-k})$ could be close to 0 or even negative. In the simulation, it is difficult to obtain acceptable estimators by using the covariance, especially when the chain size n is small.

Figure 3.6 shows the estimates of stability parameters obtained by using covariance when chain size increases, with 300 repetitions.

The mean of α estimates shows the bias towards 0.5 when the chain size is not large, regardless its true value. Besides, the variance of ψ estimates is large, for the

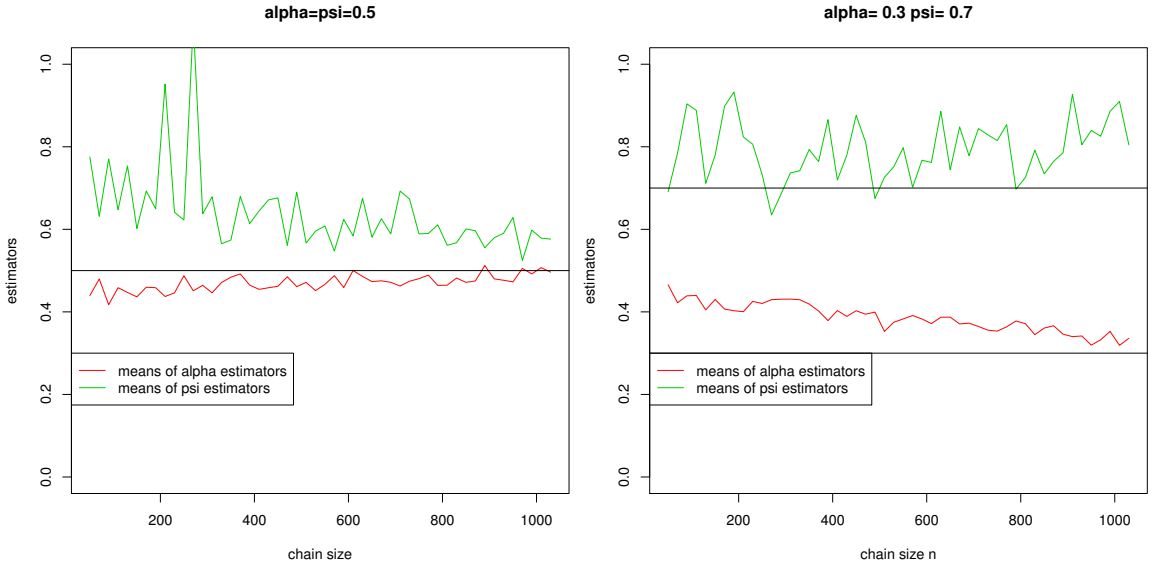

Figure 3.6: Moments estimates of α, ψ with their values shown by the black horizontal lines.

reason that $\hat{\psi}$ depends on the estimator of α. If we cannot estimate α properly, the behaviour of $\hat{\psi}$ will be also affected.

The α estimates stays around 0.5 may be because of the estimators outside (0,1), since

$$\hat{\alpha} = \frac{\text{cov}(Z_{1:n-2}, Z_{3:n})}{\text{cov}(Z_{1:n-1}, Z_{2:n})}, \qquad \hat{\psi^2} = \frac{\text{cov}(Z_{1:n-1}, Z_{2:n})}{\frac{\pi^2}{6}\hat{\alpha}} \tag{3.7}$$

are used in the simulation of Figure 3.6, where $Z_{1:n-2}$ is denoted as the sequence of $\{Z_1, Z_2 \ldots, Z_{n-2}\}$.

To avoid the estimators outside the parameter space (0,1), if either $\hat{\alpha}$ or $\hat{\psi}^2$ in Equation (3.7) is out of (0,1), we let

$$\hat{\alpha}_s = \frac{\text{cov}(Z_{s:n-2}, Z_{s+2:n})}{\text{cov}(Z_{s:n-1}, Z_{s+1:n})},$$

$$\hat{\psi}_s^2 = \frac{\text{cov}(Z_{s:n-1}, Z_{s+1:n})}{\frac{\pi^2}{6}\hat{\alpha}},$$

where $1 \leq s \leq n-k$ for some reasonable integer k.

Denote the stability estimator $\hat{\alpha} = \hat{\alpha}_t$ where $t = \arg\ \min_{1 \leq s \leq n-k}\{s : \hat{\alpha}_s \in (0,1)\}$.

$\hat{\psi} = \hat{\psi}_t$ when $t = \arg\ \min_{1 \le s \le n-k}\{s : \hat{\psi}_s \in (0,1)\}$.

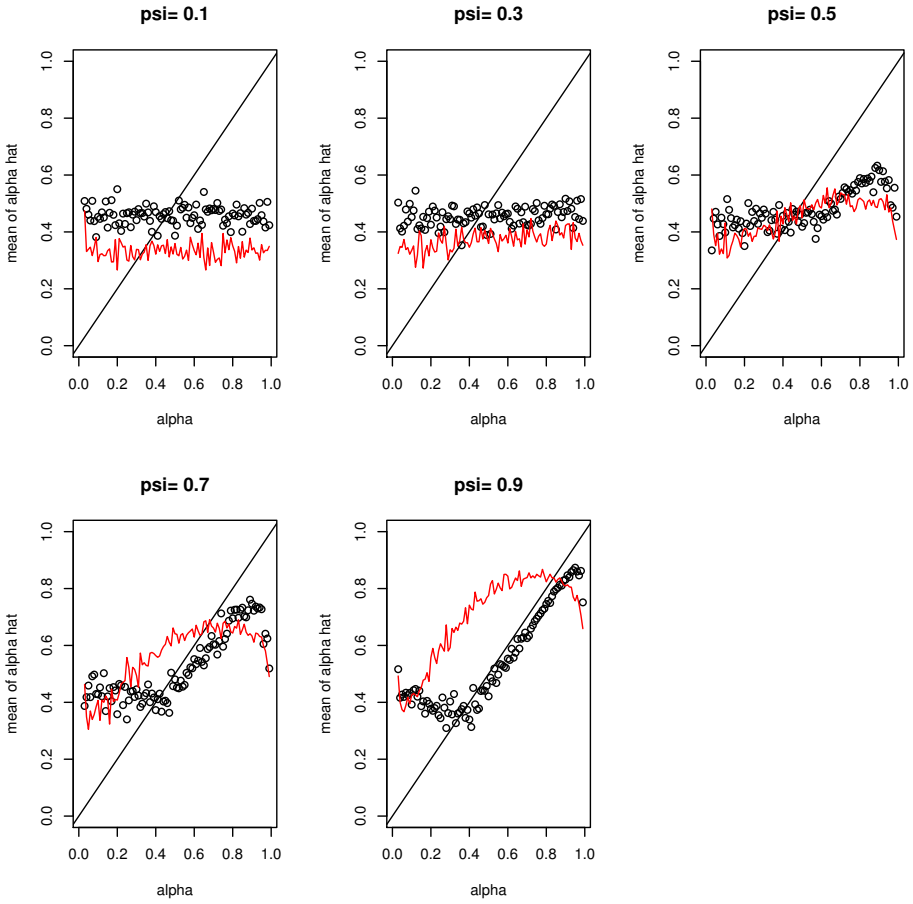

Figure 3.7: Moment estimates of the stability parameters when GEV parameters are unknown. Black dots are the means of α estimates, red curves represent the means of ψ estimates.

Figure 3.7 shows the simulation results of the moment estimator of α (black dots) and ψ (red curve) in the state space model with GEV distributed marginals with unknown parameters $(\mu = 0, \sigma = 1, \gamma = -2, n = 100)$.

The GEV parameters are estimated first, using the regression procedure stated in Section 3.2.1, ignoring the dependence. After obtaining the GEV parameters, the observations are transformed to Gumbel distributed marginals and the stability parameters are estimated.

When the chain size increases to 300 and 1000 and ψ is large $(\psi = 0.7, 0.9)$(Figure 3.8 and 3.9 respectively), the means of the α estimates are close to α. When ψ value is small, like $\psi = 0.3$, the up left plot in Figure 3.8 and 3.9), the averages of α estimators

were located at around 0.5. When $\psi = 0.5$ (the upright plot), α estimates improve a little for large value of α, but still have large variance. When $\psi = 0.7, \alpha > 0.42$ (bottom left plot) and $\psi = 0.9, \alpha > 0.11$ (bottom right plot), this estimation works okay.

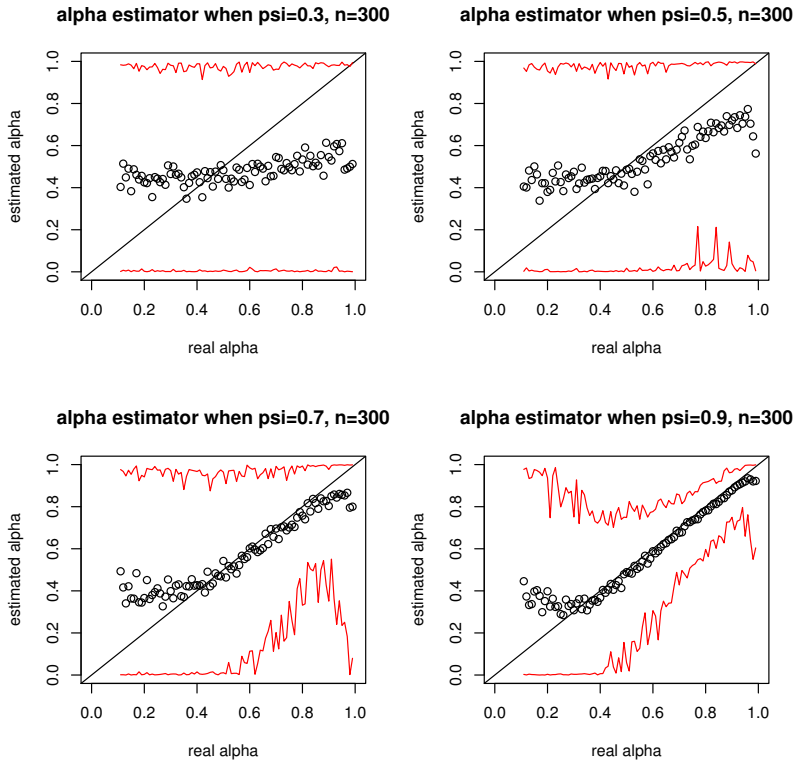

Figure 3.8: Means and 95% confidence interval of α estimators, n=300.

3.3.3 Covariance of the sum and the differences of consecutive two variables

To fully use the information of the observation sequence Z_t, to improve the stability estimation, we consider the sum and the differences of the consecutive observations like $Z_t + Z_{t-1}, Z_t - Z_{t-1}$.

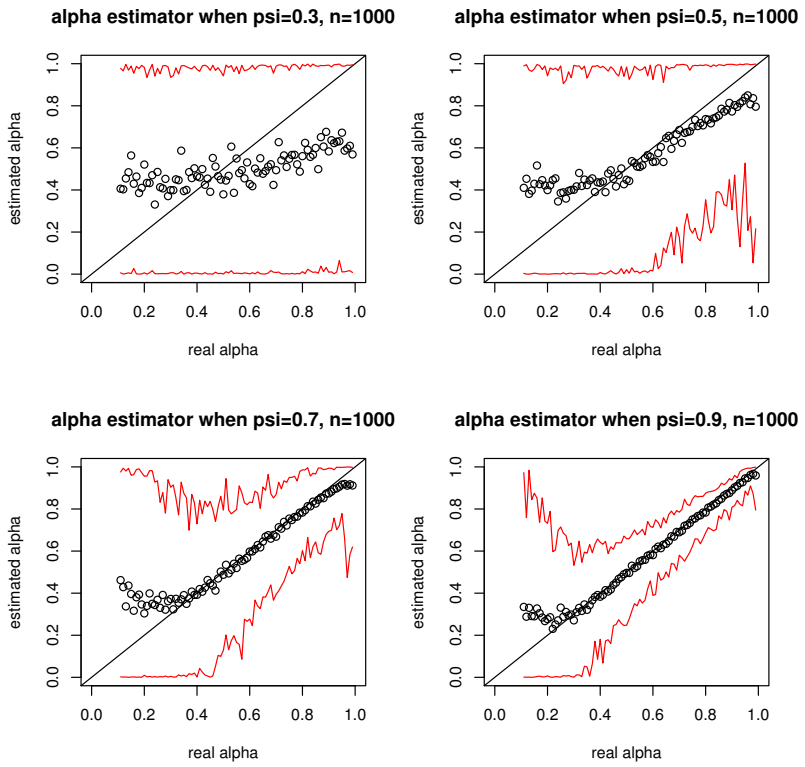

Figure 3.9: Means and 95% confidence interval of α estimators, n=1000.

Notice that

$$\begin{aligned}
\mathrm{Var}(Z_t + Z_{t-1}) &= 2\frac{\pi^2}{6}(1+\alpha\psi^2), \\
\mathrm{Cov}(Z_t + Z_{t-1}, Z_{t-1} + Z_{t-2}) &= \frac{\pi^2}{6}\left(1+\alpha^2\psi^2+2\alpha\psi^2\right), \\
\mathrm{Var}(Z_t - Z_{t-1}) &= 2\frac{\pi^2}{6}(1-\alpha\psi^2), \\
\mathrm{Cov}(Z_t - Z_{t-1}, Z_{t-1} - Z_{t-2}) &= \frac{\pi^2}{6}\left(-1-\alpha^2\psi^2+2\alpha\psi^2\right),
\end{aligned}$$

we have

$$\begin{aligned}
\alpha\psi^2 &= \frac{\mathrm{Var}(Z_t + Z_{t-1}) - \mathrm{Var}(Z_t - Z_{t-1})}{4\pi^2/6} && (3.8)\\
&= \frac{\mathrm{Cov}(Z_t - Z_{t-1}, Z_{t-1} - Z_{t-2}) + \mathrm{Cov}(Z_t + Z_{t-1}, Z_{t-1} + Z_{t-2})}{4\pi^2/6} && (3.9)\\
\alpha^2\psi^2 &= 2\alpha\psi^2 - 1 - \frac{\mathrm{Cov}(Z_t - Z_{t-1}, Z_{t-1} - Z_{t-2})}{\pi^2/6} && (3.10)\\
&= \frac{\mathrm{Cov}(Z_t + Z_{t-1}, Z_{t-1} + Z_{t-2})}{\pi^2/6} - 2\alpha\psi^2 - 1. && (3.11)
\end{aligned}$$

The covariance of the sum and the difference of the observations share some properties of the covariance of the observations.

The estimators of $\alpha\psi^2, \alpha^2\psi^2$ obtained by using sample variance and sample covariance in Equations (3.8), (3.9), (3.10), (3.11) are consistent and normally distributed (shown in Section 3.3.4). In the left plot of Figure 3.10, the simulation results are obtained by the average of Equation (3.8), (3.9) ($\hat{\alpha}\hat{\psi}^2$) and the average of Equation (3.10), (3.11) ($\hat{\alpha}^2\hat{\psi}^2$).

Use the same idea, the covariance structure of the linear combination of three consecutive observations are also used to estimate the stability parameters (right plot in Figure 3.10).

The estimation using the covariance of the linear combination of the consecutive observations is a plausible method when the covariance of observations produces unacceptable estimators.

3.3.4 Asymptotic distribution of the sample covariance

In this part, our aim is to find the asymptotic distribution of

$$\hat{r}(h) = \mathrm{cov}(Z_i, Z_{i+h}) = \frac{1}{n-h}\sum_{i=1}^{n-h}(Z_i - \bar{Z}_n)(Z_{i+h} - \bar{Z}_n),$$

where $\bar{Z}_n$ is the mean of $Z_1, \ldots, Z_n$.

To obtain the asymptotic distribution of $\hat{r}(h)$, first consider

$$r^*(h) = \frac{1}{n-h}\sum_{i=1}^{n-h}(Z_i - \gamma_e)(Z_{i+h} - \gamma_e). \qquad (3.12)$$

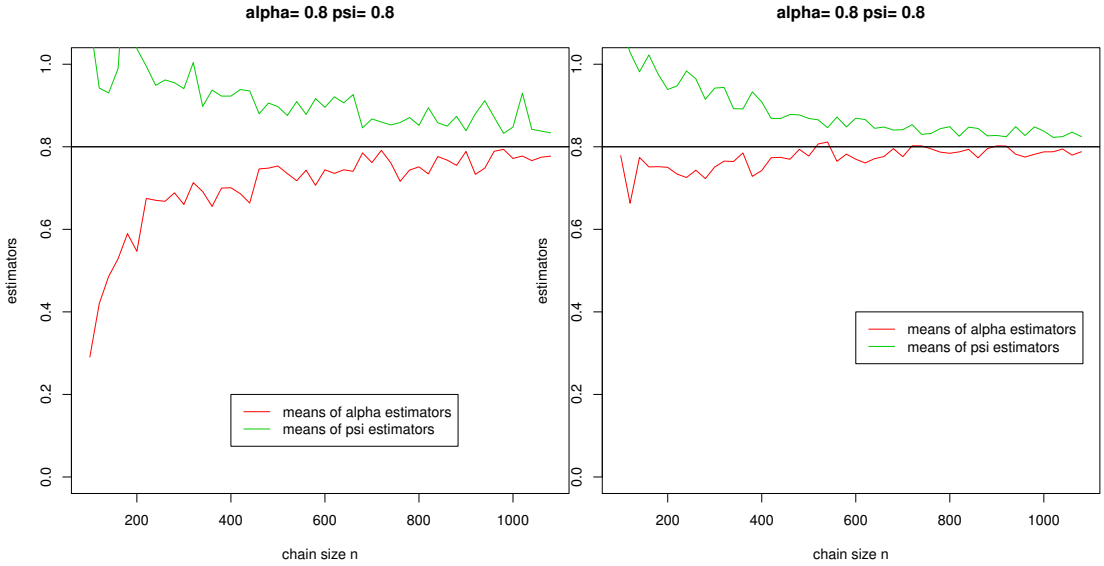

Figure 3.10: Estimates of α, ψ using the covariance structure of the sum and the difference of two (left plot) and three (right plot) consecutive observations.

If the asymptotic distribution of $r^*(h)$ is known, then the asymptotic distribution of $\hat{r}(h)$ can be obtained if $\sqrt{n}(\hat{r}(h) - r^*(h)) = o_p(1)$ as $n \to \infty$.

To simplify the problem, in the following part of this section, the sequence $\{Z_n\}$ is centred without loss of generality, which means $\{Z_n\}$ has mean zero. Once $\{Z_n\}$ is centred, since

$$Z_n = \alpha^n \psi X_0 + \alpha^n \psi \log S_1 + \cdots + \alpha\psi \log S_n + \psi \log \xi_n,$$

we assume that $X_0, \{\log S_i\}, \{\log \xi_i\}$ are all centred.

The following theorems hold for our series.

Theorem 14 (Central Limit Theorem)**.** *For stationary and centred series* $\{Z_i, 1 \leq i \leq n\}$,

$$Z_i - \alpha Z_{i-1} = \psi \log \xi_i - \alpha\psi \log \xi_{i-1} + \alpha\psi \log S_i,$$

where independent errors $\log \xi_i$ *and* $\log S_i$ *are centred exponential* $\mathcal{S}(\psi), \mathcal{S}(\alpha)$ *random variables, respectively, with* $\alpha, \psi \in (0, 1)$*, the mean of this series* $\bar{Z}_n = n^{-1}(Z_1 + \cdots + Z_n)$ *is asymptotically normal.*

$$\sqrt{n}\bar{Z}_n \xrightarrow{D} N\left(0, \frac{\pi^2}{6}\left(1 + 2\psi^2 \frac{\alpha}{1-\alpha}\right)\right).$$

Using Theorem 14, we could find the asymptotic distribution of $r^*(h)$ defined from Equation (3.12).

Theorem 15. *For the series* $\{Z_i, 1 \le i \le n\}$ *under the same conditions of Theorem 14,* $r^*(h)$ *is asymptotically normal with mean* $r(h)$ *and variance* Var_h *(in Equation* (B.12)*) and*

$$\mathrm{Var}_h = \alpha^{2h}\psi^4 \left(\frac{\pi^2}{6}\right)^2.$$

We will show that under the same conditions, $r^*(h)$ and $\hat{r}(h)$ have the same asymptotic distribution.

Theorem 16. *For the series* $\{Z_i\}$ *under the same conditions of Theorem 14,* $\hat{r}(h)$ *is asymptotically normal.*

$$\hat{r}(h) - r(h) \xrightarrow{D} N\left(0, \alpha^{2h}\psi^4 \left(\frac{\pi^2}{6}\right)^2\right).$$

In Theorem 27.4 of book [6], Billingsley considered the zero-mean sequence, if $d(n) = O_p(n^{-5})$ and the random variable has finite 12th moments, the Central Limit Theorem holds for $\bar{Z}_n$, the mean of sequence. Billlingsley pointed out that $d(n) = O_p(n^{-5})$ and $\mathrm{E}(Z_t^{12}) < \infty$ are stronger than necessary. Usually $\mathrm{E}(Z_t^4) < \infty$ or $\mathrm{E}(\bar{Z}_n^4) < \infty$ suffice instead of the condition $\mathrm{E}(Z_t^{12}) < \infty$.

The proof of Theorem 14, 15, 16 are shown in Appendix B.

As for the mean of $\{Z_i^2\}$, Central Limit Theorem holds.

$$\begin{aligned}
\mathrm{E}\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n Z_i^2\right) &= \frac{\pi^2}{6}, \\
\mathrm{Var}\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n Z_i^2\right) &= \frac{1}{n}\mathrm{Var}(Z_1^2) + \frac{2}{n^2}\sum_{h=1}^{n-1}(n-h)\,\mathrm{Cov}(Z_1^2, Z_{1+h}^2) \\
&= \frac{1}{n}\left(\left(\frac{\pi^2}{6}\right)^2\left(\frac{12}{5} + \frac{2}{n}\left(5\psi^4 - 5\psi^2 + \frac{12}{5}\right)\frac{\alpha^2}{1-\alpha^2}\left(n - \frac{1-\alpha^{2n}}{1-\alpha^2}\right)\right)\right) \\
&\quad -\frac{2}{n^2}\frac{\alpha^2}{1-\alpha^2}\psi^4\,\mathrm{E}(\log^4\xi_i)\left(n - \frac{1-\alpha^{2n}}{1-\alpha^2}\right). \qquad (3.13)
\end{aligned}$$

We have

Theorem 17. *With the same time series* $\{Z_t\}$ *as in Theorem 14, as* $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(Z_i^2 - \frac{\pi^2}{6} \right) \xrightarrow{D} N\left(0, \left(\frac{\pi^2}{6} \right)^2 \left(\frac{12}{5} + \frac{2\alpha^2}{1-\alpha^2}(5\psi^4 - 5\psi^2 + \frac{12}{5}) \right) \right).$$

To obtain the mean and variance of α estimator, the following theorem is needed.

Theorem 18. *With the same time series* $\{Z_t\}$ *as in Theorem 14, the covariance of* $r(s)$ *and* $r(h), s < h$ *goes to* $\left(\frac{\pi^2}{6}\right)^2 \psi^4 \alpha^{s+h}$ *as* $n \to \infty$.

Proof of Theorem 18. Using Equations (B.10), (B.11) in Appendix B, we have that

$$\begin{aligned}
\mathrm{Cov}(r^*(s), r^*(h)) &= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(Z_i Z_{i+s}, Z_j Z_{j+h}) \\
&= \frac{2}{n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{\min(i+s-1,n)} \mathrm{E}(Z_i Z_j Z_{i+s} Z_{j+h}) \\
&\quad + \frac{2}{n^2} \sum_{i=1}^{n-s-1} \sum_{j=i+s+1}^{n} \mathrm{E}(Z_i Z_{i+s} Z_j Z_{j+h}) \\
&\quad + \frac{1}{n} \mathrm{E}(Z_i^2 Z_{i+s} Z_{i+h}) + \frac{n-s}{n} \mathrm{E}(Z_i Z_{i+s}^2 Z_{i+s+h}) - \left(\frac{\pi^2}{6} \right)^2 \psi^4 \alpha^{s+h} \\
&\to \left(\frac{\pi^2}{6} \right)^2 \psi^4 \alpha^{s+h}.
\end{aligned}$$

□

In our simulation, Z_t is transformed from GEV distributed observation Y_t. GEV parameters are estimated first. We use $\hat{Z}_t = \frac{1}{\hat{\gamma}} \log(1 + \hat{\gamma} Y_t)$ to obtain the Gumbel distributed marginals to estimate the stability parameters. If $\hat{\gamma}$ is obtained by ignoring the dependence of $\{Y_t\}$, from the weak dependence structure, we have $\hat{\gamma} - \gamma$ goes to 0 as $n \to \infty$ and $\hat{\gamma}$ has asymptotic normality.

Denote $w_n = \hat{\gamma} - \gamma$, which is $O_p\left(\frac{\sqrt{\log\log n}}{n}\right)$, we have

$$\begin{aligned}
\hat{Z}_t &= \frac{1}{\gamma}\frac{1}{1+\frac{w_n}{\gamma}}\left(\log(1+\gamma Y_t) + \log\left(1+\frac{w_n y}{1+\gamma Y_t}\right)\right) \\
&= Z_t(1-\frac{w_n}{\gamma}) + \frac{1}{\gamma}\frac{w_n Y_t}{1+\gamma Y_t} + o_p(w_n). \\
\mathrm{Cov}(\hat{Z}_t, \hat{Z}_{t-k}) &= \mathrm{Cov}(Z_t, Z_{t-k})(1-\frac{w_n}{\gamma})^2 + O_p(w_n) \\
&= \mathrm{Cov}(Z_t, Z_{t-k}) + O_p(w_n) \\
&= \mathrm{Cov}(Z_t, Z_{t-k}) + O_p\left(\frac{\sqrt{\log\log n}}{n}\right).
\end{aligned}$$

3.3.5 Regression model

If we write model (3.2) as

$$Z_t - \gamma_e - \alpha(Z_{t-1} - \gamma_e) = q_t, \tag{3.14}$$

where $q_t = \psi \log \xi_t - \alpha\psi \log \xi_{t-1} + \alpha\psi \log S_t - (1-\alpha)\gamma_e$, which is a zero-mean sequence and

$$\mathrm{Cov}(q_t, q_{t-k}) = \begin{cases} \frac{\pi^2}{6}(1+\alpha^2-2\alpha^2\psi^2) & \text{if } |k| = 0, \\ -\frac{\pi^2}{6}\alpha(1-\psi^2) & \text{if } |k| = 1, \\ 0 & \text{if } |k| > 1. \end{cases}$$

We can estimate the stability parameters by

$$(\hat{\alpha}, \hat{\psi}) = \underset{0<\alpha,\psi<1}{\arg\min}\, q'_{2:n}\, \mathrm{cov}^{-1}(q_{2:n}) q_{2:n},$$

where $q_{2:n} = (q_2, \ldots, q_n)'$ and q_t is defined as in Equation (3.14).

The simulation result of using regression is shown in the Figure 3.11, when the chain size $n = 100$. The regression estimation gives estimator in the parameter space (0, 1), but the simulation results are not very satisfying.

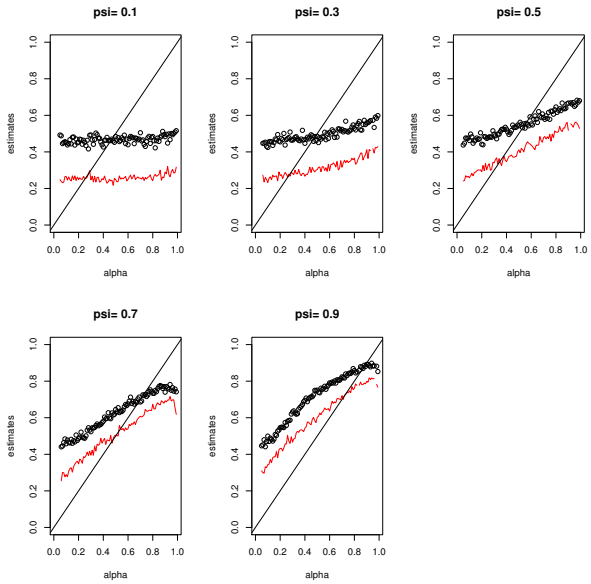

Figure 3.11: The α estimates (black dots) and ψ estimates (red curve) obtained by the regression model.

Chapter 4

Model filtering

In this chapter, we use Z_t to represent the Gumbel distributed observation and Y_t to represent the GEV distributed observation at time t in the state space model

$$\begin{cases} Y_t = \mu - \frac{\sigma}{\gamma} + \frac{\sigma}{\gamma} e^{\psi\gamma X_t} \xi_t^{\psi\gamma}, & \text{(GEV observation equation)} \\ Z_t = \psi X_t + \psi \log \xi_t, & \text{(Gumbel observation equation)} \\ X_{t+1} = \alpha X_t + \alpha \log S_{t+1}. & \text{(state equation)} \end{cases}$$

Our interest is to estimate the state X_t and the empirical filtering density $f(X_t|Y_{1:t})$, or $f(X_t|Z_{1:t})$.

4.1 Kalman filter

In 1960, Kalman [52] presented an algorithm (Kalman filter), working on linear and discrete time system, for the purpose of estimation and prediction. Kalman filter obtains the minimum mean square state error estimated by orthogonal projection.

Kalman filter returns the predictor of the state, its conditional mean and covariance, which is enough to find the conditional density, since the mean and variance characterize the Gaussian distribution.

Let the hidden state $\{X_t\}$ and the observation state $\{Y_t\}$ be random processes with zero mean. If either

(i) the random processes $\{X_t\}$ and $\{Y_t\}$ are Gaussian, or

(ii) the optimal estimator is restricted to be a linear function of the observed random variables and the loss function is the mean square error of hidden state,

then the optimal estimator of X_{t_1} given $Y_{t_0:t}$ is the orthogonal projection of X_{t_1} on a vector space of $Y_{t_0:t}$ (Theorem 2 in [52]).

Kalman filter only needs the estimated state from the previous time and the current observation to estimate the current state. No history of observations and estimated state are required.

In real life, the system is more complicated than linear model, thus Kalman filter was extended to non-linear systems. The extended Kalman filter, one of the most commonly used method for non-linear models, uses either Taylor expansion or Monte Carlo method to linearize the model for filtering purposes.

If extended Kalman filter does not produce good estimates for the non-linear system, an alternative option is unscented Kalman filter, which uses a subset of sample points located around the mean. Unscented transform is applied in the subset choosing step. More details about unscented Kalman filter can be found in Einicke and White [26], Julier and Uhlmann [50], Gustafsson and Hendeby [40].

There are many other non-linear filter methods like the second-order non-linear filter; Monte-Carlo simulation filter; single stage iteration filter; Gaussian sum filter; numerical integration filter; density based Monte-Carlo filter; rejection sampling filter; importance sampling filter and so on.

Figure 4.1 shows the estimation of the states using Kalman filter with the chain size $n = 100, \alpha = \psi = 0.8$ in a state space model with Gumbel distributed observations. The upper plot in Figure 4.1 shows the comparison of state values (solid line) and estimated states obtained by Kalman filter (broken line). To see the comparison in a clearer view, the states are ordered (the bottom plot in Figure 4.1).

In the plot with the ordered states, if the state locates in $(1, 3)$, Kalman filter works well and captures the increasing trend. However, when the state value is smaller than 3, which happens for most of the time since the state marginal distribution is Gumbel, Kalman filter returns no efficient estimators. All the large valued states, greater than 4, are under estimated. In this simulation we only have one estimate exceeding 3, which is used to estimate the state with true value around 1.

Besides, Kalman filter estimators have the unnecessary turbulence wave and it cannot estimate the trend of the time series.

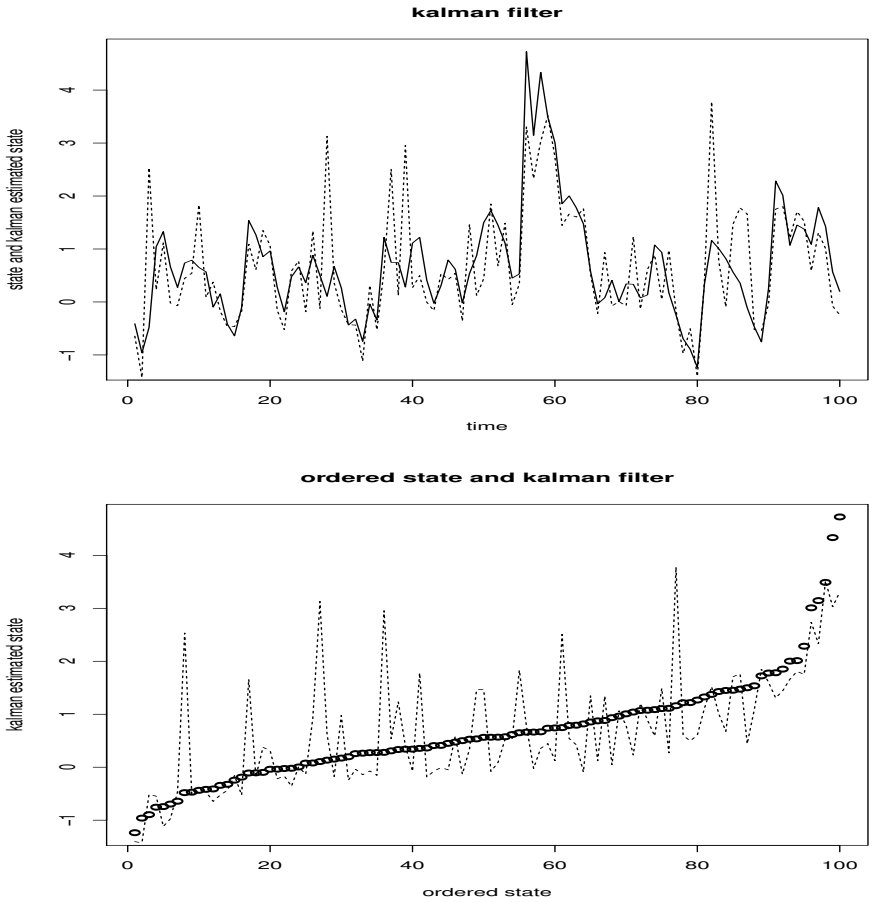

Figure 4.1: Kalman filter estimates (broken line) of the states followed Gumbel distribution.

4.2 Estimation of the hidden state

A schematic representation of a state space model would be

$$\begin{array}{ccccccccccc} X_1 & \stackrel{S_2}{\to} & X_2 & \stackrel{S_3}{\to} & X_3 & \stackrel{S_4}{\to} & X_4 & \to & \cdots & \stackrel{S_n}{\to} & X_n \\ \downarrow \xi_1 & & \downarrow \xi_2 & & \downarrow \xi_3 & & \downarrow \xi_4 & & & & \downarrow \xi_n \\ Z_1 & & Z_2 & & Z_3 & & Z_4 & & & & Z_n \end{array}$$

When α, ψ are known, the observation sequence $\{Z_t, 1 \le t \le n\}$ can be used to predict the state $\{X_t, 1 \le t \le n\}$ by estimating the errors at each time point using the error densities.

The estimation of errors is done for one pair of observations at a time. Before estimating the errors, an independent stable random sequence $\{\tilde{S}_t, 1 \le t \le n-1\}$, which follows $\mathcal{S}(\alpha)$, and an independent stable random variable sequence, $\{\tilde{\xi}_t, 1 \le t \le n\}$ which follows $\mathcal{S}(\psi)$ are generated.

At time 1 and 2, let

$$\begin{aligned} \tilde{X}_{1,1} &= \frac{Z_1}{\psi} - \log \tilde{\xi}_1, \quad \cdots, \quad \tilde{X}_{1,n} = \frac{Z_1}{\psi} - \log \tilde{\xi}_n, \\ \tilde{X}_{2,1} &= \frac{Z_2}{\psi} - \log \tilde{\xi}_1, \quad \cdots, \quad \tilde{X}_{2,n} = \frac{Z_2}{\psi} - \log \tilde{\xi}_n. \end{aligned}$$

Denote $\tilde{S}_{2,i,j} = \exp(\tilde{X}_{2,i}/\alpha - \tilde{X}_{1,j})$, where $i \ne j, 1 \le i, j \le n$. Compare $\tilde{S}_{2,i,j}$ with the generated error sequence $\{\tilde{S}_t, 1 \le t \le n-1\}$, estimate the errors by letting $\hat{S}_2 = \tilde{S}_k, \hat{\xi}_1 = \tilde{\xi}_{k_j}, \hat{\xi}_2 = \tilde{\xi}_{k_i}$ where

$$\begin{aligned} k &= \arg\min_{1 \le t \le n-1} \left\{ |\tilde{S}_t - \tilde{S}_{2,i,j}|, i \ne j, 1 \le i, j \le n \right\}, \\ k_i &= \arg\min_{1 \le i \le n} \left\{ |\tilde{S}_t - \tilde{S}_{2,i,j}|, i \ne j, 1 \le i, j \le n, 1 \le t \le n-1 \right\}, \\ k_j &= \arg\min_{1 \le j \le n} \left\{ |\tilde{S}_t - \tilde{S}_{2,i,j}|, i \ne j, 1 \le i, j \le n, 1 \le t \le n-1 \right\}. \end{aligned}$$

The state X_1, X_2 can be estimated by using $\hat{\xi}_1, \hat{\xi}_2$.

$$\hat{X}_t = \frac{Z_t}{\psi} - \log \hat{\xi}_t.$$

Before estimating the next pair of states, remove $\tilde{\xi}_{k_i}, \tilde{\xi}_{k_j}$ from the sequence $\{\tilde{\xi}_t\}$,

denote the remaining sequence as $\{\tilde{\xi}_t, 1 \le t \le n-2\}$. Also, delete $\tilde{S}_k$ from the sequence $\{\tilde{S}_t\}$, denote the remaining as $\{\tilde{S}_t, 1 \le t \le n-2\}$.

At time 3 and 4, repeat the steps at time 1 and 2, let

$$\tilde{X}_{3,1} = \frac{Z_3}{\psi} - \log \tilde{\xi}_1, \quad \cdots, \quad \tilde{X}_{3,n-2} = \frac{Z_3}{\psi} - \log \tilde{\xi}_{n-2},$$
$$\tilde{X}_{4,1} = \frac{Z_4}{\psi} - \log \tilde{\xi}_1, \quad \cdots, \quad \tilde{X}_{4,n-2} = \frac{Z_4}{\psi} - \log \tilde{\xi}_{n-2}.$$

Denote $\tilde{S}_{4,i,j} = \exp(\tilde{X}_{4,i}/\alpha - \tilde{X}_{3,j})$, where $i \neq j, 1 \le i, j \le n-2$. Compare $\tilde{S}_{4,i,j}$ with the sequence $\{\tilde{S}_t, 1 \le t \le n-2\}$, let $\hat{S}_4 = \tilde{S}_k, \hat{\xi}_3 = \tilde{\xi}_{k_j}, \hat{\xi}_4 = \tilde{\xi}_{k_i}$ where

$$k = \underset{1\le t\le n-2}{\arg\min} \left\{ |\tilde{S}_t - \tilde{S}_{4,i,j}|, i \neq j, 1 \le i, j \le n-2 \right\},$$
$$k_i = \underset{1\le i\le n-2}{\arg\min} \left\{ |\tilde{S}_t - \tilde{S}_{4,i,j}|, i \neq j, 1 \le i, j \le n-2, 1 \le t \le n-2 \right\},$$
$$k_j = \underset{1\le j\le n-2}{\arg\min} \left\{ |\tilde{S}_t - \tilde{S}_{4,i,j}|, i \neq j, 1 \le i, j \le n-2, 1 \le t \le n-2 \right\}.$$

Remove $\tilde{\xi}_{k_i}, \tilde{\xi}_{k_j}$ from the sequence $\{\tilde{\xi}_t\}$ and remove $\tilde{S}_k$ from the sequence $\{\tilde{S}_t\}$ again, denote the rest stable sequence as $\{\tilde{\xi}_t, 1 \le t \le n-4\}, \{\tilde{S}_t, 1 \le t \le n-3\}$.

The estimates $\hat{X}_3, \hat{X}_4$ can be obtained by Z_3, Z_4 and $\hat{\xi}_3, \hat{\xi}_4$. Using $\hat{X}_3, \hat{X}_2$, S_3 can be estimated by comparing $\exp(\hat{X}_3/\alpha - \hat{X}_2)$ with $\{\tilde{S}_t, 1 \le t \le n-3\}$. Let $\hat{S}_3 = \tilde{S}_{k_3}$ where

$$k_3 = \underset{1\le t\le n-3}{\arg\min} \left\{ \left| \exp(\hat{X}_3/\alpha - \hat{X}_2) - \tilde{S}_t \right| \right\}.$$

Remove $\tilde{S}_{k_3}$ from the sequence $\{\tilde{S}_t\}$ again. Denote the rest elements in the sets as $\{\tilde{S}_t, 1 \le t \le n-4\}$. Continue the same steps until all the states are estimated.

This method does not grantee that the estimated α-stable distributed errors, $\{\hat{S}_t\}$, satisfy our state equation, but the difference between $\hat{X}_{t+1}$ and $\alpha\hat{X}_t + \alpha \log \hat{S}_{t+1}$ should decrease as n increases.

Another problem with this method is that when either of the stability parameters is small, the estimators could be far away from the "true" states. The set of the generated stable distributed errors should be close to the set of the errors in the model to make this numerical method work. When any of the two stability parameters is small, the set of generated stable distributed errors can have huge differences from the set of the errors in the model due to the heavy-tail property of α-stable variable. If this is the case, the states will not be well estimated. Take $\mathcal{S}(0.1)$ as an example,

the 90-th percentile of $\mathcal{S}(0.1)$ is 3.09×10^9, while the 98th-percentile of $\mathcal{S}(0.1)$ is 4.56×10^{16}, which makes the difference of the maximum of two independent random samples from $\mathcal{S}(0.1)$ to reach values of 10^{16} when the sample size is 50.

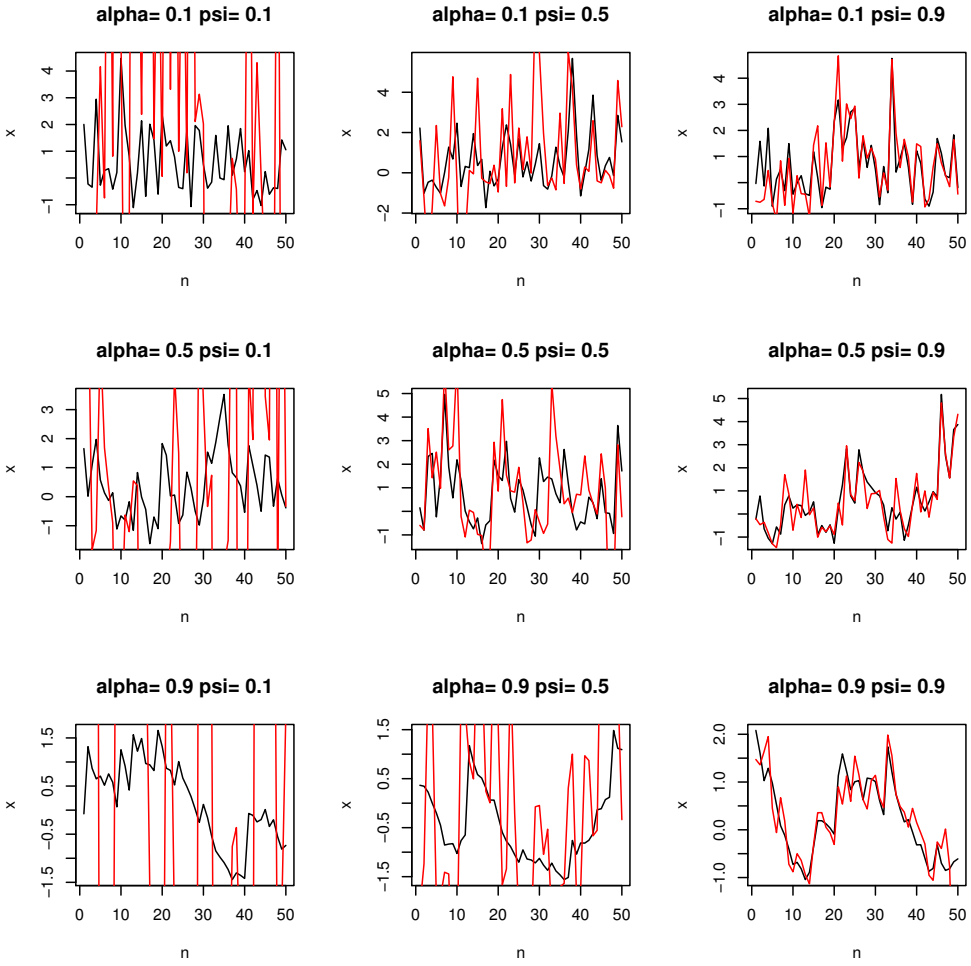

Figure 4.2: Estimates (red curve) of the hidden states (black) with n=50.

Figure 4.2, 4.3 and 4.4 show the estimation results using this method when the chain size n is $50, 100, 200$ respectively. If the stability parameter ψ takes a small value (the first column in these figures, $\psi = 0.1$), the prediction of the state sequence is very poor, due to the reason we explained before, related with the long-and heavy-tail of the stable distribution.

For large stability parameters, the simulation results in Figure 4.2, 4.3 and 4.4 show some advantages when comparing with the result of Kalman filter in Figure 4.1.

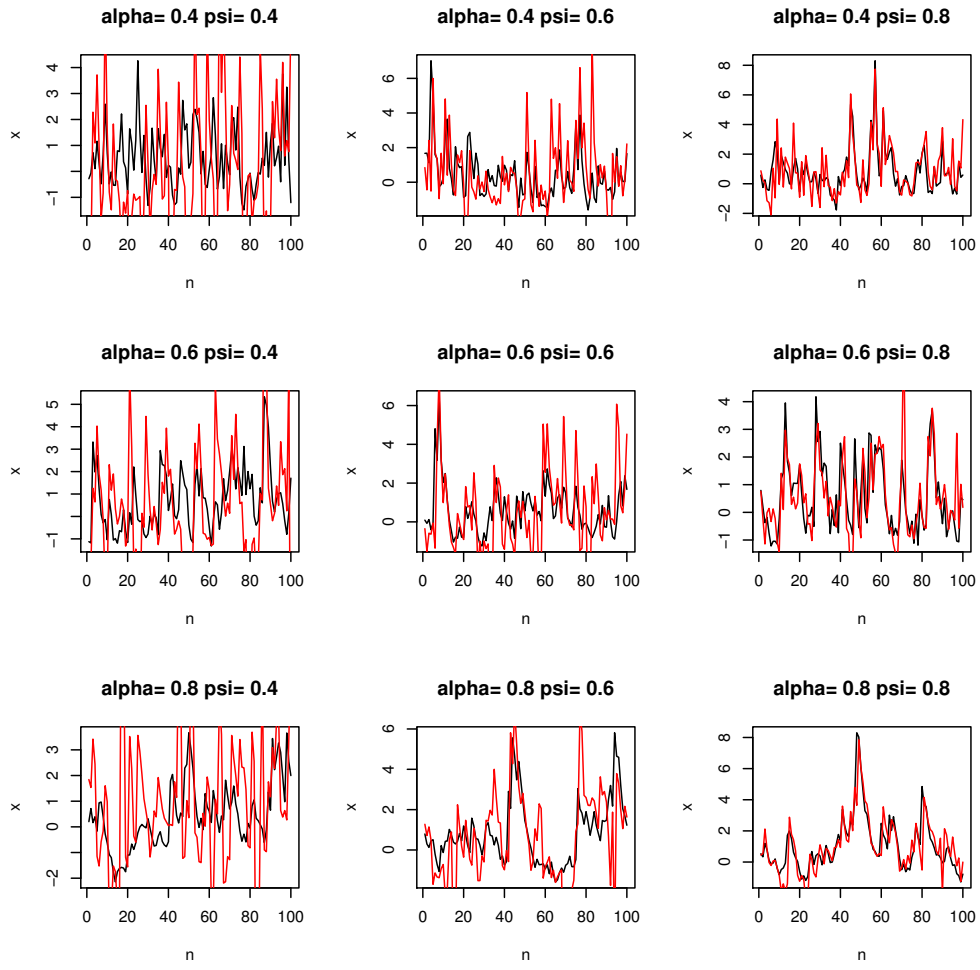

Figure 4.3: Estimates (red curve) of the hidden states (black) with n=100.

When both α and ψ are large, like the plots in the right corner of Figure 4.2, 4.3 and 4.4, $\alpha = \psi = 0.9$, the estimated states (red) are close to the true states (black). The maxima of the series, which reaches 8 in Figure 4.3 , are successfully estimated.

4.3 Particle filter

After estimating the states, now we will focus on estimating the empirical filtering density.

Particle filter is a sequential Monte Carlo technique that uses a set of particles to represent the posterior density of the state space model. It can deal with both linear

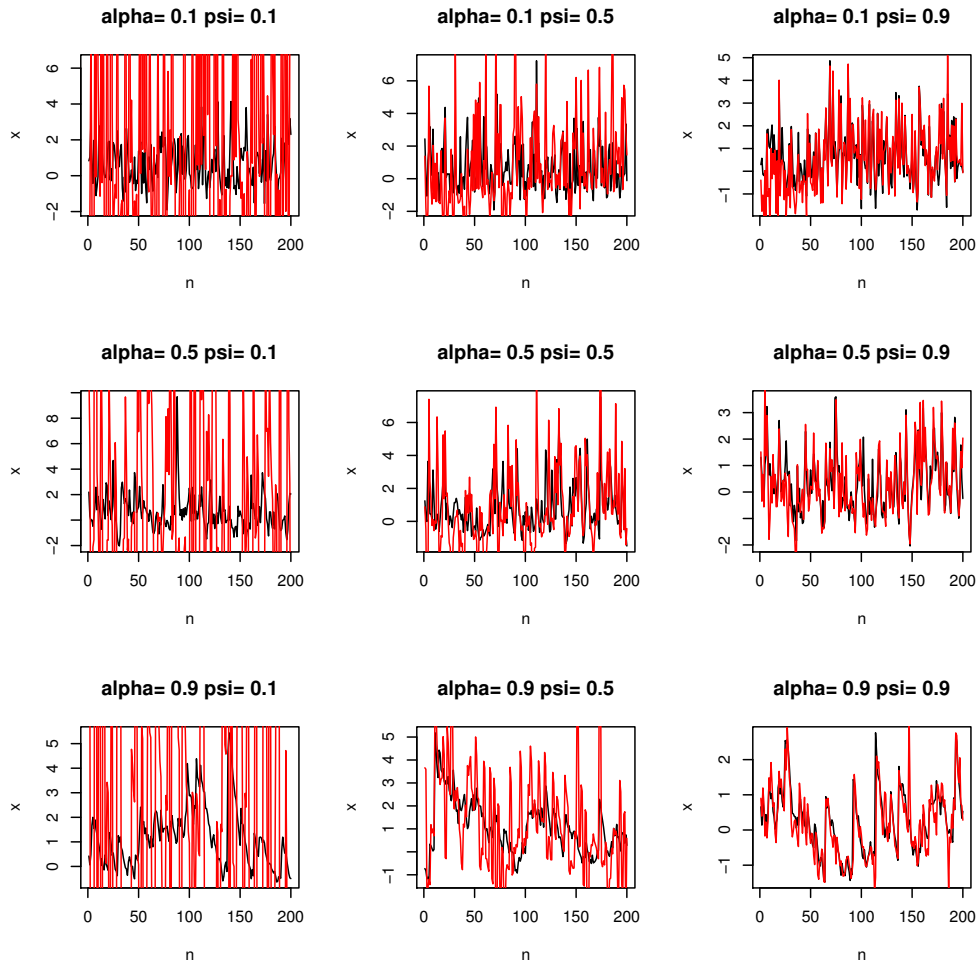

Figure 4.4: Estimates (red curve) of the hidden states (black) with n=200.

and non-linear models and any distributions for the errors.

The original particle filter method was introduced by Gordon, Salmond and Smith in their paper [36], named as Bayesian bootstrap filter. In this paper, the particles were generated by random sampling from the error distribution, weighted by the likelihood of each prior sample.

Later, Kong, Liu and Wong [58] extended this procedure to the estimation of missing data and the hidden state. Particles of the missing data or hidden states were generated by the conditional distribution, $f(X_t|Y_{t-1}, X_{t-1})$, given the past observations and the past particles. Weights of the particles were calculated by the weights of the past particles and the likelihood of each prior sample. Kong, Liu and Wong

also compared particle filter with Gibbs sampler (see [33]) and concluded that Gibbs sampler is less effective.

In Liu and Chen's paper [65], an improved sequential imputation was introduced and theoretically justified, which ensures the effectiveness of the filtering method.

There are many ways to perform particle filter, like using the sequential importance sampling filter (SIS) (see Maceachern, Clyde and Liu [67], Liu, Chen and Logvinenko [51]), bootstrap filter (see Gordon, Salmond and Smith [36], Green [38], Liu, Chen and Wong [66]), the MCMC particle filter (see Khan, Balch and Dellaert [4]), Monte Carlo filter (see Bølviken et al. [10], Kitagawa [55]) and the unscented particle filter (see Merwe and Freitas [94], Wan and Merwe[95]).

Particle filter cannot only estimate the empirical filtering density, but it is also good for the estimation of $\mathrm{E}(g(X_{1:n}))$ for any function g. For every time point, M particles $X_{1:n}^{(i)} = \{X_1^{(i)}, \ldots, X_n^{(i)}\}, 1 \leq i \leq M$ are generated from the density $f(X_{1:n}|Z_{1:n})$, then the estimator of $\mathrm{E}(g)$ is

$$\hat{E}(g) = \sum_{i=1}^{M} \frac{1}{M} g\left(X_{1:n}^{(i)}\right).$$

In the absence of information for the hidden states, a prior probability $\pi(X_{1:n}|Z_{1:n})$, depending upon the observations, can be used to generate the i-th particles $X_{1:n}^{(i)}$. Such a distribution $\pi(\cdot)$ is called the importance sampling distribution.

$$\begin{aligned}
\mathrm{E}(g) &= \int g(X_{1:n}) f(X_{1:n}|Z_{1:n}) dX_{1:n} \\
&= \int g(X_{1:n}) \cdot \frac{f(X_{1:n}|Z_{1:n})}{\pi(X_{1:n}|Z_{1:n})} \cdot \pi(X_{1:n}|Z_{1:n}) dX_{1:n} \\
&= \int g(X_{1:n}) \cdot W^*(X_{1:n}) \cdot \pi(X_{1:n}|Z_{1:n}) dX_{1:n} \\
&= E_{\pi(X_{1:n}|Z_{1:n})}\big(g(X_{1:n}) W^*(X_{1:n})\big).
\end{aligned}$$

$W^*(X_{1:n})$ is called the importance weight.

$$\hat{E}(g) = \sum_{i=1}^{M} \frac{1}{M} g\left(X_{1:n}^{(i)}\right) W_n^*(X_{0:n}^{(i)})$$

converges almost surely to $\mathrm{E}(g)$ and is unbiased (Andrieu, Doucet and Holenstein [2]).

The importance weight of $X_{1:n}^{(i)}$ is

$$W^*(X_{1:n}^{(i)}) = \frac{f(X_{1:n}^{(i)}|Z_{1:n})}{\pi(X_{1:n}^{(i)}|Z_{1:n})} = \frac{f(Z_{1:n}|X_{1:n}^{(i)})f(X_{1:n}^{(i)})}{f(Z_{1:n})\pi(X_{1:n}^{(i)}|Z_{1:n})}.$$

Notice that

$$\begin{aligned}
\mathrm{E}(g) &= \int g(X_{1:n})f(X_{1:n}|Z_{1:n})dX_{1:n} \\
&= \frac{\int g(X_{1:n})f(Z_{1:n}|X_{1:n})f(X_{1:n})dX_{1:n}}{f(Z_{1:n})} \\
&= \frac{1}{f(Z_{1:n})}\int g(X_{1:n})\frac{f(Z_{1:n}|X_{1:n})f(X_{1:n})}{\pi(X_{1:n}|Z_{1:n})}\pi(X_{1:n}|Z_{1:n})dX_{1:n} \\
&= \frac{E_{\pi(X_{1:n}|Z_{1:n})}\left(g(X_{1:n})W(X_{1:n})\right)}{\int W(X_{1:n})\pi(X_{1:n}|Z_{1:n})dX_{1:n}} \\
&= \frac{E_{\pi(X_{1:n}|Z_{1:n})}\left(g(X_{1:n})W(X_{1:n})\right)}{E_{\pi(X_{1:n}|Z_{1:n})}\left(W(X_{1:n})\right)}
\end{aligned}$$

where

$$W(X_{1:n}) = \frac{f(Z_{1:n}|X_{1:n})f(X_{1:n})}{\pi(X_{1:n}|Z_{1:n})},$$

once the independent estimators of the hidden states are generated, we have

$$\begin{aligned}
\hat{E}(g) &= \frac{\frac{1}{M}\sum_{i=1}^{M} g(X_{1:n}^{(i)})W(X_{1:n}^{(i)})}{\frac{1}{M}\sum_{i=1}^{M} W(X_{1:n}^{(i)})} \\
&= \sum_{i=1}^{M} g(X_{1:n}^{(i)})\frac{W(X_{1:n}^{(i)})}{\sum_{i=1}^{M} W(X_{1:n}^{(i)})} \\
&= \sum_{i=1}^{M} g(X_{1:n}^{(i)})\tilde{W}(X_{1:n}^{(i)}),
\end{aligned}$$

which is a biased but almost surely converges to $\mathrm{E}(g)$ under some regular conditions (Andrieu, Doucet and Holenstein [2]).

For any $t, 1 \leq t \leq n$, the normalized importance weight of the i-th particle is denoted as

$$\tilde{W}(X_{1:t}^{(i)}) = \frac{W(X_{1:t}^{(i)})}{\sum_{i=1}^{M} W(X_{1:t}^{(i)})}.$$

On the other hand, for a fixed time t, $f(Z_{1:t})$ is fixed regardless of the particle

$X_{1:t}^{(i)}, 1 \leq i \leq M$. The importance weight $W^*\left(X_{1:t}^{(i)}\right)$ is proportional to $W\left(X_{1:t}^{(i)}\right)$,

$$W^*\left(X_{1:t}^{(i)}\right) = \frac{f(Z_{1:t}|X_{1:t}^{(i)})f(X_{1:t}^{(i)})}{f(Z_{1:t})\pi(X_{1:t}^{(i)}|Z_{1:t})} \propto \frac{f(Z_{1:t}|X_{1:t}^{(i)})f(X_{1:t}^{(i)})}{\pi(X_{1:t}^{(i)}|Z_{1:t})} = W\left(X_{1:t}^{(i)}\right),$$

it would be natural to use the normalized weight of $W\left(X_{1:t}^{(i)}\right)$, i.e. $\tilde{W}\left(X_{1:t}^{(i)}\right)$ as the importance weight function.

Notice that

$$\begin{aligned} f(Z_{1:t}|X_{1:t}) &= f(Z_{1:t}|X_{1:t-1})f(Z_t|X_t), \\ f(X_{1:t}) &= f(X_{1:t-1})f(X_t|X_{t-1}), \\ \pi(X_{1:t}|Z_{1:t}) &= \pi(X_{1:t-1}|Z_{1:t-1})\pi(X_t|X_{1:t-1}, Z_{1:t}), \end{aligned}$$

thus $W\left(X_{1:t}^{(i)}\right)$ can be represented recursive as

$$\begin{aligned} W(X_{1:t}^{(i)}) &= \frac{f(Z_{1:t}|X_{1:t}^{(i)})f(X_{1:t}^{(i)})}{\pi(X_{1:t}^{(i)}|Z_{1:t})} \\ &= \frac{f(Z_{1:t-1}|X_{1:t-1}^{(i)})f(X_{1:t-1}^{(i)})}{\pi(X_{1:t-1}^{(i)}|Z_{1:t-1})} \cdot \frac{f(Z_t|X_t^{(i)})f(X_t^{(i)}|X_{t-1}^{(i)})}{\pi(X_t^{(i)}|X_{1:t-1}^{(i)}, Z_{1:t})} \\ &= W(X_{1:t-1}^{(i)}) \frac{f(Z_t|X_t^{(i)})f(X_t^{(i)}|X_{t-1}^{(i)})}{\pi(X_t^{(i)}|X_{1:t-1}^{(i)}, Z_{1:t})}. \end{aligned}$$

For simplicity, the importance weight function at time t for the i-th particle $W(X_{1:t}^{(i)})$ is denoted as $W_t^{(i)}$, so the relation between $W_t^{(i)}$ and $W_{t-1}^{(i)}, 2 \leq t$ can be written as

$$W_t^{(i)} = W_{t-1}^{(i)} \frac{f(Z_t|X_t^{(i)})f(X_t^{(i)}|X_{t-1}^{(i)})}{\pi(X_t^{(i)}|X_{1:t-1}^{(i)}, Z_{1:t})}.$$

After obtaining a reasonable importance weight function, we can perform the filtering. Here, we have used the sequential importance sampling (SIS) to estimate the empirical filtering density.

4.3.1 Sequential importance sampling

To generate $X_1^{(i)}, 1 \leq i \leq M$ from X_0 through sequential importance sampling, we first use the importance function and obtain the importance weight of $X_1^{(i)}$, then generate $X_t^{(i)}, 1 \leq i \leq M$ for time $t \geq 2$, from the particles $X_{t-1}^{(i)}$ and the observation Z_t and obtain the importance weight of $X_t^{(i)}$, $1 \leq i \leq M$. The generated particles and the standardized weights are used to estimate the conditional densities of the hidden states.

The main idea is to generate particles at time $(t+1)$ using all the information at time t, balancing the distribution of particles with different weights.

Notice that the conditional mean and the variance of the weight $W(X_{1:t}^{(i)})$ are

$$\begin{aligned} \mathrm{E}(W(X_{1:t}^{(i)})|X_{1:t}^{(i)}) &= W(X_{1:t-1}^{(i)}) \int f(Z_t|X_t^{(i)}) f(X_t^{(i)}|X_{t-1}^{(i)}) dX_t^{(i)} \\ &= W(X_{1:t-1}^{(i)}) \int f(Z_t, X_t^{(i)}|X_{t-1}^{(i)}) dX_t^{(i)} \\ &= W(X_{1:t-1}^{(i)}) f(Z_t|X_{t-1}^{(i)}), \end{aligned}$$

$$\mathrm{Var}(W(X_{1:t}^{(i)})|X_{1:t}^{(i)}) = \left(W(X_{1:t-1}^{(i)})\right)^2 \left(\int \frac{f^2(Z_t|X_t^{(i)}) f^2(X_t^{(i)}|X_{t-1}^{(i)})}{\pi(X_t^{(i)}|X_{1:t-1}^{(i)}, Z_{1:t})} dX_t^{(i)} - f^2(Z_t|X_{t-1}^{(i)}) \right),$$

whenever we choose the importance sampling distribution to be

$$\pi(X_t^{(i)}|X_{1:t-1}^{(i)}, Z_{1:t}) = f(X_t^{(i)}|X_{t-1}^{(i)}, Z_t).$$

Notice that with this choice, we have

$$\begin{aligned} & \int \frac{f^2(Z_t|X_t^{(i)}) f^2(X_t^{(i)}|X_{t-1}^{(i)})}{\pi(X_t^{(i)}|X_{1:t-1}^{(i)}, Z_{1:t})} dX_t^{(i)} \\ = & \int \frac{f^2(X_t^{(i)}|Z_t, X_{t-1}^{(i)}) f^2(Z_t|X_{t-1}^{(i)})}{f(X_t^{(i)}|X_{1:t-1}^{(i)}, Z_{1:t})} dX_t^{(i)} \\ = & f^2(Z_t|X_{t-1}^{(i)}), \end{aligned}$$

which leads to $\mathrm{Var}(W(X_{1:t}^{(i)})) = 0$, i.e., only one particle has non-zero weight.

This phenomenon is called weight degeneracy. It does not only happen for the SIS filter procedure, but also for the other filtering processes, such as Bayesian filter (see

Doucet, Godsill and Andrieu [23], Kong, Liu and Wong [58]).

To address the degeneracy problem, we can use resampling.

4.3.2 Resampling

Before doing resample, we need to define the criterion of degeneracy. The effective sample size N_{eff} is defined as

$$N_{eff}(n) = \frac{1}{E[(\tilde{W}(X_{1:t}^{(i)}))^2]}.$$

and a threshold C_{thr} is set, such that if $N_{eff}(n) \leq C_{thr}$, we assume that weight degeneracy has happened at time t.

In Liu and Chen [65], an estimator of N_{eff} is given by

$$\hat{N}_{eff}(n) = \frac{M}{\sum_{i=1}^{M} \tilde{W}(X_t^{(i)})^2}.$$

If $\hat{N}_{eff}(n) \leq C_{thr}$ at time t, resampling should be performed. To perform resampling, some particles with large weights are repeated. Denote $n_t^{(i)}$ as the repeated times of the unique particle $X_t^{(i)}$ after resampling,

$$\hat{F}(X_{0:t}|Z_{0:t}) = \sum_i \frac{n_t^{(i)}}{M} \mathrm{I}_{X_{0:t}^{(i)}}(X_{0:t})$$

is an unbiased estimator of the empirical filter distribution.

Some commonly used resampling methods are stated below.

Systematic Resampling. (Kitagawa [55])

Calculate $h_t^{(i)}$, the likelihood of each particle $X_t^{(i)}$ given the observations.

Let

$$n_t^{(i)} \sim \frac{\sum_{i=1}^{M} h_t^{(i)} I_{X_{0:t}^{(i)}}(X_{0:t})}{\sum_{i=1}^{M} h_t^{(i)}}.$$

Residual Resampling. (Doucet and Johansen [24])

Let $n_{t1}^{(i)} = \lfloor M\tilde{W}_t^{(i)} \rfloor$, the integer part of $M\tilde{W}_t^{(i)}$.

Make $n_{t2}^{(i)}$ follow the multinomial distribution with parameter M and standardized probabilities $\tilde{W}_t^{(i)} - \frac{n_{t1}^{(i)}}{M}, 1 \leq i \leq M$.

Let $n_t^{(i)} = n_{t1}^{(i)} + n_{t2}^{(i)}$.

Multinomial Resampling. (Doucet and Johansen [24])
Let $n_t^{(i)}$ be a multinomial random variable with parameters $(M, \tilde{W}_t^{(1)}, \tilde{W}_t^{(2)}, \ldots, \tilde{W}_t^{(M)})$.

4.3.3 Simulation

For simplicity, here we use the state space model with the Gumbel distributed marginals $\{Z_1, \ldots, Z_n\}$.

We tried several different importance functions and different importance weight functions.

- At time t, $1 \leq t \leq n$, generate M particles $X_t^{(l)}, 1 \leq l \leq M$, from the past particles and the present observation, like

$$X_t^{(l)} = \alpha X_{t-1}^{(l)} + \alpha \log S_t^{(l)}, \tag{4.1}$$

$$X_t^{(l)} = \frac{Z_t}{\psi} - \log \xi_t^{(l)}, \tag{4.2}$$

$$X_t^{(l)} = \frac{1}{2}\left(\alpha X_{i-1}^{(l)} + \alpha \log S_t^{(l)} + \frac{Z_t}{\psi} - \log \xi_t^{(l)}\right), \tag{4.3}$$

  where $S_t^{(l)}$ and $\xi_t^{(l)}, 1 \leq l \leq M$, are generated independently from stable distributions, $\mathcal{S}(\alpha), \mathcal{S}(\psi)$ separately.

- Define the importance weight of $X_t^{(l)}$ as the normalized weight of $\hat{W}_t^{(l)}$, which is

$$W_t^{(l)} = \frac{\hat{W}_t^{(l)}}{\sum_{l=1}^m \hat{W}_t^{(l)}},$$

  where $\hat{W}_t^{(l)}$ takes one of the following three values $W_{t-1}^{(l)} f(Z_t|X_t^{(l)}), W_{t-1}^{(l)} f(X_t^{(l)}|X_{t-1}^{(l)})$ and

$$W_{t-1}^{(l)} \frac{f(Z_t|X_t^{(l)}) f(X_t^{(l)}|X_{t-1}^{(l)})}{\frac{1}{2}\left(f(Z_t|X_t^{(l)}) + f(X_t^{(l)}|X_{t-1}^{(l)})\right)}$$

  according to the particle generating function as stated in equation (4.1), (4.2) and (4.3).

- After obtaining the particles and their corresponding weights at each time point t, check if

$$\frac{M}{\sum_{l=1}^M \left(W_t^{(l)}\right)^2} \leq C_{thr} = \frac{M}{\sum_{l=1}^M \left(W_1^{(l)}\right)^2},$$

where C_{thr} is a given constant and see if weight degeneracy has occurred and resampling is required.

In our simulation, we let $C_{thr} = \frac{M}{\sum_{l=1}^{M}\left(W_1^{(l)}\right)^2}$. If the weight degeneracy happens, we resample the particles using the cumulative distribution function of $(X_t^{(l)}, W_t^{(l)}), 1 \leq l \leq M$ at time t.

If weight degeneracy has appeared at time t,

(i). generate M independent uniformly (0,1) distributed random variables $u_1, \ldots, u_M$.

(ii). Define the u_j-quantile of the distribution of $(X_t^{(l)}, W_t^{(l)}), 1 \leq l \leq M$, for all j from 1 to M be the particles at time t with the weight $1/M$.

With the particles and their corresponding weights, the expectation of the state at time i can be estimated by $\hat{E}(X_t) = \sum_{l=1}^{M} X_t^{(l)} W_t^{(l)}$ if there is no weight degeneracy, or by the average of the particles generated after resampling.

Crisan and Doucet [17] obtained the convergence rate of the average mean square error, which is $1/M$ under certain conditions (bounded transition kernel and bounded importance function, etc...), and the almost sure convergences of the empirical distribution of the generated particles.

The simulation results of means, modes and 95% confidence intervals of the particles generated by SIS when $(\alpha, \psi) = (0.8, 0.8)$, $(0.5, 0.8)$, $(0.8, 0.5), n = 50$ are shown in Figure 4.5, 4.6, 4.7 respectively. In these figures, each row represent the simulation result of particles generated by a different importance function (Equation (4.1), (4.2), (4.3) respectively). The first column shows the comparison of the true state (black curve) and the 90% confidence interval (grey curve). The stars are the modes of the generated particles while the circles are the means. To see the comparison in a clearer view, the states are ordered and shown in a increasing trend in the second column, aiming to find out whether the particle filter can estimate the trend.

When $\alpha = \psi$ and both stability parameters are large, both importance function work well (Figure 4.5). When $\alpha < \psi$, the tail of $f(X_t|X_{t-1}; \alpha)$ is heavier than $f(X_t|Z_t; \psi)$, thus the particles generated by $f(X_t|X_{t-1}; \alpha)$ have a wider range, which reduce the chance of degeneracy. This difference can be seen in Figure 4.6, with the observation that using either $f(X_t|X_{t-1}; \alpha)$ or $\frac{1}{2}\left(f(X_t|Z_t) + f(X_t|X_{t-1})\right)$ to generate particles would result in means and modes closer to the true state than using $f(X_t|Z_t)$, especially for the large valued states. When $\alpha > \psi$, the simulation shows

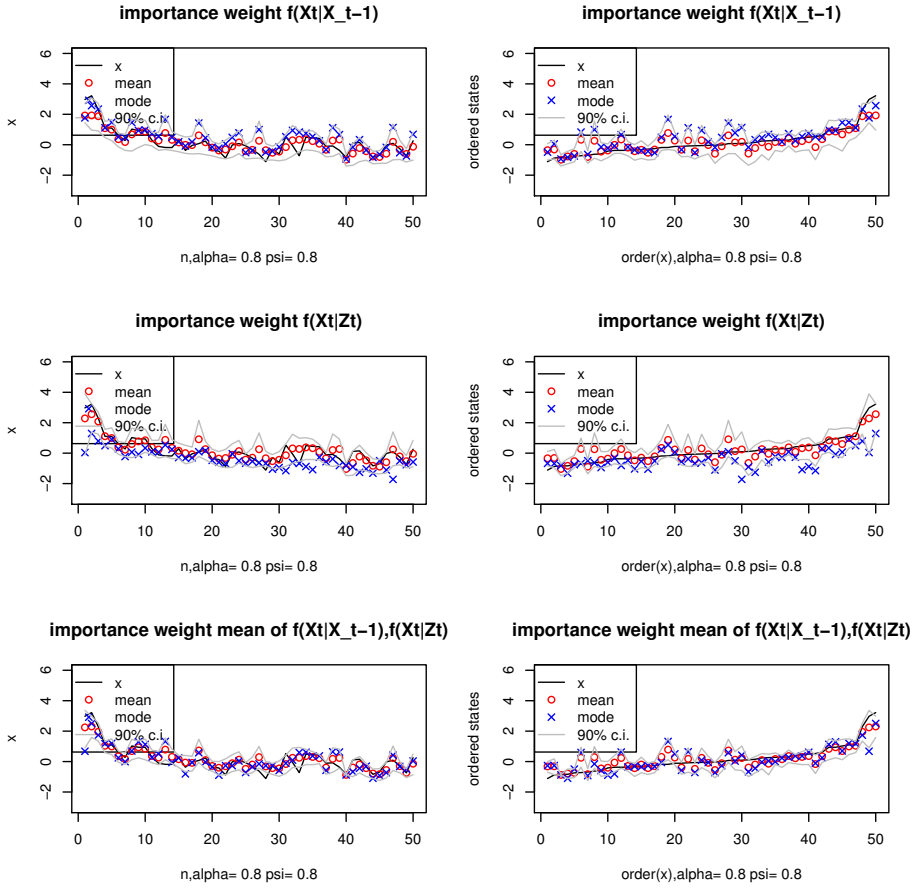

Figure 4.5: Particle filter with $(\alpha, \psi) = (0.8, 0.8)$.

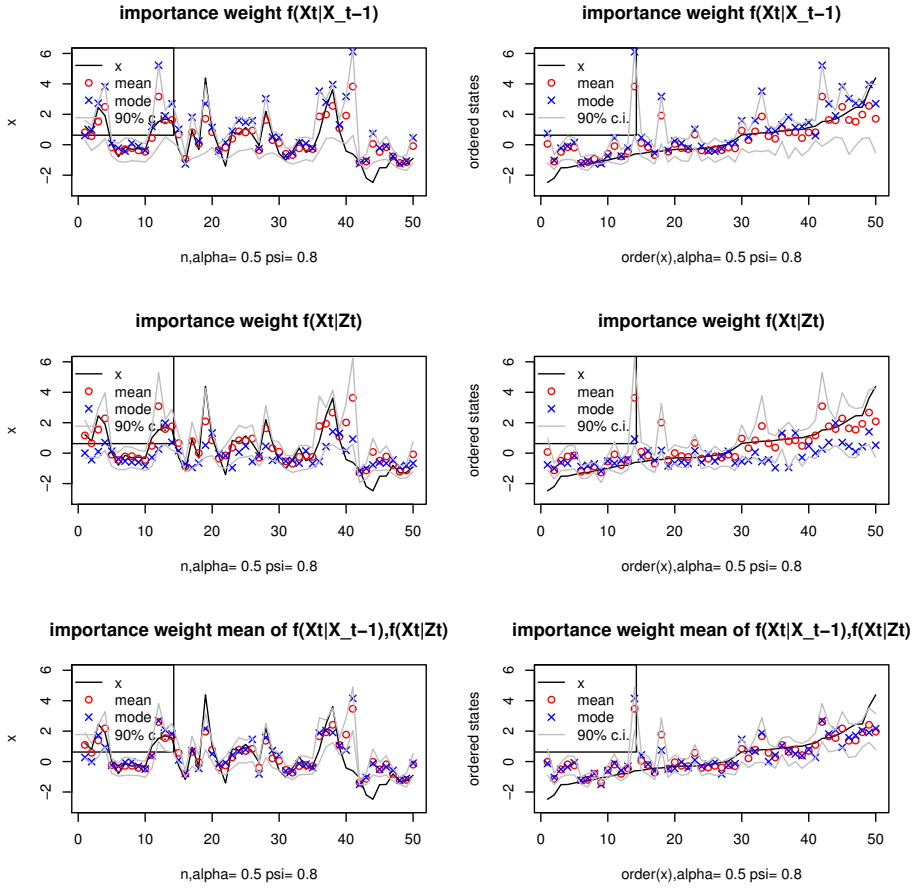

Figure 4.6: Particle filter with $(\alpha, \psi) = (0.5, 0.8)$.

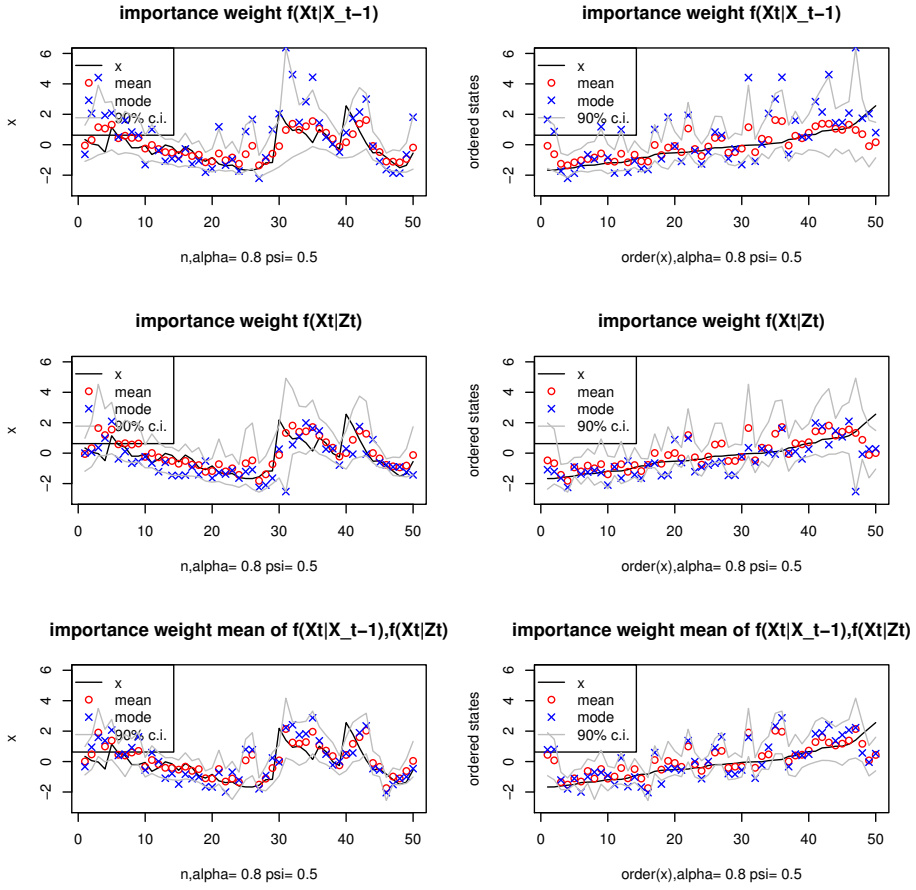

Figure 4.7: Particle filter with $(\alpha, \psi) = (0.8, 0.5)$.

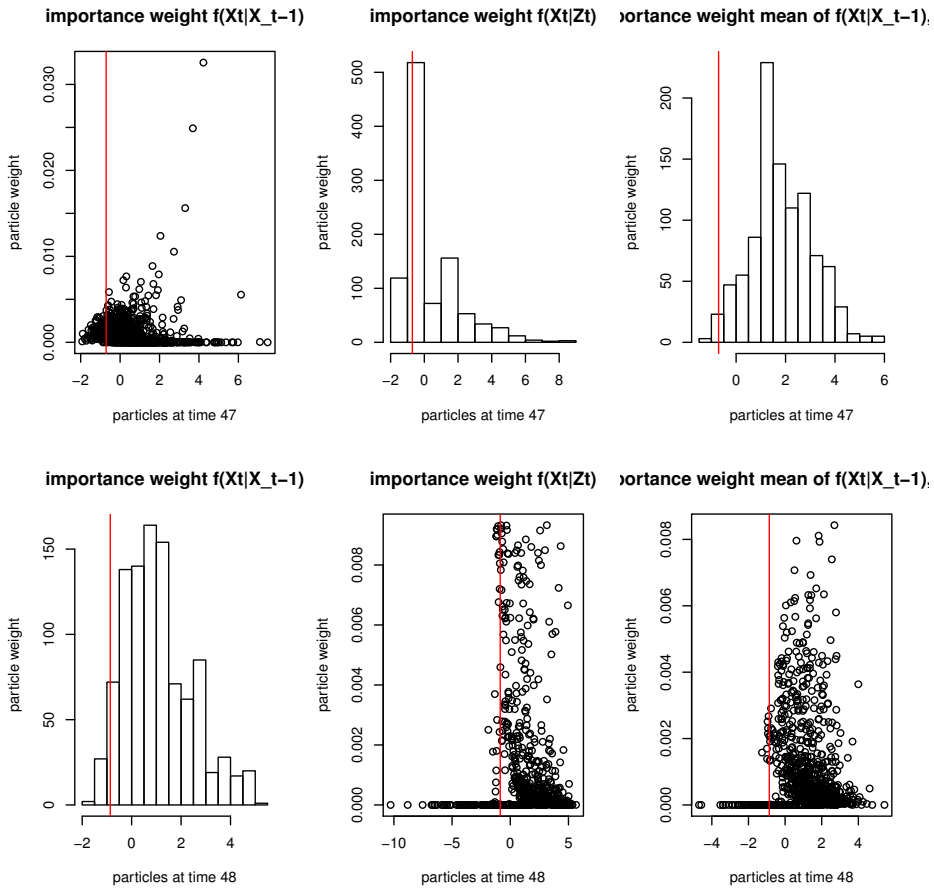

Figure 4.8: Empirical filtering density plots using different importance functions, $f(X_t|X_{t-1};\alpha)$, $f(X_t|Z_t;\psi)$, $\frac{1}{2}\left(f(X_t|Z_t)+f(X_t|X_{t-1})\right)$ from left to right.

(Figure 4.7) that using $f(X_t|Z_t;\psi)$ as the importance function performs better than using $f(X_t|X_{t-1};\alpha)$.

To see the empirical filtering density at a given time, the plots of the particles at time t=47, 48 in a state model with $\alpha = 0.8, \psi = 0.5, n = 50$ are shown in Figure 4.8. The red vertical line is the true value of the state. If weight degeneracy happens and the particles are obtained by resampling, the histogram plots are used, instead of the dot plots, to show the empirical filtering density. Each column represents the result with a different importance function.

4.3.4 Particles generated from a discrete sample space

From the simulation results of SIS particle filter, we can see that the way of generating particles affects the filtering results. Our aim is to estimate the empirical filtering density. For the given observations $X_{1:t-1}$, the joint density

$$\begin{aligned} f(Z_{1:t}, X_{1:t}) &= f(X_t|X_{1:t-1}, Z_{1:t}) f(X_{1:t-1}, Z_{1:t}) \\ &\propto f(X_t|X_{1:t-1}, Z_{1:t}) = f(X_t|X_{t-1}, Z_t). \end{aligned}$$

This makes $f(X_t|X_{t-1}, Z_t)$ an ideal importance function in particle filter.

$$\begin{aligned} f(X_t|X_{t-1}, Z_t) &= \frac{f(Z_t|X_t) f(X_t|X_{t-1})}{f(Z_t|X_{t-1})}, \\ &= e^{(\frac{1}{\alpha}-1)(X_t - \frac{Z_t}{\psi})} \frac{f_\alpha(e^{X_t/\alpha - X_{t-1}}) f_\psi(e^{Z_t/\psi - X_t})}{f_{\alpha\psi}(e^{Z_t/\alpha\psi - X_{t-1}})}, \end{aligned} \tag{4.4}$$

since

$$\begin{aligned} f(X_t|X_{t-1}) &= \frac{1}{\alpha} e^{X_t/\alpha - X_{t-1}} f_\alpha(e^{X_t/\alpha - X_{t-1}}), \\ f(Z_t|X_t) &= \frac{1}{\psi} e^{Z_t/\psi - X_t} f_\psi(e^{Z_t/\psi - X_t}), \\ f(Z_t|X_{t-1}) &= \frac{1}{\alpha\psi} e^{Z_t/\alpha\psi - X_{t-1}} f_{\alpha\psi}(e^{Z_t/\alpha\psi - X_{t-1}}). \end{aligned}$$

To carry out the simulation of $X_t \le x|Z_t, X_{t-1}$, observe that

$$\begin{aligned} \mathrm{P}(X_t \le x|Z_t, X_{t-1}) &= \mathrm{P}\left(\frac{Z_t}{\psi} - \log\xi_t \le x, \alpha X_{t-1} + \alpha \log X_t \le x\right) \\ &= \mathrm{P}\left(\log\xi_t \ge \frac{Z_t}{\psi} - x, \log S_t \le \frac{x}{\alpha} - X_{t-1}\right) \\ &= F_\alpha(e^{\frac{x}{\alpha} - X_{t-1}}) \left(1 - F_\psi(e^{\frac{Z_t}{\psi} - x})\right). \end{aligned}$$

The marginal distribution of the state variable is Gumbel, whose 99.95% confidence interval is $(-2, 10)$. Thus a discrete sample space, $\{d_i\}$, which contains M_d values evenly distributed in the interval $(-2, 10)$, can be treated as our particle sample

space. Let

$$\{d_j, 1 \le j \le M_d\} = \{-2 + \frac{12}{M_d} \times j, 1 \le j \le M_d\}. \tag{4.5}$$

For the given pair (Z_t, X_{t-1}), the value of $\mathrm{P}(X_t \le d_i | Z_t, X_{t-1})$, $1 \le i \le M_d$ can be calculated. Compare these probabilities with an uniform $(0, 1)$ distributed random variable, let the d_i, which minimizes the difference between $\mathrm{P}(X_t \le d_i | Z_t, X_{t-1})$ and the uniform random variable, be the i-th particle at time t. This method generates independent and identically distributed particles from the importance function $f(X_t | X_{t-1}, Z_t)$.

In this way, the continuous sample space is transformed to a discrete sample space. When $M_d \to \infty$, the discrete sample space is closed and dense, the difference between the distribution function of the sample on the continuous sample space and on the discrete sample space goes to zero, which means the samples generated from the discrete sample space have density $f(X_t | X_{t-1}, Z_t)$. When $M_d = 1000$, particles generated from the discrete sample space are almost the particles generate from $f(X_t | X_{t-1}, Z_t)$ rounded by two digits after decimal.

Notice that when M_d is large, the particle number M should also be large, $M >> M_d$.

The steps of particle filter with the particles generated from discrete sample space are stated below.

1. At time $t = 1$, denote

$$p_{1,j} = \mathrm{P}(X_1 \le d_j | Z_1, X_0), \quad 1 \le j \le M_d$$

   where $\{d_j\}$ is defined in (4.5).

   Generate M independent uniform (0,1) random variables $u_i, 1 \le i \le M$. For $1 \le i \le M$, define particle $X_1^{(i)} = d_s$ where

$$s = \arg \min_j |u_i - p_{t,j}|,$$

   and the importance weight $W_1^{(i)} = \frac{1}{M}, 1 \le i \le M$. Denote $(\hat{X}_1^{(i)}, \hat{W}_i^{(i)}) = (X_1^{(i)}, W_i^{(i)})$.

2. Recursive step

At time t, $2 \leq t \leq n$, for each $i, 1 \leq i \leq M$, generate M particles from $f(X_t|\hat{X}_{t-1}^{(i)}, Z_t)$ by the same way as at time $t = 1$. For a given particle $\hat{X}_{t-1}^{(i)}$ at time $t-1$, generate M independent random numbers, $u_j, 1 \leq j \leq M$, from uniform(0,1). Denote

$$p_{t,a}^{(i)} = \mathrm{P}(X_t \leq d_a|Z_t, \hat{X}_{t-1}^{(i)}), 1 \leq a \leq M_d.$$

Let particle $X_t^{(i,j)} = d_s$ where

$$s = \arg\min_a |u_j - p_{t,a}^{(i)}|,$$

and define the corresponding weight as

$$w_t^{(i,j)} = \hat{W}_{t-1}^{(i)} \frac{f(X_t^{(i,j)}|Z_t, \hat{X}_{t-1}^{(i)})}{\sum_{i=1}^{M} \hat{W}_{t-1}^{(i)} f(X_t^{(i,j)}|Z_t, \hat{X}_{t-1}^{(i)})}.$$

Standardize the weight to obtain the importance weight $W_t^{(i,j)} = \frac{w_t^{(i,j)}}{\sum_{i,j} w_t^{(i,j)}}$.

3. Now, we have particles $(X_t^{(i,j)}, W_t^{(i,j)}), 1 \leq i \leq M, 1 \leq j \leq M$. Let $\{\hat{X}_t^{(i)}\}$ be the unique values of $X_t^{(i,j)}, 1 \leq i \leq M, 1 \leq j \leq M$ and $\hat{W}_t^{(s)} = \sum_{X_t^{(i,j)} = \hat{X}_t^{(s)}} W_t^{(i,j)}$. For the set threshold C_{thr}, if $\frac{M}{\sum_i (\hat{W}_t^{(i)})^2} < C_{thr}$, resample the particles at time t, otherwise repeat the recursive step at time $t+1$.

4. Resample

   Denote the empirical distribution function of $(\hat{X}_t^{(i)}, \hat{W}_t^{(i)})$ as $\hat{F}_t$. Let particle $\hat{X}_t^{(i)} = X_t^{(s)}$ where

$$s = \arg\min_j |u_i - \hat{F}_t^{(j)}|, \quad 1 \leq i \leq M$$

   for a sequence of random numbers from $(0, 1)$. The corresponding weight $\hat{W}_t^{(i)}$ of particle $\hat{X}_t^{(i)}$ is $1/M$.

The threshold $M/\left(3\sum_{i=1}^{M}(W_2^{(i)})^2\right)$ is used as the criterion to judge whether weight degeneracy has happened in the simulation.

Figures 4.9 and 4.10 show the simulation results with $\alpha = \psi = 0.6$ and the chain size $n = 30$, $\alpha = \psi = 0.9$ and the chain size $n = 60$ respectively. Here the plots

are displayed in a different way. To reveal how the modes and means of the particles behave, the true values of hidden states are ordered and set as the line $y = x$. All the figures in the rest part of this section are displayed in this way.

The change from the continuous sample space to the discrete sample space provides an easier way to directly generate particles from the conditional density function (4.4).

A large particle number M is needed when comparing with M_d. If M is not large compared with M_d, for example, let $M_d = 1000$, i.e., the discrete particle sample points $d_i, 1 \leq i \leq 1000$, locate evenly on $[-2, 10]$ and let the particle number M be 1000 in simulation, we may have only few unique particles at some time point. Notice that for the next time point, particles are generated by the particles generated at this time and their corresponding weights are assigned proportional to the weight of particle generated at this time. This means that the number of particles generated at the next time depends on the current particles and their weights. If the unique particle number at current time is small, together with the situation that some particles have small weights, we would have the consequence that the particles generated at next time may lose information about the tail. Also the resampling step may not prevent weight degeneracy. To avoid these cases, a large number of particles are needed, which should be much larger than M_d, to better capture the tail behaviour of the hidden states.

Figure 4.11 shows the comparison of the simulation results using $M = 1000, 5000$ respectively, when $M_d = 1000$ for the same observation sequence. The modes and means of the particles are compared with the true state values, which is the line $y = x$ in the plots. We can see that most modes (crosses) and means (circles) of the generated particles are over estimated the states, i.e. on the upper side of the line, when the particle number is not large (left plot). As the particle number M increases, the modes and means are closer to the true state values.

On the other hand, we can decrease the value of M_d, instead of increasing M, to reduce the computing time without losing the advantage of generating particles from the discrete sample space. Figure 4.12 are the comparison of the simulation results when $M_d = 1000, M = 5000$, (left plot), and $M_d = 100$, i.e. $d_i = -2 + \frac{12}{100}, 1 \leq i \leq 100$, which is almost rounding the particles with one digit after decimal, $M = 5000$ (right plot) in the state space model with $\alpha = \psi = 0.7$. The left plot in Figure 4.12 is the same as the right plot in Figure 4.11, with the average absolute difference between means of the particles and true state values 0.8572, the average absolute difference

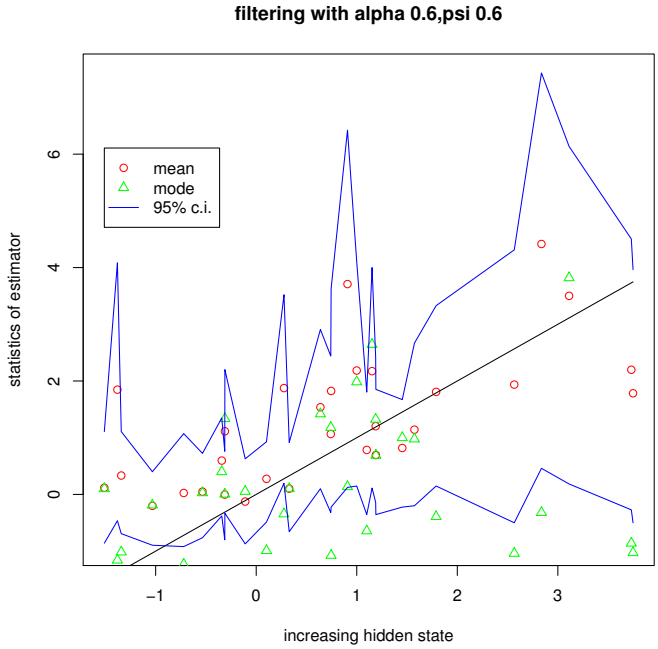

Figure 4.9: Particle filter with particles from the discrete sample space with chain size n=30. The true value of states are displayed as the line $y = x$.

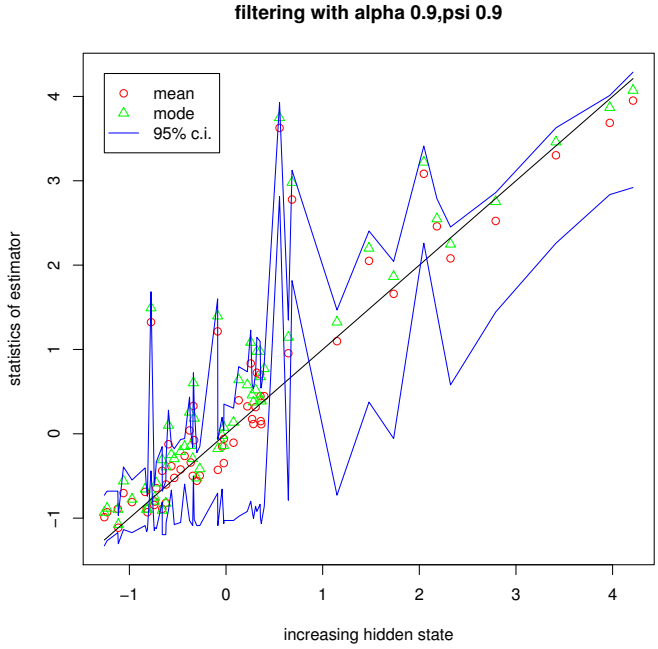

Figure 4.10: Particle filter with particles from the discrete sample space with chain size n=60. The true value of states are displayed as the line $y = x$.

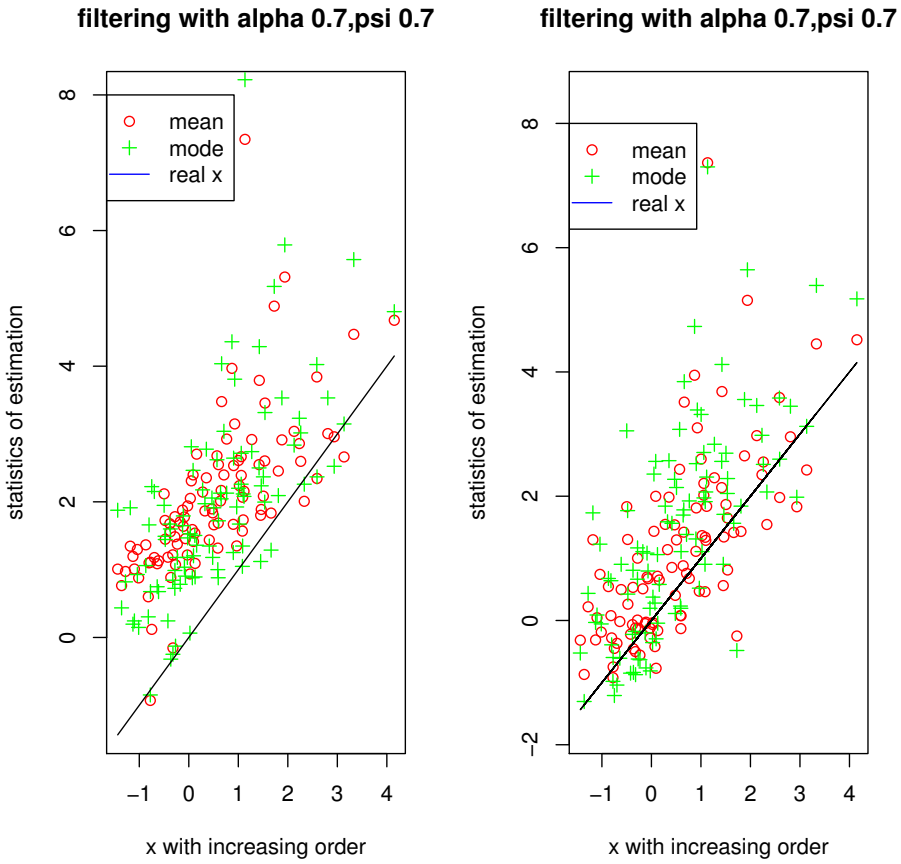

Figure 4.11: Comparison of means, modes of the states, with different discrete sample space size, M=1000(left), M=5000(right).

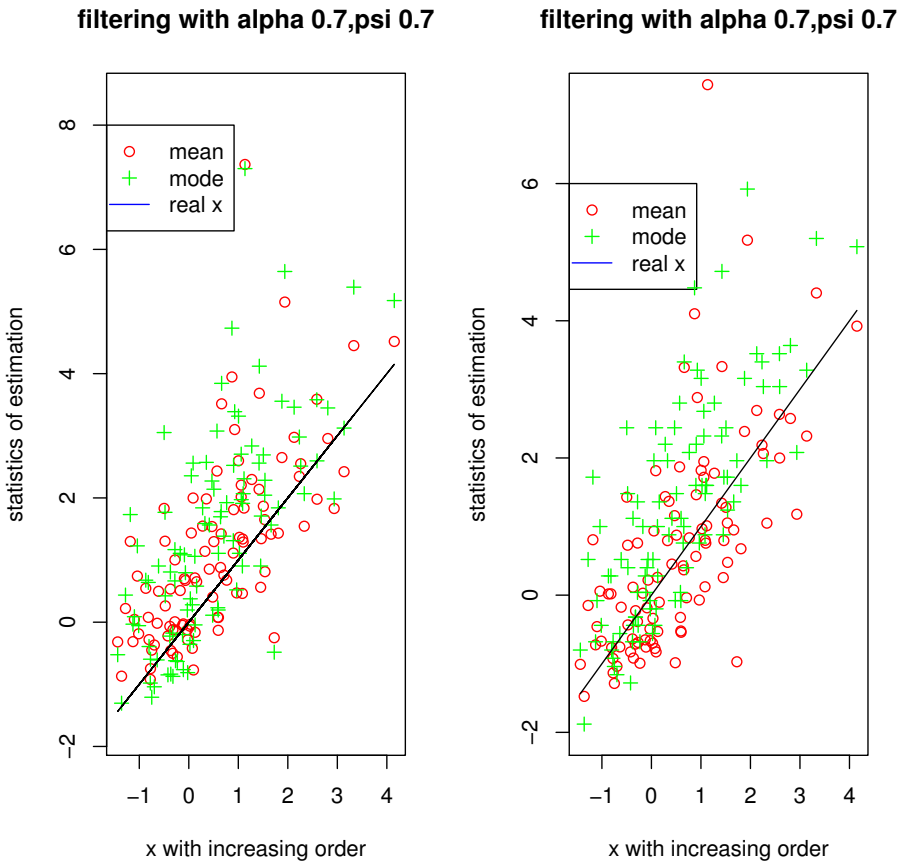

Figure 4.12: Particles generated from the discrete sample space with $M = 5000$, $M_d = 1000$ (left), $M_d = 100$(right).

between modes of the particles and true state values 1.1141. The right plot in Figure 4.12 has the average absolute difference between means of the particles and true state values 0.7934666, while the average absolute distances between the modes of particles and the real states is 0.99988.

The simulation result is also affected by the value of α, ψ. When the stability parameters are large, like $\alpha = 0.9$, and $\psi = 0.9$ (Figure 4.13), the estimation is better than when the stability parameters are small.

4.4 Auxiliary particle filter

So far, we mainly discussed the linear observation equation in the state space model with Gumbel distributed observations. For the non-linear observation equation in the

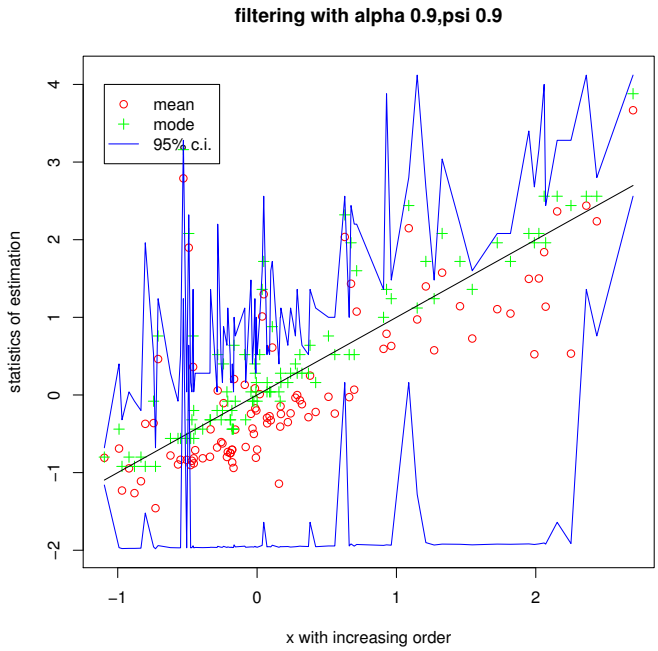

Figure 4.13: Particle filter with $M_d = 100, M = 5000$ for large stability parameters.

model with GEV distributed observations $\{Y_t\}$, another particle filter, the auxiliary particle filter, is a better option.

Auxiliary particle filter is an extend particle filter with an auxiliary variable, introduced by Pitt and Shephard in 1999 (see [80]).

Sequential importance sampling and auxiliary particle filter share the same idea: generate particles and assign weights to those particles to estimate the empirical filtering density. The difference between them is that when generating particles, auxiliary particle filter introduces an auxiliary variable to help. This auxiliary variable is related with the generated particles from SIS. It usually is the mean or the mode of the particles, or a quantity related with $X_{t+1}|X_t$.

Denote k as the auxiliary variable, define

$$f(X_{t+1}, k|Y_{t+1}) \propto f(Y_{t+1}|X_{t+1}) f(X_{t+1}|X_t^k) W_{t+1}^{(k)},$$

where X_t^k is the variable that depends on the auxiliary variable and the particles generated at time t. The sample draw from this joint density is thought to be the sample of $\hat{f}(X_{t+1}|Y_{t+1})$.

To perform an auxiliary particle filter, the following steps are done.

1. For time $t = 1$, generate m independently distributed particles $X_t^{(i)}, 1 \le i \le M$, from function $g(X_0|Y_t)$, denote the weight

$$w_t^{(i)} = \frac{f(Y_t|X_t^{(i)})f(X_t^{(i)}|X_0)}{g(X_0, k|Y_t)}, \qquad W_t^{(i)} = \frac{w_t^{(i)}}{\sum_{1\le i\le m} w_t^{(i)}}.$$

2. For time $2 \le t \le n$, generate m particles $X_t^{(i)}$ and $k^{(i)}, 1 \le i \le M$ from $g(X_t, k|Y_t)$,

$$w_t^{(i)} = \frac{f(Y_t|X_t^{(i)})f(X_t^{(i)}|X_{t-1}^{k^{(i)}})}{g(X_t^{(i)}, k^{(i)}|Y_t)}, \qquad W_t^{(i)} = \frac{w_t^{(i)}}{\sum_{1\le i\le m} w_t^{(i)}}.$$

   After obtaining the particles $(X_t^{(i)}, W_t^{(i)}), 1 \le i \le m$, do resampling if it is necessary (when weight degeneracy occurs).

The choice of function $g(X_t, k|Y_t)$ is flexible, Pitt and Shephard used the function

$$\begin{aligned} g(X_t, k|Y_t) \qquad &\propto f(Y_t|\mu_t^k)f(X_t|X_{t-1}^k)\pi^k \\ &= \int f(Y_t|\mu_t^k)dF(X_t|X_{t-1}^k)\pi^k \\ &= f(Y_t|\mu_t^k)\pi^k \end{aligned}$$

in [80], with the particle moments μ_t^k, and made the weight $w_t^i = f(Y_t|X_t^{(i)})/f(Y_t|\mu_t^{k^{(i)}})$.

For our state space model with GEV distributed marginals, the auxiliary particle filter includes the following steps.

1. Draw a random sample from the prior distribution of $X_0 \sim$Gumbel(0,1), denoted as $\{X_0^{(i)}, 1 \le i \le M\}$, with weight $W_t^{(i)} = 1/M$.

2. At time $t \ge 1$, update the particles of previous stage by letting

$$\tilde{X}_{t-1}^{(i)} = X_{t-1}^{(i)} - \theta \frac{\partial \left(Y_t - G(X_{t-1}, \tilde{\xi}_t)\right)^2}{\partial X_{t-1}}|_{X_{t-1}=X_{t-1}^{(i)}} \tag{4.6}$$

   where $Y_t = G(X_{t-1}, \xi_t)$ and $\theta \in [0, 1/10]$ is a constant.

3. Sample the new set of particles $\{X_t^{(i)}\}$ from the importance function

$$g(X_t|X_{t-1}^{(i)}; Y_{1:t}) = p(X_t|\tilde{X}_{t-1}^{(i)}).$$

4. Assign the weight

$$w_t^i = \left(W_{t-1}^{(i)}\right)^\rho \frac{f(Y_t|X_t^{(i)})p(X_t^{(i)}|X_{t-1}^{(i)})}{p(X_t|\tilde{X}_{t-1}^{(i)})}, \tag{4.7}$$

where $\rho \in [0, 1]$ is a constant. Then normalize the weight

$$W_t^{(i)} = \frac{w_t^{(i)}}{\sum_{1\le i\le M} w_t^{(i)}}.$$

To find out with which value of ρ and θ we can improve the auxiliary particle, we tried some simulations for the model with the chain size $n = 100$, $(\mu, \delta, \gamma) = (0, 1, -2)$, $\alpha = \psi = 0.8$ and compared the plots of the particles and their corresponding weights at time $t = 100$.

In Figure 4.14, $\theta = 0$ for all the plots, which means the auxiliary variable was not used. The way to simulate particles is the same as in the sequential importance sampling we did before. Its difference from SIS is that the weight function (4.7) depends on a power function of the past particle weight. The value of $\rho = 0, 0.3, 0.5, 0.7, 1$ have been used in the simulation. When $\rho = 1$, it reduces to the sequential importance sampling particle filter.

In Figure 4.14, the red vertical line is the true value of the hidden state X_n. Without the auxiliary variable ($\rho = 0$), the modes of the empirical filtering density at time n are not so close to the true value of X_{100}.

If the median of model error ξ_t is used as the auxiliary variable $\tilde{\xi}_t$ in Equation (4.6), let $\theta = 0, 0.03, 0.05, 0.07, 0.1$ and $\rho = \alpha$, the empirical filtering density at time $t = 100$, obtained by auxiliary particle filter, is shown in Figure 4.15 for the same observation sequence as used in Figure 4.14. When $\theta = 0$, i.e. generating particles the same way as the sequential importance sampling, with $\rho = \alpha$ in the weight function (4.7), the simulation result shows some improvement comparing with the SIS particle filter.

The plots showed in Figure 4.15 only have slightly differences, which means that the parameter θ may not contribute too much in auxiliary particle filtering. Compare

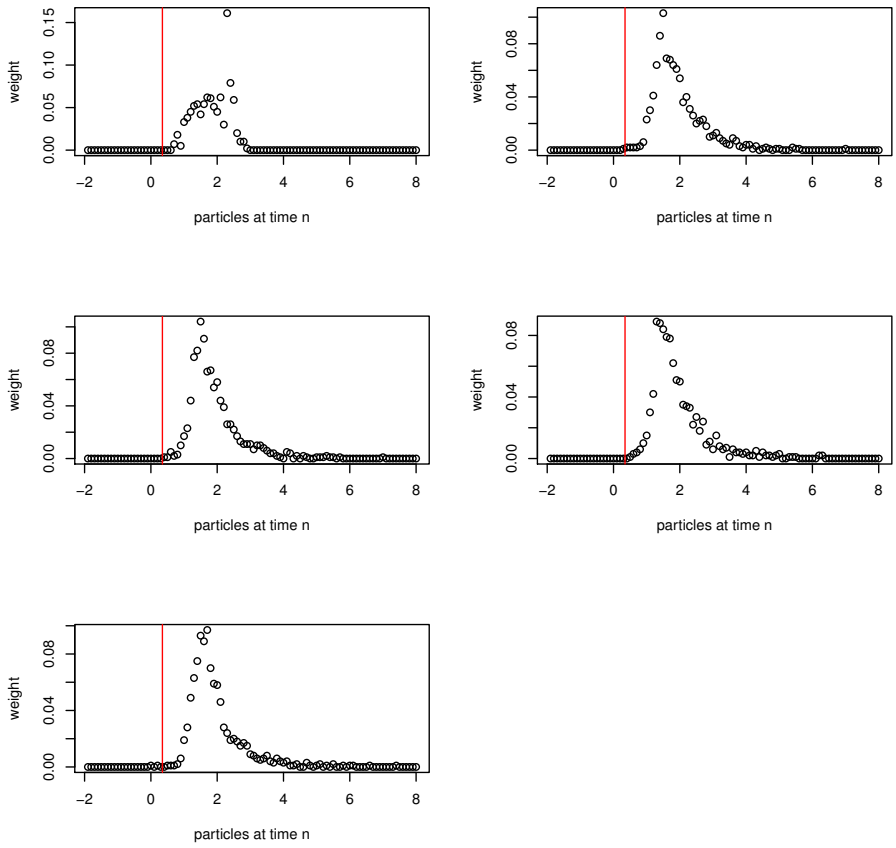

Figure 4.14: Empirical filtering density without auxiliary variable at n=100. $\theta = 0$, $\rho = 0, 0.3, 0.5, 0.7, 1$ separately (from left to right, top to bottom).

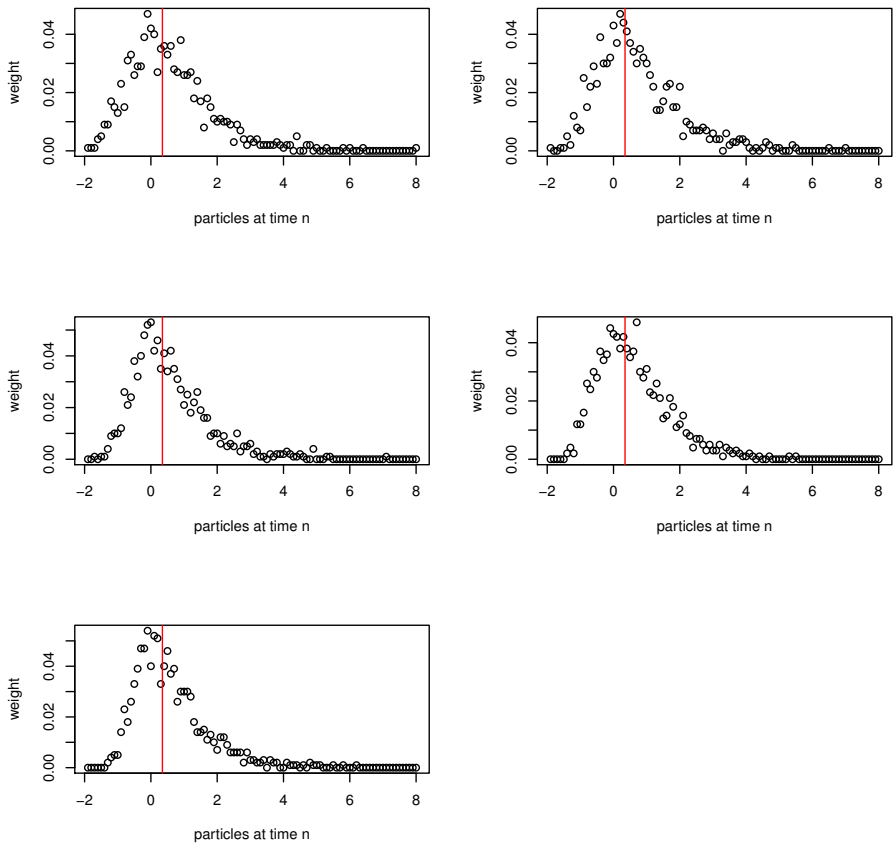

Figure 4.15: Empirical filtering density with the auxiliary variable evolved. $\theta = 0, 0.03, 0.05, 0.07, 0.1$ and $\rho = \alpha$.

all the plots in Figure 4.14 with the first plot in Figure 4.15, when $\theta = 0$, i.e. no auxiliary variable was used to generate particles, the best empirical filtering density is given by the importance weight function when $\rho = \alpha$. The improvements are mainly caused by the change of the importance weight function. Thus we consider linearization, which shares some similarities with the auxiliary particle filter on generating particles.

4.5 Plain linearization

For the GEV distributed observation Y_t, we have

$$\begin{aligned} Y_t &\approx Y_{t-1} + \frac{\partial Y_t}{\partial X_t}\Big|_{X_t=X_{t-1},\xi_t=\tilde{\xi}_t}(X_t - X_{t-1}), \\ &= Y_{t-1} + \sigma\psi e^{\psi\gamma X_{t-1}}\tilde{\xi}_t^{\psi\gamma}(X_t - X_{t-1}), \end{aligned}$$

thus

$$X_t \approx X_{t-1} + \frac{1}{\sigma\psi}e^{-\psi\gamma X_{t-1}}\tilde{\xi}_t^{-\psi\gamma}(Y_t - Y_{t-1}). \tag{4.8}$$

This approximating equation provides another way to generate particles. To obtain the importance function, we need to find the prior density $g(X_t|X_{t-1}, Y_t, Y_{t-1})$. By using the Equation (4.8) with ξ_t, we have

$$\begin{aligned} \mathrm{P}(X_t \le x|X_{t-1}, Y_t, Y_{t-1}) &= \mathrm{P}\left(\xi_t^{-\psi\gamma}(Y_t - Y_{t-1}) \le \sigma\psi e^{\psi\gamma X_{t-1}}(x - X_{t-1})\right) \\ &= \mathrm{P}\left(\xi_t \le (\psi\gamma)^{-\frac{1}{\psi\gamma}}\mathrm{e}^{-X_{t-1}}\left(\frac{x - X_{t-1}}{Y_t - Y_{t-1}}\right)^{-\frac{1}{\psi\gamma}}\right), \\ &\quad \text{if} \quad \gamma < 0, Y_t - Y_{t-1} > 0, \\ &= \mathrm{P}\left(\xi_t \ge (\psi\gamma)^{-\frac{1}{\psi\gamma}}\mathrm{e}^{-X_{t-1}}\left(\frac{x - X_{t-1}}{Y_t - Y_{t-1}}\right)^{-\frac{1}{\psi\gamma}}\right), \\ &\quad \text{if} \quad \gamma < 0, Y_t - Y_{t-1} < 0. \end{aligned}$$

The prior density when $\gamma < 0$ is

$$\begin{aligned} g(X_t|X_{t-1}, Y_t, Y_{t-1}) &= \left(-\frac{1}{\psi\gamma}\right)(\psi\gamma)^{-\frac{1}{\psi\gamma}}\mathrm{e}^{-X_{t-1}}|X_t - X_{t-1}|^{-\frac{1}{\psi\gamma}-1}|Y_t - Y_{t-1}|^{1/\psi\gamma} \\ &\quad \times f_\psi\left((\psi\gamma)^{-\frac{1}{\psi\gamma}}\mathrm{e}^{-X_{t-1}}\left(\frac{X_t - X_{t-1}}{Y_t - Y_{t-1}}\right)^{-\frac{1}{\psi\gamma}}\right). \end{aligned}$$

The corresponding importance weight for the generating Equation (4.8) is the normalized weight of

$$w_t^{(i)} = \left(w_{t-1}^{(i)}\right)^{\alpha}\frac{f(Y_t|X_t^{(i)})f(X_t^{(i)}|X_{t-1}^{(i)})}{g(X_t^{(i)}|X_{t-1}^{(i)}, Y_t, Y_{t-1})}.$$

As to the value of $\tilde{\xi}_t$ in Equation (4.8), median of $\mathcal{S}(\psi)$ is used since the α-stable distribution has no finite expectation.

To generate particles by linearization, we proceed the particle generating steps in the same way as in auxiliary particle filter. Particles are generated by the independent random variables $S_t^{(i)} \sim \mathrm{S}(\alpha)$ and the particles at the last time,

$$X_t^{(i)} = \alpha X_{t-1}^{(i)} + \alpha \log S_t^{(i)},$$

then updated by linearization. Notice that

$$\begin{aligned} Y_t &= G(X_t, \xi_t), \\ Y_t &\approx G(X_{t-1}, \xi_t) + \frac{\partial G(X_t, \xi_t)}{\partial X_t} \big|_{X_t = X_{t-1}} (X_t - X_{t-1}) \\ &= G(X_{t-1}, \xi_t) + \sigma\psi \mathrm{e}^{\psi\gamma X_{t-1}} \xi_t^{\psi\gamma} (X_t - X_{t-1}), \\ X_t &\approx X_{t-1} + \frac{1}{\sigma\psi} \mathrm{e}^{-\psi\gamma X_{t-1}} \xi_t^{-\psi\gamma} (Y_t - G(X_{t-1}, \xi_t)), \end{aligned} \tag{4.9}$$

denote the median of $\mathcal{S}(\psi)$ as $\tilde{\xi}_t$, we can obtain the updated particles by letting

$$\tilde{X}_t^{(i)} = X_{t-1}^{(i)} + \frac{Y_t - \left(\mu - \frac{\sigma}{\gamma} + \frac{\sigma}{\gamma} \mathrm{e}^{\psi\gamma X_{t-1}^{(i)}} \tilde{\xi}_t^{\psi\gamma}\right)}{\sigma\psi \mathrm{e}^{\psi\gamma X_t^{(i)}} \tilde{\xi}_t^{\psi\gamma}}.$$

The corresponding importance weight function for particle $\tilde{X}_t^i$ is the normalized weight of

$$w_t^{(i)} = w_{t-1}^{(i)} \frac{f(Z_t | X_t^{(i)}) f(X_t^{(i)} | X_{t-1}^{(i)})}{f(X_t^{(i)} | \tilde{X}_t^{(i)})}.$$

Figure 4.16 compares the density plots of particles at time $t = 39, 40$ generated by the plain linearization and the auxiliary particle filter. The left plot contains the particles and their corresponding weights at time $t = 39$, with circles represent the particles generated by plain linearization while crosses are the particles generated by auxiliary particle filter. The right plot corresponds to the comparison at time $t = 40$. The vertical line is the true state value. The simulation result shows that linearization gives better estimation of the empirical density of the states, compared with the auxiliary particle filter.

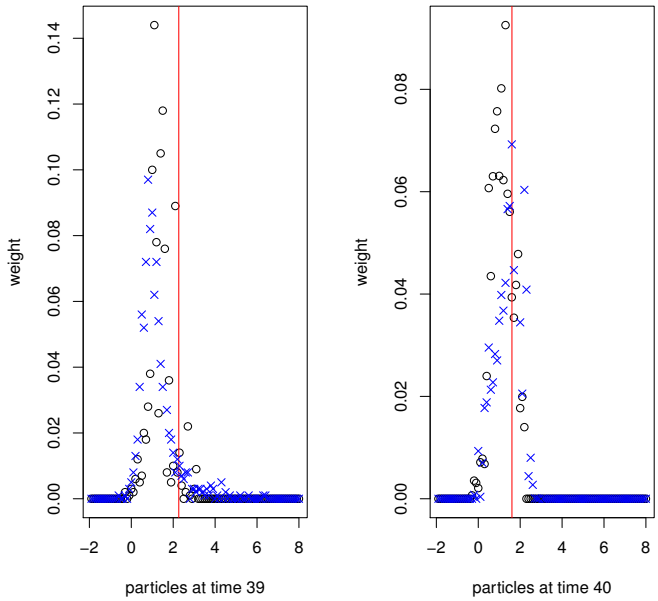

Figure 4.16: Comparison of the particles generated by the auxiliary particle filter (crosses) and the linearization (circles).

4.6 Properties of filtering

As to the qualitative properties of those discrete filters (Kalman filter, particle filter and auxiliary filter), there are many studies aimed to prove the uniform asymptotic stability (error goes to zero as time increases) or convergence result under some specific conditions but not general results.

Kalman filter, under certain conditions, such as with Gaussian state error and being uniformly completely observable and uniformly completely controllable, and as proved by Jazwinski [49], the errors of the filter stabilize as time increases and all eigenvalues of the error matrix have absolute value less than the unit. For the discrete time state space model with Gaussian original state and in finite dimensional setting, the convergence rate of the Kalman filter has been studied by Aalto [1]. The bounds of Kalman filter for a state space model with non-singular system matrix, bounded initial covariance and state error covariance matrix has been investigated in Rhudy and Gu [85].

Del Moral [22] uses martingales and semi-group techniques to analyze the asymptotic behaviour of the particle models. The convergence rate of the particle density is

obtained for the linear Gaussian filtering model.

Based on some assumptions of the prior distribution, Chopin [14] proved that the Central Limit Theorem holds for the filtering estimates produced by the particle filtering with residual resampling scheme. The stability and the asymptotic variance for a given particle filter is studied in the same paper. Later, the result that the asymptotic variance associated with some particle filters is bounded uniformly is demonstrated for some non-compact state spaces using Feynman-Kac formulas in Whiteley [96].

The properties like the convergence rate of the filtering method mentioned in this chapter will be discussed in the future.

Chapter 5

Real data analysis

The time series with Fréchet distributed marginals as stated in Equation (1.9) and the state space model (3.2) are used to analyze the weekly maximum pollution level of SO_2, CO, monitored by the monitor site 360050133 in New York city from January 1, 2016 to December 31, 2017. The data is from United States Environment Protection Agency (EPA).

The plots of the weekly maximum pollution levels are shown in the left column of Figure 5.1. There are 104 weeks, thus we have the chain size $n = 104$. With the seasonal trend removed and the dependence of the sequences ignored, the weekly maximum of SO_2 follows GEV(-0.63, 1.02, 0.04), which is a Fréchet distribution, or we can say it is close to a Gumbel distribution. The weekly maximum of CO follows GEV(-0.08, 0.15, -0.07), which is in Weibull family but is also close to a Gumbel distribution.

Since the location and scale parameters are not our main interest, we standardized the maxima of these two pollution sequences, shifted the location parameters to 0 and standardized them to scale 1 using the estimated GEV parameters. The plots of the standardized SO_2, CO weekly maximal sequences with seasonal trend removed are shown in the second column of Figure 5.1.

In the following analysis, all the sequences mentioned are standardized and with the seasonal trend removed.

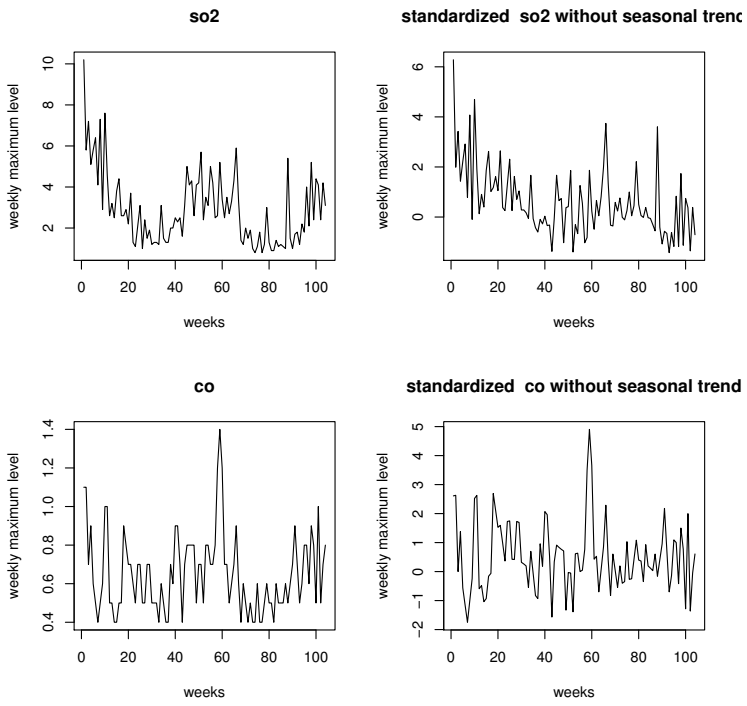

Figure 5.1: Weekly maximum of SO_2, CO from 2016 to 2017.

	SO_2		CO	
	α estimate	γ estimate	α estimate	γ estimate
Yule-Walker	0.2274209	0.9413648	0.4033022	1.0810085
AR(1)	0.2702623	0.6632850	0.4080463	0.6521737
Linear programming	0.2273278	1.0595494	0.3994137	1.2121325
Recursive Hill	0.7372432	1.0961531	0.6181514	1.2419783
Recursive Fan	0.4336363	1.2097567	0.6325037	1.4164079
K-W metric	0.3612906	1.0007979	0.2689868	1.1736196

Table 5.1: Estimates of α, γ for the transformed weekly maximum pollution levels of SO_2, CO.

5.1 Time series model

We assume the weekly maximum pollutant, SO_2, CO, are Gumbel distributed. To fit the time series model with Fréchet distributed marginals, the exponential function of these pollution data are used. With the standardized Gumbel distributed assumption, the marginal distribution of these exponential sequences follow Fréchet(0, 1, 1) distribution.

All the estimation procedures discussed in Chapter 2 are implemented to estimate the Fréchet tail parameter γ and the stability parameter α for the transformed weekly maxima. The estimation results are shown in Table 5.1.

For the exponential function of the SO_2 weekly maxima sequence, the recursive method using K-W metric provides the γ estimate closest to 1, Yule-Walker estimation and linear programming come the second, recursive Hill estimation also works okay. The AR(1) model gives the γ estimate far away from the other estimates, which coincides with the conclusion obtained in Chapter 2 that AR(1) estimates have large variance for γ. For α estimates, the recursive Hill method returns a large estimator when comparing with the other estimations. Our guess is that α is between 0.22 and 0.43.

For the time series of the exponential function of CO weekly maxima, the γ estimates are around 1.2 for most of the estimations, except Yule-Walker estimation and AR(1) model. Notice that the CO weekly maxima has Weibull marginal distribution, with the GEV tail parameter -0.07 when ignoring the dependence structure, to fit the time series model (1.9), we assumed it is a Gumbel distributed sequence, which may result that the γ here does not equal to 1. Our estimation results are acceptable here. We guess γ locates between 1.2 and 1.4. For α estimates, Yule-Walker estimation,

	SO_2		CO	
	α estimate	ψ estimate	α estimate	ψ estimate
Moments	0.9470124	0.5205703	0.1207518	1.690595
Regression	0.9473785	0.5205869	0.4037104	0.99999

Table 5.2: Estimates of α, ψ in the state space model with weekly maximum pollution levels of SO_2, CO.

AR(1) model and linear regression give the result around 0.4, while the recursive Hill and recursive Fan's estimation have the estimator around 0.6. Only K-W metric gives the α estimator less than 0.3.

5.2 State space model

Since the marginal distributions of the weekly maximum of SO_2, CO pollution levels are close to Gumbel, we can use the state space model with Gumbel distributed marginals (3.2) to fit them. Moment estimation and the regression model discussed in Chapter 3 are applied to those observations and the results are stated in Table 5.2.

For the weekly maximum of SO_2, the estimates obtained by the moment estimation and the regression model are similar, thus we believe the parameter α in this model is close to 0.95 and the ψ parameter is close to 0.52. As discussed in Chapter 3, when the stability parameters are large, the estimates obtained by these methods are more reliable and have smaller biases than the estimates obtained from a model with small stability parameters. Thus we believe the stability estimates for the weekly maximum of SO_2 in Table 5.2 are good.

To obtain the empirical filtering density, we tried SIS procedure with different importance functions. The α parameter is close to 1, thus using the smaller stability parameter ψ to generate particles can obtain particles in a wider range and decrease the chance of weight degeneracy. The means and 95% confidence intervals of the states using different importance functions are shown in the Figure 5.2. We can see the 95% confidence intervals of the states (grey broken lines in the left plot) generated by $f(X_{t-1}|X_t;\alpha)$ do not show too much changes as the observations change, while the 95% confidence intervals of the states generated by $f(X_t|Z_t;\psi)$ (grey broken lines in the right plot) changes when the observations increase or decrease.

With the particles and their corresponding weights, we can also obtain the estimator of the state functions.

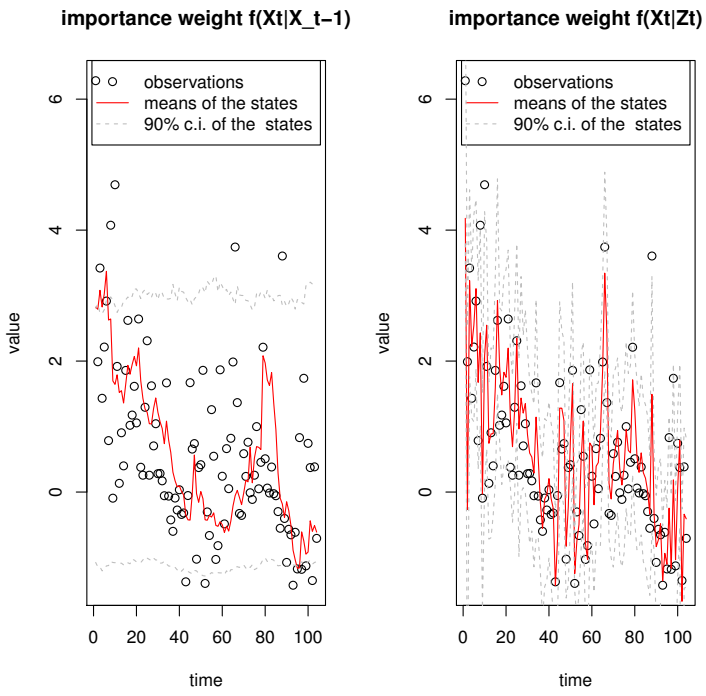

Figure 5.2: Means and 95% confidence intervals of the states using SIS in the model with weekly maximum pollutant SO_2.

As to the weekly maximum of CO, we cannot obtained a meaningful estimator of ψ, even by letting $\hat{\psi} = \hat{\psi}_k$ where $k = \arg\ \min_{1 \le s \le n/2}\{s; 0 < \hat{\psi}_s < 1\}$,

$$\hat{\psi}_s^2 = \frac{\mathrm{cov}(Z_{s:n-1}, Z_{s+1:n})}{\frac{\pi^2}{6}\hat{\alpha}},$$

with half of the observations been eliminated. To apply the regression model, we need reasonable initial values for stability parameters. The initial value outside the parameter space makes the estimation results obtained by the regression model not reliable.

In the moment estimation, the α estimates are obtained first, which is in the parameter space (0,1). We can use this α estimate to obtain ψ. Use the equation $\mathrm{Cov}(Z_t - \alpha Z_{t-1}, Z_{t-1} - \alpha Z_{t-2}) = -\alpha(1-\psi^2)\frac{\pi^2}{6}$ we have $\hat{\psi} = 0.8843999$. We used the Monte Carlo procedure stated in Section 3.3.1 (estimate one stability parameter when the other is known), let $\alpha = 0.1207518$ to estimate ψ and the hidden states. The ψ estimator obtained by this numerical method is 0.9972798. The simulation results stated in 3.3.1 show that when α is small, the ψ estimates may be over estimated. Here we assume that ψ takes the value approximate 0.95. Since the stability parameter α is small and ψ is large, the observations are heavily depend on the hidden states. In the estimation of the states, the states value are very close to the observations, thus the estimation result of the states and the filtering results are skipped here.

Chapter 6

Conclusion and future work

6.1 Conclusion

The estimation of a non-linear time series with Fréchet distributed marginals and the α-stable distributed errors, the estimation and filtering of a state space model with GEV distributed marginals and α-stable distributed errors are considered in this manuscript.

For the time series

$$X_t = X_{t-1}^{\alpha} S_t^{\frac{\alpha}{\gamma}},$$

where $\{S_t\}$ is a α-stable distributed error sequence following $\mathrm{S}(\alpha)$ with $\alpha \in (0,1)$, $\{X_t\}$ is a Fréchet$(0,1,\gamma), \gamma > 0$, distributed sequence for $t \in \mathbb{Z}$. The sequence $\{X_t\}$ satisfies strong mixing condition. Yule-Walker estimation, autoregressive model and linear programming can be applied to the sequence $\{\log X_t\}$ to estimate the stability parameter α and the Fréchet tail parameter γ. However, these estimation methods may return estimates outside the parameter space $(0,1)$. Besides this, the γ estimates obtained through the autoregressive model would have large variances. Thus, three recursive methods are developed to improve the estimates.

Since the sequence $\{X_t, 1 \leq t \leq n\}$ satisfies the strong mixing condition and the extremal index goes to 1 as $n \to \infty$, Hill estimation is considered. With an initial estimator of γ, in a second stage, the error term $S_t(\alpha) = X_t^{\hat{\gamma}/\alpha} X_{t-1}^{-\hat{\gamma}}$ can be treated as a function of α. Using the heavy-tail property of the α-stable distributed error sequence $\{S_t\}$, the estimator of α can be obtained by minimizing the distance between

α and the Hill estimator of α using the sequence $\{S_t(\alpha)\}$ (as stated in Equation (2.9)). The γ estimator is updated by minimizing the distance between γ and the Hill estimator of γ from the heavy tailed sequence $\{X_t^{1/\hat{\alpha}} X_{t-1}^{-\gamma}\}$, with the tail parameter $\gamma\hat{\alpha}$ (in Equation (2.10)). The α estimates and γ estimates are updated recursive until numerical convergence.

This method takes the advantage of the heavy tailed property of the α-stable sequence, returns estimates within the parameter space. The other advantage of the recursive Hill estimation is that the α estimates outperform Yule-Walker as well as the regression estimators when α is small, since the tail of the α-stable distribution is thicker when the stability parameter is smaller.

The asymptotic density of the order statistic of the sequence $\{X_t^{\hat{\gamma}/\alpha} X_{t-1}^{-\hat{\gamma}}\}$, where $\hat{\gamma}$ is the moment estimator of γ, is studied to gain insight regarding the reasonableness of using $\hat{\gamma}$ instead of γ in the updating step.

In addition to the heavy tailed property, the α-stable distributed error sequence has another property, stable property. Using the similar idea as the recursive Hill estimation, a recursive Fan's estimation is proposed to update the estimates. In the update step, Fan's estimation, using the sum of two independent α-stabled distributions is still α-stable distributed to estimate the stability parameter, is applied to the unobserved error sequence, to get the distance between α and α estimates before optimization. With reasonable initial estimates (e.g. a moment estimate of γ and a Yule-Walker estimate of α), the convergence rate of recursive Fan's estimator is $O_p(1/n)$.

Besides the moment estimator of γ, Kantorovich-Wasserstein metric provides another direction of estimation procedure. Since the sequence satisfies the strong mixing condition and the marginal distribution is known, minimizing the distance between the empirical distribution function of the observations and the Fréchet distribution function can provide an estimator of γ. Using this estimator, the distance function between the Laplace transform of $X_t X_{t-1}^{-\alpha}$ and $S_t^{\alpha/\gamma}$ can be minimized to obtain the estimator of α. These estimates can be updated recursive. The asymptotic convergence rate of $\hat{\alpha}$ to α depends on the convergence rate of $\hat{\gamma}$, which is faster than $\sqrt{\frac{\log\log n}{n}}$.

As for the estimation of the state space model, first, we considered the linear state space model with Gumbel distributed marginals, for simplicity. To transform the GEV distributed observations to Gumbel, GEV parameters are required. A recursive regression model is proposed, stated in Theorem 13, for the heavy tailed marginals.

This model uses the property that the density at the neighbourhood of lower bound or upper bound of the support set goes to infinity, thus works well when the sample size is small.

After estimating the GEV parameters, the GEV distributed observations can be transformed to Gumbel and the state space model becomes linear. However, it is difficult to use the observation sequence to estimate two stability parameters at the same time. Thus, we considered the case that one stability parameter is known.

For the state space model with Gumbel distributed marginals

$$Z_t = \psi X_t + \psi \xi_t,$$
$$X_t = \alpha X_{t-1} + \alpha \log S_t,$$

where $\xi_t \sim \mathrm{S}(\psi)$ and $S_t \sim \mathrm{S}(\alpha)$ are independent α-stable distributed errors and $\alpha, \psi \in (0, 1)$, if α is known, in addition to the moment estimation, I proposed a Monte Carlo procedure to estimate ψ. This procedure uses the similar idea as the particle filter, generating particles at each time point using the known α and assigning weights to the particles by the moment estimation of ψ. The mode of the particles at each time point can be thought as the estimator of the state. The stability parameter can be estimated after obtaining the state estimates. When comparing with the moment estimation, this method gives better estimation result especially when α is small.

If ψ is know, another Monte Carlo procedure is proposed, which has a slight difference as we performed before. A random sample is generated from $\mathrm{S}(\psi)$, as the error sequence used in the model. The model structure is used to input the error to a most plausible time point according the generated random sample. This procedure works well for large ψ, but not effect for small ψ since the generated sample set would be far away from the real errors due to the heavy tail of $\mathrm{S}(\psi)$.

For the case that both stability parameters are unknown, Yule-Walker estimation and regression model are used. Yule-Walker estimates have the asymptotic normality, however the bias is large when the size of the series is small. There is no procedure that works well when the stability parameters are small, due to the weakly dependent structure.

After estimation, we did model filtering, estimating the states and the empirical filtering densities. For the state estimation, two random samples are generated, one from $\mathrm{S}(\alpha)$ and the other from $\mathrm{S}(\psi)$, treated as the errors used in the state space model. The same idea to estimate ψ when α is known is used here. I proposed a Monte Carlo

procedure to estimate the errors and the states based on the model structure. This procedure works better than Kalman filter. However, the quality of the estimates is very poor when either the stability parameters is small.

For the empirical filtering density estimation, three filtering procedures are applied. The first one is the sequential importance sampling (SIS), with three different importance functions and importance weight functions, built on the state space model with Gumbel distributed marginals. To generate good particles and reduce the probability of weight degeneracy, an importance function that can generate particles in a wider range should be chosen. This depends on the value of α, ψ. Theoretically, $f(X_t|X_{t-1}, Z_t)$ is an ideal importance function, however there is not practical way to sample directly from it. We used the distribution function $F(X_t|X_{t-1}, Z_t)$ instead, transforming the continuous particle sample space into a discrete one and performed the SIS.

The second and third filtering procedures, auxiliary particle filter and linearization, are applied to the state space model with GEV distributed marginals. The auxiliary particle filter includes two constants and one auxiliary variable, where one constant and the auxiliary variable are used in the particle generating step, the other constant is used as the power in the weight function. The linearization uses SIS and Taylor's expansion with the median of the error term to generate particles. Linearization works better than the auxiliary particle filter in the simulation.

The time series and the state space model with Gumbel distributed marginals are applied to two air pollution data in New York city. All the estimation methods and some of the filtering methods discussed in the manuscript are performed and compared.

6.2 Future work

There are many interesting problems that need to be addressed for state space models with GEV distributed marginals. For the linear state space model with Gumbel distributed marginals, I think it is still worth to answer the question of whether there are better options to estimate the stability parameters than the Yule-Walker estimation.

For the state space model with GEV distributed marginals, the GEV parameters

are estimated without considering the dependence structure of the observation sequence in this manuscript, which increases the error of stability estimates. A better estimation should consider the effect of the stability parameters when estimating the GEV parameters. We will consider other optimization methods in the future.

Some Monte Carlo procedures are used to estimate one stability parameters in the state space mode. The asymptotically properties of the estimates should also be addressed in the future.

There are issues with filtering that should be addressed. All the asymptotic properties are shown by simulation result in this manuscript. The result of convergence rates, the bounded property, the asymptotic variances of the filtering in this kind of models haven't been found in the literature.

In addition to the state space model with GEV distributed marginals and the α-stable distributed errors, the estimation of other models like the GEV-M3 model in Kunihama et al. [60] and the estimation of max-stable model in Naveau and Poncet [71] can be considered.

State space models for extreme value data are needed in many fields, like in economics and in applied science area, including physical, biological and social, that models with heavy tailed errors can be applied.

Bibliography

[1] A. Aalto. Convergence of discrete-time Kalman filter estimate to continuous time estimate. *International Journal of Control*, 89(4):668–679, 2016.

[2] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Satatistical Society*, 3(72):269–342, 2010.

[3] S. R. Asmussen. Steady-state properties of GI/G/1. applied probability and queues. *Stochastic Modelling and Applied Probability*, 51:266–301, 2003.

[4] Z. K. T. Balch and F. Dellaert. An MCMC-based particle filter for tracking multiple interacting targets. *Computer Vision: Lecture Notes in Computer Science*, 3024:279–290, 2004.

[5] R. H. Berk and D. H. Jones. Goodness-of-fit test statistics that dominate the Kolmogorov statistics. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 47(1):47–59, 1979.

[6] P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York, Third edition, 1995.

[7] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. *Time Series Analysis: Forecasting and Control*. Prentice Hall, 1976.

[8] R. C. Brandley. Basic properties of strong mixing conditions, a survey and some open questions. *Probability Surveys*, 2:107–144, 2005.

[9] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, New York, 2nd edition, 2009.

[10] E. Blviken, P. J. Acklam, N. Christophersen, and J.-M. Strdal. Monte Carlo filters for nonlinear state estimation. *Automatica*, 37:177–183, 2001.

[11] J. M. Chambers, C. L. Mallows, and B. W. Stuck. A method for simulation stable random variables. *Journal of the American Statistical Association*, 71(354):340–344, 1976.

[12] F. Chamú Morales. *Estimation of max-stable processes using Monte Carlo methods with applications to financial risk assessment.* PhD dissertation, Department of Statistics, University of North Carolina, Chapel Hill, 2005.

[13] X. Chen, L. P. Hansen, and M. Carrasco. Nonlinearity and temporal dependence. *Journal of Econometrics*, 155:155–169, 2010.

[14] N. Chopin. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *The Annals of Statistics*, 32(6):2385–2411, 2004.

[15] S. Coles. *Introduction to statistical modeling of extreme values.* Springer, London, 2001.

[16] P. Cowpertwait and A. Metcalfe. *Introductory Time Series with R.* Springer, 2009.

[17] D. Crisan and A. Doucet. Convergence of sequential Monte Carlo methods. Technical report, Cambridge University, 2000. CUED/F-INFENG/TR381.

[18] M. E. Crovella and M. S. Taqqu. Estimating the heavy tail index from scaling properties. *Methodology and Computing in Applied Probability*, 1:55–57, 1999.

[19] M. J. Crowder. A multivariate distribution with Weibull components. *Journal of the Royal Statistical Society: Series B*, 51:93–108, 1989.

[20] R. A. Davis. Stable limits for partial sums of dependent random variables. *The Annals of Probability*, 11(2), 1983.

[21] H. Dehling and W. Philipp. Almost sure invariance principles for weakly dependent vecter-valued random variables. *The Annals of Probability*, 10(3):689–701, 1982.

[22] P. del Moral. *Feynman-Kac Formulae.* Springer, New York, 2004.

[23] A. Doucet, S. J. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.

[24] A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: fifteen years later. Technical report, Department of Statistics, University of British Columbia, 2008.

[25] J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods.* Oxford University Press, 2001.

[26] G. A. Einicke and L. B. White. Robust extended Kalman filtering. *IEEE Transactions on Signal Processing*, 47(9):2596–2599, 1999.

[27] M. Falk, J. Hüsler, and R. D. Reiss. *Laws of Small Numbers: Extremes and Rare Events*. Springer, 2010.

[28] Z. Fan. Parameter estimation of stable distributions. *Communications in Statistics-Theory and Methods*, 35:245–255, 2006.

[29] W. Feller. *An Introduction to Probability Theory and its Applications*, volume 2. John Wiley & Sons, Inc, New York, 1966.

[30] R. A. Fisher and L. H. C. Tippett. On the estimation of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24:180–190, 1928.

[31] A. L. Fougéres, J. P. Nolan, and H. Rootzén. Models for dependant extremes using stable mixtures. *Scandinavian Journal of Statistics*, 36(1):42–59, 2009.

[32] M. Fréchet. Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathematique Cracovie*, 6:93–116, 1927.

[33] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[34] J. D. Gibbons. *Nonparametric statistical inference*. Marcel Dekker Inc, New York, 1992.

[35] J. Gil-Peláez. Note on the inversion theorem. *Biometrika*, 38(3-4):481–482, 1951.

[36] N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE proceedings Part F, radar and signal processing*, 140(2):107–113, 1993.

[37] C. Gourieroux and A. Monfort. *Time Series and Dynamic Models*. Cambridge University Press, 1997.

[38] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995.

[39] E. J. Gumbel. *Statistics of Extremes*. Columbia University Press, New York, 1958.

[40] F. Gustafsson and G. Hendeby. Some relations between extended and unscented Kalman filters. *Signal Processing, IEEE Transactions*, 60(2):545–555, 2012.

[41] P. Hall, L. Peng, and Q. Yao. Moving-maximum models for extrema of time series. *Journal of Statistical Planning and Inference*, 103:51–63, 2002.

[42] J. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

[43] G. Hardy. *Divergent series*. Oxford at the Clarendon Press, 1949.

[44] P. Hartman and A. Wintner. On the law of the iterated logarithm. *American Journal of Mathematics*, 63:169–176, 1941.

[45] P. Hougaard. A class of multivariate failure time distributions. *Biometrika*, 73:671–678, 1986.

[46] J. Hüsler. Extreme values and rare events of non-stationary random sequences. *Dependence in Probability and Statistics*, pages 438–456, 1986.

[47] J. Hüsler. Extreme values of non-stationary random sequences. *Journal of Applied Probability*, 23:937–950, 1986.

[48] J. Hüsler. A note on exceedances and rare events of non-stationary sequences. *Journal of Applied Probability*, 30(4):877–888, 1993.

[49] A. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.

[50] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, pages 401–422, 2004.

[51] S. Jun, R. Chen, and T. Logvinenko. A theoretical framework for sequential importance sampling with resampling. *Sequential Monte Carlo Methods in Practise; Statistics for Engineering and Information Science*, pages 225–246, 2001.

[52] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

[53] M. Kanter. Stable densities under change of scale and total variation inequalities. *The Annals of Probability*, 3(4):697–707, 1975.

[54] B. Kedem and K. Fokianos. *Regression Models for Time Series Analysis*. Wiley, 2002.

[55] G. Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996.

[56] A. N. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Giornale dellIstituto Italiano degli Attuari*, 4:83–91, 1933.

[57] A. N. Kolmogorov and Y. Rozanov. On strong mixing conditions for stationary Gaussian processes. *Theory of Probability and Its Applications*, 5:204–208, 1960.

[58] A. Kong, J. S. Liu, and W. H. Wong. Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, 89(425):278–288, 1994.

[59] A. J. Koning and L. Peng. Goodness-of-fit tests for a heavy tailed distribution. *Econometric Institute Report*, 44, 2005.

[60] T. Kunihama, Y. Omorib, and Z. Zhang. Efficient estimation and particle filter for max-stable processes. *Journal of time series analysis*, 33:61–80, 2010.

[61] M. Leadbetter and H. Rootzén. On extreme values in stationary random fields. *Stochastic processes and related topics*, pages 275–285, 1988.

[62] M. R. Leadbetter and H. Rootzén. *Extremes and Related Properties of Random Sequences and Series*. Springer Verlag, New York, 1983.

[63] M. R. Leadbetter and H. Rootzén. Special invited paper: Extremal theory for stochastic processes. *The Annals of Probability*, 16(2):431–478, 1988.

[64] S. Ling and L. Peng. Hill's estimator for the tail index of an ARMA model. *Journal of Statistical Planning and Inference*, 123:279–293, 2004.

[65] J. S. Liu and R. Chen. Blind deconvollution via sequential imputations. *Journal of the American Statistical Association*, 90(430):567–576, 1995.

[66] J. S. Liu, R. Chen, and W. H. Wong. Rejection control and sequential importance sampling. *Journal of the American Statistical Association*, 93:1022–1031, 1998.

[67] S. N. Maceachern, M. Clyde, and J. S. Liu. Sequential importance sampling for nonparametric Bayes models: The next generation. *Canadian Journal of Statistics*, 27:251–267, 1999.

[68] D. L. McLeish. Invariance principles for dependent random variables. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, pages 165–178, 1975.

[69] J. Mikusinski. On the function whose Laplace transform is $e^{-t^{\alpha}}$. *Studia Mathematica*, 18(XVIII):191–198, 1959.

[70] J. Nakajima, T. Kunihama, Y. Omori, and S. Frühwirth-Schnatter. Generalized extreme value distribution with time-dependence using the AR and MA models in state space form. *Computational Statistics and Data Analysis*, 56(11):3241–3259, 2012.

[71] P. Naveau and P. Poncet. State-space models for maxima precipitation. *Journal de la Societe Francaise de Statistique et Revue de Statistique Appliquee*, 148:107–120, 2007.

[72] G. Neil. *The Nature of Mathematical Modeling*. Cambridge University Press, 2000.

[73] J. P. Nolan. Numerical calculation of stable densities and distribution functions. *Stochastic Models*, 13(4):759–774, 1997.

[74] J. P. Nolan. Parameterizations and modes of stable distributions. *Statistics & Probability Letters*, 38:187–195, 1998.

[75] J. P. Nolan. An algorithm for evaluating stable densities in Zolotarev's (m) parameterization. *Mathematical and Computer Modelling*, 29:229–233, 1999.

[76] J. P. Nolan. Maximum likelihood estimation and diagnostics for stable distributions. *Lévy Process*, pages 379–400, 2001.

[77] L. Pereira and H. Ferreira. Limiting crossing probabilities of random fields. *Journal of Applied Probability*, 43(3):884–891, 2006.

[78] J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131, 1975.

[79] M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599, 1999.

[80] K. Pitt Michael and N. Shephard. Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association*, 94:590–599, 1999.

[81] S. Resnick and C. Stric. Asymptotic behaviour of Hill's estimator for autogregressive data. *Communications in Statistics. Stochastic Models*, 13:703–721, 1997.

[82] S. Resnick and C. Stric. Tail index estimation for dependent data. *The Annals of Applied Probability*, 8(4):1156–1183, 1998.

[83] S. I. Resnick. *Extreme values, regular variation, and point processes*. Springer-Verlag, New York, 1987.

[84] M. K. Reznik. The law of the iterated logarithm for some classes of stationary processes. *Theory of Probability and its Application*, 13(4):606–621, 1968.

[85] M. Rhudy and Y. Gu. Online stochastic convergence analysis of the Kalman filter. *International Journal of Stochastic Analysis*, 2013.

[86] H. Rootzén, M. R. Leadbetter, and L. de Haan. Tail and quantile estimation for strongly mixing stationary sequences. Technical Report 292, Center for Stochastic Processes, Department of Statistics, University of North Carolina, 1990.

[87] M. Rosenblatt. A central limit theorem and a strong mixing condition. *National Academy of Sciences*, 42(1):43–47, 1956.

[88] G. Stoica. A not on the rate of convergence in the strong law of large numbers for martingales. *Journal of Mathematical Analysis and Applications*, 381:910–913, 2011.

[89] J. A. Tawn. Modelling multivariate extreme value distributions. *Biometrika*, 77:245–253, 1990.

[90] J. L. Teugels. *The Class of Subexponential Distributions*. Annals of Probability, 1975.

[91] E. Titchmarch. *The Theory of the Riemann Zeta-Function*. Oxford University Press, 2nd edition, 1939.

[92] G. Toulemonde, A. Guillou, and P. Naveau. Particle filtering for Gumbel distributed daily maxima of methane and nitrous oxide. *Environmetrics*, 24(1):51–62, 2012.

[93] G. Toulemonde, A. Guillou, P. Naveau, M. Vrac, and F. Chevallier. Autoregressive models for maxima and their applications to CH_4 and N_2O. *Environmetrics*, 4(21):189–207, 2010.

[94] R. van der Merwe, J. F. G. de Freitas, A. Doucet, and E. Wan. *The unscented particle filter*. Tech. Rep. CUED/F- INFENG/TR 380, Cambridge Unversity, 2001.

[95] E. Wan and R. van der Merwe. *The unscented Kalman filter, Kalman Filtering and Neural Networks*. Wiley, New York, 2001.

[96] N. Whiteley. Stability properties of some particle filters. *The Annals of Applied Probability*, 23(6):2500–2537, 2013.

[97] M. B. Wilk and R. Gnanadesikan. Probability plotting methods for the analysis of data. *Biometrika*, 55(1):1–17, 1968.

[98] V. M. Zolotarev. *One-dimensional stable distributions*. American Mathematical Society, 1986.

Appendix A

Proof of Theorem 13

To show Theorem 13, we need the following lemmas:

Lemma 19. *The maximum of the sample $Y_{(n)}$ from the independent GEV(μ, σ, γ) sequence converges in probability to the upper bound $\mu - \frac{\sigma}{\gamma}$ when $\gamma < -1$.*

Lemma 20. *In Equation* (3.6)*,*

$$\eta_{n,i} = \frac{\gamma \epsilon_{n,i}}{\frac{i}{n+1} \log \frac{i}{n+1}} + o_p(\epsilon_{n,i}) - \frac{\mu - \frac{\sigma}{\gamma} - Y_{(n)}}{\mu - \frac{\sigma}{\gamma} - Y_{(i)}} + o_p\left(\frac{\mu - \frac{\sigma}{\gamma} - Y_{(n)}}{\mu - \frac{\sigma}{\gamma} - Y_{(i)}}\right)$$

where

$$\epsilon_{n,i} = F(Y_{(i)}) - \frac{i}{n+1}.$$

Lemma 21. *For* $1 \leq i < j \leq n$,

$$\begin{aligned}
&\mathrm{E}(\epsilon_{n,i}) = 0, \\
&\mathrm{Var}(\epsilon_{n,i}) = \frac{in + i - i^2}{(n+1)^2(n+2)}, \\
&\mathrm{Cov}(\epsilon_{n,i}, \epsilon_{n,j}) = \frac{in + i - ij}{(n+1)^2(n+2)} + \frac{ij}{(n+1)^2}.
\end{aligned}$$

As $\eta_{n,i}$ depends on the ratio of the order statistics $\frac{\mu - \frac{\sigma}{\gamma} - Y_{(n)}}{\mu - \frac{\sigma}{\gamma} - Y_{(i)}}$, we have

Lemma 22. *Using Taylor expansion of two-dimensional equation,*

$$\begin{aligned}\frac{\mu-\frac{\sigma}{\gamma}-Y_{(n)}}{\mu-\frac{\sigma}{\gamma}-Y_{(i)}} = \quad & \frac{E_n}{E_i}-\frac{\partial F^{-1}(\frac{n}{n+1})}{E_i}\epsilon_{n,n}+\frac{E_n\partial F^{-1}\left(\frac{i}{n+1}\right)}{E_i}\epsilon_{n,i}-\frac{\partial F^{-1}\left(\frac{i}{n+1}\right)\partial F^{-1}(\frac{n}{n+1})}{2E_i^2}\epsilon_{n,i}\epsilon_{n,n}\\ &+\frac{E_n(\partial F^{-1}(\frac{n}{n+1}))^2}{E_i^3}\epsilon_{n,i}^2-\frac{\partial^2 F^{-1}(\frac{n}{n+1})}{2E_i}\left(\epsilon_{n,n}^2-\mathrm{E}(\epsilon_{n,n}^2)\right)\\ &+\frac{E_n\partial^2 F^{-1}\left(\frac{i}{n+1}\right)}{2E_i}\left(\epsilon_{n,i}^2-\mathrm{E}(\epsilon_{n,i}^2)\right)+o_p(\epsilon_{n,i}),\end{aligned}$$

Now we need to prove these lemmas. The proof of lemma 19 comes first:

Proof of Lemma 19. The density function of $\mathrm{GEV}(\mu,\sigma,\gamma)$ is

$$f(y)=\frac{1}{\sigma}(1+\gamma\frac{y-\mu}{\sigma})^{-1-1/\gamma}e^{-(1+\gamma\frac{y-\mu}{\sigma})^{-1/\gamma}}.$$

When $y\to\mu-\frac{\sigma}{\gamma}$, which is $1+\gamma\frac{y-\mu}{\sigma}\to 0$, we have $f(y)\to\infty$. This means no matter how small the sample size n is, there exists a sample closed enough to the upper bound $\mu-\frac{\sigma}{\gamma}$, thus $Y_{(n)}\xrightarrow{p}\mu-\frac{\sigma}{\gamma}$, since

$$\mathrm{P}\left(\left|Y_{(n)}-\mu+\frac{\sigma}{\gamma}\right|>\varepsilon\right)=e^{-n\varepsilon^{-1/\gamma}\sigma^{-1/\gamma}(-\gamma)^{-1/\gamma}}\to 0$$

when $\gamma<0$.

□

Proof of Lemma 20. The empirical distribution function is an unbiased and consistent estimator of the true distribution, and converges uniformly to the true distribution, also has normality (see chapter 2.9 in [34]) for the independent sequence. By the definition of empirical distribution function, $Y_{(i)}, 1\le i\le n$, is seen as the $\frac{i}{n+1}$-th quantile of the sample sequence,

$$F(Y_{(i)})=\mathrm{P}(Y\le Y_{(i)})=\exp(-(1+\gamma\frac{Y_{(i)}-\mu}{\sigma})^{-1/\gamma})=\frac{i}{n+1}+\epsilon_{n,i},\tag{A.1}$$

Perform the same transformation to both sides of the last Equation (A.1), we have

$$
\begin{aligned}
&-\frac{1}{\gamma}\log(1+\gamma\frac{Y_{(i)}-\mu}{\sigma}) = \log\left(-\log(\frac{i}{n+1}+\epsilon_{n,i})\right),\\
&-\frac{1}{\gamma}\log(Y_{(n)}-Y_{(i)}) - \frac{1}{\gamma}\log\left(\frac{1+\gamma\frac{Y_{(i)}-\mu}{\sigma}}{Y_n-Y_{(i)}}\right) = \log\left(-\log(\frac{i}{n+1})\right) - \frac{\epsilon_{n,i}}{\frac{i}{n+1}\log\frac{i}{n+1}} + o_p(\epsilon_{n,i}),\\
&\log(Y_{(n)}-Y_{(i)}) + \log\left(-\frac{\gamma}{\sigma}\frac{\mu-\frac{\sigma}{\gamma}-Y_{(i)}}{Y_{(n)}-Y_{(i)}}\right) = -\gamma\log\left(-\log\frac{i}{n+1}\right) + \frac{\gamma\epsilon_{n,i}}{\frac{i}{n+1}\log\frac{i}{n+1}} + o_p(\epsilon_{n,i}),\\
&\log(Y_{(n)}-Y_{(i)}) = -\log\left(-\frac{\gamma}{\sigma}\right) - \gamma\log\left(-\log\frac{i}{n+1}\right) + \eta_{n,i},
\end{aligned}
$$

which is $y_i = -\log(-\frac{\gamma}{\sigma}) - \gamma x_i + \eta_{n,i}$ with (x_i, y_i) defined in Theorem 13 and

$$
\begin{aligned}
\eta_{n,i} &= \frac{\gamma\epsilon_{n,i}}{\frac{i}{n+1}\log\frac{i}{n+1}} + o_p(\epsilon_{n,i}) - \log\left(\frac{\mu-\frac{\sigma}{\gamma}-Y_{(i)}}{Y_{(n)}-Y_{(i)}}\right)\\
&= \frac{\gamma\epsilon_{n,i}}{\frac{i}{n+1}\log\frac{i}{n+1}} + o_p(\epsilon_{n,i}) - \frac{\mu-\frac{\sigma}{\gamma}-Y_{(n)}}{\mu-\frac{\sigma}{\gamma}-Y_{(i)}} + o_p\left(\frac{\mu-\frac{\sigma}{\gamma}-Y_{(n)}}{\mu-\frac{\sigma}{\gamma}-Y_{(i)}}\right).
\end{aligned}
$$

□

If the error term $\{\eta_{n,i}, 1 \le i \le n-1\}$ is an independent sequence, which has zero mean and constant variance, it is an ordinary linear regression problem. Those $n-1$ points

$$(\log(-\log(\frac{i}{n+1})), \log(Y_{(n)}-Y_{(i)})), 1 \le i \le n-1$$

can be used to fit the line $y = -\gamma x - \log(\frac{-\gamma}{\sigma})$. Thus the estimator of GEV parameters could be obtained.

However, the error term here is not independent, also has mean and variance depend on the GEV parameters and the order i. To obtain the mean and covariance of $\eta_{n,i}$, we need the result of Lemma 21.

Proof of Lemma 21. To calculate the mean and variance of $\epsilon_{n,i}$, we need to consider the densities of the order statistics. The maximum of the independent sample, denoted as $Y_{(n)}$, has the density

$$f_{Y_{(n)}}(y) = nf(y)F(y)^{n-1}.$$

The density of the i-th smallest sample $Y_{(i)}$ is

$$\begin{aligned} f_{Y_{(i)}}(y) = & \lim_{\varepsilon \to 0} \frac{\partial \, \mathrm{P}\left(Y_{(i)} \in [y, y+\varepsilon)\right)}{\partial \varepsilon} \\ = & \; nf(y) \frac{(n-1)!}{(i-1)!(n-i)!} F(y)^{i-1}(1-F(y))^{n-i}. \end{aligned} \tag{A.2}$$

With the density of $Y_{(i)}$, the moments of $F(Y_{(i)})$ can be obtained.

$$\begin{aligned} \mathrm{E}(F(Y_{(i)})) &= \int F(y) n f(y) \frac{(n-1)!}{(i-1)!(n-i)!} F(y)^{i-1}[1-F(y)]^{n-i} dy \\ &= \int_0^1 n \frac{(n-1)!}{(i-1)!(n-i)!} u^i (1-u)^{n-i} du \quad (u = F(y)) \\ &= \frac{i}{n+1}, \end{aligned}$$

$$\begin{aligned} \mathrm{E}(F^2(Y_{(i)})) &= \int F^2(y) n f(y) \frac{(n-1)!}{(i-1)!(n-i)!} F(y)^{i-1}[1-F(y)]^{n-i} dy \\ &= \int_0^1 n \frac{(n-1)!}{(i-1)!(n-i)!} u^{i+1} (1-u)^{n-i} du \quad (u = F(y)) \\ &= \frac{i(i+1)}{(n+1)(n+2)}, \end{aligned}$$

thus we have $\mathrm{E}(\epsilon_{n,i}) = 0$ and

$$\mathrm{Var}\,(\epsilon_{n,i}) = \mathrm{E}(F(Y_{(i)}) - \frac{i}{n+1})^2 = \frac{in + i - i^2}{(n+1)^2(n+2)}.$$

For the jointly density of $Y_{(i)}$ and $Y_{(j)}$, let $1 \le i < j < n, y_1 \le y_2 < \mu - \frac{\sigma}{\gamma}$,

$$\begin{aligned} f_{Y_{(i)},Y_{(j)}}(y_1, y_2) = & \; n(n-1) f(y_1) f(y_2) \frac{(n-2)!}{(i-1)!(n-i-1)!} F^{i-1}(y_1) \\ & \frac{(n-i-1)!}{(j-i-1)!(n-j)!} \left(F(y_2) - F(y_1)\right)^{j-i-1} \left(1 - F(y_2)\right)^{n-j}. \end{aligned}$$

Thus

$$\begin{aligned}\mathrm{E}(F(Y_{(i)})F(Y_{(j)})) &= \frac{i(j+1)}{(n+1)(n+2)},\\ \mathrm{Cov}(F(Y_{(i)}),F(Y_{(j)})) &= \frac{in+i-ij}{(n+1)^2(n+2)}.\end{aligned}$$

Specifically,

$$\begin{aligned}\mathrm{E}(F(Y_{(i)})F(Y_{(n)})) &= \frac{i}{(n+2)},\\ \mathrm{Cov}(F(Y_{(i)}),F(Y_{(n)})) &= \frac{i}{(n+1)^2(n+2)}.\end{aligned}$$

□

The moments of $Y_{(i)}$ is needed to obtain the Taylor expansion of $\frac{\mu-\frac{\sigma}{\gamma}-Y_{(n)}}{\mu-\frac{\sigma}{\gamma}-Y_{(i)}}$.

Proof of Lemma 22. The order statistic $Y_{(i)}$ can be represented by $\epsilon_{n,i}$. Since $Y_{(i)} = F^{-1}(F(Y_{(i)}))$, where

$$F^{-1}(y) = \mu - \frac{\sigma}{\gamma} + \frac{\sigma}{\gamma}\left(\log\frac{1}{y}\right)^{-\gamma}, \tag{A.3}$$

thus

$$\begin{aligned}Y_{(i)} = \quad & F^{-1}\left(\frac{i}{n+1}\right) + \partial F^{-1}\left(\frac{i}{n+1}\right)\left(F(Y_{(i)}) - \frac{i}{n+1}\right)\\ & + \frac{\partial^2 F^{-1}(\frac{i}{n+1})}{2}\left(F(Y_{(i)}) - \frac{i}{n+1}\right)^2 + o_p\left(\left(F(Y_{(i)}) - \frac{i}{n+1}\right)^2\right)\\ = \quad & F^{-1}\left(\frac{i}{n+1}\right) + \partial F^{-1}\left(\frac{i}{n+1}\right)\epsilon_{n,i} + \frac{\partial^2 F^{-1}(\frac{i}{n+1})}{2}\epsilon_{n,i}^2 + o_p(\epsilon_{n,i}^2),\end{aligned}$$

where $\partial F^{-1}, \partial^2 F^{-1}$ represent the first and second order derivative of F^{-1} in Equation (A.3). Combined with lemma 21,

$$\mathrm{E}(Y_{(i)}) = F^{-1}\left(\frac{i}{n+1}\right) + \frac{\partial^2 F^{-1}(\frac{i}{n+1})}{2}\,\mathrm{Var}(\epsilon_{n,i}) + o_p\left(\mathrm{Var}(\epsilon_{n,i})\right),$$

so

$$Y_{(i)} - \mathrm{E}(Y_{(i)}) = \partial F^{-1}\left(\frac{i}{n+1}\right)\epsilon_{n,i} + \frac{\partial^2 F^{-1}(\frac{i}{n+1})}{2}\left(\epsilon_{n,i}^2 - \mathrm{Var}(\epsilon_{n,i})\right) + o_p\left(\epsilon_{n,i}^2 + \mathrm{Var}(\epsilon_{n,i})\right).$$

Denote

$$\begin{aligned} E\left(\mu-\frac{\sigma}{\gamma}-Y_{(i)}\right) &= -\frac{\sigma}{2}\frac{(r+1)(\log\frac{n+1}{i})^{-\gamma-2}-\log(\frac{n+1}{i})^{-\gamma-1}}{i}\frac{n+1-i}{n+2} \\ &\quad -\frac{\sigma}{\gamma}\left(\log\frac{n+1}{i}\right)^{-\gamma}+o_p\left(\frac{in+i-i^2}{(n+1)^2(n+2)}\right) \\ &= E_i. \end{aligned}$$

Under the condition of $\gamma<-1$,

$$\begin{aligned} E_n = E\left(\mu-\frac{\sigma}{\gamma}-Y_{(n)}\right) = & -\frac{\sigma}{2}\frac{(r+1)(\log\frac{n+1}{n})^{-\gamma-2}-\log(\frac{n+1}{n})^{-\gamma-1}}{n(n+2)} \\ & -\frac{\sigma}{\gamma}(\log\frac{n+1}{n})^{-\gamma}+o_p\left(\frac{n}{(n+1)^2(n+2)}\right) \\ = & \ o_p\left(\left(\frac{1}{n}\right)^{\min(-\gamma,2)}\right). \end{aligned}$$

Using Taylor expansion we have

$$\begin{aligned} \frac{\mu-\frac{\sigma}{\gamma}-Y_{(n)}}{\mu-\frac{\sigma}{\gamma}-Y_{(i)}} = & \ \frac{E_n}{E_i}+\frac{1}{E_i}\left(\mathrm{E}(Y_{(n)})-Y_{(n)}\right)-\frac{E_n}{E_i^2}\left(\mathrm{E}(Y_{(i)})-Y_{(i)}\right)+\frac{E_n}{E_i^3}\left(\mathrm{E}(Y_{(i)})-Y_{(i)}\right)^2 \\ & -\frac{1}{2E_i^2}\left(\mathrm{E}(Y_{(n)})-Y_{(n)}\right)\left(\mathrm{E}(Y_{(i)})-Y_{(i)}\right)+o_p(\left(\mathrm{E}(Y_{(i)})-Y_{(i)}\right)^2 \\ & +\left(\mathrm{E}(Y_{(n)})-Y_{(n)}\right)\left(\mathrm{E}(Y_{(i)})-Y_{(i)}\right)+\left(\mathrm{E}(Y_{(n)})-Y_{(n)}\right)^2) \\ = & \ \frac{E_n}{E_i}-\frac{\partial F^{-1}(\frac{n}{n+1})}{E_i}\epsilon_{n,n}+\frac{E_n\partial F^{-1}\left(\frac{i}{n+1}\right)}{E_i}\epsilon_{n,i}-\frac{\partial F^{-1}\left(\frac{i}{n+1}\right)\partial F^{-1}(\frac{n}{n+1})}{2E_i^2}\epsilon_{n,i}\epsilon_{n,n} \\ & +\frac{E_n(\partial F^{-1}(\frac{n}{n+1}))^2}{E_i^3}\epsilon_{n,i}^2-\frac{\partial^2F^{-1}(\frac{n}{n+1})}{2E_i}\left(\epsilon_{n,n}^2-\mathrm{E}(\epsilon_{n,n}^2)\right) \\ & +\frac{E_n\partial^2F^{-1}\left(\frac{i}{n+1}\right)}{2E_i}\left(\epsilon_{n,i}^2-\mathrm{E}(\epsilon_{n,i}^2)\right)+o_p(\epsilon_{n,i}). \end{aligned}$$

Since $\frac{\partial F^{-1}\left(\frac{i}{n+1}\right)}{E_i}, \frac{\partial^2F^{-1}\left(\frac{i}{n+1}\right)}{E_i}$ is bounded, $E_n\to 0$ and $|\epsilon_{n,i}|<1$ as $n\to\infty$, thus $\frac{\mu-\frac{\sigma}{\gamma}-Y_{(n)}}{\mu-\frac{\sigma}{\gamma}-Y_{(i)}}$ can be simplified to $\frac{E_n}{E_i}+o_p(\epsilon_{n,i})+o_p(\epsilon_{n,n})$.

□

Appendix B

Proof of Theorem 14, 15, 16

Since

$$d(n) = \sup |\,\mathrm{P}(A \cap B) - \mathrm{P}(A)\,\mathrm{P}(B)| \leq \sup |\,\mathrm{P}(B|A) - \mathrm{P}(B)|,$$

which means for any value z, Z_1, we need to prove that

$$\mathrm{P}(Z_{n+1} \leq z|Z_1) - \mathrm{P}(Z_{n+1} \leq z) = O_p(n^{-5}).$$

To obtain the asymptotic normality of $\bar{Z}_n$, we need to show $d(n) = O_p(n^{-5})$.

Proof of Theorem 14. From Lemma 12 we have

$$\mathrm{P}(Z_{n+1} \leq z|Z_1) - \mathrm{P}(Z_{n+1} \leq z) = O_p(n\alpha^n)$$

and

$$n\alpha^n = o_p(n^{-5}),$$

$\mathrm{Cov}(Z_i, Z_{i+h}) = \alpha^h \psi^2 \frac{\pi^2}{6}$,

$$\begin{aligned} n\,\mathrm{Var}(\bar{Z}_n) &= \frac{1}{n}\sum_{i,j=1}^{n} \mathrm{Cov}(Z_i, Z_j) \\ &= \mathrm{Var}(Z_i) + \frac{2}{n}\sum_{i=1}^{n-1}(n-i)\alpha^i\psi^2\frac{\pi^2}{6} \\ &= \frac{\pi^2}{6} + 2\psi^2\frac{\pi^2}{6}\left((n+1)\frac{\alpha(1-\alpha^{n-1})}{1-\alpha} - \left(\frac{\alpha^2(1-\alpha^{n-1})}{1-\alpha}\right)'\right) \\ &= \frac{\pi^2}{6}\left(1 + 2\psi^2\frac{\alpha}{1-\alpha} - 2\psi^2\frac{\alpha(1-\alpha^n)}{n(1-\alpha)^2}\right) \\ &< \infty. \end{aligned}$$

Using the Theorem 27.4 in [6],

$$\sqrt{n}\bar{Z}_n \to N\left(0, \frac{\pi^2}{6}\left(1 + 2\psi^2\frac{\alpha}{1-\alpha}\right)\right). \tag{B.1}$$

□

Theoretically, this conclusion is true. However for this result to be valid in practical uses, it is worth to notice that the size of n should be in the thousands to make $\mathrm{P}(Z_{n+1} \le z|Z_1) - \mathrm{P}(Z_{n+1} \le z)$ to converge to zero faster than n^{-5} when α is large, say $\alpha > 0.1$. When α is small, this conclusion is okay for practical purpose.

Proof of Theorem 15. For the asymptotic property of $r^*(h)$, consider the sequence

$$Z_1Z_{1+h}, Z_2Z_{2+h}, Z_3Z_{3+h}, \ldots.$$

The dependence coefficient $d^*(n)$ of this sequence is smaller than $d(n-h)$,

$$\begin{aligned} & d^*(n) \\ = \; & \sup\left|\,\mathrm{P}(Z_{n+1}Z_{n+1+h} \le z, Z_1Z_{1+h} \le z_1) - \mathrm{P}(Z_{n+1}Z_{n+1+h} \le z)\,\mathrm{P}(Z_1Z_{1+h} \le z_1)\right| \\ \le \; & d(n-h). \end{aligned}$$

$\{Z_t\}$ is a stationary sequence with mean $\alpha^h\psi^2\frac{\pi^2}{6}$,

$$\begin{aligned}\mathrm{E}(Z_nZ_{n+h}) &= E\left(Z_n(\alpha^hZ_n+V_{\alpha,\psi,n}-\alpha^h\psi\log\xi_n)\right)\\ &= \alpha^h\frac{\pi^2}{6}-\alpha^h\psi^2(\frac{1}{\psi^2}-1)\frac{\pi^2}{6}\\ &= \alpha^h\psi^2\frac{\pi^2}{6}.\end{aligned}$$

The fourth moment of $\log S_i$ is finite, so

$$r^*(h)=\frac{1}{n}\sum_{i=1}^{n}Z_iZ_{i+h}$$

is asymptotically normal with mean $r(h)$.

To obtain the variance of $r^*(h)$,

$$\begin{aligned}\mathrm{Var}(r^*(h)) &= \frac{1}{n^2}\sum_{i,j=1}^{n}\mathrm{Cov}(Z_iZ_{i+h},Z_jZ_{j+h})\\ &= \frac{1}{n}\mathrm{Var}(Z_iZ_{i+h})+\frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathrm{Cov}(Z_iZ_{i+h},Z_jZ_{j+h})\\ &= \frac{1}{n}\mathrm{Var}(Z_iZ_{i+h})+\frac{1}{n^2}\sum_{i=1}^{n-h}\mathrm{E}(Z_iZ_{i+h}^2Z_{i+2h})\\ &\quad+\frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{s=1}^{\min(h-1,n-i)}\mathrm{E}(Z_iZ_{i+s}Z_{i+h}Z_{i+s+h})\\ &\quad+\frac{2}{n^2}\sum_{i=1}^{n-h-1}\sum_{s=h+1}^{n-i}\mathrm{E}(Z_iZ_{i+h}Z_{i+s}Z_{i+s+h})-\frac{n-1}{n}\psi^4\alpha^{2h}\left(\frac{\pi^2}{6}\right)^2 \quad \text{(B.2)}\end{aligned}$$

we need the value $\mathrm{Var}(Z_iZ_{i+h}),\mathrm{E}(Z_iZ_{i+h}^2Z_{i+2h}),\mathrm{E}(Z_iZ_{i+s}Z_{i+s+p}Z_{i+s+p+q})$ for positive h,s,p,q.

If $h=1$, $\frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{s=1}^{\min(h-1,n-i)}\mathrm{E}(Z_iZ_{i+s}Z_{i+h}Z_{i+s+h})$ is zero.

Note that

$$Z_{i+h}=\alpha^hZ_i-\alpha^h\psi\log\xi_i+\alpha^h\psi\log(S_{\alpha^h\psi}),$$

where $S_{\alpha^h\psi} =: S_{i+1}\dots S_{i+h}^{1/\alpha^h}\xi_{i+h}^{1/\alpha^{h-1}} \sim S(\alpha^h\psi, 1)$ and independent of Z_i, ξ_i,

$$\begin{aligned}\mathrm{Var}(Z_iZ_{i+h}) &= \mathrm{Var}\left(Z_i(\alpha^hZ_i - \alpha^h\psi\log\xi_i + \alpha^h\psi\log(S_{\alpha^h\psi})\right)\\ &= \mathrm{Var}(\alpha^hZ_i^2) + \mathrm{Var}(\alpha^h\psi Z_i\log\xi_i) + Var(\alpha^h\psi Z_i\log S_{\alpha^h\psi})\\ &= \psi^4\alpha^{2h}\,\mathrm{E}(\log^4\xi_i) + \alpha^{2h}\left(\frac{\pi^2}{6}\right)^2(\psi^2+\frac{2}{5}) \qquad \text{(B.3)}\end{aligned}$$

since all the variables we considered here $(Z_i, \{\xi_i\}, \{S_i\})$ are centred, so the covariance part equals zero and

$$\begin{aligned}\mathrm{Var}(Z_i^2) &= \mathrm{E}(Z_i^4) - \left(\mathrm{E}(Z_i^2)\right)^2 = \frac{7}{5}\left(\frac{\pi^2}{6}\right)^2,\\ \mathrm{Var}(Z_i\log\xi_i) &= \mathrm{Var}((\psi\log\xi_i + \psi X_i)\log\xi_i)\\ &= \psi^2\,\mathrm{Var}(\log^2\xi_i) + \psi^2\left(\frac{\pi^2}{6}\right)^2(\frac{1}{\psi^2}-1)\\ &= \psi^2\,\mathrm{E}(\log^4\xi_i) - \left(\frac{\pi^2}{6}\right)^2(\frac{1}{\psi^2}-1).\end{aligned}$$

Before calculate $\mathrm{E}(Z_iZ_{i+h}^2Z_{i+2h})$ and $\mathrm{E}(Z_iZ_{i+s}Z_{i+s+p}Z_{i+s+p+q})$ we need

$$\mathrm{E}(Z_i\log\xi_i) = \psi(\frac{1}{\psi^2}-1)\frac{\pi^2}{6}, \qquad \text{(B.4)}$$

$$\mathrm{E}(Z_i\log^3\xi_i) = \psi\,\mathrm{E}(\log^4\xi_i), \qquad \text{(B.5)}$$

$$\mathrm{E}(Z_i^2\log^2\xi_i) = \psi^2\,\mathrm{E}(\log^4\xi_i) + \psi^2(\frac{1}{\psi^2}-1)\left(\frac{\pi^2}{6}\right)^2, \qquad \text{(B.6)}$$

$$\mathrm{E}(Z_i^3\log\xi_i) = \psi^3\,\mathrm{E}(\log^4\xi_i) + 3\psi^3(\frac{1}{\psi^2}-1)\left(\frac{\pi^2}{6}\right)^2. \qquad \text{(B.7)}$$

Similar, decompose Z_{i+h}, Z_{i+2h} as:

$$\begin{aligned}Z_{i+h} &= \alpha^hZ_i - \alpha^h\psi\log\xi_i + \psi\log\xi_{i+h} + \alpha^h\psi\log(S_{1,\alpha^h}),\\ Z_{i+2h} &= \alpha^{2h}Z_i - \alpha^{2h}\psi\log\xi_i + \psi\log\xi_{i+2h} + \alpha^{2h}\psi\log(S1_{\alpha,h}) + \alpha^h\psi\log(S_{2,\alpha^h}),\end{aligned}$$

where

$$S_{1,\alpha^h} = S_{i+1}\dots S_{i+h}^{1/\alpha^{h-1}} \sim \mathcal{S}(\alpha^h), \qquad \text{(B.8)}$$

$$S_{2,\alpha^h} = S_{i+1+h}\dots S_{i+2h}^{1/\alpha^{h-1}} \sim \mathcal{S}(\alpha^h), \qquad \text{(B.9)}$$

$S_{1,\alpha^h}, S_{2,\alpha^h}$ are mutually independent of $Z_i, \{\xi_i\}$.

By using Equations (B.8), (B.9) and (B.4)–(B.7), we obtain

$$\begin{aligned}
& \mathrm{E}(Z_i Z_{i+h}^2 Z_{i+2h}) \\
= \quad & \mathrm{E}\Big(Z_i \left(\alpha^h Z_i - \alpha^h \psi \log \xi_i + \psi \log \xi_{i+h} + \alpha^h \psi \log S_{1,\alpha^h}\right)^2 \\
& \times \left(\alpha^{2h} Z_i - \alpha^{2h} \psi \log \xi_i + \psi \log \xi_{i+2h} + \alpha^{2h} \psi \log S1_{\alpha,h} + \alpha^h \psi \log S_{2,\alpha^h}\right)\Big) \\
= \quad & E\left(Z_i (\alpha^h Z_i - \alpha^h \psi \log \xi_i + \psi \log \xi_{i+h} + \alpha^h \psi \log S_{1,\alpha^h})^2\right. \\
& \left.\times \left(\alpha^{2h} Z_i - \alpha^{2h} \psi \log \xi_i + \alpha^{2h} \psi \log S_{1,\alpha^h}\right)\right) \\
= \quad & \alpha^{4h} \mathrm{E}(Z_i^4) - 3\alpha^{4h} \psi \,\mathrm{E}(Z_i^3 \log \xi_i) + 3\alpha^{4h} \psi^2 \,\mathrm{E}(Z_i^2 \log^2 \xi_i) \\
& - \alpha^{4h} \psi^3 \,\mathrm{E}(Z_i \log^3 \xi_i) + 3\alpha^{4h} \psi^2 \,\mathrm{E}(Z_i^2 \log^2 S_{1,\alpha^h}) - 3\alpha^{4h} \psi^3 \,\mathrm{E}(Z_i \log \xi_i \log^2 S_{1,\alpha^h}) \\
& + \alpha^{2h} \psi^2 \,\mathrm{E}(Z_i^2 \log^2 \xi_{i+h}) - \alpha^{2h} \psi^3 \,\mathrm{E}(Z_i \log \xi_i \log^2 \xi_{i+h})\Big) \\
= \quad & -\alpha^{4h} \psi^4 \,\mathrm{E}(\log^4 \xi_i) + \alpha^{4h} \left(\frac{\pi^2}{6}\right)^2 \left(\frac{12}{5} - 6\psi^2 + 3\psi^4\right) + \alpha^{2h} \left(\frac{\pi^2}{6}\right)^2 \left(\psi^2 + 2\psi^4\right).
\end{aligned} \tag{B.10}$$

$$\begin{aligned}
& \mathrm{E}(Z_i Z_{i+s} Z_{i+s+p} Z_{i+s+p+q}) \\
= \quad & \mathrm{E}\,(\ Z_i(\alpha^s Z_i - \alpha^s \psi \log \xi_i + \psi \log \xi_{i+s} + \alpha^s \psi \log S_{1,\alpha^s}) \\
& \times(\alpha^{s+p} Z_i - \alpha^{s+p} \psi \log \xi_i + \psi \log \xi_{i+s+p} + \alpha^{s+p} \psi \log S_{1,\alpha^s} + \alpha^p \psi \log S_{2,\alpha^p}) \\
& \times(\alpha^{s+p+q} Z_i - \alpha^{s+p+q} \psi \log \xi_i + \psi \log \xi_{i+s+p+q} + \alpha^{s+p+q} \psi \log S_{1,\alpha^s} + \\
& \qquad \alpha^{p+q} \psi \log S_{2,\alpha^p} + \alpha^q \psi \log S_{3,\alpha^q})\ ) \\
= \quad & \mathrm{E}\,(\ Z_i(\alpha^s Z_i - \alpha^s \psi \log \xi_i + \alpha^s \psi \log S_{1,\alpha^s}) \\
& \times(\alpha^{s+p} Z_i - \alpha^{s+p} \psi \log \xi_i + \alpha^{s+p} \psi \log S_{1,\alpha^s} + \alpha^p \psi \log S_{2,\alpha^p}) \\
& \times(\alpha^{s+p+q} Z_i - \alpha^{s+p+q} \psi \log \xi_i + \alpha^{s+p+q} \psi \log S_{1,\alpha^s} + \alpha^{p+q} \psi \log S_{2,\alpha^p})\ ) \\
= \quad & \alpha^{3s+2p+q} \,\mathrm{E}\,(\ Z_i^4 - 3\psi Z_i^3 \log \xi_i + 3\psi^2 Z_i^2 \log^2 \xi_i - \psi^3 Z_i \log^3 \xi_i + 3\psi^2 Z_i^2 \log^2 S_{1,\alpha^s} \\
& - 3\psi^3 Z_i \log \xi_i \log^2 S_{1,\alpha^s}\ ) + \alpha^{s+2p+q} \,\mathrm{E}\left(\psi^2 Z_i^2 \log^2 S_{2,\alpha^p} - \psi^3 Z_i \log \xi_i \log^2 S_{2,\alpha^p}\right) \\
= \quad & -\alpha^{3s+2p+q} \psi^4 \,\mathrm{E}(\log^4 \xi_i) + \alpha^{3s+2p+q} \left(\frac{\pi^2}{6}\right)^2 \left(\frac{12}{5} - 6\psi^2 + 3\psi^4\right) \\
& + \alpha^{s+2p+q} \left(\frac{\pi^2}{6}\right)^2 \left(\psi^4 (1 + \frac{2}{\alpha^{2p}})\right)
\end{aligned} \tag{B.11}$$

where $S_{1,\alpha^s} \sim \mathcal{S}(\alpha^s), S_{2,\alpha^p} \sim \mathcal{S}(\alpha^p), S_{3,\alpha^q} \sim \mathcal{S}(\alpha^q)$ and are mutually independent with

each other and independent of $Z_i, \{\xi_i\}$.

By combining the Equations (B.2), (B.3), (B.10) and (B.11), we obtain

$$
\begin{aligned}
& \operatorname{Var}(r^*(h)) \\
= \quad & \frac{1}{n}\left(\psi^4\alpha^{2h}\,\mathrm{E}(\log^4\xi_i) + \alpha^{2h}\left(\frac{\pi^2}{6}\right)^2(\psi^2+\frac{2}{5})\right) \\
& +\frac{n-h}{n^2}\left(-\alpha^{4h}\psi^4\,\mathrm{E}(\log^4\xi_i) + \alpha^{4h}\left(\frac{\pi^2}{6}\right)^2\left(\frac{12}{5}-6\psi^2+3\psi^4\right) + \alpha^{2h}\left(\frac{\pi^2}{6}\right)^2(\psi^2+2\psi^4)\right) \\
& +\frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{s=1}^{\min(n-i,h-1)}\left(-\alpha^{2s+2h}\psi^4\,\mathrm{E}(\log^4\xi_i) + \alpha^{2s+2h}\left(\frac{\pi^2}{6}\right)^2\left(\frac{12}{5}-6\psi^2+3\psi^4\right)\right. \\
& \qquad \left. +\alpha^{2h}\left(\frac{\pi^2}{6}\right)^2\psi^4 + 2\alpha^{2s}\psi^4\left(\frac{\pi^2}{6}\right)^2\right) \\
& +\frac{2}{n^2}\sum_{i=1}^{n-h-1}\sum_{s=1+h}^{n-i}\left(-\alpha^{2s+2h}\psi^4\,\mathrm{E}(\log^4\xi_i) + \alpha^{2s+2h}\left(\frac{\pi^2}{6}\right)^2\left(\frac{12}{5}-6\psi^2+3\psi^4\right)\right. \\
& \qquad \left. +\alpha^{2s+h}\left(\frac{\pi^2}{6}\right)^2\psi^4 + 2\alpha^{2h}\psi^4\left(\frac{\pi^2}{6}\right)^2\right) - \frac{n-1}{n}\psi^4\alpha^{2h}\left(\frac{\pi^2}{6}\right)^2 \qquad \text{(B.12)}
\end{aligned}
$$

with

$$
\begin{aligned}
&\sum_{i=1}^{n-1}\sum_{s=1}^{\min(n-i,h-1)}\alpha^{2s+2h} = \frac{\alpha^{2h+2}}{1-\alpha^2}\left(n-1-(n-h+1)\alpha^{2h-2}-\frac{\alpha^2}{1-\alpha^2}(1-\alpha^{2h-4})\right), \\
&\sum_{i=1}^{n-1}\sum_{s=1}^{\min(n-i,h-1)}\alpha^{2h} = (h-1)(n-\frac{h}{2})\alpha^{2h}, \\
&\sum_{i=1}^{n-1}\sum_{s=1}^{\min(n-i,h-1)}\alpha^{2s} = \frac{\alpha^2}{1-\alpha^2}\left(n-1-\frac{\alpha^2}{1-\alpha^2}+\frac{\alpha^{2h-2}}{1-\alpha^2}+(n-h+1)\alpha^{2h-2}\right), \\
&\sum_{i=1}^{n-h-1}\sum_{s=1+h}^{n-i}\alpha^{2s+2h} = \frac{\alpha^{4h+2}}{1-\alpha^2}\left(n-h+1-\frac{\alpha^2}{1-\alpha^2}(1-\alpha^{2(n-h-1)})\right), \\
&\sum_{i=1}^{n-h-1}\sum_{s=1+h}^{n-i}\alpha^{2h} = \frac{(n-h)(n-h-1)}{2}\alpha^{2h}, \\
&\sum_{i=1}^{n-h-1}\sum_{s=1+h}^{n-i}\alpha^{2s} = \frac{\alpha^{2h+2}}{1-\alpha^2}\left(n-h+1-\frac{\alpha^2}{1-\alpha^2}(1-\alpha^{2(n-h-1)})\right).
\end{aligned}
$$

The variance of $r^*(h)$ has finite boundaries,

$$\lim_{n\to\infty} \mathrm{Var}(r^*(h)) = \alpha^{2h}\psi^4 \left(\frac{\pi^2}{6}\right)^2,$$

so

$$r^*(h) \to N\left(r(h), \alpha^{2h}\psi^4 \left(\frac{\pi^2}{6}\right)^2\right).$$

□

Proof of Theorem 16.

$$\begin{aligned}
r^*(h) - \hat{r}(h) &= \frac{1}{n}\sum_{i=1}^{n} Z_i Z_{i+h} - \frac{1}{n}\sum_{i=1}^{n-h}(Z_i - \bar{Z}_n)(Z_{i+h} - \bar{Z}_n) \\
&= \frac{1}{n}\sum_{i=n-h+1}^{n} Z_i Z_{i+h} + \frac{1}{n}\bar{Z}_n\left(\sum_{i=1}^{n-h} Z_{i+h} + \sum_{i=1}^{n-h} Z_i - (n-h)\bar{Z}_n\right) \\
&= \frac{1}{n}\sum_{i=n-h+1}^{n} Z_i Z_{i+h} + \frac{1}{n}\bar{Z}_n(\sum_{i=h+1}^{n-h} Z_i + h\bar{Z}_n). \\
\sqrt{n}(r^*(h) - \hat{r}(h)) &= \frac{1}{\sqrt{n}}\sum_{i=n-h+1}^{n} Z_i Z_{i+h} + \sqrt{n}\bar{Z}_n(\frac{1}{n}\sum_{i=h+1}^{n-h} Z_i + \frac{h}{n}\bar{Z}_n).
\end{aligned}$$

$\frac{1}{\sqrt{n}}\sum_{i=n-h+1}^{n} Z_i Z_{i+h}$ is $o_p(1)$ since

$$\frac{1}{\sqrt{n}}\sum_{i=n-h+1}^{n} \mathrm{E}(Z_i Z_{i+h}) \to 0.$$

By Theorem 14, we have $\sqrt{n}\bar{Z}_n$ is asymptotically normal (see Equation (B.1)), so $\sqrt{n}\bar{Z}_n(\frac{1}{n}\sum_{i=h+1}^{n-h} Z_i + \frac{h}{n}\bar{Z}_n)$ is $O_p(n^{-1/2})$.

$\sqrt{n}(r^*(h) - \hat{r}(h)) = o_p(1)$ as $n \to \infty$, so $\hat{r}(h) \to N\left(r(h), \alpha^{2h}\psi^4\left(\frac{\pi^2}{6}\right)^2\right)$.

□