On Parametric Modeling of Bivariate Extreme Value Distributions

Zhichen Zhao

Master's thesis
2012:E15

Lund University

Faculty of Science
Centre for Mathematical Sciences
Mathematical Statistics

Centrum Scientiarum Mathematicarum

Organization	Document name
LUND UNIVERSITY **Centre for Mathematical Sciences** **Mathematical Statistics**	**MASTER'S THESIS**
	Date of issue **2012-05-31**
	Internal Number: **LUNFMS-3037-2012**
Author **Zhichen Zhao**	**2012:E15**

Title and subtitle

On Parametric Modeling of Bivariate Extreme Value Distributions

Abstract

Extreme value theory is a branch of probability theory and statistics which deals with large values in a dataset. It has become more widespread in the past decade as a tool for risk management in different areas. The theory can be used by banks to estimate extreme investment losses, enables insurance companies to price their products and aids the government to budget for possible storms, earthquakes and other natural disasters.

Generally, there are two approaches to studying the distribution of extreme values namely, Block Maxima and Peaks Over Threshold. In the univariate case these approaches respectively lead to the Generalized Extreme Value Distributions (GEV) with location, scale and shape parameters and the Generalized Pareto Distributions (GPD) with shape and scale parameters. However, in most practical applications the data is high dimensional which requires the extension of the above mentioned approaches to the corresponding bivariate and multivariate distributions. In this thesis two bivariate extreme value models are considered namely, generalized symmetric mixed model and generalized symmetric logistic model. These two models were suggested by Tajvidi [2] in 1996 but their statistical properties have not been explored yet. In particular, we are interested in studying how the dependence between margins is affected by the change of parameters in each model. We study maximum likelihood estimation of the parameters and investigate the strength of dependence relationship by using Kendall's τ and Spearman's ρ.

The historical data on daily return of IBM and Apple stock prices are chosen for analysis in this thesis as they both are among world's leading computer companies. We show how the theory can be used for risk management and study the distribution of one stock's return level conditioned on the extreme return levels of the other stock.

Key words

Classification system and/or index terms (if any)

Supplementary bibliographical information	Language

Signature Zhichen zhao **Date** 31/5/12

On Parametric Modeling of Bivariate Extreme Value Distributions

Zhichen Zhao

Abstract

Extreme value theory is a branch of probability theory and statistics which deals with large values in a dataset. It has become more wide spread in the past decade as a tool for risk management in different areas. The theory can be used by banks to estimate extreme investment losses, enables insurance companies to price their products and aids the government to budget for possible storms, earthquakes and other natural disasters.

Generally, there are two approaches to study the distribution of extreme values namely, Block Maxima and Peaks Over Threshold. In the univariate case these approaches respectively lead to the Generalized Extreme Value Distributions (GEV) with location, scale and shape parameters and the Generalized Pareto Distributions (GPD) with shape and scale parameters. However, in most practical applications the data is multi-dimensional which requires the extension of the above mentioned approaches to the corresponding bivariate and multivariate distributions. In this thesis two bivariate extreme value models are considered namely, generalized symmetric mixed model and generalized symmetric logistic model. These two models were suggested by Tajvidi in 1996 but their statistical properties have not been explored yet. In particular, we are interested in studying how the dependence between margins is affected by the change of parameters in each model. We study maximum likelihood estimation of the parameters and investigate the strength of dependence relationship by using Kendall's τ, Spearman's ρ and Pickands dependence function.

The historical data on daily return of IBM and Apple stock prices are chosen for analysis in this thesis as they both are among world's leading computer companies. We show how the theory can be used for risk management and study the distribution of one stock prices' extreme returns conditional on the extreme returns of the other stock.

Acknowledgements

Firstly, I want to give my sincere thanks to my supervisor Nader Tajvidi. I become interested in extreme values theory when I took his univariate extreme values course. His careful and clear explanation about this theory make me interested in this field and make up my mind to write a thesis about this. Even though I had lots of questions during the thesis writing, he always be patient and kind. His encouragements and teaching have helped me to focus on the thesis and get a clearer vision about my future. Thus, I would really like to thank Nader again for his precious and countless help.

Also, I want to thank my parents who have supported me all the time. I have got some failures since I was born but they always been supportive. They never force me to be the one who can bring fames to parents but always encourage me to be a confident, happy and healthy person. For me, they are the best parents in the world.

At last, I want to thank those who have ever helped me during these two years in Sweden. Their help makes me feel loved even if I am far away from home.

Contents

1 Introduction of Extreme value theory

1.1 Basis Concepts

One of the basis concepts in probability and statistics theory is random variable X, which represents a quality whose outcome is uncertain. The set of possible outcomes of X, denoted as Ω, is the sample space. Another concept is called probability distribution, which assigns probabilities to consider. A random variable X is said to be a discrete random variable if its sample space is discrete: like $\Omega = \{0, 1, 2, ...\}$. In this case the probability distribution is determined by the probability mass function, which takes the form

$$f(x) = Pr\{X = x\}$$

for each value of x in Ω. Also, there exists continuous random variables whose sample space Ω is continuous. In continuous scale, we can take probability distribution function, defined as

$$F(x) = Pr\{X \leqslant x\}$$

for each x in Ω. If the distribution function F is differentiable, it is also useful to define the probability density function of X as

$$f(x) = \frac{dF}{dx}$$

A multivariate random variable is a vector of random variables

$$X = \begin{bmatrix} x_1 \\ \cdots \\ x_k \end{bmatrix}.$$

Each of components x_i is a random variable in its own right, but to know properties of X requires information about how every variable in X influence the other variables. Generalizing the single variable case, the joint distribution function of X is defined by

$$F(x) = \{X_1 \leqslant x_1, ..., X_k \leqslant x_k\}$$

where $x = (x_1, ..., x_k)$. When the X_i are continuous random variables, the joint density function is given by

$$f(x) = \frac{\partial^k F}{\partial x_1, ..., \partial x_k}$$

Now, we constrain discussions to the case of continuous random variable whose probability density function exist and belongs to the parametric density functions

$$\mathcal{F} = \{f(x; \theta) : \theta \in \Theta\}.$$

The parameter θ may be a scalar, such as $\theta = p$ in binomial family or be a vector of parameters, such as $\theta = (\mu, \sigma)$ in normal family.

A general method to estimate the unknown parameter θ within a family $\mathcal{F}$ is maximum likelihood. Each value of θ gives a model in $\mathcal{F}$ that attaches different probabilities to the observed data. The probability of observed data as a function of θ is called likelihood function. Values of θ that have high likelihood correspond to models which give high probability to the observed data. The principle of maximum likelihood estimation is to adopt the model with greatest likelihood, since of all the models are under consideration, this is the one that assigns highest probability to the observed data. The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

It is often more convenient to take logarithms and work with the log-likelihood function

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta)$$

1.2 Introduction to Generalized Extreme Value (GEV) Model

The model which represents the cornerstone of extreme value focuses on $M_n = \max\{X_1, ..., X_n\}$, where $X_1, ..., X_n$ is a sequence of independent random variables having a common distribution function F. In applications, X_i may represent hourly measured sea level or daily stock prices, so that M_n may represent the maximum seal level in a day or the maximum stock price in a week.

In theory, the distribution of M_n can be derived exactly for all values of n:

$$\begin{aligned} P(M_n \leqslant x) &= P(X_1 \leqslant x, ..., X_n \leqslant x) \\ &= P(X_1 \leqslant x) \cdots P(X_n \leqslant x) \\ &= F^n(x) \end{aligned}$$

However this is not very useful in practice, since the distribution F is mostly unknown or even if we know F, the computations will be too complicated. Thus, we proceed by looking at the behavior of F^n as $n \to \infty$. But this alone is not enough: for any $z < z_+$, where $z_+ = \sup\{F(z) < 1\}$, then $F^n \to 0$ as $n \to \infty$, so that the distribution of M_n degenerate to a point mass on z_+. This difficulty is avoided by allowing a linear renormalization of the variable M_n:

$$M_n^* = \frac{M_n - b_n}{a_n}$$

for sequences of constants $\{a_n > 0\}$ and b_n. We therefore seek limit distribution for M_n^* with the choices of a_n and b_n rather than M_n.

Theorem 1.1 If there exist sequences of constants for $\{a_n > 0\}$ and b_n such that

$$Pr\{\frac{(M_n - b_n)}{a_n} \leqslant z\} \to G(z)$$

as $n \to \infty$, where G is a non-degerate distribution function. Then G belongs to one of the following families:

$$\text{Fréchet}: \quad G(z) = \begin{cases} 0 & z \leqslant b, \\ \exp\{-(\frac{z-b}{a})^{-\alpha}\} & z > b, \end{cases} \quad \alpha > 0$$

$$\text{Weibull:} \quad G(z) = \begin{cases} \exp\{-[-(\frac{z-b}{a})^{\alpha}]\} & z < b, \\ 1 & z \geqslant b, \end{cases} \quad \alpha > 0$$

Gumble: $G(z) = \exp\{-\exp\{-(\frac{z-b}{a})\}\}$, $-\infty < z < \infty$.
with parameters $a > 0, b$.

Theorem 1.1 implies that, if the M_n can be stabilized, the corresponding M_n^* has a limiting distribution that must result in one of the three types of extreme value distributions.

A better analysis is offered by a reformulation of the models in Theorem 1.1. It is straightforward to show that the Fréchet, Weibull and Gumbel families can be combined into a single family of models having distribution functions:

$$G(z) = \exp\{-[1 + \gamma(\frac{z-\mu}{\sigma})]_+^{-\frac{1}{\gamma}}\}$$

where the parameters satisfy $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \gamma < \infty$. This is the Generalized Extreme Value (GEV) distribution with three parameters: a location parameter μ, a scale parameter σ and a shape parameter γ.

Thus, the Theorem 1.1 can be restated as:

Theorem 1.2 If there exist sequences of constants for $\{a_n > 0\}$ and b_n such that

$$Pr\{\frac{(M_n - b_n)}{a_n} \leqslant z\} \to G(z)$$

as $n \to \infty$, for G is a non-degenerate distribution function. Then G is a member of the GEV family

$$G(z) = \exp\{-[1 + \gamma(\frac{z-\mu}{\sigma})]^{-\frac{1}{\gamma}}\}$$

defined on $\{z : 1 + \gamma\frac{z-\mu}{\sigma} > 0\}$, where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \gamma < \infty$.

To fit extreme value models, data $X_1, ..., X_n$ are blocked into sequences of observations with length n, generating a series of block maxima $M_1, ..., M_m$. Often, the block maxima are chosen to correspond to a time period. If the period is too long, we will have fewer observations which gives large variance in GEV estimation. On the other hand, if the period is too short, we have poor approximation of $F^n(x)$ by GEV which leads to large bias in estimation. Thus, we need to check the validity of our GEV models after we decide the block length.

Estimation of extreme quantiles of the periodical maximum distribution can be obtained by inverting GEV distribution:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\gamma}[1-(-\log(1-p))^{\gamma}], & \text{for} \quad \gamma \neq 0 \\ \mu - \sigma\log[-\log(1-p)], & \text{for} \quad \gamma = 0 \end{cases} \tag{1.1}$$

where $G(z_p) = 1 - p$. In common terminology, z_p is the return level associated with the return period $1/p$, since to a reasonable degree of accuracy, the level z_p is expected to be exceeded on average once every $1/p$ period.

1.3 Inference for the GEV Distribution

Under the assumption that $Z_1, ..., Z_m$ are independent variables having the GEV distribution, the log-likelihood for the GEV parameter when $\gamma \neq 0$ is

$$\ell(\mu,\sigma,\gamma) = -m\log\sigma - (1+\frac{1}{\sigma})\sum_{i=1}^{m}\log[1+\gamma(\frac{z_i-\mu}{\sigma})] - \sum_{i=1}^{m}[1+\gamma(\frac{z_i-\mu}{\sigma})]^{-\frac{1}{\gamma}}$$

provided that

$$1+\gamma(\frac{z_i-\mu}{\sigma}) > 0, \text{for} \quad i = 1, ..., m$$

The case $\gamma = 0$ requires separate treatment using the Gumbel limit of the GEV distribution. This leads to the log-likelihood

$$\ell(\mu,\sigma) = -m\log\sigma - \sum_{i=1}^{m}(\frac{z_i-\mu}{\sigma}) - \sum_{i=1}^{m}\{-(\frac{z_i-\mu}{\sigma})\}.$$

Maximization of these above log-likelihood functions with respect to the parameter vector (μ, σ, γ) leads to the maximum likelihood estimates with respect to the entire GEV family. For any given dataset the maximization is straight forward using standard numerical optimization algorithm.

By substitution of the maximum likelihood estimates of the GEV parameters into (1.1) the maximum likelihood estimate z_p for $0 < p < 1$, the $1/p$ return level, is obtained as

$$\hat{z}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}}[1-(-\log(1-p))^{-\hat{\gamma}}] & \text{for} \quad \hat{\gamma} \neq 0 \\ \hat{\mu} - \hat{\sigma}\log(-\log(1-p)) & \text{for} \quad \hat{\gamma} = 0 \end{cases}$$

Furthermore, by the delta method,

$$Var(\hat{z}_p) \approx \nabla z_p^T V \nabla z_p$$

where V is the variance-covariance matrix of $(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ and

$$\begin{aligned} \nabla z_p^T &= [\frac{\partial z_p}{\partial \mu}, \frac{\partial z_p}{\partial \sigma}, \frac{\partial z_p}{\partial \gamma}] \\ &= [1, -\gamma^{-1}(1-(-\log(1-p))^{-\hat{\gamma}}), \sigma\gamma^{-2}(1-(-\log(1-p))^{-\hat{\gamma}}) - \\ &\sigma\gamma^{-1}(-\log(1-p))^{-\gamma}\log(-\log(1-p))] \end{aligned}$$

evaluated at $(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$.

Numerical evaluation of the profile likelihood for any of individual parameters μ, σ or γ is straightforward. For example, to obtain the profile likelihood for γ, we fix $\gamma = \gamma_0$, and maximize the log-likelihood with respect to the remaining parameters μ and σ. This is repeated for a range value of γ_0. The corresponding maximized values of the log-likelihood constitute the profile log-likelihood for γ. In particular, we can obtain confidence intervals for any specified return level z_p which requires reparameterization of the GEV model. Then z_p is one of the model parameters, after the profile log-likelihood is obtained. Reparameterization is proceeded by:

$$\mu = z_p + \frac{\sigma}{\gamma}[1 - (-\log(1-p))^{-\gamma}]$$

so that replacement of μ has desired effects of expressing the GEV model in terms of the parameter (z_p, σ, γ).

The model assessment can be made with the observed data by the use of probability plots and quantile plots. A probability plot is a comparison of the empirical and fitted distribution functions. With ordered block maximum data $z_{(1)} \leqslant z_{(2)} \leqslant, ..., z_{(m)}$, the empirical distribution function evaluated at $z_{(i)}$ is given by

$$\tilde{G}(z_{(i)}) = \frac{i}{m+1}$$

The model estimates are

$$\hat{G}(z_{(i)}) = \exp\{-[1 + \hat{\gamma}\frac{z_{(i)} - \hat{\mu}}{\hat{\sigma}}]^{-\frac{1}{\hat{\gamma}}}\}$$

If the GEV model works well,

$$\tilde{G}(z_{(i)}) \approx \hat{G}(z_{(i)}) \quad i = 1, ..., m$$

so a probability plot, consisting of the points

$$\{(\tilde{G}(z_{(i)}), \hat{G}(z_{(i)})), i = 1, ..., m\}$$

should lie close to the unit diagonal.

A weakness of the probability plot for extreme value models is that both $\tilde{G}(z_{(i)})$ and $\hat{G}(z_{(i)})$ are bound to approach 1 as $z_{(i)}$ increases, while the accuracy of a model can achieved by checking large values of z.

The deficiency can be avoided by the quantile plot, consisting of the points

$$\{(\hat{G}^{-1}(\frac{i}{m+1}), z_{(i)}), i = 1, ..., m\}$$

where

$$\hat{G}(\frac{i}{m+1}) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}}[1 - (-\log(\frac{i}{m+1}))^{-\hat{\gamma}}]$$

Departures from linearity in the quantile plot indicate model failure.

1.4 Introduction to Generalized Pareto (GPD) Model

Modeling only block maxima is a wasteful approach to extreme values analysis. If an entire time series of observations is available, then a more efficient use of data can be achieved by the following approach.

Let $X_1, X_2, ...$ be a sequence of independent and identically distributed random variables, having marginal distribution function F. It is nature to regard those of the X_i that exceed high threshold u as extreme values. Denoting an arbitrary term in the X_i sequence by X, it follows that a description of the stochastic behavior of extreme events is given by the conditional probability

$$P(X > u + x | X > u) = \frac{P(X > u + x)}{P(X > u)} \quad x > 0$$

If the parent distribution F were known, the distribution of threshold exceedances would also be known. By the use of GEV models as an approximation to the distribution, we get a new distribution namely, Generalized Pareto Distribution.

Theorem 1.3 Let $X_1, X_2, ...$ be a sequence of independent random variables with common distribution function F, and let

$$M_n = \max\{X_1, ..., X_n\}$$

Denote the arbitrary term in the X_i sequence by X, and suppose that F satisfies Theorem 1.2 so that for large n we have

$$Pr\{M_n \leqslant z\} \approx G(z),$$

where

$$G(z) = \exp\{-[1 + \gamma\frac{z - \mu}{\sigma}]^{-\frac{1}{\gamma}}\}$$

for some μ, $\sigma > 0$ and γ. Then, for large enough u, the distribution function of $(X - u)$, conditional on $X > mu$, is approximately

$$H(x) = 1 - (1 + \gamma\frac{x}{\tilde{\sigma}})^{-\frac{1}{\gamma}} \tag{1.2}$$

defined on $\{x : x > 0 \text{ and } (1 + \gamma\frac{x}{\tilde{\sigma}}) > 0\}$, where $\tilde{\sigma} = \sigma + \gamma(u - \mu)$.

If $\gamma = 0$,

$$H(x) = 1 - \exp(-\frac{x}{\sigma}), \quad x > 0$$

The family defined by Equation (1.2) is called the Generalized Pareto family.

Proof of Theorem 1.3 can be found in an outline of the proof but here we give a rough proof as following [4] :

By the assumption of Theorem 1.1, for large enough n,

$$F^n(z) \approx \exp\{-[1 + \gamma(\frac{z - \mu}{\sigma})]^{-\frac{1}{\gamma}}\}$$

Hence,

$$n \log F(z) \approx -[1+\gamma(\frac{z-\mu}{\sigma})]^{-\frac{1}{\gamma}}$$

For large values of z, a Taylor expansion implies that

$$\log F(z) \approx -(1-F(z))$$

gives

$$1-F(u) \approx \frac{1}{n}[1+\gamma(\frac{u-\mu}{\sigma})]^{-\frac{1}{\gamma}}$$

for large u. Similarly, we have

$$1-F(u+x) \approx \frac{1}{n}[1+\gamma(\frac{u+x-\mu}{\sigma})]^{-\frac{1}{\gamma}}$$

Hence,

$$\begin{aligned} Pr\{X > u+x | X > u\} &\approx \frac{n^{-1}[1+\gamma\frac{u+x-\mu}{\sigma}]^{-\frac{1}{\gamma}}}{n^{-1}[1+\gamma\frac{u-\mu}{\sigma}]^{-\frac{1}{\gamma}}} \\ &= [\frac{1+\gamma\frac{u+x-\mu}{\sigma}}{1+\gamma\frac{u-\mu}{\sigma}}]^{-\frac{1}{\gamma}} \\ &= [1+\frac{\gamma x}{\tilde{\sigma}}]^{-\frac{1}{\gamma}} \end{aligned}$$

Where $\tilde{\sigma} = \sigma + \gamma(u-\mu)$.

An important point here is that, like the choice of number of observations in each block, the choice of threshold also needs to be done carefully. A low threshold is likely to violate the asymptotic basis of model, leading to bias. A high threshold generate fewer excesses leading to a large variance.

Like the GEV model, it is more convenient to interpret extreme value models in terms of quantiles or return levels, rather than individual parameter values. So, suppose that a generalized Pareto distribution with parameters σ and γ is a suitable model for exceedances of a threshold u by a variable X. That is, for $x > u$,

$$Pr(X > x | X > u) = [1+\gamma\frac{x-u}{\sigma}]^{-\frac{1}{\gamma}}$$

follows

$$Pr(X > x) = \zeta_u[1+\gamma\frac{x-u}{\sigma}]^{-\frac{1}{\gamma}}$$

where $\zeta_u = Pr\{X > u\}$. Hence, the level x_m that is exceeded on average once every m observations is the solution of

$$\zeta_u[1+\gamma\frac{x_m-u}{\sigma}]^{-\frac{1}{\gamma}} = \frac{1}{m}$$

if $\gamma \neq 0$.

Rearanging,

$$x_m = u + \frac{\sigma}{\gamma}[(m\zeta_u)^\gamma - 1]$$

provided m is sufficiently large to ensure that $x_m > u$.

If $\gamma = 0$

$$x_m = u + \sigma \log(m\zeta_u).$$

x_m is the m-observations return level.

1.5 Inference for the GPD distribution

Having determined a threshold, the parameters of generalized Pareto distribution can be estimated by maximum likelihood. Suppose that the values $y_1, ..., y_k$ are the k exceedances of a threshold u.

For $\gamma \neq 0$ the log-likelihood is

$$\ell(\sigma, \gamma) = -k \log \sigma - (1 + \frac{1}{\gamma}) \sum_{i=1}^{k} \log(1 + \frac{\gamma y_i}{\sigma})$$

if $(1 + \sigma^{-1}\gamma y_i) > 0$ for $i = 1, ..., k$; otherwise, $\ell(\sigma, \gamma) = -\infty$.

If $\gamma = 0$, the log-likelihood function is obtained as

$$\ell(\sigma, \gamma) = -k \log \sigma - \sigma^{-1} \sum_{i=1}^{k} y_i.$$

Estimation of return level requires the use of estimated parameter values σ and γ which corresponds to maximum likelihood estimation, but an estimate of ζ_u can be naturally made by

$$\hat{\zeta_u} = \frac{k}{n}.$$

Standard errors or confidence intervals for x_m can be derived by the delta method. The properties of binomial distribution gives $Var(\hat{\zeta_u}) \approx \frac{\hat{\zeta_u}(1-\hat{\zeta_u})}{n}$, so the complete covariance matrix for $(\hat{\zeta_u}, \hat{\sigma}, \hat{\gamma})$ is approximately

$$V = \begin{bmatrix} \frac{\hat{\zeta_u}(1-\hat{\zeta_u})}{n} & 0 & 0 \\ 0 & v_{1,1} & v_{1,2} \\ 0 & v_{2,1} & v_{2,2} \end{bmatrix}$$

where $v_{i,j}$ denotes the (i, j) term of the covariance matrix of $\hat{\sigma}$ and $\hat{\gamma}$. Hence, by the delta method,

$$Var(\hat{x_m}) \approx \nabla x_m^T V \nabla x_m$$

where

$$\nabla x_m^T = [\frac{\partial x_m}{\partial \zeta_u}, \frac{\partial x_m}{\partial \sigma}, \frac{\partial x_m}{\partial \gamma}]$$
$$= [\sigma m^\gamma \zeta_u^{\gamma-1}, \gamma^{-1}((m\zeta_u)^\gamma - 1), -\sigma\gamma^{-2}((m\zeta_u)^\gamma - 1) + \sigma\gamma^{-1}(m\zeta_u)^\gamma \log(m\zeta_u)]$$

with evaluated $(\hat{\zeta_u}, \hat{\sigma}, \hat{\gamma})$.

For accessing the quality of fitted generalized Pareto model, probability plots and quantile plots are still useful. Assuming a threshold u, ordered threshold excesses $y_{(1)} \leqslant ... \leqslant y_{(k)}$ and an estimated model $\hat{H}$, the probability plot consists of the pairs

$$\{(\frac{i}{k+1}, \hat{H}(y_{(i)})), i = 1, .., k\},$$

where

$$\hat{H}(y) = 1 - (1 + \frac{\hat{\gamma} y}{\hat{\sigma}})^{-\frac{1}{\hat{\gamma}}}$$

provided $\hat{\gamma} \neq 0$.

If $\hat{\gamma} = 0$,

$$\hat{H}(y) = 1 - \exp(-\frac{y}{\hat{\sigma}})$$

Assumming $\hat{\gamma} \neq 0$, the quantile plot consists of the pairs

$$\hat{H}^{-1}(y) = u + \frac{\hat{\sigma}}{\gamma}[y^{-\hat{\gamma}} - 1]$$

2 Multivariate Extreme Value Distribution

2.1 Copula

Before we discuss the multivariate extreme models, the idea of the copula will be introduced.

Consider a random vector $(X_1, X_2, ..., X_d)$. Suppose its margins are continuous, i.e. the marginal cumulative distribution functions $F_i(x) = P(X_i \leqslant x)$ are continuous functions. By applying the probability integral transform to each component, the random vector can be seen that

$$(U_1, U_2, ..., U_d) = (F_1(X_1), F_2(X_2), ..., F_d(X_d))$$

has uniform margins.

The copula of $(X_1, X_2, ..., X_d)$ is defined as the joint cumulative function of $(U_1, U_2, ..., U_d)$:

$$C(u_1, u_2, ..., u_d) = P(U_1 \leqslant u_1, U_2 \leqslant u_2 ..., U_d \leqslant u_d)$$

The copula C contains all information on the dependence structure between the components of $(X_1, X_2, ..., X_d)$ whereas the marginal cumulative distribution functions F_i

contain in all information on the marginal distributions.

The definition of copula can be defined as follow: $C : [0,1]^d \to [0,1]$ is a d-dimension copula if:

$$C(u_1, ..., u_{i-1}, 0, u_{i+1}, ..., u_d) = 0$$
$$C(1, ..., 1, u, 1, ..., 1) = u$$

C is d-increasing, i.e. for each hyper rectangle $B = \times_{i=1}^{d}[x_i, y_i] \subseteq [0,1]^d$ the C-volume of B is non-negative:

$$\sum_{z \in x_{i=1}^{d}(x_i, y_i)} (-1)^{N(z)} C(Z) \geqslant 0$$

where the $N(z) = \#\{k : z_k = x_k\}$.

For bivariate case which will be discussed in this thesis, for every u, v in $I = [0,1]$,

$$C(u,0) = C(0,v) = 0$$
$$C(u,1) = u \quad \text{and} \quad C(1,v) = v$$

For every u_1, u_2, v_1, v_2 in I with $u_1 \leqslant u_2$ and $v_1 \leqslant v_2$,

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geqslant 0$$

The following theorem shows the Fréchet-Hoeffding bounds.

Theorem 2.1 Let C be a d-dimension copula. Then for $u_1, ..., u_d$ in I^d

$$\max\{u_1 + ... + u_d - (d-1), 0\} \leqslant C(u_1, ..., u_d) \leqslant \min\{u_1, ..., u_d\}$$

For bivariate case, we have

Theorem 2.2 Let C be a two dimension copula. Then for every (u,v) in I^2,

$$\max\{u + v - 1, 0\} \leqslant C(u,v) \leqslant \min\{u, v\}$$

The Sklar's theorem guarantees that if we a joint distribution H with margins F and G, we can find a bivariate copula C.

Sklar's Theorem In bivariate case, let H be a joint distribution function with margins F and G. Then there exists a copula such that

$$H(x,y) = C(F(x), G(y))$$

2.2 Max-stable and Max-infinitely divisible

From univariate extreme values, we know that for positive integer k there exists vectors $\alpha_k > 0$ and β_k such that $a_n^{-1}a_{nk} \to \alpha_k$ and $a_n^{-1}(b_{nk} - b_n) \to \beta_k$ as $n \to \infty$. Also, as we have $F^{nk}(a_{nk}x + b_{nk}) \to G(x)$.

Then we obtain:

$$G^k(\alpha_k x + \beta_k) = G(x) \quad x \in R^d \tag{2.1}$$

with $G(x)$ is a d-variable distribution. We can find vectors $\alpha_k > 0$ and β_k such that (2.1) is true and we call this max-stable. The meaning of (2.1) is if we can find $Y, Y_1, Y_2, ...$ are independent random vector with distribution function G, we have

$$\alpha_k^{-1}(\max\{Y_1, Y_2, ..\} - \beta_k) \overset{D}{=} Y \quad k = 1, 2, ...$$

A result of (2.1) is that if $G^{1/k}$ is a distribution function for every positive integer k, then G is max-infinitely divisible. It means in particular, we can find a measure, μ, on $[-\infty, \infty]$, such that

$$G(x) = \exp\{-\mu([-\infty, x]^c)\}, \quad x \in [-\infty, \infty].$$

To learn more about max-stable, it is convenient if we use Unit Fréchet margins. Let G be a d-variable distribution function with $Y_1, ..., Y_d$, G_j is the jth margin of G and G_j^q is the quantile function of G_j, that is $G_j^q(p) = x$ where $0 < p < 1$. We have

$$G_j = \exp\{-(1 + \gamma_j \frac{x_j - \mu_j}{\sigma_j})_+^{-\frac{1}{\gamma_j}}\} \quad x_j \in R^d$$
$$G_j^q(e^{-1/z_j}) = \mu_j + \sigma_j \frac{z_j^{\gamma_j} - 1}{\gamma_j} \quad z_j > 0$$

and if we let G_* be a d-variable function with $(-1/\log(G_1), ..., -1/\log(G_d))$, then we get

$$G_*(z) = G\{G_1^q(e^{-1/z_1}), ..., G_d^q(e^{-1/z_d})\} \quad z > 0$$
$$G(x) = G_*\{-1/\log(G_1(x_1)), ..., -1/\log(G_d(x_d))\}.$$

Because we have $G_j^k(\alpha_{k,j}x_j + \beta_{k,j}) = G_j(x_j)$ for any positive integer k and $j = 1, ..., d$, so not only does G_* have max-stable margins, it is also max-stable as well.

$$G_*^k(kz) = G_*(z) \quad z \in R^d; k = 1, 2, ...$$

Let μ_* be an measure of the extreme value distribution with Unit Fréchet margins, then we have

$$-\log G_*(z) = \mu_*([0, \infty]/[0, z]) \quad z \in [0, \infty) \tag{2.2}$$

leading to

$$\mu([q, \infty)/[q, x]) = -\log G(x) = -\log G_*(z) = \mu_*([0, \infty)/[0, z]) \quad x \in [q, \infty] \quad z \in [0, \infty]$$

As a result of above discussions, a multivariate extreme value distribution can be given with Unite Fréchet margins [6] .

2.3 Measures of dependence for bivariate extreme value distribution

Let $\{(x_1, y_1), (x_2, y_2), .., (x_n, y_n)\}$ denote a sample of observations of continuous random variables. If c denotes the number of concordant pairs , that is $(x_i - x_j)(y_i - y_j) > 0$, and d denotes the number of discordant paris, that is $(x_i - x_j)(y_i - y_j) < 0$. Then the Kendall's τ for these observations can be defined as

$$\frac{c-d}{c+d}$$

If these observations are denoted by two vectors (X, Y), $(X_1 Y_1)$ and (X_2, Y_2) be independent and identically distributed vectors with joint distribution H. Then the Kendall's τ can be written as

$$\tau = \tau_{X,Y} = P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\}$$

$-1 \leqslant \tau \leqslant 1$. Also, it can be written as an integration of the copula,

$$\begin{aligned}\tau &= 4 \int\int_{[0,1]^2} C(u, v) dC(u, v) - 1 \\ &= 4E\{C(u, v)\} - 1.\end{aligned}$$

To know more about the dependence, we now let (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) be three independent and identically distributed vectors with joint distribution H. Then the Spearman's ρ can be defended as

$$\rho_s = 3(P\{(X_1 - X_2)(Y_1 - Y_3) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_3) < 0\}).$$

Like Kendall's τ, it can be written as

$$\begin{aligned}\rho_s &= 12 \int\int_{[0,1]^2} uv dC(u, v) - 3 \\ &= 12E\{uv\} - 3\end{aligned}$$

$-1 \leqslant \rho_s \leqslant 1$.

For a bivariate extreme value distribution $G(x, y)$ with Unit Fréchet margins, from (2.2) we know the $G(x, y)$ can be written as

$$G(x, y) = \exp\{-\mu([0, (x, y)]^c)\}$$

The measure μ can be shown as

$$\mu([0, (x, y)]^c) = (\frac{1}{x} + \frac{1}{y}) A(\frac{x}{x+y}) \tag{2.3}$$

where we call the function $A(\frac{x}{x+y})$ as Pickands Dependence function.

Let $t = \frac{x}{x+y}$, the properties of $A(t)$ are:

(1) $A(0) = A(1) = 1$;

(2) $\max\{t, 1-t\} \leqslant A(t) \leqslant 1$;
(3) $A(t)$ is convex in $t \in [0,1]$.

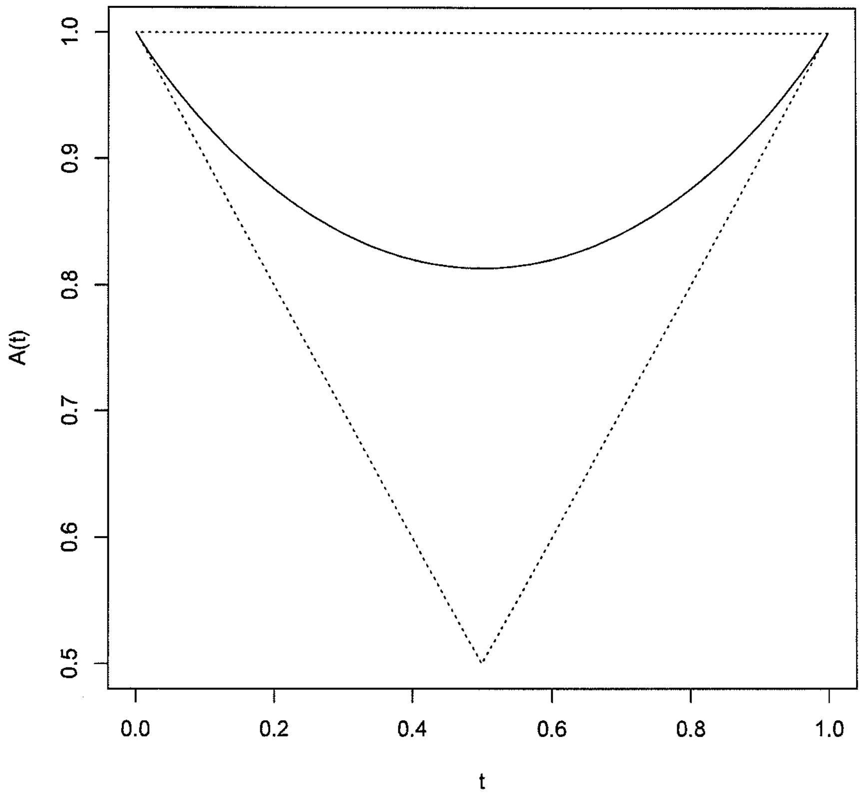


Figure 1: Pickands Dependence

The strength of dependence increases if the dependence function get closer the the lower bound, and we have independent case if $A(t) = 1$ for t $\in$[0,1], complete dependence if $A(\frac{1}{2}) = \frac{1}{2}$. The upper straight line shows the independent case while the lower bound shows the complete dependent case.

The Kendall's τ and Spearman's ρ can also be achieved by the integration of $A(t)$ [1] :

$$\tau = \int_0^1 \frac{t(1-t)}{A(t)} dA'(t) \tag{2.4}$$

$$\rho_s = 12 \int_0^1 [A(t)+1]^{-2} dt - 3 \tag{2.5}$$

2.4 Bivariate Extreme Value Models

For general bivariate extreme value distribution G with Unit Fréchet margins G_1 and G_2 and Pickands dependence function A. The distribution can be shown as:

$$G(x,y) = \exp\{\log(G_1G_2)A(\frac{\log G_2}{\log(G_1G_2)})\}.$$

The measure μ of $G(x,y)$ can also be written as

$$\mu([0,(x,y)]^c) = \int_N \max\{\frac{a^{(1)}}{x}, \frac{a^{(2)}}{y}\}S(da)$$

where $||\cdot||$ is the norm in R^2, N $= \{(a^{(1)}, a^{(2)})|||(a^{(1)}, a^{(2)})|| = 1\}$.

If we let $x \to \infty$ or $y \to \infty$, we get Unit Fréchet which leading to the expression

$$\begin{cases} \int_N a^{(1)}S(da) = 1 \\ \int_N a^{(2)}S(da) = 1. \end{cases}$$

Then, we have

$$\begin{aligned} \mu([0,(x,y)]^c) &= (\frac{1}{x}+\frac{1}{y})(\frac{xy}{x+y})\int_N \max\{\frac{a^{(1)}}{x}, \frac{a^{(2)}}{y}\}S(da) \\ &= (\frac{1}{x}+\frac{1}{y})\int_N \max\{\frac{ya^{(1)}}{x+y}, \frac{xa^{(2)}}{x+y}\}S(da) \\ \Longrightarrow \quad A(t) &= \int_N \max\{(1-t)a^{(1)}, ta^{(2)}\}S(da) \end{aligned}$$

with $t = \frac{x}{x+y}$.

3 Two new Bivariate Extreme Value Models

Following parametric models, generalized symmetric mixed model and generalized symmetric logistic model, are firstly suggested by Tajvidi [3] and formulated under the assumption of having Unit Fréchet margins. The parameters of these two models will decide how distribution functions look like which means the dependence of X and Y in the distribution change with the values of k and p.

Bivariate extreme values distributions with Unit Fréchet margins can be written as

$$G(x,y) = \exp\{-\mu([0,(x,y)]^c)\}.$$

As (2.4) and (2.5) show, the dependence measures, Kendall's τ and Spearman's ρ, can be formulated with the dependence function $A(t)$. Thus, to get the dependence relationship between X and Y, $A(t)$ where $t = \frac{x}{x+y}$ is required.

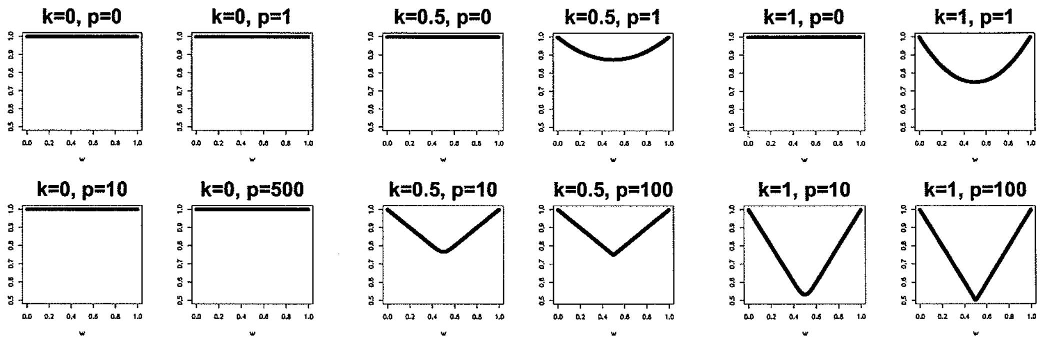
An extreme value model can be equivalently expressed by giving either the measure μ_* in (2.2) or the dependence function defined in (2.3). And in both extreme value models of this thesis, they will be introduced by showing their measures and dependence functions.

3.1 Generalized Symmetric Mixed Model

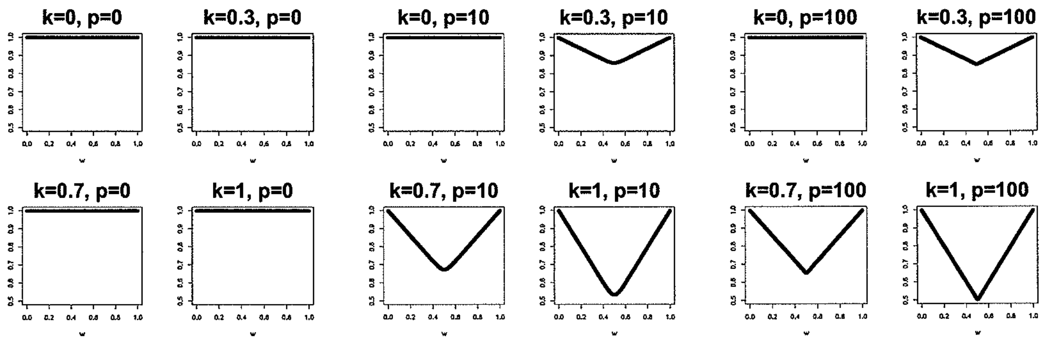
$$\mu([0,(x,y)]^c) = \frac{1}{x} + \frac{1}{y} + k(\frac{1}{x^p + y^p})^{1/p} \quad 0 \leqslant k \leqslant 1, \quad p \geqslant 0$$

$$A(t) = 1 - \frac{k}{(t^{-p} + (1-t)^{-p})^{1/p}} \quad 0 \leqslant k \leqslant 1, \quad p \geqslant 0$$

Pictures below show how $A(t)$ changes with parameters p and k:



(a) k=0 (b) k=0.3 (c) k=0.7



(d) p=0 (e) p=5 (f) p=800

Figure 2: $A(t)$ changes with k and p

The information of this model and above figures tell:

For $k = 0$ or $p = 0$ we have independence case.

For $k = 1$ and $p = \infty$ we have complete dependence.

For neither $k = 0$ nor $p = 0$, the increase of either k or p value results in more dependent relationship.

It would be really helpful if Kendall's τ and Spearman's ρ can be written as a function of p and k. However as the the integration function for $A(t)$ of τ, ρ_s, like we have shown in (2.4) and (2.5), is too complex to get a specific result. Adopting the idea of Riemann integration, that is if we want to get the answer of

$$\int_a^b f(x)dx \quad a < b$$

and we cannot work it out, then it can be calculated by approximation [7] . Let $a = x_0 < ... < x_n = b$, we have n equal length intervals with $x_k = x_{k-1} + k\frac{b-a}{n}$ for $k = 1, ..n$, the approximation is made by

$$\sum_{k=1}^{n} f(x_k^*)\frac{b-a}{n}$$

where $x_k^* \in [x_{k-1}, x_k]$. Here, we take the midpoint in the kth interval, which leads $x_k^* = \frac{1}{2}(x_{k-1} + x_k) = a + \frac{1}{2}(2k-1)\frac{b-a}{n}$ and the approximation is

$$\frac{b-a}{n}\sum_{i=1}^{n} f(a + \frac{1}{2}(2k-1)\frac{b-a}{n}).$$

Applying this approximation an approximated surface of how ρ_s changes with k and p is obtained.

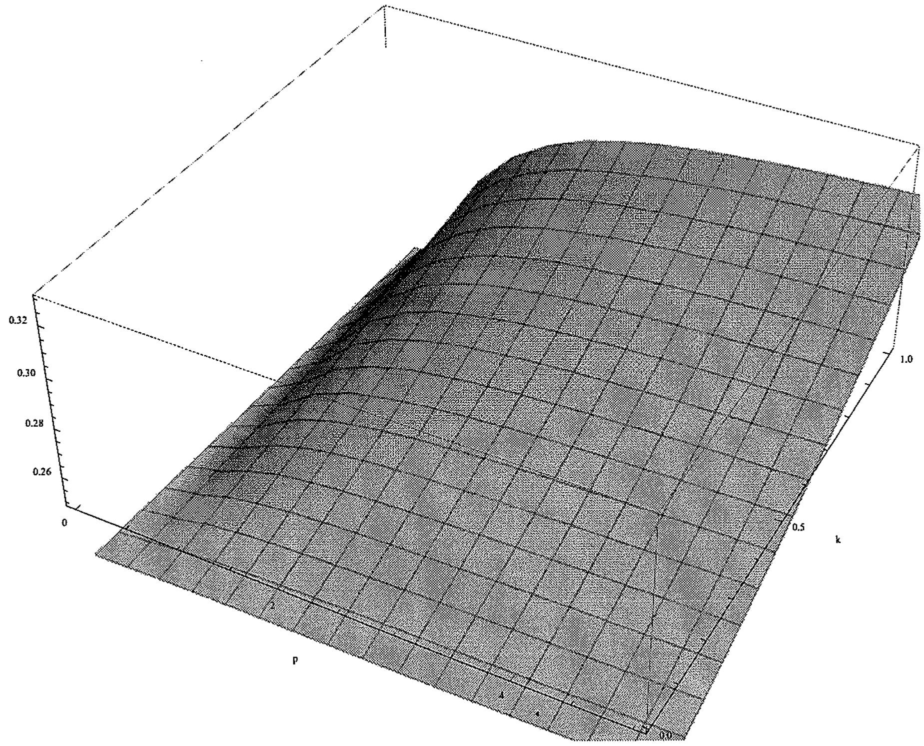


Figure 3: Spearman's ρ changes with k and p of generalized symmetric mixed model

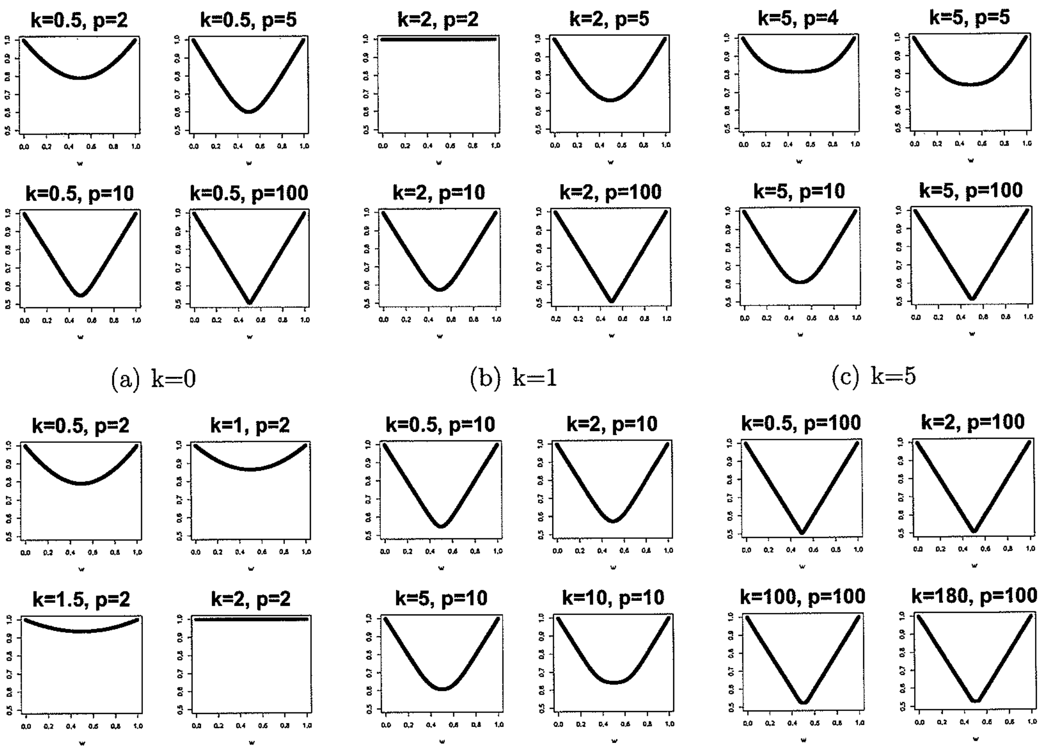
From the picture, the ρ_s will increase if k or p increase.

As a conclusion above, either k or p grows the dependent relationship of X and Y grows.

3.2 Generalized Symmetric Logistic Model

$$\mu([0,(x,y)]^c) = (\frac{1}{x^p} + \frac{1}{y^p} + \frac{k}{(xy)^{p/2}})^{1/p} \quad 0 < k \leqslant 2(p-1), \quad p \geqslant 2$$
$$A(t) = (t^p + (1-t)^p + k(t(1-t))^{p/2})^{1/p} \quad 0 < k \leqslant 2(p-1), \quad p \geqslant 2$$

Pictures below show how A(t) changes with parameters k and p:



(a) k=0 (b) k=1 (c) k=5

(d) p=2 (e) p=5 (f) p=800

Figure 4: $A(t)$ changes with k and p

The information of the model and above figures tell:

For $k = 2$ gives the symmetric logistic model.

For $k = 2$ and $p = 2$ we have independence case.

For $k = 2$ and $p = \infty$ we can obtain complete dependence.

When we set k, the larger of p the larger of dependence function. When we set p, the more k differs from p, the larger of dependence function. But if the p is large enough, k can only influence the dependent relationship a little.

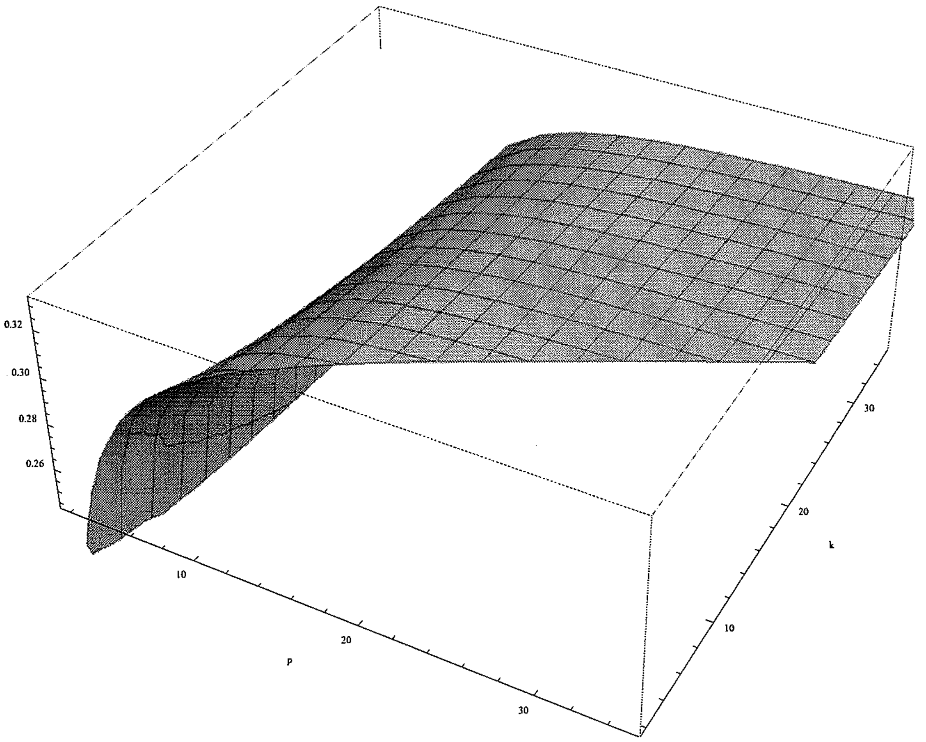


Figure 5: Spearman's ρ changes with k and p of generalized symmetric logistic model

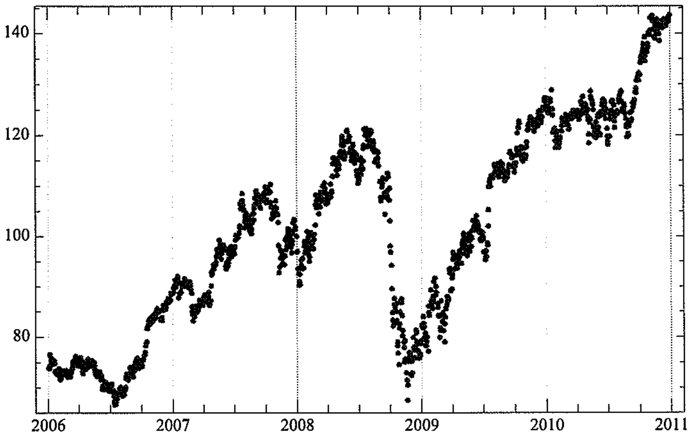
The figure shows if the p increases the ρ_s increases and if p is large enough k can only influence the dependence a little.

4 Application

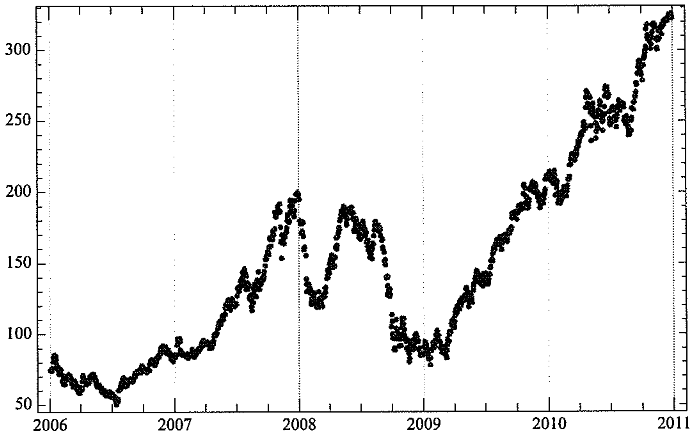
4.1 Data Introduction

The two datasets which will be studied in this section are the daily stock prices from 2006 to 2011 of two most well known computer companies in the world namely, IBM and Apple. As they both are world's leading computer companies, it is interesting to see how their stock prices depend on each other.

Figures 6(a) and 6(b) show the individual daily stock prices of these companies from 2006 to 2011



(a) IBM stock price



(b) Apple stock price

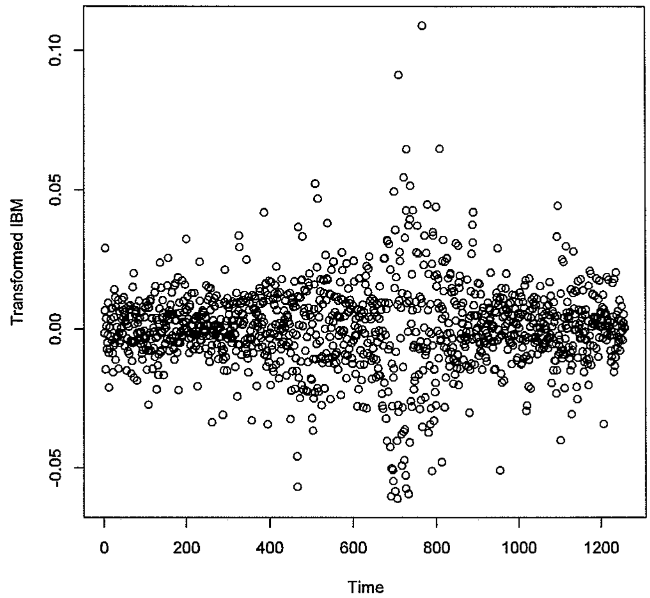
Figure 6: Original stock prices of IBM and Apple

These figures show that both IBM and Apple stock prices have two peaks around 2008, a sudden downwards at the end of 2008 and a comparatively stable increase after 2009. But they both have their own trends, for example during 2010, IBM seems to have a comparable stable stock price while Apple keeps its up-going trend.

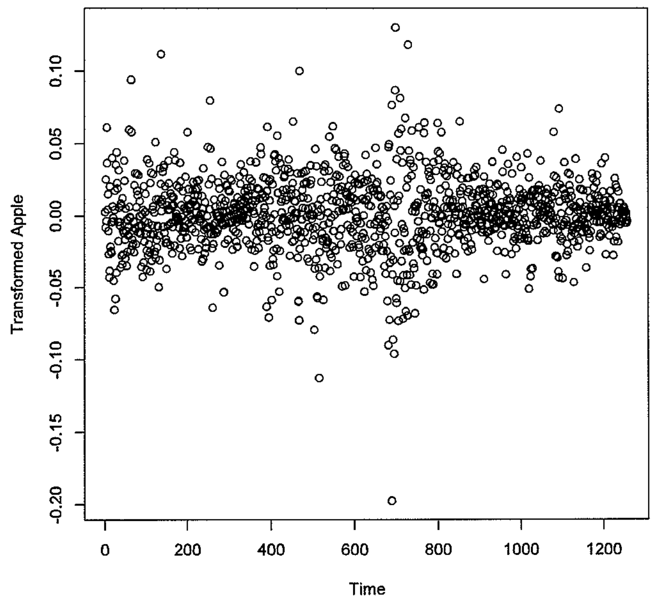
For further analysis, data needs to be transformed to a stationary set by using the transformation :

$$\ln(\frac{x_i}{x_{i-1}}) \quad i = 2, ..., n \tag{4.1}$$

where x_i are stock prices of IBM or Apple and n is the length of data and the transformed dataset tells us the returns of daily stock prices. After transformation we obtain a stationary series which are depicted in Figure 7(a) and 7(b) for IBM and Apple stock prices, respectively



(a) IBM



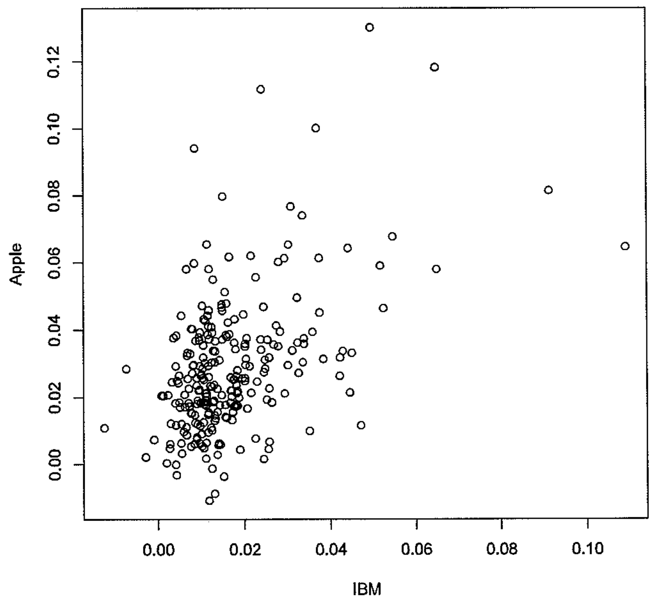
(b) Apple

Figure 7: Transformed data of IBM and Apple

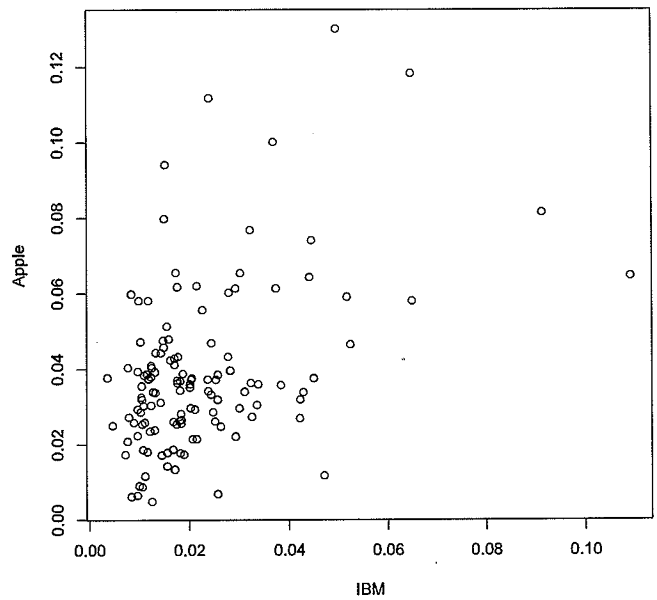
Both of the transformed datasets vary around y=0 and are assumed to follow a stationary process.

There are 5 workdays in each week. Thus, 5 seems to be a reasonable length to get the maxima in each block. For comparison, however, we also take 10 days as a block length to compare with the 5-day block maxima.

For visual investigation of dependence in transformed and blocked data, we present plots of IBM extreme returns against Apple extreme returns below.



(a) extreme returns with 5 observation in each block



(b) extreme returns with 10 observations in each block

Figure 8: IBM extreme returns against Apple extreme returns

4.2 Maximum-Likelihood estimation of parameters

To fit these extreme returns to our bivariate extreme value models, the estimation method called Inference Function for Margins (IFM) [8] is adopted. For multivariate distributions $F(Y_1, ..., Y_d)$ with parameters $(\alpha_1, ..., \alpha_d, \theta)$ which can be associated to univariate marginal distributions $F_i(Y_i; \alpha_i)$. Instead of maximizing the log-likelihood function to fit all parameters together, we maximize the log-likelihood function

$$\ell(\alpha_i) = \sum_{k=1}^{n} \log f(y_i; \alpha_i) \quad i = 1, .., d$$

(n is the size of random vector y_i) to get $\tilde{\alpha_i}$ first and then maximize the pseudo log-likelihood function

$$\ell(\theta; \tilde{\alpha_i}, ..., \tilde{\alpha_d}) = \sum_{i=1}^{d} \log c(F_i(y_i; \tilde{\alpha_i}), ..., F_i(y_d; \tilde{\alpha_d}); \theta)$$

where c is the density copula, to get the $\tilde{\theta}$.

In this thesis, we are going to fit parameters of two extreme value models where each model has Unit Fréchet margins. Following the IFM method, the first step is to get the estimated parameters $(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ for margins. After transforming datasets to Unit Fréchet distributions, the transformed datasets are used to fit our k and p by maximizing pseudo log-likelihood functions.

Thus the first step is to fit these data into univariate extreme models. Actually, there can be a time trend in local parameters which enables local parameters to change with time.

Hence, tests for different models are addressed:

Model 1: $\mu = \beta_0$
Model 2: $\mu = \beta_0 + \beta_1 t$

If 2(negloglik(model 1)-negloglik(model 2)) $< \chi^2(0.95, 1)$ we accept model 1 with no time trend in location parameter.

Estimated value	μ β_0	β_1	σ	γ	nllh
Model 1(5 obs)	0.011351		0.009268	0.039803	-782.4639
Model 2 (5 obs)	0.010274	0.002209	0.009240	0.037993	- 783.1285
Model 1(10 obs)	0.014905592		0.007601846	0.281810192	-395.3201
Model 2 (10 obs)	0.013429	0.002899	0.007464	0.305914	-396.7967

Table 1: Local parameter tests for transformed IBM

Estimated value	μ β_0	β_1	σ	γ	nllh
Model 1(5 obs)	0.01967		0.01524	0.02065	- 656.0937
Model 2 (5 obs)	0.023407	-0.007154	0.015119	0.015887	-658.7624
Model 1(10 obs)	0.02911		0.01542	0.05025	- 324.116
Model 2 (10 obs)	0.03615	-0.01357	0.01470	0.06556	-329.0732

Table 2: Local parameter tests for transformed Apple

Test result	IBM	Apple
5 obs in each block	Accept Null Hypothesis	Reject Null Hypothesis
10 obs in each block	Accept Null Hypothesis	Reject Null Hypothesis

Table 3: Test results for time time trend of GEV local parameters

As a conclusion, transformed IBM dataset has no time trend in location parameters and transformed Apple dataset has a time trend in location parameters. The time trend has been defined as $t = \frac{i}{m}$, $i = 1, 2, ..., m$ and m is the length of transformed dataset. Those estimated models can be shown as

$$\text{IBM} : G_X(x) = \exp\{-(1 + \gamma_x \frac{x - \mu_x}{\sigma_x})^{-\frac{1}{\gamma_x}}\}$$
$$\text{APPLE} : G_Y(y) = \exp\{-(1 + \gamma_y \frac{y - \beta_{0y} - \beta_{1y}t}{\sigma_y})^{-\frac{1}{\gamma_y}}\}$$

Check the estimated GEV models for 5 and 10 block length with probability plots and quantile plots:

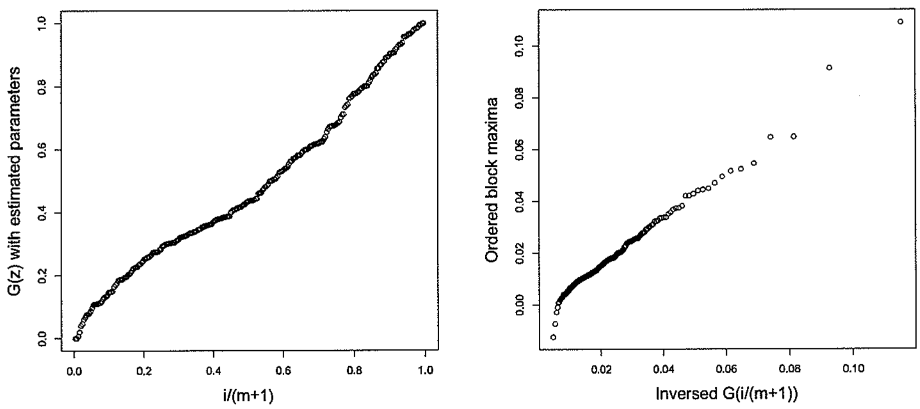


Figure 9: Probability plot and quantile plot for IBM extreme returns' model (5 obs)

	μ				
Estimated value	β_0	β_1	σ	γ	nllh
Model 1(5 obs)	0.01967		0.01524	0.02065	- 656.0937
Model 2 (5 obs)	0.023407	-0.007154	0.015119	0.015887	-658.7624
Model 1(10 obs)	0.02911		0.01542	0.05025	- 324.116
Model 2 (10 obs)	0.03615	-0.01357	0.01470	0.06556	-329.0732

Table 2: Local parameter tests for transformed Apple

Test result	IBM	Apple
5 obs in each block	Accept Null Hypothesis	Reject Null Hypothesis
10 obs in each block	Accept Null Hypothesis	Reject Null Hypothesis

Table 3: Test results for time time trend of GEV local parameters

As a conclusion, transformed IBM dataset has no time trend in location parameters and transformed Apple dataset has a time trend in location parameters. The time trend has been defined as $t = \frac{i}{m}$, $i = 1, 2, ..., m$ and m is the length of transformed dataset. Those estimated models can be shown as

$$\text{IBM}: G_X(x) = \exp\{-(1 + \gamma_x \frac{x - \mu_x}{\sigma_x})^{-\frac{1}{\gamma_x}}\}$$

$$\text{APPLE}: G_Y(y) = \exp\{-(1 + \gamma_y \frac{y - \beta_{0y} - \beta_{1y} t}{\sigma_y})^{-\frac{1}{\gamma_y}}\}$$

Check the estimated GEV models for 5 and 10 block length with probability plots and quantile plots:

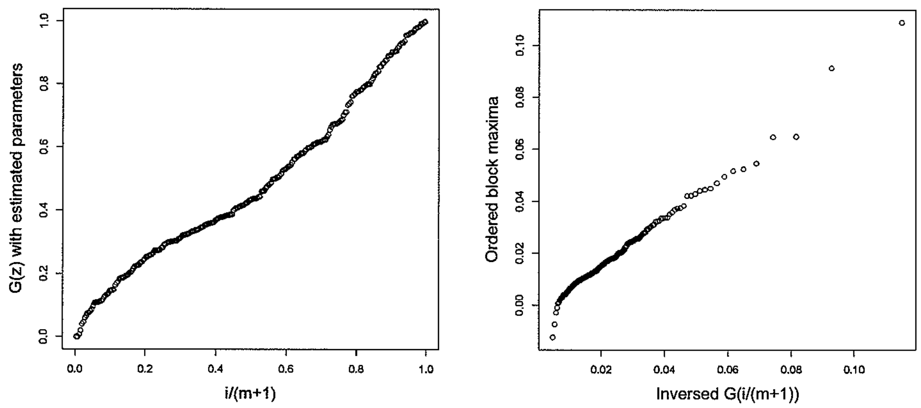


Figure 9: Probability plot and quantile plot for IBM extreme returns' model (5 obs)

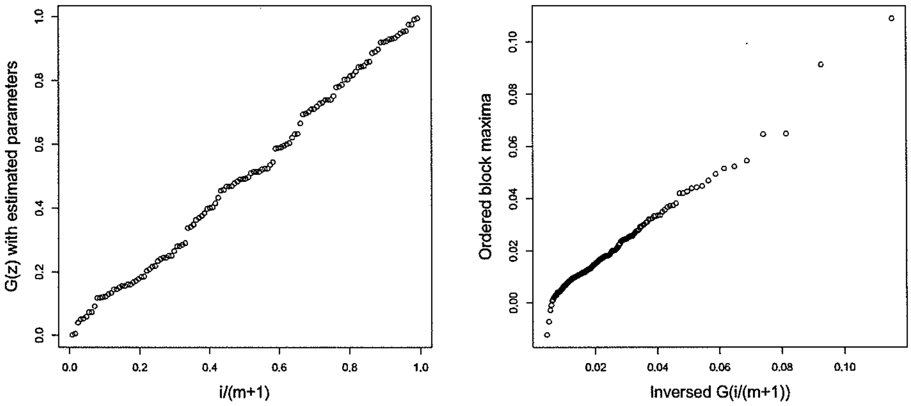


Figure 10: Probability plot and quantile plot for IBM extreme returns' model (10 obs)

For non-stationary variables, the lack of homogeneity in the distributional assumption leads to the modification

$$\tilde{z}_t = \frac{1}{\sigma}\log\{1 + \gamma\frac{z_t - \mu(t)}{\sigma}\}$$

each have the standard Gumbel distribution. And the probability plots consists of

$$\{(\frac{i}{m+1}, \exp\{-\exp\{-\tilde{z}_{(i)}\}\}), i = 1, ..., m\}$$

while the quantile plots consist of

$$\{(\tilde{z}_{(i)}, -\log\{-\log\{\frac{i}{m+1}\}\}), i = 1, ..., m\}.$$

We have probability plots and quantile plots for GEV models of Apple's extreme returns.

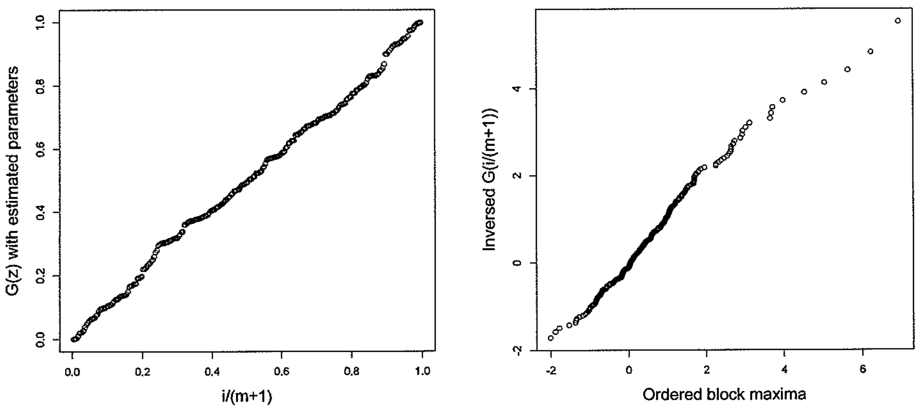


Figure 11: Probability plot and quantile plot for APPLE extreme returns' model (5 obs)

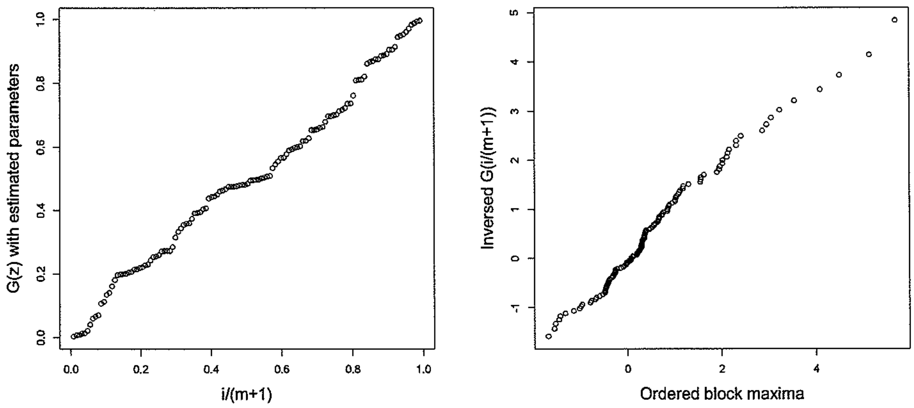


Figure 12: Probability plot and quantile plot for APPLE extreme returns' model (10 obs)

Figures above suggest that our GEV models are reasonable.

Then, we let

$$\begin{aligned} z_1 &= (1+\gamma_x \frac{x-\mu_x}{\sigma_x})^{\frac{1}{\gamma_x}} \\ z_2 &= (1+\gamma_y \frac{y-\beta_{0y}-\beta_{1y}t}{\sigma_y})^{\frac{1}{\gamma_y}} \end{aligned}$$

to transfer these datasets into Unit Fréchet.

Finally, by maximizing the pseudo log-likelihood functions with estimated parameters from margins, parameters of generalized symmetric mixed model and generalized symmetric logistic model k and p can be achieved.

Estimated value	k	p
Mixed model	1.0000000	0.6242569
Logistic model	2.696930	3.007469

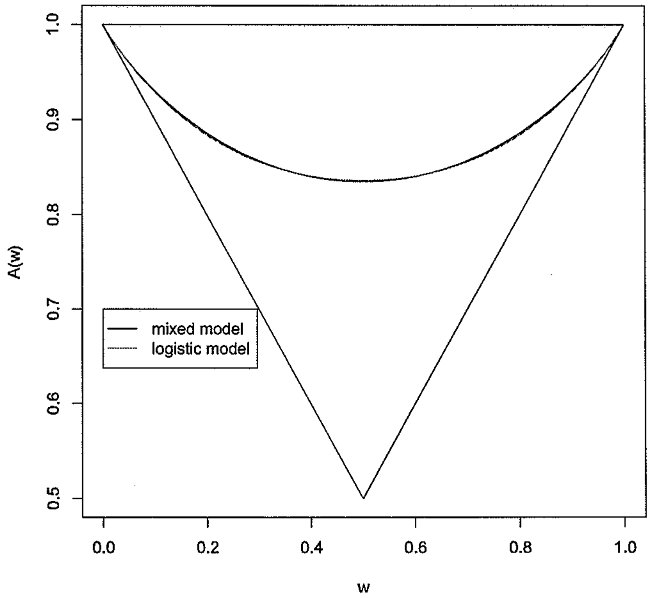
Table 4: Estimated parameters (5 observations in each block)

Estimated value	k	p
Mixed model	1.000000	0.6132411
Logistic model	3.233722	3.135627

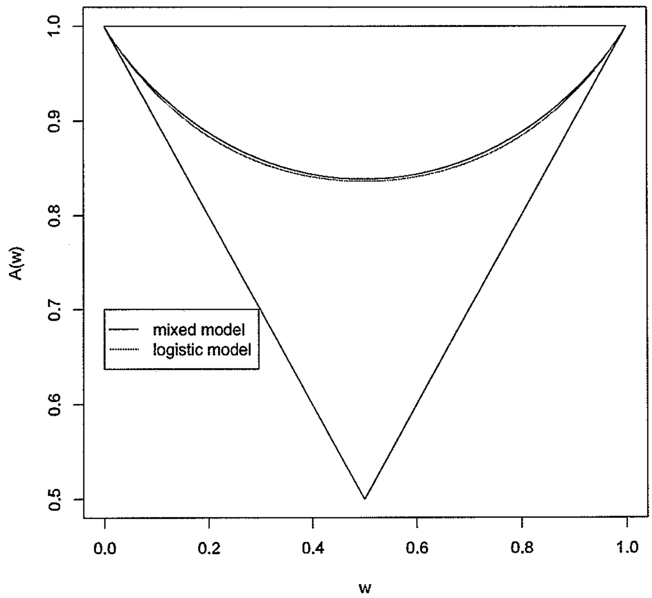
Table 5: Estimated parameters (10 observations in each block)

4.3 Further study

With all estimated parameters, it is possible to see extreme returns' dependent relationship between IBM and APPLE. Pickands dependence functions can be plotted as below:



(a) Dependence functions of two models(5 obs)

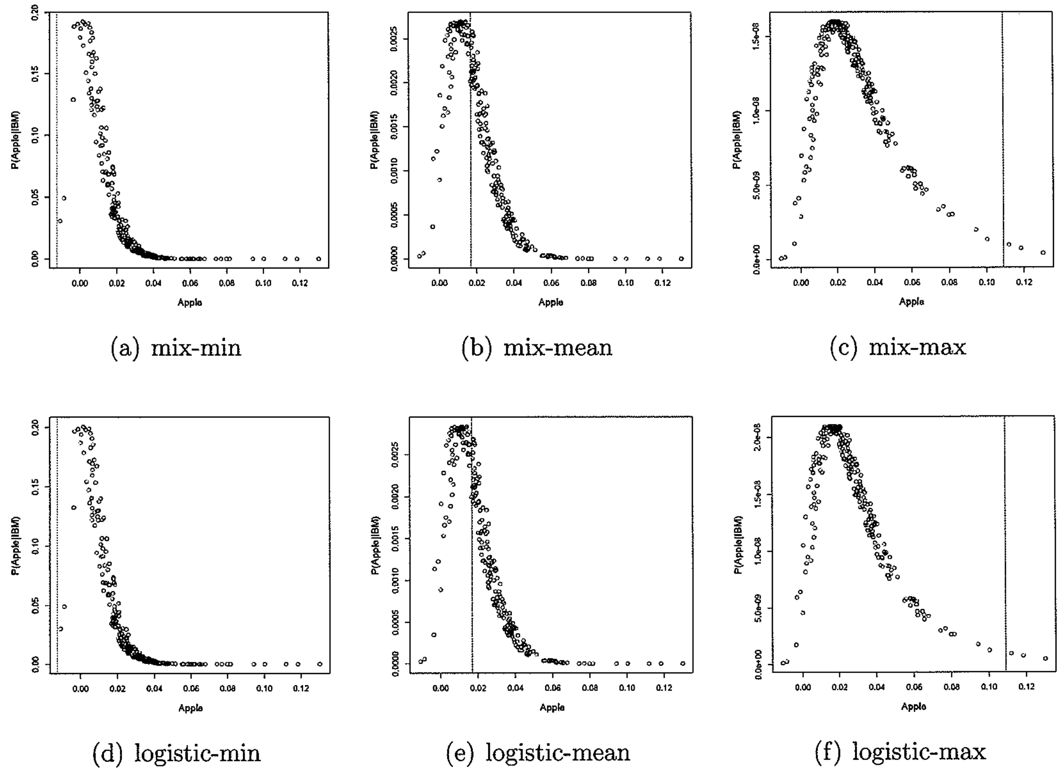


(b) Dependence functions of two models(10 obs)

Figure 13: Pickands dependence functions

It has been shown, firstly, the dependence functions follow with each other which ensures that these two new bivariate extreme models are proper. Secondly, the locations of depend functions tell us the extreme reruns of IBM and Apple are dependent with each other.

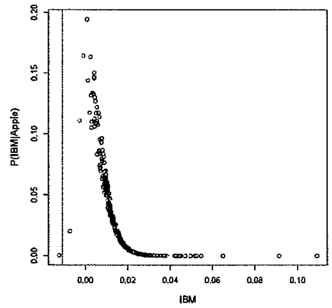
This analysis can be extended in the following way. In practice, people may be interested in conditional distributions (IBM| Apple) and (Apple| IBM) , that is, if one stock prices' extreme returns are known we would like to see how the other stock prices' extreme returns vary. Below, we set the minimum, mean and maximum extreme returns of one stock prices and see how the other stock prices' returns change in different models.



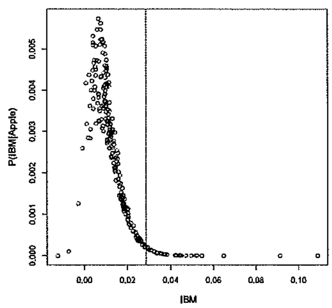
(a) mix-min (b) mix-mean (c) mix-max

(d) logistic-min (e) logistic-mean (f) logistic-max

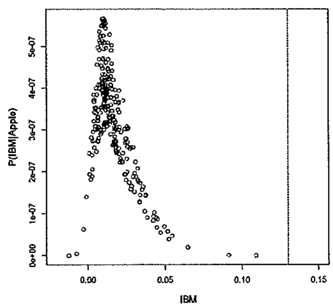
Figure 14: $f(Apple|IBM)$ (5 obs)



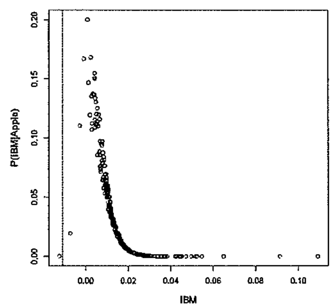
(a) mix-min



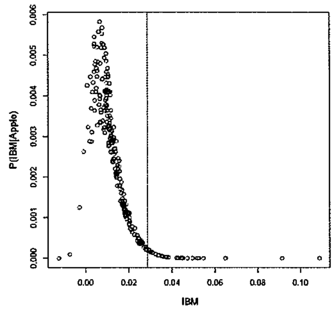
(b) mix-mean



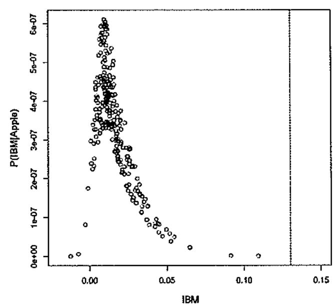
(c) mix-max



(d) logistic-min

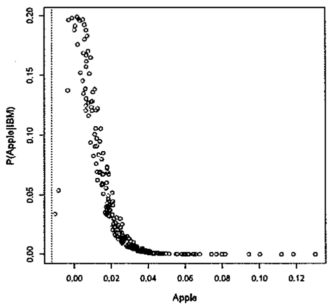


(e) logistic-mean

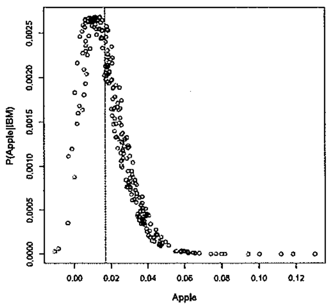


(f) logistic-max

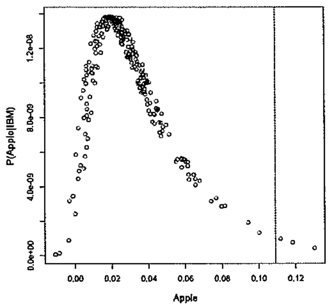
Figure 15: $f(IBM|Apple)$ (5 obs)



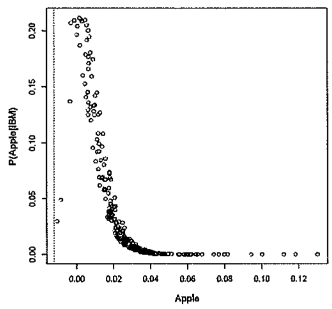
(a) mix-min



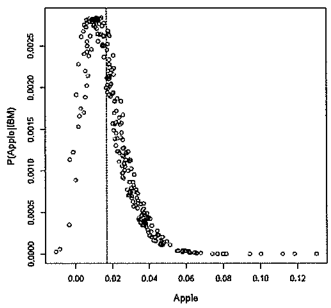
(b) mix-mean



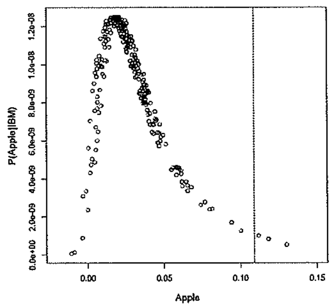
(c) mix-max



(d) logistic-min

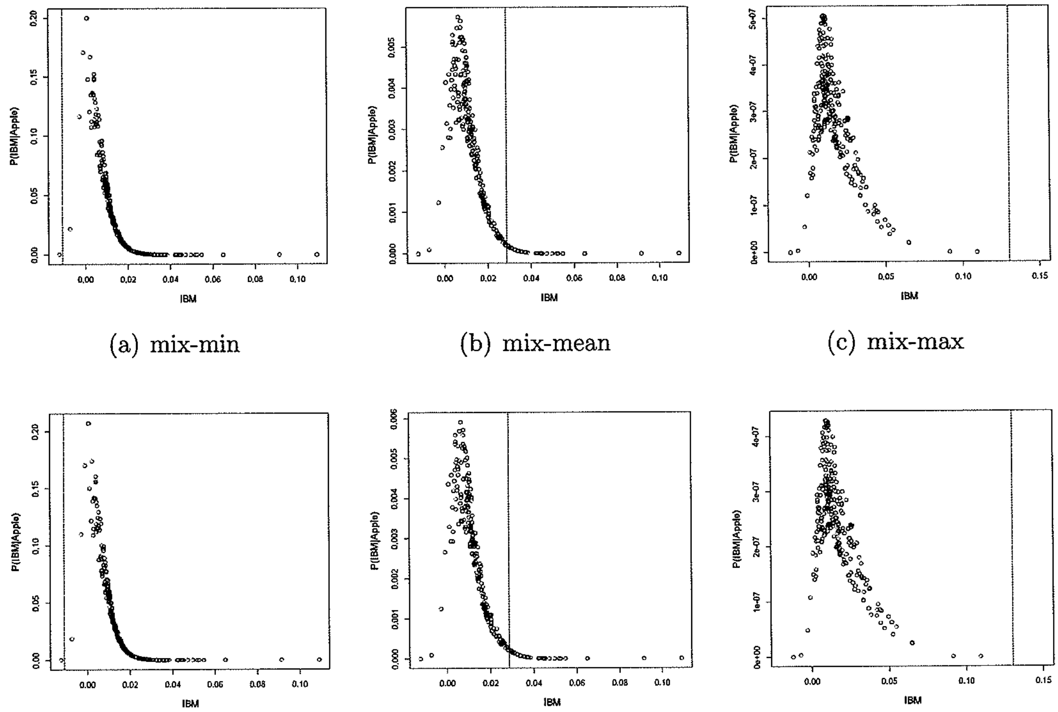


(e) logistic-mean



(f) logistic-max

Figure 16: $f(Apple|IBM)$ (10 obs)



(a) mix-min (b) mix-mean (c) mix-max

(d) logistic-min (e) logistic-mean (f) logistic-max

Figure 17: $f(IBM|Apple)$ (10 obs)

It seems that both block sizes and both parametric models result in similar conclusions. This demonstrates the validity of our two bivariate extreme models once again.

For $f(Apple|IBM)$, the conditional probability is slightly flat at the top and varies slower. Also, there are more cases that Apple stock prices' returns lie around 0.02. Furthermore, if we set the mean of IBM extreme returns, the mean of the conditional distribution lies around the mean of IBM extreme returns. It means IBM's extreme returns depend on Apple's.

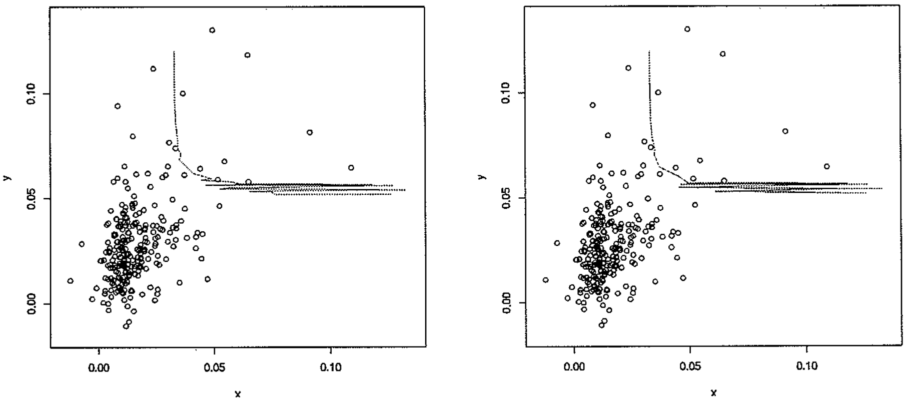
For $f(IBM|Apple)$, the conditional probability has a sharp peak and are more sensitive with IBM's extreme returns. Also, there are more cases that the IBM stock prices' extreme returns lie around 0.01. If we set the mean of Apple extreme returns, the mean of the conditional distribution lies away from the mean of Apple extreme returns. It means Apple's extreme returns hardly depend on IBM's

Another thing to be investigated is quantile plots. In many distributions, quantile plots are drawn to analysis the distributed datasets. It means for

$$G(X \leqslant x_q, Y \leqslant y_q) = p, \quad 0 < p < 1. \tag{4.2}$$

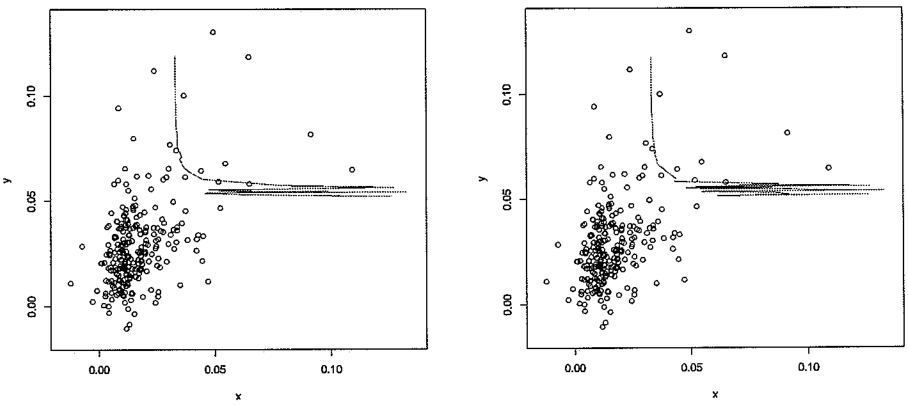
we want to find a series of (x_q, y_q) satisfying (4.2) which give us a curve. The points outside quantile curves of our bivariate extreme models mean: based on the distribution G, the extreme returns (x, y) in every 5 or 10 days be larger than (x_q, y_q) occur with the probability less or equal than $1 - p$.

This thesis takes $p = 90\%$ as the quantile to fit a quantile plot.



(a) 90%quantile plot for mixed model (b) 90% quantile plot for logistic model

Figure 18: 90% quantile plots for two models(5 obs in each block)



(a) 90%quantile plot for mixed model (b) 90% quantile plot for logistic model

Figure 19: 90% quantile plots for two models(10 obs in each block)

The points above quantile curves for both models are few, it means that the large returns which result in large change of stock prices for both IBM and Apple will not happen with probability over 10%.

The similar results in quantile plots of mixed and logistic models prove the validity of these models again.

5 Conclusions and Further Work

By fitting estimated model parameters and plotting the Pickands dependence functions, conditional distributions and quantile plots, we get similar results of the same dataset for generalized symmetric mixed model and generalized symmetric logistic model. This demonstrates that these two models are valid and reasonable. The influence of dependence in these two models are measured parameters k and p. Thus, once the estimated k and p is calculated the dependence is known.

For IBM and Apple stock prices' extreme returns, it is reasonable to say that they are dependent with each other. But on the other hand, from the conditional distributions it is obvious to see that IBM depends on Apple more than Apple depends on IBM. In a word, they are dependent with each other but the dependent relationship is asymmetric.

Even though we can study the dependent relationship in extreme returns with dependence function $A(t)$, it would be interesting to study Kendall's τ and Spearman's ρ to see how exactly the k and p influence these measures of dependence.

References

[1] P.Capereaa, A.-L. Fougeres and c.Genest(1997), A nonparametric estimation procedure for bivariate extreme value copulas

[2] Hall Peter and Tajvid nader, Distribution and Dependence-Function Estimation for Bivariate Extreme-Value Distributions. Bernoulli 6(5), 835-844 (2000).

[3] Tajvidi Nader, Characterisation and Some Statistical Aspect of Univariate an Multivariate Generalised Pareto distribution(1996)

[4] Stuart Coles, An introduction of statistical modeling of extreme values(2001)

[5] Alec Stephenson, package evd in R

[6] J.Beirlant,Y.Goegebeur, JlSegers and J.Teugels , Statistics of Extremes: Theory and Applications

[7] Martha L. Abell and James P. Braselton, Mathematica By Example

[8] Harry Joe and James J. X, The Estimation Method of Inference Functions for Margins for Multivariate Models

A R codes for applications

The R codes used to estimate two new models parameters are shown below:

```
#x_IBM , y_APPLE original data, n is the length of datasets.

# change them  to stationary
xs<-log(x_IBM[2:n]/x_IBM[1:(n-1)])
ys<-log(y_APPLE[2:n]/y_APPLE[1:(n-1)])

#take the block maxima with 5 or 10 observations in each block. And plot them togethe
l=5;# block maxima
m<-ceiling(n/l)
x<-rep(0,m)
y<-rep(0,m)
for(i in 1:m-1){
x[i]<-max(xs[(1+l*(i-1)):(l*i)])
}
x[m]<-max(xs[(l*m-l):n-1])
for(i in 1:m-1){
y[i]<-max(ys[(1+l*(i-1)):(l*i)])
}
y[m]<-max(ys[(l*m-l):n-1])
plot(x,y,xlab = "IBM",ylab = "Apple")

#Fitting GEV models with extreme values and get one with time trend in location param
trend<-(1:m)/m
m1<-fgev(x)
m1t<-fgev(x,nsloc=trend)
-2*(logLik(m1)-logLik(m1t)) < qchisq(0.95,1)
m2<-fgev(y)
m2<-fgev(y,nsloc=trend)
-2*(logLik(m2)-logLik(m2t)) < qchisq(0.95,1)

#probability plot and quantile plot
#IBM
z<-sort(x)
prpx<-c(1:m)/(m+1);
prpy<- exp(-(1+m1$estimate[3]*(z-m1$estimate[1])/m1$estimate[2])^(-1/m1$estimate[3]))
plot(prpx,prpy,xlab = "i/(m+1)", ylab = "G(z) with estimated parameters",cex.lab=1.5)
#quantile plots
qupx<-m1$estimate[1]-m1$estimate[2]/m1$estimate[3]*(1-(-log(prpx))^(-m1$estimate[3]))
plot(qupx,z,xlab = "Inversed G(i/(m+1))", ylab = "Ordered block maxima",cex.lab=1.5)
#for APPLE
t<-(1:m)/m
ty<-1/m2$estimate[4]*log(1+m2$estimate[4]*(y-m2$estimate[1]-m2$estimate[2]*t)/m2$esti
z<-sort(ty)
prpy<-exp(-exp(-z))
plot(prpx,prpy,xlab = "i/(m+1)", ylab = "G(z) with estimated parameters",cex.lab=1.5)
```

```
qupy<- -log(-log(prpx))
plot(z,qupy,xlab = "Ordered block maxima", ylab = "Inversed G(i/(m+1))",cex.lab=1.5)



#tansform to univariate Frechet margins.
xu<-(1+m1$estimate[3]*(x-m1$estimate[1])/m1$estimate[2])^(1/m1$estimate[3])
yu<-(1+m2$estimate[4]*(y-m2$estimate[1]-m2$estimate[2]*t)/m2$estimate[3])^(1/m2$estim

#Get our parameters in extreme value models by maximizing pseudo log- likelihood func
#negloglikelihood for mix model
wx<-(xu+yu)/xu
wy<-(xu+yu)/yu
nllmix<-function(par) {
k<-par[1]
p<-par[2]
-sum(log(1/xu^3/yu^3*exp(-(1/xu+1/yu)*(1-k*(wx^p+wy^p)^(-1/p)))*(wx^p+wy^p)^(-2*(1+p)
}
est.mix<-function(start){
optim(start,nllmix,method="L-BFGS-B",lower=c(0,0),upper=c(1,Inf))
}
#negloglikelihood for logistic model
nlllog<-function(par){
k<-par[1]
p<-par[2]
exy<-(xu^(-p)+yu^(-p)+k*((xu*yu)^(-p/2)))^(1/p)
    fxy<-(exp(-exy)*xu^(p-1)*yu^(p-1)*exy*(k^2*xu^p*yu^p*(exy-1)+4*(xu*yu)^p*(exy-1+p
if(par[1]>2*(par[2]-1))
 {cat("Warning 2")
1e09}
else if (any(fxy < 1e-12))
     {cat("Warning 1")
      1e09}
      else
    -sum(log(fxy))
}
est.log<-function(start){
 optim(start,nlllog,method="L-BFGS-B",lower=c(0,2),upper=c(Inf,Inf))
}



#Dependece function of two extreme value models with estimated parameters
#10obs,k=1,p=0.6132411
#5obs k=1,p=0.5971535
A_mix<-function(x,k=1,p=0.6132411){
1- k/(x^(-p) +(1-x)^(-p))^(1/p)
}
```

```
#10 obs, k= 3.233722 ,p=3.135627
#5 obs k=2.774460,p=2.974842
A_log<-function(x,k= 3.233722 ,p=3.135627){
(x^p+(1-x)^p+k*(x*(1-x))^(p/2))^(1/p)
}
curve(A_mix,0,1,xlim=c(0,1),ylim=c(0.5,1),xlab="w",ylab="A(w)")
par(new=TRUE)
curve(A_log,0,1,xlim=c(0,1),ylim=c(0.5,1),col="red",xlab="",ylab="")
legend(list(x=0,y=0.7), legend = c("mixed model","logistic model"), col=1:2,lty=1, me

#Conditional density functions ( they work in the same way, here we show one conditio
Hai_mix<-function(x,y,kh,ph){
gevx<-exp(-(1 + m1$estimate[3]*((x-m1$estimate[1])/m1$estimate[2]))^(-1/m1$estimate[3
gevy<-exp(-(1 + m2$estimate[4]*(y-m2$estimate[1]-m2$estimate[2]*t)/m2$estimate[3])^(-
xu_s<-(1+m1$estimate[3]*(x-m1$estimate[1])/m1$estimate[2])^(1/m1$estimate[3])
yu_s<-(1+m2$estimate[4]*(y-m2$estimate[1]-m2$estimate[2]*t)/m2$estimate[3])^(1/m2$est
wx_s<-(xu_s+yu_s)/xu_s
wy_s<-(xu_s+yu_s)/yu_s

(1/xu_s^3/yu_s^3*exp(-(1/xu_s+1/yu_s)*(1-kh*(wx_s^ph+wy_s^ph)^(-1/ph)))*(wx_s^ph+wy_s
}
plot(y,Hai_mix(min(x),y,1,0.6132411),xlab="Apple",ylab="P(Apple|IBM)",cex.lab=1.2)
abline(v=min(x),col="blue")




#Quantile plots for two extreme models
G<-function(x,y,k,p){
xu<-(1+m1$estimate[3]*(x-m1$estimate[1])/m1$estimate[2])^(1/m1$estimate[3])
yu<-(1+m2$estimate[4]*(y-m2$estimate[1]-m2$estimate[2]*t)/m2$estimate[3])^(1/m2$estim
exp(-(1/xu+1/yu-k*(1/(xu^p+yu^p))^(1/p)))
}
H<-function(x,y,k,p){
xu<-(1+m1$estimate[3]*(x-m1$estimate[1])/m1$estimate[2])^(1/m1$estimate[3])
yu<-(1+m2$estimate[4]*(y-m2$estimate[1]-m2$estimate[2]*t)/m2$estimate[3])^(1/m2$estim
exp(-(1/xu^p+1/yu^p+k/(xu*yu)^(p/2))^(1/p))
}
xax<-seq(-0.015,0.135,by=0.0001)#1-6 70%  11-15 90%
yax<-seq(-0.015,0.135,by=0.0001)
#mix10obs,k=1,p=0.6132411
#mix5obs k=1,p=0.5971535
#10 obs, k=3.233722 ,p=3.135627
#5 obs k=2.774460,p=2.974842
#tmp<-outer(xax,yax,G,k=1,p=0.6132411)
tmp<-outer(xax,yax,H,k=3.233722 ,p=3.135627)
q<-0.90
I<-which(abs(tmp-q)<1e-5)
```

```
xx<-rep(xax,length(yax))
yy<-rep(yax,each=length(xax))
xd<-rep(0,length(I))
yd<-rep(0,length(I))
for(i in 1:length(I)){
xd[i]<-xx[I[i]]
yd[i]<-yy[I[i]]
}# xd,yd is the quantile line
plot(x,y,xlim=c(-0.015,0.135),ylim=c(-0.015,0.135))
lines(xd,yd,col="red")
```

Master's Theses in Mathematical Sciences 2012:E15
ISSN 1404-6342

LUNFMS-3037-2012

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/