Order-statistics-based Inferences for Censored Lifetime Data and Financial Risk Analysis

A thesis submitted for the degree of
Doctor of Philosophy

by

Zhuo Sheng

Department Mathematical Sciences

School of Information Systems, Computing and Mathematics

Brunel University, London

July 2013

Abstract

This thesis focuses on applying order-statistics-based inferences on lifetime analysis and financial risk measurement. The first problem is raised from fitting the Weibull distribution to progressively censored and accelerated life-test data. A new order-statistics-based inference is proposed for both parameter and confidence interval estimation.

The second problem can be summarised as adopting the inference used in the first problem for fitting the generalised Pareto distribution, especially when sample size is small. With some modifications, the proposed inference is compared with classical methods and several relatively new methods emerged from recent literature.

The third problem studies a distribution free approach for forecasting financial volatility, which is essentially the standard deviation of financial returns. Classical models of this approach use the interval between two symmetric extreme quantiles of the return distribution as a proxy of volatility. Two new models are proposed, which use intervals of expected shortfalls and expectiles, instead of interval of quantiles. Different models are compared with empirical stock indices data.

Finally, attentions are drawn towards the heteroskedasticity quantile regression. The proposed joint modelling approach, which makes use of the parametric link between the quantile regression and the asymmetric Laplace distribution, can provide estimations of the regression quantile and of the log linear heteroskedastic scale simultaneously. Furthermore, the use of the expectation of the check function as a measure of quantile deviation is discussed.

Certificate of Originality

"I hereby certify that the work presented in this thesis is my original research and has not been presented for a higher degree at any other university or institute".

.....

Zhuo Sheng

Acknowledgements

I would like to thank many people who have helped me through the completion of this dissertation. First,I would like to express my sincere gratitude to my supervisor Dr Keming Yu, who is patient, honest, and inspiring when dealing with me, throughout the four years of research. It would have been very difficult to achieve whatever little I have managed to achieve without his guidance, advise, and, sometimes, a little motivational push that keeps my foot on the accelerator.

My thanks also goes to Prof Bing Xing Wang and Prof James W. Taylor, for providing valuable support and assistance throughout different phases of research and especially during the final revision period. I would also like to extend sincere gratitude to Matthias Glafke and Raheem Jabbar Thaher, my fellow PhD students at the Department of Mathematical Sciences of Brunel University, for the generous help I have received from them.

The two examiners of my thesis, Dr Shaomin Wu and Dr Veronica Vinciotti, have given me a lot of valuable comments and suggestions. I thank them for their guidance and critical opinions.

Last but not the least, I would like to thank the ones closest to me: my parents, Da Quan Sheng and Yu Ying Dong, and my girlfriend, Ruobing Ai. They have been a constant source of support emotional, moral and of course financial during my postgraduate years, and this thesis would certainly not have existed without them. it is to them that this thesis is dedicated.

Author's Publication

1. **Sheng, Z.**, Yu, K. and Wang, B.X. (2012), Point and interval estimation for the generalised Pareto distribution with small samples, *Submitted to Statistics and Its Interface, Special Issue on "Extreme Values"*

2. **Sheng, Z.** and Yu, K (2012), Volatility forecast using expected shortfall and expectile, *Working paper*

3. Wang, B.X., Yu, K. and **Sheng, Z.** (2013), Inference on Weibull constant stress accelerated life-testing model under progressively censoring, *Submitted to IEEE Transactions on Reliability*

4. **Sheng, Z.** and Yu, K (2013), Joint modelling of regression quantiles and heteroskedasticity with asymmetric Laplace distribution, *Working paper*

Contents

List of Figures

List of Tables

Acronym

ALT accelerated life test

ARCH autoregressive conditional heteroskedasticity

BC$_a$ bias-corrected and accelerated bootstrap confidence intervals

CAViaR conditional autoregressive VaR

c.d.f. cumulative distribution function

CI confidence interval

CP coverage probability

CSALT constant stress ALT

EPM elemental percentile method

EVD extreme value distribution

EVT Extreme Value Theory

EWMA exponentially weighted moving average

GARCH generalized autoregressive conditional heteroskedasticity

GPD generalised Pareto distribution

ISE integrated squared error

LAD least absolute deviation

LME likelihood moment estimation

MDA maximum domain of attraction

MAD mean absolute deviation

MLE maximum likelihood estimation

MOM method of moments

MSE mean squared error

NEW1 the proposed new inference in Chapter 2

NEW2 the proposed new inference in Chapter 3

OLS ordinary least square

p.d.f. probability density function

PERC percentile bootstrap confidence intervals

POT Peak Over Threshold

PWM probability weighted moments

QD quantile deviation

RMSE root mean square error

RV Realised Volatility

VaR value at risk

ZS method introduced by Zhang and Stephens (2009) for estimating the generalised Pareto distribution

Chapter 1

Introduction

Order statistics are among the most fundamental tools in non-parametric statistics. The earliest order-statistics-based inference was studied by Karl Pearson in 1902, according to Wilks (1948). In the past century, numerous order statistics based theories and applications have been created and studied, which makes it impossible to summarise in the length of a thesis. Title of this thesis employs the term 'Order-statistics-based inference' to unify different contents of related researches under a categorical and summary framework. The research scopes of this thesis are given in details in this chapter. First of all, we begin with a brief overview of some most basic concepts of the order statistics.

Order statistics of a random sample $X_1, , ...X_n$ are, in short words, the sample value re-sorted in ascending order, which are denoted as $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$. The sample maximum and minimum then can be denoted as the last and the first order statistics, respectively:

$$X_{1:n} = \min\{X_1, , ...X_n\}, \qquad X_{n:n} = \max\{X_1, , ...X_n\}.$$

Furthermore, many common statistics can be derived from order statistics or be viewed as a special case. Median, quartiles are essentially special order statistics; the sample range, which is defined as

$$\text{Range} = X_{n:n} - X_{1:n},$$

is also a special order statistics.

Order statistics are also used in analysing and estimating distributions. For instance, the L-moments (Hosking, 1990) are essentially linear combinations of order statistics,

which provides an alternative view to summarise the distribution.

Studying of the limiting distribution of the largest order statistics by Fisher and Tippett (1928) has become part of the foundations of the modern extreme value theory. Moreover, if viewed as special cases of order statistics, the quantiles also provides a unique angle of analysing distributions, not to mention the large amount of theoretical and practical research results is based on this concept.

The scopes of the thesis involve several different applications of order-statistic-based inferences, ranged from distribution estimation for lifetime data to measurement of financial risk. To this end, the following sections of this chapter introduce and review related topics with the aim of creating links between research scopes and order-statistics-based inferences.

1.1 Lifetime analysis with order-statistics

Lifetime analysis is an important topic in many areas, such as engineering, health and medical, biology, and social science. In different contexts, it is often referred variously as survival analysis, failure time analysis or reliability analysis. These studies are categorised as *time-to-event* data analysis in some literatures, where the term 'event' refers to significant occurrence or result, such as failure or death. Note that the time scale in this subject are not necessarily real or chronological time (Lawless, 1982), as it could refer to other measurements of usage or durability that are often encountered in engineering, for example, mileage of auto-mobiles, number of hits of keyboard buttons, *etc.* Nonetheless, consider lifetime variable T, then the lifetime distribution of T is a fundamental subject in lifetime analysis. For continuous models, the lifetime or failure time function

$$F(t) = \mathbb{P}(T \leq t) = \int_0^t f(x)\, dx,$$

where $F(t)$ and $f(t)$ denote the cumulative distribution function (c.d.f.) and probability density function (p.d.f.) respectively, and t is usually defined over the interval $[0, +\infty)$. The survival function is given by:

$$S(t) = \mathbb{P}(T \geq t) = \int_t^\infty f(x)\, dx,$$

which is defined as the probability of an object surviving to at least time t. Common lifetime distributions include the exponential distribution, Weibull distribution, and extreme value distribution, *etc.*. For more types of lifetime distributions and their

details, see Lawless (1982), Miller et al. (1981).

Let T_i $(i = 1, ..., n)$ be observations of the lifetime variable T, registered as the time of 'event' taking place. Sorting in increasing order then $T_{1:n} \leq T_{2:n} \leq ... \leq T_{n:n}$ is order statistics of the lifetime. However, in reality or experiments, the actual number of observed failure is often smaller than the number of object at the beginning of the observation for various reasons. That is, the order statistic of the last observation is $T_{m:n}$, where $m < n$. This problem is closely related with three topics: censoring, accelerated life-test, and progressive censoring, which are especially the focuses as part of this thesis. We briefly explore and review these topics in the following part of this section.

A sample is said to be *censored* if the number of actual observations is smaller than the number of objects at the beginning of the observation. In lifetime analysis, samples are often *right censored*, that is, some data points above certain value is unobserved. For example, a sample order statistics $T_{1:n}, ..., T_{m:n}$, where $m < n$, is right-censored. On contrary, *left censoring* is that some data points below certain value is unknown, *i.e.*, the first observed data point is $T_{k:n}$, where $k > 1$ is known, but $T_{1:n}, ..., T_{(k-1):n}$ are unobserved. Right censoring is very common in lifetime data, as it is often impractical or impossible to obtain full data set.

Furthermore, another set of censoring types that usually concerned in lifetime analysis are the Type-I and Type-II censoring, of which definitions are as following (Miller et al., 1981):

- Type-I: let t_c be the fixed censoring time such that events beyond that time is not observed. Namely, only data points $T_{i:n} \leq t_c$, $i = 1, ...n$ are available in the sample set.
- Type-II: let $m < n$ be fixed, then the observation stops when the mth sample $T_{m:n}$ is obtained.

To our best knowledge, Gross and Clark (1975) is the first book to discuss parametric and non-parametric models for both complete and censored lifetime data.

Another problem that is often confronted in practice is that, samples sometimes are withdrawn from the sample pool before the event is observed, due to various reasons (*i.e.*, patients quit experiment prematurely, testing items are accidentally damaged or contaminated rendering results invalid). This problem is defined and discussed as *progressive censoring* in many literatures. Balakrishnan and his coordinators have made considerable contributions in this area of research, see Balakrishnan and Aggarwala

(2000), Ng et al. (2002), Ng et al. (2004), Balakrishnan et al. (2004), Balakrishnan and Xie (2007), for examples. Other contributors in the specific area of progressive censoring for lifetime data include (but not limited to): Cohen (1963), and more recently, Wang and Yu (2009), Wang et al. (2010), *etc.*

For lifetime data, both Type-I and Type-II right censoring arise in practice or experiment, and often are due to the reason that it is impractical or inefficient to keep observing until defined events occur for all samples. Thus the observation has to stop after a certain period of time (Type-I), or after a certain amount of events being registered (Type-II). In engineering, a technique called *accelerated life-test* is frequently used to shorten the length of an experiment by putting samples under 'stress', which is beyond the normal or standard level, and shortens their life, hence 'to quickly obtain data which, properly modelled and analysed, yield desired information on product life or performance under normal use' (Nelson, 2009). For example, Nelson (1972) described the results of a life-test of a type of electronic insulating fluid under different voltage stress, ranging from 26 to 38 kilovolts (kV), while the 'normal' voltage is 20kV. The accelerated life-tests for Weibull distribution and extreme value distribution are firstly studied by Meeker and Nelson (1975), and for normal and log-normal distributions by Kielpinski and Nelson (1975). Research in this area has been further extended by many, for example, Miller and Nelson (1983), Nelson (1980), Yang (1994), and Huang (2011), among others.

Sometimes the data censoring problem and the accelerated tests problem are combined in practice. Hence, Chapter 2 of this thesis is dedicated to addressing estimation problems of Weibull distribution for accelerated life-tests under progressive Type-II right censoring scheme.

1.2 Extreme value theory

In this section we briefly introduce the Extreme Value Theory, which is a branch of statistics that fulfils the needs of modelling extreme events and extreme probabilities in many disciplines, and that emerged from research of the limiting distribution of the largest order statistics in a sample as the sample size increases to infinity (Fisher and Tippett, 1928). The analysis of extreme values has been widely used in areas such as research of nature phenomenon and nature disasters, reliability analysis in engineering, *etc.* In recent decades there is also increased use of EVT in finance, especially financial risk management (for example, see McNeil, 1999). Generally, there are two approaches in the framework of EVT for practical extreme value analysis, which are

the block of maximum approach and the peak over threshold approach. The first approach is mainly based on the extreme value distributions, while the second approach focuses on utilising the generalised Pareto distribution. We briefly discuss the two approaches from a distributional point of view in the following sections.

1.2.1 The extreme value distributions and the generalised extreme value distribution

The idea of the block of maximum approach is as follows. Assume a block (subsample) of i.i.d. samples $\{X_1, X_2, ..., X_n\}$ with distribution function $F(x)$, and define the maxima $M_n = \max(X_1, X_2, ..., X_n)$. Hence, it is straightforward that M_n has distribution $F^n(x) = \prod_{i=1}^{n} F(x_i)$. If there are sequences of real value constants $a_n > 0$ and b_n, such that the normalised maxima of $(M_n - b_n)/a_n$ converges to some non-degenerate distribution function H. That is:

$$\mathbb{P}\{\frac{M_n - b_n}{a_n} \leq x\} = F^n(a_n x + b_n) \to H(x), \qquad \text{as } n \to \infty.$$

Then the random variable X, with distribution function F, belongs to the maximum domain of attraction of H, denoted as $F \in \text{MDA}(H)$.

Fisher and Tippett (1928) and Gnedenko (1948) introduce the following class of distributions, which are referred as the standard extreme value distributions, thus H belongs to one of the following three distributions:

$$\begin{aligned}
&\text{Gumbel (type I):} && F(x) = e^{-e^{-x}}, x \in \mathbb{R}, \\
&\text{Fréchet (type II):} && F(x;\alpha) = \begin{cases} e^{-x^{-\alpha}}, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad \alpha > 0, \\
&\text{Weibull (type III):} && F(x;\alpha) = \begin{cases} e^{-(-x)^{\alpha}}, & x \leq 0 \\ 1, & x > 0 \end{cases} \quad \alpha > 0.
\end{aligned}$$

Jenkinson (1955) suggested a one parameter representation for the above three distributions, which is called the generalised extreme value distribution that has *c.d.f.*: $H(x;\xi)$ with parameter ξ:

$$H(x;\xi) = \begin{cases} \exp\left(-(1+\xi x)^{1/\xi}\right) & \text{if } \xi \neq 0, \\ \exp\left(-e^{-x}\right) & \text{if } \xi = 0, \end{cases}$$

where x is such as $1 + \xi x > 0$ and ξ is a shape parameter. The generalised extreme value distribution is obtained by setting $\xi = 0$ for the Gumbel distribution, $\xi = \alpha^{-1}$

for the Fréchet distribution, and $\xi = -\alpha^{-1}$ for the Weibull distribution. Furthermore, the $\xi = \alpha^{-1} > 0$ (Fréchet) case corresponds to heavy-tailed distributions, such as Cauchy, Student t and Pareto distributions; the $\xi = 0$ case corresponds to so-called thin-tailed distributions, of which the tail decays exponentially, such as normal distribution; the $\xi = -\alpha^{-1} < 0$ (Weibull) case corresponds to distributions that have "no tail", that is, have finite right points. Refer to the book of De Haan and Ferreira (2006) for more details about the block of maximum approach and related extreme value distributions.

1.2.2 The generalised Pareto distribution

In the extreme value theory, the generalised Pareto distribution is used to model exceedances over certain thresholds, or in other words, data on the tail of the underlying distribution. This method of exceedances is often referred to as the Peak Over Threshold method. Assuming a random variable Y with distribution function $F_Y(y)$, the definition of exceedances is the values of Y above threshold u, namely $X = Y - u$. In the EVT, instead of assessing $F_Y(y)$ directly, we are more interested in estimating the distribution of the exceedances

$$F_u(x) = \mathbb{P}\{Y - u \leq x | Y \geq u\}, \qquad 0 \leq x \leq y_{\mathrm{F}} - u,$$

where x is the excess and $y_{\mathrm{F}} \leq \infty$ denotes the right end points of F_Y. It follows that

$$F_u(x) = \frac{F(u+x) - F(u)}{1 - F(u)} = \frac{F_Y(y) - F(u)}{1 - F(u)}.$$

Balkema and De Haan (1974) along with Pickands (1975) showed that, when u is large, for a large class of distribution functions, the conditional excess distribution function F_u is well approximated by the generalised Pareto distribution,

$$F_u(x) \approx F_{\mathrm{GPD}}(x; \sigma, \xi), \text{ as } u \to \infty,$$

where $F_{\mathrm{GPD}}(x; \sigma, \xi)$ is the cumulative distribution function

$$F_{\mathrm{GPD}}(x; \sigma, \xi) = \begin{cases} 1 - (1 + \frac{\xi}{\sigma} x)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - e^{-x/\sigma} & \text{if } \xi = 0, \end{cases}$$

for $x \in [0, y_{\mathrm{F}} - u)$ if $\xi \geq 0$ and $x \in [0, -\frac{\sigma}{\xi}]$ if $\xi < 0$, where ξ is the shape parameter and σ is the scale parameter. Note that in some literature the shape parameter ξ has an opposite sign and denoted as $\kappa = -\xi$. If $\xi > 0$ the GPD is heavy-tailed on

the right side. If $\xi = 0$ then the distribution reduces to the exponential distribution and the right tail decays exponentially. When $-1 \leq \xi \leq -0.5$, the distribution has finite right end-points, which is sometimes referred to as 'short-tailed'. Furthermore, when $\xi = -1$, the GPD actually becomes a uniform $\mathrm{U}(0, \sigma)$ distribution. The kth central moment of the GPD exists only if $\xi < 1/k$. For example, when it has a shape parameter $\xi \geq 1/2$, $\mathrm{var}(Y) = +\infty$ and the second central moment no longer exists. For more details about properties of the GPD and its parameters, see Hosking and Wallis (1987).

Moment-based methods such as method of moments and probability weighted moments are widely used for estimating GPD. Both methods are adopting the fact that the empirical moments from samples should be in according with theoretical moments in some way. The MOM estimators for the GPD parameters are given by

$$\hat{\xi}_{\texttt{MOM}} = (\bar{X}^2/s^2 - 1)/2$$
$$\hat{\sigma}_{\texttt{MOM}} = \bar{X}(\bar{X}^2/s^2 + 1)/2,$$

where $\bar{X}$ and s^2 are sample's mean and variance, respectively.

The PWM estimators for the GPD are

$$\hat{\xi}_{\texttt{PWM}} = \bar{X}^2/(\bar{X} - 2b) - 2$$
$$\hat{\sigma}_{\texttt{PWM}} = 2\bar{X}t/(\bar{X} - 2b),$$

where

$$b = n^{-1} \sum_{i=1}^{n} \frac{n-i}{n-1} X_{i:n}.$$

Castillo and Hadi (1997) proposed the so-called elemental percentile method. Let $X_{i:n}$ and $X_{j:n}$ $(i \neq j)$ be two distinct order statistics from sample of the GPD and re-parameterising $\alpha = \xi/\sigma$, then the EPM estimator $\hat{\alpha}_{\texttt{EPM}}$ is obtained by solving the following equation:

$$c_i \log(1 + \alpha X_{i:n}) = c_j \log(1 + \alpha X_{j:n})$$

where $c_i = \log(1 - b_{i:n}) < 0$ with $b_{i:n} = i/(n+1)$. Then, the shape parameter is estimated as

$$\hat{\xi}_{\texttt{EPM}} = -\log(1 + \hat{\alpha}_{\texttt{EPM}} X_{i:n})/c_i,$$

and consequently $\hat{\sigma} = \hat{\xi}/\hat{\alpha}$. Note that values of these estimators vary when different pairs of $X_{i:n}$ and $X_{j:n}$ are chosen. Castillo & Hadi suggest that obtain initial estimations for $\hat{\sigma}(i,j)$ and $\hat{\xi}(i,j)$ and use medians from each family $\{\hat{\sigma}(i,j)\}_{i \neq j}$ and $\{\hat{\xi}(i,j)\}_{i \neq j}$ as the final estimators. When the sample size is large, computing each

and every pair could be too time-consuming, so certain re-sampling scheme is also recommended.

Zhang (2007) proposed an alternative method for fitting the GPD, the likelihood moment estimation. The LME contains an auxiliary parameter h, and the estimator of $\alpha = \xi/\sigma$ is obtained by solving the following equation

$$n^{-1}\sum_{i=1}^{n}(1+\alpha X_{i:n})^b - (1-h)^{-1} = 0, \quad \alpha > -X_{n:n}^{-1},$$

where $b = hn/\sum_{i=1}^{n}\log(1+\alpha X_{i:n})$. It follows that the shape parameter ξ is estimated by

$$\hat{\xi} = \frac{1}{n}\sum_{i=1}^{n}\log(1+\hat{\alpha}X_i).$$

The shape estimator is $\hat{\sigma} = \hat{\xi}/\hat{\alpha}$. $h = -1/2$ is recommended as default value.

Zhang and Stephens (2009) provided another estimator (referred as ZS) for the parameters of the GPD which uses a procedure similar to Bayesian method to estimate α

$$\hat{\alpha}_{\text{ZS}} = \sum_{j=1}^{m} a_j w(a_j),$$

where

$$a_j = 1/X_{n:n} + \left(1 - \sqrt{\frac{m}{j-0.5}}\right)/(3X_{\frac{n+2}{4}:n}),$$
$$w(a_j) = -\frac{1}{\sum_{i=1}^{m}\exp^{g(a_i)-l(a_j)}},$$

with $g(a) = n(\log(a/\xi) - \xi - 1)$. After obtaining $\hat{\alpha}_{\text{ZS}}$, estimators $\hat{\xi}_{\text{ZS}}$ and $\hat{\sigma}_{\text{ZS}}$ are obtained same as in the LME method.

In Chapter 3, methods introduced above are discussed and reviewed. Also, in Chapter 3 we focuse on point and interval estimation problems of the GPD, using an order-statistic-based inference.

1.3 Quantile regression and order-statistic-based financial risk measures

Assume random variable X follows a c.d.f. F, then for $0 \leq p \leq 1$, the inverse of the c.d.f., $F^{-1}(p)$, is also called the quantile function of X, and it is defined as:

$$F^{-1}(p) = \inf\{x : F(x) \leq p\}.$$

We denote the p-th quantile as $Q(p)$, which is the unique solution to $F^{-1}(p) = x$. On the other hand, quantile can be seen as a special type of order statistics, for example, the p-th $(0 < p < 1)$ quantile of a sample $X_1, ...X_n$ can be estimated as the $\lfloor np \rceil$-th order statistic from $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$, where $\lfloor \cdot \rceil$ indicates rounding up to the nearest integer.

The p-th population quantile can also be obtained via the following optimisation problem:

$$\min_{Q} \mathbb{E}\left[\rho_p(y_t - Q(p))\right],$$

where $\rho_p(u) = u(p - I[u \leq 0])$ ($I[\cdot]$ indicates the indicator function) is the so-called check function.

Koenker and Bassett (1978) (see also Koenker, 2005) introduce the quantile regression, by introducing the linear relationship

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t, \qquad t = 1, ..., T,$$

where y_t are the dependent variables, $\mathbf{x}_t$ are the explanatory variables, $\boldsymbol{\beta}$ are unknown parameters, and ε_t is an error term which usually has no distributional assumption other than that the pth quantile of ε_t is zero and ε_t has finite variance.

The p-th $(0 < p < 1)$ conditional quantile of y_t given $\mathbf{x}_t$ is then in the expression as:

$$Q_{y_t|\mathbf{x}_t}(p) = \mathbf{x}_t'\boldsymbol{\beta}(p),$$

where $\boldsymbol{\beta}(p)$ is the parameter vector depends on p.

Then the regression parameters $\boldsymbol{\beta}(p)$ for the p-th quantile can be defined as the solution to the following minimisation problem:

$$\min_{\boldsymbol{\beta}} \mathbb{E}\left[\rho_p(y_t - \mathbf{x}_t'\boldsymbol{\beta})\right],$$

where $\rho_p(u) = u(p - I[u \leq 0])$ is the so-called check function, and $I[\cdot]$ is an indicator

function. The p-th regression quantile estimator $\hat{\boldsymbol{\beta}}(p)$ is obtained by solving the sample analog minimisation problem

$$\min_{\boldsymbol{\beta}} S(\boldsymbol{\beta}) \equiv \frac{1}{T}\sum_{t=1}^{T} \rho_p(y_t - \mathbf{x}_t'\boldsymbol{\beta}).$$

Quantile regression has become a rewarding methodology for regression analysis in many areas. In this thesis we mainly focus on inference and applications of quantile regression in the field of financial risk evaluating and measurement. In the proceeding subsection brief overviews of popular risk measures that are based on quantile or order statistics are given.

1.3.1 Volatility

By definition, volatility of a financial asset is the variance of its return distribution, denoted as $\sigma_t^2 = \text{var}(r_t)$, where $r_t = ln(S_t/S_{t-1})$ is the log return at time t. When assuming the average return is constant for a short period m and there is no auto-correlation between successive price 'shocks', the Realised Volatility is possibly the best estimation of variance, defined as

$$RV_t = \frac{1}{m-1}\sum_{t=1}^{m} \epsilon_t^2,$$

where $\epsilon_t = r_t - \bar{r}$ and $\bar{r}$ is the mean return.

Popular estimation methods for volatility include (but not limited to), for example, the ARCH/GARCH family of models, implied volatility models, *etc.*. There also exists group of order-statistic-based methods that based on the range or the *intra-quantile* range (interval between two symmetric extreme quantiles) for volatility estimation and forecasting.

The most important and most discussed range-based volatility estimator is developed by Parkinson (1980):

$$V_t = \frac{1}{4n\log 2}\sum_{t=1}^{n} \left(\log H_t - \log L_t\right)^2,$$

where H_t and L_t denote the highest and the lowest values in the t-th time period, respectively. Numerous researches have been carried out since Parkinson has shown that range-based volatility estimators are better than log or absolute return volatility in not only the ideal world that exists on paper or from simulated data, but also in

many realistic circumstances (see, for examples, Garman and Klass, 1980, Rogers and Satchell, 1991, Yang and Zhang, 2000, and Alizadeh et al., 2002, among others).

Pearson and Tukey (1965) proposed a measure of the standard deviation σ of a distribution, which is expressed as a proportion to the difference of the two symmetric extreme quantiles, or in other words, the inter-quantile range, of that distribution:

$$\text{Standard Deviation} = \frac{Q(1-p) - Q(p)}{C(p)}$$

where $p \in (0, 1)$ and $Q(p)$ is the p-th population quantile, and the value of denominator $C(p)$ depends on p. They found that for $p = 0.01, 0.025, 0.05$ the corresponding values for $C(\theta)$ are given by 4.65, 3.92 and 3.25 respectively. Taylor (2005) greatly improved Pearson's original model by replacing the interval between quantiles with the interval between value at risk measures. More details on this aspect are given in Chapter 4.6.

1.3.2 Value at risk

The $100(1-p)\%$ value at risk is defined as the threshold value of loss at probability level p, which is essentially the p-th quantile of the return distribution. In most industry application the value of p is chosen to be either 1% or 5%, which is required by the Basel II Accord published in 2004 by the Basle Committee on Banking Supervision.

One of the most significant improvements made for VaR is possibly the conditional autoregressive VaR (CAViaR) models introduced by Engle and Manganelli (2004). Different structures of the CAViaR models bring more versatile characteristics to the VaR forecast, such as capturing the asymmetric 'leverage effect' of impact of negative and positive news on return process, which is commonly accepted in financial area.

1.3.3 Expected shortfall

The $100p\%$ expected shortfall is defined as the conditional tail expectation that the value of r_t exceeds $\text{VaR}(p)$

$$\text{ES}_t(p) = \mathbb{E}[r_t | r_t < \text{VaR}_t(p)].$$

Amongst recent researches (e.g., Acerbi and Tasche (2002a), Acerbi and Tasche (2002b) and Tasche (2002)), the ES is considered as a better risk measure than

the VaR, as it is a coherent risk measure, while VaR is not, and it contains more information from extreme cases that beyond VaR.

1.4 Thesis outline

The outline of the thesis is as follow. Chapter 2 introduces a new inference on accelerated life-testing model based on Weibull distribution under Type-II progressive censoring. Constant stress procedures based on parametric lifetime distributions and models are often used for accelerated life testing in product reliability experiments. Maximum Likelihood Estimation is the typically statistical inference method. This chapter presents a new inference of parameters on Weibull constant stress accelerated life-testing model with progressively Type-II right censoring (including ordinary Type-II right censoring). We employ the two-parameter Weibull life distribution with a scale parameter that is a log-linear function of stress. The new inference for both life distribution parameters and the log-linear function coefficients is provided. Exact confidence intervals for these parameters are also explored. Theoretic properties, numerical examples of new estimators and their comparison with MLE via simulation tests and a real data example are given.

Chapter 3 extends the inference, with some modification, to the estimation problems of generalised Pareto distribution under small samples. The GPD is often used in the extreme value theory framework, and the sample size is often small, which increases the difficulty of accurate estimation. The estimation becomes even more difficult for extremely high quantiles of the GPD when it is heavy-tailed. The performance of classical methods, such as MLE and moments-based methods, usually suffers under these conditions. This chapter presents a method of estimation for the GPD that does not depend on asymptotic theory or bootstrapping. Exact confidence intervals and generalised confidence intervals as introduced by Weerahandi (1993, 2004) for parameters and quantiles of the GPD are also explored. We compare the proposed method with classical methods and several new methods in recent literature under this topic, via extensive simulation experiments. A real data example is also given.

In Chapter 4.6 we study distribution-free approaches for volatility forecasts, which are established based on the variance estimation using a pair of symmetric extreme quantiles model, as introduced by Pearson and Tukey (1965). Recent improvements of this approach include Taylor (2005), in which CAViaR models are used for quantile forecasting, and a more general regression form predictor is considered. We propose two new approaches with the idea of using not just one single pair of quantiles, but

'tail expectations' such as expected shortfall and expectile. Thereafter, in empirical studies we compare different models, including classic parametric models, for short period volatility forecasts of stock indices. Furthermore, we also consider multiple regression models that combine two predictors, to see if it improves forecast performance.

In Chapter 5 we focus on the joint modelling problem of heteroskedastic quantile regression with the assumption that the error term has a specific distribution. Quantile regression is known to be more robust against heteroskedacity, and its asymptotic properties have been studied under certain conditions. Standard quantile regression inference often assume a linear relationship between the variability and the independent variables, and the regression parameters for the location component and for the scale component are usually estimated separately. We introduce a new inference procedure that can obtain quantile regression parameters simultaneously by assuming that the error term follows an asymmetric Laplace distribution. Furthermore, we explore the idea of a statistical dispersion measure for the quantile, which can be seen as the analog of standard deviation for the ordinary least square regression.

And finally, in Chapter 6 we summarise the main results of the research and propose recommendations for possible future research directions.

Note that each chapter is written to be read independently with self contained notations and definitions, and where reference is made to another chapter the connection is clearly explained.

Chapter 2

Inference on Weibull constant stress accelerated life-testing model under progressive Type-II censoring

Constant stress procedure based on parametric lifetime distribution and model is often used for accelerated life testing in product reliability experiment. Maximum likelihood estimation (MLE) is the typically statistical inference method. This chapter presents a new inference method for Weibull constant stress accelerated life test with progressively Type-II right censoring (including ordinary Type-II right censoring). An two-parameter Weibull life distribution with a scale parameter that is a log-linear function of stress is used. New estimates for both life distribution parameters and the log-linear function coefficients are provided. Exact confidence intervals for these parameters are also explored. Numerical comparison of new estimates with MLE shows that the proposed new inference method is very promising,

2.1 Introduction

In many industrial fields it is requested for lots of products to operate for a long period of time. Accompanied with that, it is very important to give reliability in relation to the lifetime of products. In such cases, however, life testing under a normal stress can lead to a lengthy procedure with expensive cost. As a means to cope with these problems, the study of accelerated life test has been developed. The ALT can

quickly yield information on the lifetime distribution of products by inducing early failure with stronger stress than normal. The results obtained at the accelerated conditions are analyzed in terms of a model to relate life length to stress; they are extrapolated to the design stress to estimate the life distribution. One important way in ALT is constant stress ALT . The problem of modeling data from the CSALT and making inferences from such data have been studied by many authors. For example, Vander Wiel and Meeker (1990) studied accuracy of approx confidence bounds for Weibull CSALT model. Yang (1994) considered optimum 4-level CSALT plans under location-scale family of distributions. Watkins (1994) discussed likelihood method for fitting Weibull CSALT models. Wang and Kececioglu (2000) further studied this issue and gave an efficient algorithm to fit the Weibull CSALT model. Tang et al. (1999) discussed an optimum CSALT plan for two-parameter exponential distribution. René Van Dorp and Mazzuchi (2005) discussed Bayes inference for ALT. Leon et al. (2007) discussed Bayesian modeling of CSALT with random effects. Watkins and John (2008) discussed the maximum likelihood estimates for CSALT with terminated by type-II censoring at one of the stress levels. Pascual (2008) studied the planning of CSALT in the presence of competing risks under Weibull distributions. Ma and Meeker (2010) discussed strategy for planning CSALT with small sample sizes. Liu and Tang (2010) considered CSALT for repairable systems with multiple independent risks, and derived an accelerated life test plans. Tang and Liu (2010) proposed a sequential CSALT, and discussed its inference procedure and test plan. Yu and Chang (2012) applied Bayesian model to average quantile estimation for CSALT. Liu (2012) discussed the model and plan for CSALT with dependent failure modes. Nelson (2009) provided some excellent information on past and current developments in the area.

Progressive censoring is a generalised form of censoring which includes the conventional right censoring as a special case. Compared to the conventional censoring, however, it provides higher flexibility to the experimenter in the design stage by allowing the removal of test units at non-terminal time points and thus, it proves to be highly efficient and effective in utilizing the available resources. Another advantage of progressive censoring is that the degeneration information of the test units is obtained from those removed units. For these reasons, we consider a more general censoring scheme called progressive type-II censoring. Progressive type-II censoring is a method which enables an efficient exploitation of the available resources by continual removal of a prespecified number of surviving test units at each failure time. Balakrishnan et al. (2007) derived point and interval estimation for a simple step-stress model with type-I censoring. Gouno et al. (2004) and Balakrishnan and Han (2009) discussed the optimal step-stress ALT plans under progressive Type-I censoring. Wang and Yu

(2009) discussed the optimal step-stress ALT plans under progressive Type-II censoring. Wang (2010) derived interval estimation for exponential progressive Type-II censored step-stress ALT. A book dedicated completely to progressive censoring was published by Balakrishnan and Aggarwala (2000). Moreover, Balakrishnan (2007) gave an excellent and extensive review for the progressive censoring methodology.

In this chapter, we choose the two-parameter Weibull distribution for modelling CSALT data. The two-parameter Weibull distribution is one of the most popular distributions for modelling lifetime data. It has shape and scale parameters that gives it the flexibility for fitting various reliability curves. The hazard function derived from Weibull can describe increasing, decreasing or constant failure rate. Also, Weibull analysis provides reasonable accurate results for small samples. However, one of the disadvantage of Weibull is that the asymptotic convergence to normality for this distribution of the MLE is slow (for example, see Gupta and Kundu, 2001). More discussion is given in following sections of this chapter.

Under a combination of CSALT and progressive Type-II censoring, the sample size is typical not big, even small, so that large-sample based inference methods such as MLE-based asymptotic unbiased estimate and asymptotic normal confidence intervals (CI) may not be suitable, even misleading. In this chapter, we consider CSALT with progressive Type-II censoring and provides new inference for parameter estimation and CIs. The advantages of our method are (1) small-sample based confidence intervals are promising whereas MLE-based confidence intervals are based on asymptotic normality with large sample, (2) progressive censoring scheme which generalizes the censoring schemes considered for accelerated life test in literature, (3) the log-linear model for shape parameter includes the exponential life distribution as a special case. The Weibull CSALT model considered is under the following assumptions:

A 2.1 For any stress level x_i, the lifetime distribution of a test unit is Weibull with c.d.f.

$$F_i(t) = 1 - \exp\left(-(t/\theta_i)^{\beta}\right), \quad t > 0, \tag{2.1}$$

where $\beta > 0$ is the shape parameter, $\theta_i > 0$ is the scale parameter.

Furthermore, the density function of Weibull is

$$f_i(t) = \frac{\beta}{\theta_i}\left(\frac{t}{\theta_i}\right)^{\beta-1} \exp\left(-(t/\theta_i)^{\beta}\right), \qquad t > 0.$$

A 2.2 The stress-life relationship is given by

$$\log(\theta_i) = \alpha_0 + \alpha_1 x_i. \tag{2.2}$$

where α_0 and α_1 are unknown parameters.

The log-linear model above for the scale parameter includes the exponential life distribution as a special case which was widely studied in the literature.

Let x_0 be the designed stress level, and let $x_1 < x_2 < ... < x_k$ be the k accelerated stress levels. Suppose that n_i test units are placed at stress level x_i. Prior to the experiment, a number r_i $(< n_i)$ is fixed and the progressive censoring scheme $R_i = (R_{i,1}, R_{i,2}, \ldots, R_{i,r_i})$ with $R_{i,j} \geq 0$ and $\sum_{j=1}^{r_i} R_{i,j} + r_i = n_i$ is specified. At the first failure time $T_{i,1}$, $R_{i,1}$ units are randomly removed from the remaining $n_i - 1$ surviving units. At the second failure time $T_{i,2}$, $R_{i,2}$ units are randomly removed from the remaining $n_i - 2 - R_{i,1}$. The test continues until the r_ith failure time T_{i,r_i}. At failure time T_{i,r_i}, all remaining units are removed. When $R_{i,j} = 0$, $i = 1, ..., k$, $j = 1, ..., r_i - 1$, then $R_{i,r_i} = n - r_i$ which corresponds to the conventional constant stress accelerated life-testing with Type-II censoring scheme.

We aim to compare new method to MLE, so Section 2 outlines the MLE under the constant-stress model and progressively censored scheme. Sections 3 details the new estimating method and estimation properties. Section 4 focuses on new interval estimation for unknown parameters and their functions such as reliability function. Section 5 illustrates the numerical performance of the new method and comparison with MLE. Section 6 concludes.

2.2 Maximum Likelihood Estimation

At each stress level x_i $(i = 1, ..., k)$, there are $r_i (r_i < n_i)$ observations for the lifetime variable $T_{i,j}$ $(j = 1, ..., r_i)$ and progressive censoring scheme $R_i = (R_{i,1}, R_{i,2}, ..., R_{i,r_i})$. In total let $t = \{t_{i,j} : i = 1, ..., k; j = 1, .., r_i\}$ be the observed values of lifetime $\{T_i\}_{i=1}^k$.

Therefore, based on the likelihood function

$$L(\beta, \alpha_0, \alpha_1 | t) = \prod_{i=1}^{k} \prod_{j=1}^{r_i} \frac{\beta}{\theta_i^\beta} t_{i,j}^{\beta-1} \exp\left(- \sum_{i=1}^{k} \sum_{j=1}^{r_i} (R_{i,j} + 1)(t_{i,j}/\theta_i)^\beta \right),$$

with $\log(\theta_i) = \alpha_0 + \alpha_1 x_i$, and $\sum_{j=1}^{r_i} R_{i,j} + r_i = n_i$, we have the log-likelihood function as

$$\ell(\beta, \alpha_0, \alpha_1) \quad \propto \quad \sum_{i=1}^{k} r_i \log(\beta) + (\beta - 1) \sum_{i=1}^{k} \sum_{j=1}^{r_i} \log(t_{i,j}) - \alpha_0 \beta \sum_{i=1}^{k} r_i - \alpha_1 \beta \sum_{i=1}^{k} r_i x_i$$

$$-\sum_{i=1}^{k}\sum_{j=1}^{r_i}(R_{i,j}+1)t_{i,j}^{\beta}\exp(-\alpha_0\beta-\alpha_1\beta x_i)$$

Hence the MLEs $\hat{\beta}_M, \hat{\alpha}_{0,M}, \hat{\alpha}_{1,M}$ of the parameters $\beta, \alpha_0, \alpha_1$ are the solutions of the following equations:

$$\begin{aligned}
&\frac{1}{\beta}+\frac{\sum_{i=1}^{k}\sum_{j=1}^{r_i}\log(t_{i,j})}{\sum_{i=1}^{k}r_i}-\\
&\qquad\frac{\sum_{i=1}^{k}\sum_{j=1}^{r_i}(R_{i,j}+1)t_{i,j}^{\beta}\log(t_{i,j})\exp(-\alpha_1\beta x_i)}{\sum_{i=1}^{k}\sum_{j=1}^{r_i}(R_{i,j}+1)t_{i,j}^{\beta}\exp(-\alpha_1\beta x_i)}=0,\\
&\sum_{i=1}^{k}r_i x_i\sum_{i=1}^{k}\sum_{j=1}^{r_i}(R_{i,j}+1)t_{i,j}^{\beta}\exp(-\alpha_1\beta x_i)-\\
&\qquad\sum_{i=1}^{k}r_i\sum_{i=1}^{k}\sum_{j=1}^{r_i}(R_{i,j}+1)x_i\,t_{i,j}^{\beta}\exp(-\alpha_1\beta x_i)=0,\\
&\alpha_0=\frac{1}{\beta}\log\left(\sum_{i=1}^{k}\sum_{j=1}^{r_i}(R_{i,j}+1)t_{i,j}^{\beta}\exp(-\alpha_1\beta x_i)\right)-\frac{1}{\beta}\log\left(\sum_{i=1}^{k}r_i\right).
\end{aligned}$$

Note that

$$\begin{aligned}
\frac{\partial^2\ell}{\partial\beta^2} &= -\frac{\sum_{i=1}^{k}r_i+\sum_{i=1}^{k}\sum_{j=1}^{r_i}(R_{i,j}+1)x_i\left(t_{i,j}/\theta_i\right)^{\beta}\left[\log\left(t_{i,j}/\theta_i\right)\right]^2}{\beta^2},\\
\frac{\partial^2\ell}{\partial\alpha_0^2} &= -\beta^2\sum_{i=1}^{k}\sum_{j=1}^{r_i}(R_{i,j}+1)\left(\frac{t_{i,j}}{\theta_i}\right)^{\beta},\\
\frac{\partial^2\ell}{\partial\alpha_1^2} &= -\beta^2\sum_{i=1}^{k}\sum_{j=1}^{r_i}(R_{i,j}+1)x_i^2\left(\frac{t_{i,j}}{\theta_i}\right)^{\beta},\\
\frac{\partial^2\ell}{\partial\beta\partial\alpha_0} &= -\sum_{i=1}^{k}r_i+\sum_{i=1}^{k}\sum_{j=1}^{r_i}(R_{i,j}+1)\left(\frac{t_{i,j}}{\theta_i}\right)^{\beta}+\\
&\qquad\sum_{i=1}^{k}\sum_{j=1}^{r_i}(R_{i,j}+1)\left(\frac{t_{i,j}}{\theta_i}\right)^{\beta}\log\left(\frac{t_{i,j}}{\theta_i}\right)^{\beta},\\
\frac{\partial^2\ell}{\partial\beta\partial\alpha_1} &= -\sum_{i=1}^{k}r_i x_i+\sum_{i=1}^{k}\sum_{j=1}^{r_i}(R_{i,j}+1)x_i\left(\frac{t_{i,j}}{\theta_i}\right)^{\beta}+\\
&\qquad\sum_{i=1}^{k}\sum_{j=1}^{r_i}(R_{i,j}+1)x_i\left(\frac{t_{i,j}}{\theta_i}\right)^{\beta}\log\left(\frac{t_{i,j}}{\theta_i}\right)^{\beta},
\end{aligned}$$

$$\frac{\partial^2 \ell}{\partial \alpha_0 \partial \alpha_1} = -\beta^2 \sum_{i=1}^{k} \sum_{j=1}^{r_i} (R_{i,j}+1) x_i \left(\frac{t_{i,j}}{\theta_i}\right)^{\beta}.$$

Numerical solutions of these estimators will be studied in Section 5. The Fisher-information matrix is often used to calculate the covariance matrices associated with MLE. Here the observed Fisher-information matrix for $(\beta, \alpha_0, \alpha_1)$ is given by

$$\mathcal{I}(\hat{\beta}_M, \hat{\alpha}_{0,M}, \hat{\alpha}_{1,M}) = \begin{pmatrix} -\frac{\partial^2 \ell}{\partial \beta^2} & -\frac{\partial^2 \ell}{\partial \beta \partial \alpha_0} & -\frac{\partial^2 \ell}{\partial \beta \partial \alpha_1} \\ -\frac{\partial^2 \ell}{\partial \beta \partial \alpha_0} & -\frac{\partial^2 \ell}{\partial \alpha_0^2} & -\frac{\partial^2 \ell}{\partial \alpha_0 \partial \alpha_1} \\ -\frac{\partial^2 \ell}{\partial \beta \partial \alpha_1} & -\frac{\partial^2 \ell}{\partial \alpha_0 \partial \alpha_1} & -\frac{\partial^2 \ell}{\partial \alpha_1^2} \end{pmatrix}_{(\hat{\beta}_M, \hat{\alpha}_{0,M}, \hat{\alpha}_{1,M})}.$$

2.3 New inference method

To make things clear in this section we first consider the case with the known shape parameter β and propose new estimators for parameters α_0, α_1 and $\theta_0 = \exp(\alpha_0 + \alpha_1 x_0)$, then extend it to the estimation of β together.

2.3.1 The known shape parameter case

When parameter β is known, let

$$S_i = \sum_{j=1}^{r_i} (R_{i,j}+1)\, T_{i,j}^{\beta},\ i = 1, 2, ..., k.$$

Then it is well known that $2S_i/\theta_i^{\beta}$ follows the χ^2 distribution with $2r_i$ degrees of freedom (proof can be found, *e.g.*, in Wang and Yu, 2009).

According to the property of log-Gamma distribution,

$$\begin{aligned} \mathbb{E}[\log(S_i) - \beta \log(\theta_i)] &= \psi(r_i), \\ \text{var}[\log(S_i) - \beta \log(\theta_i)] &= \psi'(r_i), \end{aligned}$$

where $\psi(x) = d \log(\Gamma(x))/dx$, $\psi'(x) = d^2 \log(\Gamma(x))/dx^2$.

Therefore we consider the following regression model:

$$\mathbb{E}(U_i) = \beta \log(\theta_i) = \alpha_0 \beta + \alpha_1 \beta x_i,\ \text{var}(U_i) = \psi'(r_i),$$

where $U_i = \log(S_i) - \psi(r_i)$.

According to Gauss-Markov theorem the unbiased estimators for (α_0, α_1) are respectively given by

$$\begin{aligned} \tilde{\alpha}_0 &= \frac{GH - IM}{\beta(EG - I^2)} \\ \tilde{\alpha}_1 &= \frac{EM - IH}{\beta(EG - I^2)}, \end{aligned} \tag{2.3}$$

where $\mathbb{E} = \sum_{i=1}^{k}[\psi'(r_i)]^{-1}$, $I = \sum_{i=1}^{k}[\psi'(r_i)]^{-1} x_i$, $G = \sum_{i=1}^{k}[\psi'(r_i)]^{-1} x_i^2$, $H = \sum_{i=1}^{k}[\psi'(r_i)]^{-1} U_i$, $M = \sum_{i=1}^{k}[\psi'(r_i)]^{-1} x_i U_i$.

Further, we have

$$\mathrm{var}(\tilde{\alpha}_0) = \frac{G}{\beta^2(EG - I^2)}, \mathrm{var}(\tilde{\alpha}_1) = \frac{E}{\beta^2(EG - I^2)}, \mathrm{cov}(\tilde{\alpha}_0\,, \tilde{\alpha}_1) = -\frac{I}{\beta^2(EG - I^2)}.$$

Therefore the scale parameter θ_0 at designed stress level x_0 could be estimated by $\tilde{\theta}_0 = \exp(\tilde{\alpha}_0 + \tilde{\alpha}_1 x_0)$.

Under assumptions A 2.1 and A 2.2, we obtain the following results for the estimation of θ_0, which are essentially the same as the Theorem 5(1) and the Theorem 6(1) from Wang and Yu (2009) for the multi-level of stress model.

Theorem 2.1 *$\tilde{\alpha}_0$ and $\tilde{\alpha}_1$ defined in equation (2.3) are the unbiased estimators of α_0 and α_1, and let $D_i = [G - (x_0 + x_i)I + x_0 x_i E]/[\beta\psi'(r_i)(EG - I^2)]$, then*

(1) if $r_i + D_i > 0$ $(i = 1, 2, ..., k)$, then the expectation of $\tilde{\theta}_0$ exists but $\tilde{\theta}_0$ is a biased estimator of θ_0. However, an unbiased estimator of θ_0 is thus given by

$$\tilde{\theta}_{0U} = \tilde{\theta}_0 \exp\left(\sum_{i=1}^{k} D_i \psi(r_i)\right) \prod_{i=1}^{k} \frac{\Gamma(r_i)}{\Gamma(r_i + D_i)}. \tag{2.4}$$

Furthermore, if $r_i + 2D_i > 0$ $(i = 1, 2, ..., k)$, then the variance of $\tilde{\theta}_{0U}$ exists and is given by

$$Var(\tilde{\theta}_{0U}) = \left(\prod_{i=1}^{k} \frac{\Gamma(r_i)\,\Gamma(r_i + 2D_i)}{\Gamma^2(r_i + D_i)} - 1\right) \theta_0^2.$$

(2) if $r_i + 2D_i > 0, (i = 1, 2, \ldots, k)$, then $\widetilde{\theta}_{0U}$ has a smaller mean squared error than that of $\tilde{\theta}_0$.

Proof of Theorem 2.1 can also be found in Wang and Yu (2009), as in proofs of Theorems 5 and 6 in their paper.

In summary, in contrary to MLE whose estimators are asymptotic unbiased with

asymptotic variances, we have obtained exact unbiased estimators of parameters $(\alpha_0, \alpha_1, \theta_0)$ with exact variances of estimators.

2.3.2 The unknown shape parameter case

Now we consider the case with unknown shape parameter β.

For each $i = 1, 2, ..., k$, $j = 1, 2, ..., r_i,$, let

$$S_{i,j} = \sum_{l=1}^{j}(R_{i,l}+1)T_{i,l}^{\beta} + [n_i - \sum_{l=1}^{j}(R_{i,l}+1)]T_{i,j}^{\beta},$$

Similar to the argument for S_i in Section 2.3.1, we have

$$W_i(\beta) = 2\sum_{j=1}^{r_i-1}\log\left(\frac{S_{i,r_i}}{S_{i,j}}\right) \sim \chi^2(2r_i - 2),\ i = 1, 2, ..., k$$

and $W_i(\beta)$ is a strictly monotone function of β.

Notice that $W_1(\beta), ..., W_k(\beta)$ are independent, because each $W_i(\beta)$ is calculated independently using samples under the ith stress level, $T_{i,\cdot}$ $(i = 1, ..., k)$, and there is no overlapping among the total k groups of samples. Thus we define

$$W(\beta) = 2\sum_{i=1}^{k}\sum_{j=1}^{r_i-1}\log\left(\frac{S_{i,r_i}}{S_{i,j}}\right) \sim \chi^2(2\sum_{i=1}^{k} r_i - 2k). \tag{2.5}$$

The mode of $\chi^2(2\sum_{i=1}^{k} r_i - 2k)$ distribution is $2\sum_{i=1}^{k} r_i - 2k - 2$, thus $W(\beta)/(2\sum_{i=1}^{k} r_i - 2k - 2)$ converges with probability one to 1. Then let $\tilde{\beta}$ be the estimator of of the shape parameter β from the solution of the following equation:

$$\sum_{i=1}^{k}\sum_{j=1}^{r_i-1}\log\left(\frac{S_{i,r_i}}{S_{i,j}}\right) = \sum_{i=1}^{k} r_i - k - 1. \tag{2.6}$$

Due to the strictly increasing function of β, the equation (2.6) has exactly one unique solution. Let $\hat{\beta}$ be the solution of the equation (2.6). Then plugging $\hat{\beta}$ in (2.3) and (2.4), we obtain the following estimators $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\theta}_0)$ of $(\alpha_0, \alpha_1, \theta_0)$:

$$\hat{\alpha}_0 = \frac{G\hat{H} - I\hat{M}}{\hat{\beta}(EG - I^2)}, \tag{2.7}$$

$$\hat{\alpha}_1 = \frac{E\hat{M} - I\hat{H}}{\hat{\beta}(EG - I^2)}, \tag{2.8}$$

$$\hat{\theta}_0 = \exp\left(\hat{\alpha}_0 + \hat{\alpha}_1 x_0 + \sum_{i=1}^{k} \hat{D}_i \psi(r_i)\right) \prod_{i=1}^{k} \frac{\Gamma(r_i)}{\Gamma(r_i + \hat{D}_i)}, \tag{2.9}$$

where $\hat{U}_i = \log\left(\sum_{j=1}^{r_i}(R_{i,j}+1)\, T_{i,j}^{\hat{\beta}}\right) - \psi(r_i)$, $\hat{H} = \sum_{i=1}^{k}[\psi'(r_i)]^{-1}\hat{U}_i$, $\hat{M} = \sum_{i=1}^{k}[\psi'(r_i)]^{-1} x_i \hat{U}_i$, $\hat{D}_i = [G - (x_0 + x_i)I + x_0 x_i E]/[\hat{\beta}\psi'(r_i)(EG - I^2)]$.

The estimators $(\hat{\beta}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\theta}_0)$ of $(\alpha_0, \alpha_1, \theta_0)$ given by (6)–(9) are alternative estimators of the parameters $(\beta, \alpha_0, \alpha_1, \theta_0)$. We shall study the finite sample properties of the proposed estimators in Section 5.

2.4 Interval estimation of unknown estimators

In this section, we will first obtain an exact confidence interval for the shape parameter, then derive the generalized confidence intervals for other parameters and some important quantities of the Weibull distribution at designed stress level x_0, such as its mean, quantiles and reliability function.

2.4.1 Exact confidence interval for the shape parameter

Consider the pivotal quantity $W(\beta)$. From (2.5), it is of great importance to note that $W(\beta)$ is a function of β only and *does not depend on other parameters*. Hence, we obtain an exact confidence interval for the shape parameter β as follows.

Theorem 2.2 *Suppose* $(T_{i,1}, ..., T_{i,r_i}), i = 1, 2, ..., k$ *are progressively Type II censored samples from the Weibull constant stress accelerated life testing with the progressive censoring scheme* $(R_{i,1}, R_{i,2}, \ldots, R_{i,r_i}), i = 1, 2, ..., k$. *Then, for any* $0 < \tau < 1$,

$$\left[W^{-1}\left\{\chi^2_{1-\tau/2}\left(2\sum_{i=1}^{k} r_i - 2k\right)\right\},\ W^{-1}\left\{\chi^2_{\tau/2}\left(2\sum_{i=1}^{k} r_i - 2k\right)\right\}\right]$$

is a $1 - \tau\%$ *confidence interval for the shape parameter* β, *where* $\chi^2_\tau(v)$ *is the upper* τ *percentile of the* χ^2 *distribution with* v *degrees of freedom and, for* $w > 0$, $W^{-1}(w)$ *is the solution in* β *of the equation* $W(\beta) = w$.

2.4.2 Generalized confidence intervals for other parameters

We now derive generalized confidence intervals for other parameters and some important quantities of the Weibull distribution at designed stress level x_0.

Let

$$V_1 = \frac{\sum_{i=1}^{k}[\psi'(r_i)]^{-1}(G - x_i I)\,\log(2S_i)}{EG - I^2} - \alpha_0\beta, \tag{2.10}$$

$$V_2 = \frac{\sum_{i=1}^{k}[\psi'(r_i)]^{-1}(x_i E - I)\,\log(2S_i)}{EG - I^2} - \alpha_1\beta. \tag{2.11}$$

Then

$$V_1 = \frac{\sum_{i=1}^{k}[\psi'(r_i)]^{-1}(G - x_i I)\,\log(T_i)}{EG - I^2}, \tag{2.12}$$

$$V_2 = \frac{\sum_{i=1}^{k}[\psi'(r_i)]^{-1}(x_i E - I)\,\log(T_i)}{EG - I^2}, \tag{2.13}$$

where $T_i = 2S_i/\theta_i^{\beta} \sim \chi^2(2r_i)$. It is obvious from (2.12) and (2.13) that the distributions of V_1 and V_2 do not depend on any unknown parameters. Thus V_1 and V_2 are pivotal quantities.

Note that $W(\beta)$ is a strictly increasing function of β, then the equation $W(\beta) = W$ has the unique solution $g(W, T)$, where $W \sim \chi^2(2\sum_{i=1}^{k} r_i - 2k)$. In addition, from (2.10) and (2.11), we have

$$\alpha_0 = \frac{\sum_{i=1}^{k}[\psi'(r_i)]^{-1}(G - x_i I)\,\log(2S_i)}{\beta(EG - I^2)} - \frac{V_1}{\beta}, \tag{2.14}$$

$$\alpha_1 = \frac{\sum_{i=1}^{k}[\psi'(r_i)]^{-1}(x_i E - I)\,\log(2S_i)}{\beta(EG - I^2)} - \frac{V_2}{\beta}. \tag{2.15}$$

According to the substitution method given by Weerahandi (1993, 2004), we substitute $g(W, T)$ for β in the expression for α_0, α_1 in (2.14) and (2.15) and obtain the following generalized pivotal quantities for the parameters α_0, α_1:

$$Y_0 = \frac{\sum_{i=1}^{k}[\psi'(r_i)]^{-1}(G - x_i I)\,\log(2s_i)}{g(W,t)(EG - I^2)} - \frac{V_1}{g(W,t)}, \tag{2.16}$$

$$Y_1 = \frac{\sum_{i=1}^{k}[\psi'(r_i)]^{-1}(x_i E - I)\,\log(2s_i)}{g(W,t)(EG - I^2)} - \frac{V_2}{g(W,t)}, \tag{2.17}$$

where $s_i = \sum_{j=1}^{r_i}(R_{i,j} + 1)\,t_{i,j}^{g(W,t)}$.

Notice that Y_0 and Y_1 reduce to α_0 and α_1 when $T = t$ respectively, and the distributions of Y_0 and Y_1 are free of any unknown parameters, thus Y_0 and Y_1 is indeed a generalized pivotal quantities. If $Y_{0,\tau}$ and $Y_{1,\tau}$ denote the upper τ percentiles of Y_0 and Y_1, then $[Y_{0,1-\tau/2}, Y_{0,\tau/2}]$ and $[Y_{1,1-\tau/2}, Y_{1,\tau/2}]$ are the $1-\tau$ generalized confidence intervals for α_0 and α_1 respectively.

The percentiles of Y_0 and Y_1 can be obtained from (2.16) and (2.17) using the following Monte Carlo simulation algorithm.

Step 1: For a given data set (n, m, t, R), generate $W \sim \chi^2\left(2\sum_{i=1}^{k} r_i - 2k\right)$, $T_1 \sim \chi^2(2r_1)$, ..., $T_k \sim \chi^2(2r_k)$ independently. Using these values, compute $g(W, t)$, V_1 and V_2 from equations $W(\beta) = W$, (2.12) and (2.13).

Step 2: In terms of (2.16) and (2.17), compute the values of Y_0 and Y_1.

Step 3: Repeat the steps 1 and 2 a large number of times, say, $m_1(\geq 10,000)$ times. The m_1 values of Y_0 and Y_1 can be obtained respectively.

Step 4: Arrange all Y_0 and Y_1 values in ascending order respectively: $Y_{0,1} < Y_{0,2} < ... < Y_{0,m_1}$ and $Y_{1,1} < Y_{1,2} < ... < Y_{1,m_1}$. Then the τ percentile of Y_0 and Y_1 are estimated by $Y_{0,\tau m_1}$ and $Y_{1,\tau m_1}$ respectively.

Now note that the mean, pth quantile $(0 < p < 1)$ and reliability function of the Weibull distribution at designed stress level x_0 are given by $\mu = \theta_0\Gamma(1 + 1/\beta)$, $t_p = \theta_0[-\log(1-p)]^{1/\beta}$ and $R(t_0) = \exp[-(t_0/\theta_0)^{\beta}]$ respectively. Along the same lines as the derivation of Y_0 and Y_1 for the parameters α_0 and α_1, we obtain the following generalized pivotal quantities Y_2, Y_3 and Y_4 for μ, x_p and $R(x_0)$ respectively:

$$Y_2 = e^{Y_0 + Y_1 x_0}\Gamma\left(1 + \frac{1}{g(W,t)}\right), \tag{2.18}$$

$$Y_3 = e^{Y_0 + Y_1 x_0}\left[-\log(1-p)\right]^{\frac{1}{g(W,t)}}, \tag{2.19}$$

$$Y_4 = \exp\left[-(t_0 e^{-Y_0 - Y_1 x_0})^{g(W,t)}\right]. \tag{2.20}$$

Let $Y_{2,\tau}, Y_{3,\tau}, Y_{4,\tau}$ denote the upper τ percentiles of Y_2, Y_3, Y_4 respectively. Then $Y_{2,\tau}, Y_{3,\tau}, Y_{4,\tau}$ are the $1-\tau$ upper confidence limits for μ, t_p and $R(t_0)$, respectively. Just as in the cases of Y_0 and Y_1, the percentiles of Y_2, Y_3, Y_4 can be obtained by Monte Carlo simulations.

Although for given $(T_{i,1}, ..., T_{i,r_i}), i = 1, 2, ..., k$, the distributions of Y_0, Y_1, Y_2, Y_3, Y_4 do not depend on any unknown parameters, the coverage probabilities of their generalized confidence intervals may depend on nuisance parameters. We study the performance of coverage probabilities of these confidence intervals by simulation. Such simulation results are reported in Section 5.

Remark: It is easy to prove that $\hat{\beta}(\hat{\alpha}_0 - \alpha_0)$, $\hat{\beta}(\hat{\alpha}_1 - \alpha_1)$ and $\hat{\beta}(\hat{\alpha}_0 + \hat{\alpha}_1 x_0 - \log(t_p))$ are pivotal quantities for the parameters α_0, α_1 and t_p respectively. Hence, confidence intervals for these parameters can alternatively be obtained based on these pivotal quantities. The percentiles of these pivotal quantities can also be obtained by Monte

Carlo simulations. It is worth noting that there is no pivotal quantity for μ.

2.5 Simulation Study

To evaluate and compare the performance of the MLE method and proposed estimators with the proposed alternative method (NEW1), we performed simulation tests with data generated via various scenarios. We consider different number of stress levels ($k = 2, 3, 4$ for simulation design scenario 1-3, 4-6 and 7-9, respectively), combined with different censoring schemes (for example, progressive and non-progressive censoring). Details of the simulation design scenarios are summarised in Table 2.1. For each scenario, 10,000 replicates of random samples were generated from the Weilbull distribution as specified in (1), with three different parameter settings: (1) $(\beta, \alpha_0, \alpha_1) = (0.5, 5, -1)$, (2) $(\beta, \alpha_0, \alpha_1) = (1, 5, -1)$, (3) $(\beta, \alpha_0, \alpha_1) = (3, 5, -1)$, respectively.

Then Tables 2.2 to 2.4 compare the percentage-biases and -MSE of parameter estimators form NEW1 method with those from the MLE method under different simulation scenarios, with respect to three different parameter settings. The percentage-biases and the percentage-MSE are defined as follow:

$$\text{percentage} - \text{biases} = \frac{100\%}{n} \sum_{i=1}^{n} \frac{\hat{\xi}_i - \xi}{\xi},$$
$$\text{percentage} - \text{MSE} = \frac{100\%}{n} \sum_{i=1}^{n} \frac{(\hat{\xi}_i - \xi)^2}{\xi^2},$$

where $\hat{\xi}$ denotes the estimated value and ξ denotes the true value. Findings from Table 2.2, 2.3 and 2.4 are listed as follows:

- Generally speaking, the relative-bias and -MSE of the NEW1 method for β is significantly smaller than that of the MLE method. The β estimator of the NEW1 is almost unbiased and very accurate. MLE's β estimator is slightly over-estimate, as bias are all positive.

- For α_0, both methods have small percentage-bias and -MSE. The performance difference between MLE and NEW1, in terms of bias and MSE, is very close. The MSE of $\hat{\alpha}_0$ decreases, as the true value of β increases, namely, the right tail of the Weibull distribution becomes "thinner". For example, when $\beta = 0.5$, the percentage-MSE of MLE and NEW1 are between $0.087 \sim 0.035$ interval; when $\beta = 1$, the interval reduced to $0.022 \sim 0.009$; when $\beta = 2$, the interval

reduced to $0.006 \sim 0.002$.

- For α_1, the NEW1 estimates has smaller percentage-bias, while the MLE estimator tends to over-estimate. The percentage-MSE for both methods are about the same, and significantly decreases as the true value of β increases.
- For θ_0, the MLE estimator still tends to over-estimate, while the NEW1 estimator seems slightly under-estimate θ_0 for most cases. The MLE has much larger percentage-MSE than NEW1 method, especially when the true value of β is small. *E.g.*, in Table 2.2, the percentage-MSE for θ_0 for simulation scheme 2 are: 84.562 for MLE, 12.960 for NEW1. Also, similar to estimators for α_0 and α_1, estimation bias and error for $\hat{\theta}_0$ significantly decrease as true value of β increases. For example, when $\beta = 2$, the percentage-MSE for θ_0 for simulation scheme 2 are: 0.180 for MLE, 0.146 for NEW1 (in Table 2.4).
- Overall, as the number of stress levels increasing leads to larger sample size, estimation bias and MSE decreases as sample size increases.

To sum up, simulation for parameter estimations of the Weibull distribution shows that, in terms of estimation bias and error, the performance of the NEW1 method is significantly better than of the MLE method. Also, performance of both methods are somewhat sensitive to the value of shape parameter β of the Weibull distribution. Smaller value of β leads to less accurate results, as the Weibull distribution becomes more heavy-tailed.

We also compare the estimation of CIs from MLE method and the NEW1 method. 1000 replicates of progressively Type II censored samples were generated from Weibull distribution with parameters $(\beta, \alpha_0, \alpha_1) = (1, 5, -1)$, for simulation design scenarios 1, 4 and 7 (non-progressive censoring) and scenarios 3, 6 and 9 (progressive censoring), respectively. We calculate the 95% CI that based on MLE method and the NEW1 inference, for different estimators. The average interval lengths and coverage probabilities were reported to compare performance of the two methods. Table 2.5 summarises results from simulation design scenarios 1, 3, 4, 6, 7, 9.

Results of the simulation tests for the 95% confidence intervals from MLE method and the NEW1 method are summarised as follows:

- Overall, MLE method for the 95% CI has smaller interval length than the NEW1 method does, especially when the sample size is small. For simulation scenario 1, the interval length for μ of MLE is 644.29, while interval length for μ of NEW1 is 1531.015 (in Table 2.5), which is more than twice longer than the MLE.

- However, the MLE CI estimates are significantly poorer than the NEW1's in terms of coverage probability (CP). For the 95% CI, the CP of MLE are all lower than the expected confidence level for all tested statistics. On the other hand, the NEW1's generalised CI estimates have a CP around ±1% of the nominal 95% confidence level. For example, for small sample size such as simulation scheme 1, although MLE has smaller interval length, the CP are all under 90%, but NEW1's generalised CI has CP between 94.5% to 95%.

2.6 A real example and its analysis

Nelson (1975) presented some data on the times to breakdown of a type of electrical insulting fluid subject to various constant voltage stresses. The purpose of the experiment was to estimate the distribution of time to breakdown at 20 kilovolt.

Table 2.1: Table of the simulation design scenarios

Scenario No.	$x_1, ..., x_k$	$n_1, ..., n_k$	$r_1, ..., r_k$	$R_1, ..., R_k$
1	(0.5, 1)	(20, 10)	(12, 6)	$R_1 = (0, ..., 0, 8)$ $R_2 = (0, ..., 0, 4)$
2	(0.5, 1)	(20, 10)	(12, 6)	$R_1 = (8, 0, ..., 0)$ $R_2 = (4, 0, ..., 0)$
3	(0.5, 1)	(20, 10)	(12, 6)	$R_1 = (4, 0, ..., 0, 4)$ $R_2 = (2, 0, ..., 0, 2)$
4	(0.5, 0.75, 1)	(20, 15, 10)	(12, 9, 6)	$R_1 = (0, ..., 0, 8)$ $R_2 = (0, ..., 0, 6)$ $R_3 = (0, ..., 0, 4)$
5	(0.5, 0.75, 1)	(20, 15, 10)	(12, 9, 6)	$R_1 = (8, 0, ..., 0)$ $R_2 = (6, 0, ..., 0)$ $R_3 = (4, 0, ..., 0)$
6	(0.5, 0.75, 1)	(20, 15, 10)	(12, 9, 6)	$R_1 = (4, 0, ..., 0, 4)$ $R_2 = (3, 0, ..., 0, 3)$ $R_3 = (2, 0, ..., 0, 2)$
7	(0.5, 0.75, 1, 1.25)	(30, 20, 15, 10)	(18, 12, 9, 6)	$R_1 = (0, ..., 0, 12)$ $R_2 = (0, ..., 0, 8)$ $R_3 = (0, ..., 0, 6)$ $R_4 = (0, ..., 0, 4)$
8	(0.5, 0.75, 1, 1.25)	(30, 20, 15, 10)	(18, 12, 9, 6)	$R_1 = (12, 0, ..., 0)$ $R_2 = (8, 0, ..., 0)$ $R_3 = (6, 0, ..., 0)$ $R_4 = (4, 0, ..., 0)$
9	(0.5, 0.75, 1, 1.25)	(30, 20, 15, 10)	(18, 12, 9, 6)	$R_1 = (6, 0, ..., 0, 6)$ $R_2 = (4, 0, ..., 0, 4)$ $R_3 = (3, 0, ..., 0, 3)$ $R_4 = (2, 0, ..., 0, 2)$

Table 2.2: Percentage-bias and Percentage-MSE of MLE estimates and the NEW1 method's estimates. Samples generated with $(\beta, \alpha_0, \alpha_1) = (0.5, 5, -1)$. $x_0 = 0$, true value of $\theta_0 = \exp(5) \approx 148.413$, 10000 replicates

	percentage-bias							
Scenario	MLE				NEW1			
No.	β	α_0	α_1	θ_0	β	α_0	α_1	θ_0
1	0.165	-0.006	0.238	1.822	-0.004	0.013	0.046	-0.049
2	0.102	0.007	0.218	2.036	-0.011	-0.011	-0.013	-0.006
3	0.140	-0.005	0.191	1.693	-0.005	-0.003	-0.021	-0.109
4	0.104	-0.015	0.077	1.679	0.000	0.003	-0.009	-0.049
5	0.067	0.001	0.104	1.884	-0.004	-0.006	-0.005	0.007
6	0.085	-0.005	0.110	1.742	-0.006	0.001	0.011	-0.044
7	0.059	-0.005	0.056	0.492	0.000	0.004	0.008	-0.006
8	0.040	0.003	0.061	0.550	-0.002	-0.002	0.000	-0.012
9	0.050	-0.001	0.069	0.535	-0.002	0.002	0.017	0.005
	percentage-MSE							
	MLE				NEW1			
	β	α_0	α_1	θ_0	β	α_0	α_1	θ_0
1	0.107	0.086	4.407	55.653	0.059	0.087	4.358	5.728
2	0.053	0.087	4.430	84.562	0.035	0.087	4.406	12.960
3	0.083	0.083	4.306	44.638	0.048	0.083	4.280	4.361
4	0.053	0.085	4.065	50.759	0.038	0.085	4.076	6.066
5	0.029	0.085	4.108	70.623	0.023	0.086	4.136	11.685
6	0.041	0.083	4.039	64.457	0.031	0.083	4.051	10.532
7	0.025	0.035	1.321	3.425	0.021	0.035	1.346	1.363
8	0.015	0.035	1.302	3.532	0.014	0.035	1.318	1.408
9	0.021	0.035	1.338	3.628	0.018	0.035	1.359	1.477

For the purpose of illustrating the methods presented in this chapter, two Type II progressively censored samples have been randomly generated from the $n_1 = 11$ and $n_2 = 15$ observations recorded at 30 and 36 kilovolts in Nelson (1975) respectively. The observations and the progressive censored plans are reported in Table 2.6. The design stress level $x_0 = 20$kv. Parameter estimation and confidence interval estimation results are shown in Table 2.7 and 2.8, respectively.

The estimates of the new method for the parameter θ_0 and for the mean time to breakdown μ largely depart from the estimates of the MLE. For example, the mean time to breakdown estimated using the new method proposed in this chapter is 8613.56, which is approximately 40% shorter than the value estimated by MLE, 14740.47. Note that in the simulation tests we report that MLE tends to overestimate θ_0, by a percentage up to approximately one third. Hence, in this data, the mean time to breakdown estimated by MLE are possibly also overestimated.

For the CI estimations, results from the real data agree with those from simulations.

Table 2.3: Samples generated with $(\beta, \alpha_0, \alpha_1) = (1, 5, -1)$. $x_0 = 0$, true value of $\theta_0 = \exp(5) \approx 148.413$, 10000 replicates

	percentage-bias							
Scenario	MLE				NEW1			
No.	β	α_0	α_1	θ_0	β	α_0	α_1	θ_0
1	0.162	-0.007	0.079	0.255	-0.006	0.002	-0.017	-0.034
2	0.107	0.005	0.113	0.336	-0.007	-0.004	-0.001	-0.034
3	0.140	0.000	0.105	0.298	-0.005	0.001	0.000	-0.031
4	0.101	-0.008	0.026	0.249	-0.003	0.000	-0.017	-0.028
5	0.064	0.002	0.057	0.312	-0.007	-0.002	0.005	-0.018
6	0.085	-0.003	0.048	0.283	-0.005	0.000	-0.002	-0.020
7	0.060	-0.003	0.027	0.096	0.000	0.002	0.003	0.001
8	0.035	0.001	0.031	0.120	-0.006	-0.001	0.000	-0.009
9	0.049	-0.003	0.018	0.097	-0.003	-0.001	-0.009	-0.012
	percentage-MSE							
	MLE				NEW1			
	β	α_0	α_1	θ_0	β	α_0	α_1	θ_0
1	0.106	0.022	1.067	1.118	0.058	0.022	1.062	0.623
2	0.055	0.022	1.112	1.331	0.036	0.022	1.105	0.670
3	0.082	0.021	1.087	1.222	0.048	0.021	1.075	0.643
4	0.053	0.021	1.032	1.145	0.037	0.021	1.034	0.646
5	0.028	0.021	1.036	1.249	0.023	0.022	1.037	0.679
6	0.041	0.021	1.027	1.211	0.031	0.021	1.028	0.668
7	0.025	0.009	0.327	0.290	0.021	0.009	0.332	0.232
8	0.015	0.009	0.334	0.312	0.013	0.009	0.339	0.241
9	0.021	0.009	0.335	0.290	0.018	0.009	0.335	0.232

The new method has larger confidence interval length for all parameters and the mean time to breakdown. This is reasonable, considering that simulation tests have shown the asymptotic CI based on MLE has poor coverage probability below the nominal level, thus it also has shorter CI length in general.

2.7 Chapter Summary

In this chapter, we have considered a constant stress ALT model with Weibull distribution when the data are progressively censored. We have derived the estimators of unknown parameters, exact confidence interval of shape parameter and generalized CIs of other parameters. One possible disadvantage of the method we proposed for the generalised confidence interval estimation is that, the Monte Carlo simulation algorithm as described in Section 2.4.2 is computational extensive. A large number of Monte Carlo simulations, m, is recommended to ensure convergence, but it may seriously slow down the computation. Nonetheless, the method and theoretic

Table 2.4: Samples generated with $(\beta, \alpha_0, \alpha_1) = (2, 5, -1)$. $x_0 = 0$, true value of $\theta_0 = \exp(5) \approx 148.413$, 10000 replicates

	percentage-bias							
Scenario	MLE				NEW1			
No.	β	α_0	α_1	θ_0	β	α_0	α_1	θ_0
1	0.163	-0.003	0.045	0.053	-0.005	0.002	-0.003	-0.003
2	0.106	0.002	0.056	0.084	-0.009	-0.002	-0.002	-0.014
3	0.145	0.000	0.054	0.068	0.000	0.000	0.002	-0.007
4	0.099	-0.004	0.013	0.062	-0.004	0.000	-0.010	-0.009
5	0.065	0.000	0.023	0.065	-0.006	-0.002	-0.004	-0.015
6	0.087	-0.002	0.019	0.058	-0.003	-0.001	-0.005	-0.008
7	0.063	-0.002	0.010	0.016	0.003	0.000	-0.002	0.000
8	0.037	0.000	0.014	0.027	-0.004	-0.001	-0.001	-0.007
9	0.052	-0.002	0.008	0.018	-0.001	-0.001	-0.006	-0.007
	percentage-MSE							
	MLE				NEW1			
	β	α_0	α_1	θ_0	β	α_0	α_1	θ_0
1	0.104	0.006	0.290	0.163	0.057	0.005	0.270	0.138
2	0.056	0.005	0.280	0.180	0.036	0.006	0.278	0.146
3	0.086	0.005	0.275	0.169	0.050	0.005	0.273	0.142
4	0.051	0.005	0.263	3.017	0.036	0.005	0.254	0.135
5	0.029	0.005	0.253	0.158	0.023	0.005	0.254	0.133
6	0.042	0.005	0.265	0.166	0.031	0.005	0.263	0.144
7	0.026	0.002	0.080	0.056	0.021	0.002	0.081	0.054
8	0.015	0.002	0.081	0.058	0.013	0.002	0.083	0.055
9	0.021	0.002	0.082	0.056	0.018	0.002	0.083	0.054

Table 2.5: Average CP and interval length (in parentheses) of 95% CI estimation. Samples generated $(\beta, \alpha_0, \alpha_1) = (1, 5, -1)$, $x_0 = 0$, and $\log(\theta_0) = 5$. 1000 replicates

S/N	MLE				NEW1			
	β	α_0	α_1	$\log(\theta_0)$	β	α_0	α_1	$\log(\theta_0)$
1	0.966 (1.048)	0.887 (2.524)	0.884 (3.553)	0.887 (2.574)	0.944 (0.934)	0.945 (3.174)	0.95 (4.497)	0.945 (3.228)
3	0.970 (0.961)	0.906 (3.613)	0.913 (627.152)	0.909 (2.642)	0.955 (0.858)	0.941 (4.437)	0.953 (924.62)	0.954 (3.208)
4	0.969 (0.792)	0.925 (2.658)	0.928 (3.676)	0.925 (2.559)	0.952 (0.753)	0.952 (3.068)	0.943 (4.258)	0.952 (2.926)
6	0.957 (0.533)	0.922 (3.689)	0.932 (678.736)	0.932 (2.690)	0.943 (0.692)	0.953 (4.205)	0.951 (883.011)	0.953 (3.096)
7	0.964 (0.581)	0.928 (1.733)	0.933 (2.127)	0.928 (1.736)	0.965 (0.568)	0.944 (1.899)	0.953 (2.35)	0.944 (1.886)
9	0.967 (0.533)	0.919 (2.168)	0.916 (326.839)	0.916 (1.795)	0.953 (0.521)	0.939 (2.355)	0.942 (370.504)	0.944 (1.965)

Table 2.6: Data of the times to breakdown of a type of electrical insulting fluid subject to various constant voltage stresses (Nelson, 1975).

Voltage level	n_i	R_i	Breakdown times
30 kv	11	(2, 0, ..., 0, 2)	7.74, 17.05, 21.02, 43.40, 47.30, 139.07, 144.12
36 kv	15	(4, 0, ..., 0, 1)	0.35, 0.96, 1.69, 1.97, 2.58, 2.71, 3.67, 3.99, 5.35, 13.77

Table 2.7: Parameter estimations of Weibull distribution of the lifetime of the electrical insulting fluid subject.

	$\hat{\beta}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\theta}_0$	$\log(\hat{\theta}_0)$	$\hat{\mu}$
MLE	1.02	19.54	-0.50	14868.87	9.61	14740.47
NEW1	0.93	19.84	-0.50	8342.39	9.03	8613.56

Table 2.8: 95% confidence interval for estimators of Weibull distribution of the lifetime of the electrical insulting fluid subject.

	MLE		NEW1	
	lower boundary	upper boundary	lower boundary	upper boundary
β	0.93	1.11	0.64	1.37
α_0	14.23	24.86	13.98	26.48
α_1	-0.65	-0.34	-0.70	-0.33
$\log(\theta_0)$	7.42	11.79	7.44	12.56

results are new, totally different from MLE-based inference. The numerical analysis and comparison are promising, even for small sample and different censoring rates or schemes.

Chapter 3

Point and interval estimation for the generalised Pareto distribution with small samples

In extreme value theory, the generalised Pareto distribution is used to model data on the tail. Since only a proportion of the data is used, the effective data size for fitting the GPD is often small. Also, statistical properties, especially tail behaviour of the GPD varies depending on its shape parameter. Performances of most existing methods are inconsistent for different ranges of the shape parameter. In this chapter, we introduce a new method that provides consistently unbiased point estimations, even for very small sample size. This method also provides an inference for confidence interval estimation, which is accurate in terms of coverage probability, combined with reasonable interval length as shown in simulation tests.

3.1 Introduction

In recent years the generalised Pareto distribution has drawn increasingly attention from researchers because of its application in the extreme value theory (see Castillo and Hadi, 1997; Zhang, 2007; Zhang and Stephens, 2009 and Hüsler et al., 2011, amongst others). The development of EVT is motivated by needs of measuring extreme tail probabilities and tail quantiles, which, in fields such as the research of disasters, nature phenomenon, insurance and finance, are often the major concern of risk-related problems. The essential idea of EVT is, as stated in Diebold et al. (2000), estimating extreme quantiles and probabilities by fitting models to a set

of data using only the data on the tail. Consequently, since only a proportion of data from the original dataset is used, small sample size is often confronted in tail modelling. Hence, we emphasise the aspect of small samples on fitting the generalised Pareto distribution in this chapter.

In the extreme value theory, generalised Pareto distribution is used to model exceedances over certain thresholds, or in other words, data on the tail of the underlying distribution. This method of exceedances is often referred to as the peak over threshold method. In the POT approach, rather than assessing $F_Y(y)$ directly, it is more focused on estimating the distribution of the exceedances. Consider the following assumptions

A 3.1 Assuming random variable Y has distribution function $F_Y(y)$, the definition of exceedances is the values of Y above threshold u, namely $X = Y - u$. The probability function of the exceedance is given by

$$F_u(x) = \mathbb{P}\{Y - u \leq x | Y \geq u\}, \qquad 0 \leq x \leq y_{\mathrm{F}} - u,$$

where x is the excess and $y_{\mathrm{F}} \leq \infty$ denotes the right end points of F_Y.

A 3.2 $F_u(x)$ can be expressed as a function of F_Y and the

$$F_u(x) = \frac{F_Y(u+x) - F_Y(u)}{1 - F_Y(u)} = \frac{F_Y(y) - F_Y(u)}{1 - F_Y(u)}. \tag{3.1}$$

Based on these assumptions, Balkema and De Haan (1974) along with Pickands (1975) introduced the following results

Theorem 3.1 (Pickands-Balkema-de Haan) *when u is large, for a large class of underlying distribution functions, the conditional excess distribution function F_u is well approximated by the generalised Pareto distribution,*

$$F_u(x) \approx F_{\mathrm{GPD}}(x; \sigma, \xi), \ \ \textit{as } u \to \infty. \tag{3.2}$$

$F_{\mathrm{GPD}}(x; \sigma, \xi)$ is the cumulative distribution function

$$F_{\mathrm{GPD}}(x; \sigma, \xi) = \begin{cases} 1 - (1 + \frac{\xi}{\sigma} x)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-x/\sigma} & \text{if } \xi = 0 \end{cases} \tag{3.3}$$

for $x \in [0, y_{\mathrm{F}} - u)$ if $\xi \geq 0$ and $x \in [0, -\frac{\sigma}{\xi}]$ if $\xi < 0$, where ξ is the shape parameter and σ is the scale parameter. Note that sometimes the shape parameter ξ is denoted

as $\kappa = -\xi$. If $\xi > 0$ the GPD is heavy-tailed on the right side. If $\xi = 0$ then the distribution reduces to the exponential distribution and the right tail decays exponentially. When $-1 \leq \xi \leq -0.5$, the distribution has finite right end-points, which is sometimes referred to as 'short-tailed'. Furthermore, when $\xi = -1$, the GPD actually becomes a uniform $\mathrm{U}(0, \sigma)$ distribution. The kth central moment of the GPD exists only if $\xi < 1/k$. For example, when it has a shape parameter $\xi \geq 1/2$, $\mathrm{var}(X) = +\infty$ and the second central moment no longer exists. For more details about properties of the GPD and its parameters, see Hosking and Wallis (1987).

Furthermore, the p-th quantile of GPD is given by the inverse of (3.3)

$$Q_{\mathrm{GPD}}(p; \sigma, \xi) = F^{-1}_{\mathrm{GPD}}(p; \sigma, \xi) = \frac{\sigma}{\xi}\left((1-p)^{-\xi} - 1\right). \tag{3.4}$$

To simplify, we will drop the extra notion and write $Q(p) = Q_{\mathrm{GPD}}(p; \sigma, \xi)$ henceforth where no confusion should be caused.

In practice, the final goal of employing the EVT is still estimating extreme probabilities or quantiles of the original distribution of Y, but not only the exceedances X. Hence, by substituting F_u in (3.1) with the GPD and approximating $F_Y(u)$ by $(N-n)/N$, where N is the total number of observations and n is the number of observations over the threshold u, the original distribution $F_Y(y)$ is expressed with respect to the GPD parameters as

$$F_Y(y) = 1 - \frac{n}{N}\left(1 + \frac{\xi}{\sigma}(y)\right)^{-1/\xi}. \tag{3.5}$$

The p^*-th quantile of Y, $F_Y^{-1}(p^*)$ is simply given by inverting (3.5)

$$F_Y^{-1}(p^*) = u + \frac{\sigma}{\xi}\left(\left(\frac{N}{n}(1-p^*)\right)^{-\xi} - 1\right). \tag{3.6}$$

Comparing (3.4) and (3.6) we will obtain that

$$F_Y^{-1}(p^*) = u + Q\left(\frac{N}{n}p^*\right). \tag{3.7}$$

The expression in (3.6) represents one of the main goals of the EVT: when p^* is very small, instead of estimating extreme p^*-th quantile of F_Y, which could be very difficult, one can estimate less-extreme $(\frac{N}{n}p^*)$-th quantiles of $F_{\texttt{GPD}}$, assuming $N >> n$.

3.1.1 Existing problems

Since in POT approach, only a proportion of the original sample is used for fitting the GPD, the sample size of the excess, n, is often small. Generally, the sample size n for modelling the GPD is determined by threshold u. According to (3.2), a high threshold generally provides better approximation of the data on the tail with smaller bias. However, higher u also leads to smaller sample size n, which may increase estimation variance and bring other problems to the GPD estimations. Furthermore, for some estimators, their asymptotic properties based on large samples may not be valid when n is too small. Consequently, interval estimations that are based on asymptotic theory or bootstrap approaches become less efficient and less accurate. Unfortunately, in practice the GPD fitting is often confronted with small samples, although definitions of 'small' and 'large' sample are somehow vague. We illustrate this via a simple example: in finance, a typical one-year dataset of daily returns (or losses) normally consists of around 250 samples (as there are around 250 trading days per year), which might not be very large but at least not 'small'. If 10% data on the tail is selected from the original dataset, there are only 25 samples available for fitting the GPD; even if 30% data is selected there are still only 75 samples. It is possible to increase the sample size by using a longer sampling period, but it is not always wise as older data could be out-of-date. Data from other fields of research might contain larger samples, nonetheless, as not the entire dataset is used, the effective sample size for estimating the GPD is often much smaller. In this chapter, we focus on sample size $n \leq 50$.

As described in previous section, some of the distributional properties of the GPD vary distinctively as value of the shape parameter ξ changes, an ideal estimator for GPD should be reliable and consistent regardless of its 'shape', namely, the tail behaviour of the underlying distribution. However, performance of some methods are heavily impacted by the value of ξ, or even become invalid.

For examples, the MLE method does not exist for $\xi < -1$ since it can have no local maximum. The MOM estimates do not exist for $\xi \geq 0.5$ and the PWM estimates do not exist for $\xi \geq 1$, since the second and first moments of the GPD become infinity in each case, respectively. Although Hosking and Wallis (1987) suggested that it should be sufficient to restrict to $-0.5 < \xi < 0.5$ when considering reliability, Castillo and Hadi (1997) made a good line of argument that a good estimator for the GPD should cover a wider domain of ξ for both practical and theoretical reasons. The EPM method they proposed for parameter estimations are reliable under a wider domain of ξ (from -2 to 2), but it does not outperform classical methods when the value

of ξ falls into $(-0.5, 0.5)$. The LME method proposed by Zhang (2007) has high asymptotic efficiency when r is close to its optimal value, which equals to the true value of $-\xi$. However, the LME only supports for $r < 1/2$, which corresponds to $\xi > -0.5$. Also, some methods such as MOM and PWM sometimes conflict with data (see, *e.g.,* Castillo and Hadi (1997); Dupuis and Tsao (1998)). That is, for $\xi < 0$, the supported domain of x is $0 < x < -\sigma/\xi$, however, invalid estimation of σ and ξ may be given so that some observations can be out of this domain. This is referred to as non-feasible solution or invalid estimates issue in some literature.

Although some methods may be preferred for certain range of ξ, it is typical that the tail behaviour of the underlying data is unknown prior to modelling in reality. In this chapter, we elaborate on the reliability with regards to ξ, ranging from -1 to 1, of different methods, including the method we proposed.

3.1.2 A brief review of existing methods

For fitting the GPD, the maximum likelihood estimation has been considered by many (including Davison, 1984; Smith, 1985, Smith, 1987; Hosking and Wallis, 1987; Grimshaw, 1993; *etc.*). However, Hosking and Wallis suggested that unless sample size is larger or equal to 500, the method of moment and the probability weighted moment are more reliable for $\xi \in (-0.5, 0.5)$. Castillo and Hadi (1997) proposed the so-called elemental percentile method which is reliable for a wider domain of ξ (from -2 to 2), but it does not out-perform classic methods as long as value of ξ falls into $(-0.5, 0.5)$. Zhang (2007) proposed an alternative method for fitting the GPD, the likelihood moment estimation. The LME contains an auxiliary parameter r, which needs to be determined without knowing ξ (or with preliminary estimation of ξ from other methods). Zhang showed that the optimal value of r should be close to the true value of $-\xi$, but $r = 1/2$ is recommended if no information about ξ is available. Zhang and Stephens (2009) provided another estimator (referred as ZS henceforth) of parameters of the GPD which uses a procedure similar to Bayesian method. The ZS estimators are fast to compute and their simulation tests show that it is stable when ξ is in the domain of $[-0.5, 1]$ for sample size $n = \{50, 500\}$, and generally outperform MLE, MOM, PWM and LME in most cases.

Compared with point estimations, confidence interval estimation for the GPD are less intensively discussed in previous literature, and there exist obvious difficulties of constructing such interval estimates explicitly. Hence, a common strategy is to find approximate confidence intervals with coverage probabilities equal to the target confidence level, theoretically. One well-known approach is approximations based on

the asymptotic normal distribution. The $1-\tau$ CI of parameter θ is constructed in the form of

$$\hat{\theta} - z_{\tau/2}\sqrt{\hat{v}(\hat{\theta})} < \theta < \hat{\theta} + z_{1-\tau/2}\sqrt{\hat{v}(\hat{\theta})},$$

where z_τ is the τ-th quantile of the standard normal distribution, and $\hat{v}(\hat{\theta})$ is the asymptotic variance of estimator $\hat{\theta}$. Further more, the asymptotic variance can be obtained via either $\hat{v}(\hat{\theta}) = 1/\mathcal{I}(\hat{\theta})$ or $\hat{v}(\hat{\theta}) = 1/\mathcal{J}(\hat{\theta})$, where $\mathcal{I}(\hat{\theta})$ and $\mathcal{J}(\hat{\theta})$ are the expected and observed Fisher information, respectively. The asymptotic variance derived from the expected Fisher information of some estimators can be found in Smith (1987) (for MLE), Hosking and Wallis (1987) (for MOM and PWM), and Zhang (2007) (for LME), respectively. Example of CI estimation based on observed information can be found in Davison and Smith (1990)) amongst others. CI for quantile estimators can be obtained also using asymptotic variance via the Delta method (Rao, 1973). However, CI estimates based on asymptotic theory may suffer under small samples, as asymptotic properties may not be the same as stated. For small samples, profile log-likelihood approach may be preferred (*e.g.*, VENZON and MOOLGAVKAR, 1988; Davison and Smith, 1990). The profile log-likelihood approach is based on the fact that the relative likelihood $L(\theta)/L(\hat{\theta})$ follows the χ^2 distribution with 1 degree of freedom. Gilli and Këllezi (2006) gave examples of profile log-likelihood CI estimates for some financial criteria based on the GPD estimation. Tajvidi (2003) compared sampling based approaches (*i.e.* jackknife and bootstrap) with profile likelihood approaches for heavy-tailed GPD and found the later is better for both small and large samples. In this chapter, we adapt the inference in Wang et al. (2010) and propose an approach that does not depend on asymptotic normality for calculating the exact and generalised confidence intervals for the GPD.

In this chapter, we propose an alternative approach for fitting the GPD. The new method provides consistent unbiased estimations for parameters, as well as accurate exact and generalised confidence interval estimations in terms of coverage probability.

Results of simulation tests for different methods and for CI estimations are presented in Section 3.3. An example using real-world data (same data has been used in Castillo and Hadi, 1997 and Zhang and Stephens, 2009) will be discussed in Section 3.4, and finally summary and conclusions in the last section.

3.2 Exact inference for the GPD

3.2.1 Point estimation

Re-parametrise the GPD with $\alpha = \xi/\sigma$, for $\xi \neq 0$, and assume $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$ be an ordered statistics of n observations from the GPD. The proposed estimator for α is then given via some transformation from the c.d.f. (3.3) as follows. First, let

$$\begin{aligned} V_{(i)} &= -\log\left(1 - F_{\text{GPD}}(X_{i:n}; \alpha, \xi)\right) \\ &= \frac{1}{\xi}\log\left(1 + \alpha X_{i:n}\right), \qquad i = 1, 2, ..., n, \end{aligned}$$

be a sequence of ascending order statistics of samples from the standard exponential distribution . $V_{(i)}$ is not independent, so let

$$W_i = (n - i + 1)(V_{(i)} - V_{(i-1)}), \qquad i = 1, 2, ..., n$$

be the ith normalised spacing, with $V_{(0)} = 0$. Then $W_1, ..., W_n$ are random variables from independent standard exponential distributions (proof can be found in Viveros and Balakrishnan, 1994, Appendix B, in which the property of normalised spacings between order statistics of exponential samples is used). Then let

$$\begin{aligned} D_i &= \sum_{j=1}^{i} W_j \\ &= -\frac{1}{\xi}\left(\sum_{j=1}^{i}\log\left(1 + \alpha X_{j:n}\right) + (n - i)\log\left(1 + \alpha X_{i:n}\right)\right), \qquad (3.8) \\ &\quad i = 1, 2, ..., n; \end{aligned}$$

and

$$\begin{aligned} U_{i:n} &= \frac{D_i}{D_n} \\ &= \frac{\sum_{j=1}^{i}\log\left(1 + \alpha X_{j:n}\right) + (n - i)\log\left(1 + \alpha X_{i:n}\right)}{\sum_{j=1}^{n}\log\left(1 + \alpha X_{j:n}\right)}, \qquad (3.9) \\ &\quad i = 1, 2, ..., n - 1. \end{aligned}$$

$\{U_{i:n}\}_{i=1}^{n-1}$ can be equivalently regarded as an induced order statistics of an i.i.d. sample which has a standard uniform U$(0, 1)$ distribution. Note that $U_{i:n}$ is also an ancillary statistic that only depends on parameter α. Hence the mean of $\{U_{i:n}\}_{i=1}^{n-1}$

converges with probability one to 1/2. Let

$$\bar{U}(\alpha) = \frac{1}{n-1}\sum_{i=1}^{n-1} U_{i:n}.$$

Then $\hat{\alpha}$ is determined by solving

$$\bar{U}(\alpha) = \frac{1}{2}. \tag{3.10}$$

Since $V_{(i)}$ can be seen as order statistics from standard expornential distribution, it has sample mean $\frac{1}{n}\sum_{i=1}^{n} V_{(i)} = 1$. This leads to an estimator for the shape parameter ξ, given estimator $\hat{\alpha}$:

$$\hat{\xi} = \frac{1}{n}\sum_{i=1}^{n} \log(1 + \hat{\alpha} X_i). \tag{3.11}$$

Note that (3.11) is identical with the log-likelihood function for ξ as in MLE. The method we proposed is referred as NEW2 henceforth.

We notice that the idea behind (3.10) is similar to Hüsler et al. (2011), which assume that the probabilities of the GPD given data $(X_1, X_2, ..., X_n)$ can also be regarded as uniformly U$(0, 1)$ distributed, and then estimates of α can be obtained by solving the equation

$$\frac{1}{n}\sum_{i=1}^{n} F_{\texttt{GPD}}(X_i; \hat{\alpha}, \hat{\xi}) = \frac{1}{2},$$

along with the profile log-likelihood function (3.11). The essential difference between NEW2 estimates and the approach in Hüsler *et. al* is that, for the NEW2 estimates one does not have to assume the profile log-likelihood in priori, since $U_{i:n}$ is an ancillary statistic that does not depend on any parameters other than α. Actually, there are several different ways to obtain $\hat{\xi}$ given $\hat{\alpha}$, and through simulations we found the profile log-likelihood estimates for $\hat{\xi}$ in (3.11) are the most accurate.

Thus, our estimators for the GPD parameters are finally given by combining (3.10) and (3.11). And the pth quantile of the GPD is estimated by either $\hat{Q}(p; \hat{\sigma}, \hat{\xi})$ or $\hat{Q}(p; \hat{\alpha}, \hat{\xi})$, where $Q(p)$ is shown as in (3.4).

3.2.2 Exact and generalised confidence intervals estimations

Confidence interval for estimators of α and ξ can be derived from (3.10) with the IE inference. Note that the mean of n independent uniformly U$(0, 1)$ distributed random variables follows the Bates distribution (Bates, 1955), which has the p.d.f.

as

$$f_X(x;n) = \frac{n}{2\,(n-1)!} \sum_{k=0}^{n} (-1)^k \binom{n}{k} (nx-k)^{n-1} \operatorname{sgn}(nx-k),$$

for $x \in (0,1)$, and

$$\operatorname{sgn}(nx-k) = \begin{cases} -1 & nx < k \\ 0 & nx = k \\ 1 & nx > k. \end{cases}$$

Recall from the last section, given data $X = x$, it is obvious that the sample distribution of $\bar{U}(\alpha)$ is the Bates distribution with size $n-1$ (noted as $\bar{U}(\alpha) \sim \text{Bates}(n-1)$). Let μ_L and μ_U be lower and upper boundaries such that the probability of the value of $\bar{U}(\alpha)$ falls between these boundaries are $1-\tau$, i.e.

$$\mathbb{P}(\mu_L < \bar{U}(\alpha) < \mu_U) = 1 - \tau.$$

Let $\text{Bates}_\tau^{-1}(n)$ represents the τ-th quantile of the Bates distribution with sample size n, then $\mu_L = \text{Bates}_{\tau/2}^{-1}(n-1)$ and $\mu_U = \text{Bates}_{1-\tau/2}^{-1}(n-1)$, respectively. Since $\bar{U}(\alpha)$ is an pivot quantity and monotonic, it is straightforward to show that

$$\mathbb{P}(A(\mu_L;x) < \alpha < A(\mu_U;x)) = 1 - \tau,$$

where $A(\mu;x) = \alpha$ is the inverse function of $\bar{U}(\alpha) = \mu$, given value of observations $X = x$. That is, the exact $(1-\tau)\%$ confidence interval for α is

$$\left(A(\mu_L;x), \quad A(\mu_U;x)\right), \tag{3.12}$$

However, μ_L and μ_U are difficult to calculated analytically from the Bates distribution, especially as n grows large. On the other hand, it is more convenient to obtain these values via Monte Carlo simulation. For example, generate large amount (say, $m \geq 2000$) of random samples $\boldsymbol{\mu}$ from $\text{Bates}(n-1)$ distribution, and the τ-th population quantiles, $\text{Bates}_\tau^{-1}(n-1)$, can be well approximated by the τ-th sample quantiles of $\boldsymbol{\mu}$. Then the interval boundaries as in (3.12) can be obtained, using these simulated sample quantiles.

$$T = 2D_n = 2\sum_{j=1}^{n} \log\left(1 + \alpha X_{j:n}\right)^{\frac{1}{\xi}}.$$

As shown in Wang et al. (2010), T has a χ^2 distribution with $2n$ degrees of freedom, regardless of values of α and X. Then consider the following generalised pivot

quantity for parameter ξ:

$$Z = \frac{2\sum_{j=1}^{n} \log\left(1 + A(\mu; x)X_{j:n}\right)}{T}, \tag{3.13}$$

where $A(\mu; x)$ is previously defined. Note that the value of Z will reduce to ξ if $\mu = 1/2$. As there is no unknown parameters in (3.13), Z is a generalised pivotal quantity for ξ as defined in Weerahandi (1993) and Weerahandi (2004), and its distribution can be approximated via Monte Carlo simulations, as follows:

1. generate random samples $\mathbf{t} \sim \chi^2(2n)$, from the χ^2 distribution with 2n degrees of freedom, with the same sample size m as of random samples $\boldsymbol{\mu}$;
2. replace μ and T in (3.13) with $\boldsymbol{\mu}$ and $\mathbf{t}$, respectively. This produces a simulated random sample set of Z, denoted as $\mathbf{z}$. Hence, the τ-th sample quantile of $\mathbf{z}$ can be used to approximate the τ-th population quantile of Z, denoted as Z_τ;
3. the generalised $(1-\tau)\%$ confidence interval of ξ is given by:

$$\left(Z_{\tau/2} \quad , \quad Z_{1-\tau/2}\right).$$

The generalised confidence interval for the quantile of the GPD, $Q_{\text{GPD}}(p; \alpha, \xi)$, can be obtained via the following generalised pivotal quantity:

$$S = Q_{\text{GPD}}\left(p; A(\mu; x), Z\right). \tag{3.14}$$

The population quantiles of S can be derived along the same line: replacing μ and Z in (3.14) with $\boldsymbol{\mu}$ and $\mathbf{z}$, which are previously defined, to obtain the Monte Carlo samples of S. Then the τ-th population quantile of S, denoted as S_τ, can be approximated by the τ-th sample quantile of the generated sample. Hence, the $(1-\tau)\%$ generalised confidence interval of $Q_{\text{GPD}}(p; \alpha, \xi)$ is given by

$$(S_{\tau/2}, \quad S_{1-\tau/2}).$$

In the next section, we perform simulation tests to compare and analysis performance of the NEW2 inference with performance of other methods, for both point estimation and interval estimation.

3.3 Simulation study

We conduct simulation experiments to compare parameter estimations from methods mentioned in previous sections. The data is generated from the GPD as follow: sample size $n = \{50, 30, 15\}$, scale parameter $\sigma = 1$ and shape parameter ξ ranged from -1 to 1 by 0.05, respectively; for each case there are 5000 replicates. Since the scope of the study is for small sample sizes, the MLE method is omitted from tests due to the fact that it has been proved inefficient when sample size is small, as stated in previous section. Totally six methods are considered for point estimations, including MOM, PWM, EPM, LME, ZS, which have all been introduced in Section 3.1, and the new method we proposed (NEW2). MOM and PWM are listed as benchmarking methods, and the POT package (Ribatet, 2011) in the statistical programming language R is used for parameter and quantile estimation. For the EPM, the sampling scheme 3 in Sec. 2.2 of Castillo and Hadi (1997) is applied to eliminate non-feasible solution. And for the LME method, since we assume no preliminary estimation or information of the data is available and it is only fair to all tested models, we set $r = -1/2$ as recommended by Zhang (2007) for all cases. Results are then used for quantile estimations for different quantile levels.

We also tested the confidence interval estimation procedure proposed in this chapter and it is compared with several other models. Details and results are summarised in Section 3.3.3.

3.3.1 Parameter estimation

First we tested performance of different models for estimating the parameters. The main purpose of this test is to examine the performance of different methods for the GPD with different tail behaviours. Estimation bias and root mean square error are calculated to assess performance of each method. To save space, instead of presenting results for all cases, we select several cases with respect to different sample size and $\xi = \{-1, -0.5, -0.25, 0.0.25, 0.5, 0.75, 1\}$ in Table 3.1 and 3.2. Also, bias and RMSE of $\hat{\sigma}$ and $\hat{\xi}$ from EPM, LME, ZS and NEW2 methods are plotted against ξ in Figure 3.1 and 3.2, respectively. Results from MOM and PWM are not included in figures due to their poor performance under extreme heavy-tailed situations, also because similar graphic evidence can be found in other papers, such as Zhang and Stephens (2009), Fig. 1&2 and Hüsler et al. (2011), Fig. 2-5. The results can be summarised as follows:

1. The NEW2 method is consistently unbiased when estimating σ and ξ, even for

very small sample size $n = 15$, with reasonable RMSE. For example, among the $n = 15$ cases, the NEW2 method's estimation bias of $\hat{\xi}$ range from -0.019 to 0.007, while the ZS method's estimation bias of $\hat{\xi}$ range from -0.109 to 0.155, for $-1 \leq \xi \leq 1$.

2. For $\hat{\sigma}$, the NEW2 estimator has the smallest RMSE for very heavy-tailed situations such as $\xi = 0.75, 1$; for $\hat{\xi}$, the ZS estimator has the smallest RMSE for almost all cases. The RMSEs of the NEW2 method do not stand out on top every time, but the difference can be concluded as only marginal, and, at least, it is generally better than moment based methods (MOM and PWM).

3. Overall, the RMSEs for both parameters of all tested methods gradually increase as the GPD becomes more heavy-tailed ($\xi > 0$). On the other hand, for methods other than NEW2, their estimation for both parameters become more biased when true value of ξ departs from 0, especially when sample size becomes smaller. This can be observed from Figure 3.1 and 3.2.

4. For the EPM method, the bias and the RMSE of the shape estimator $\hat{\xi}$ increase substantially as ξ increases. But overall it provides better estimations than moment-based methods and confirms the conclusion in Castillo and Hadi (1997).

We believe it is safe to draw conclusions that, for parameter estimations, the NEW2 method can be considered as the best, since it is the most consistent estimators for small sample sizes and different tail behaviours, with overall smallest bias and reasonable estimation error.

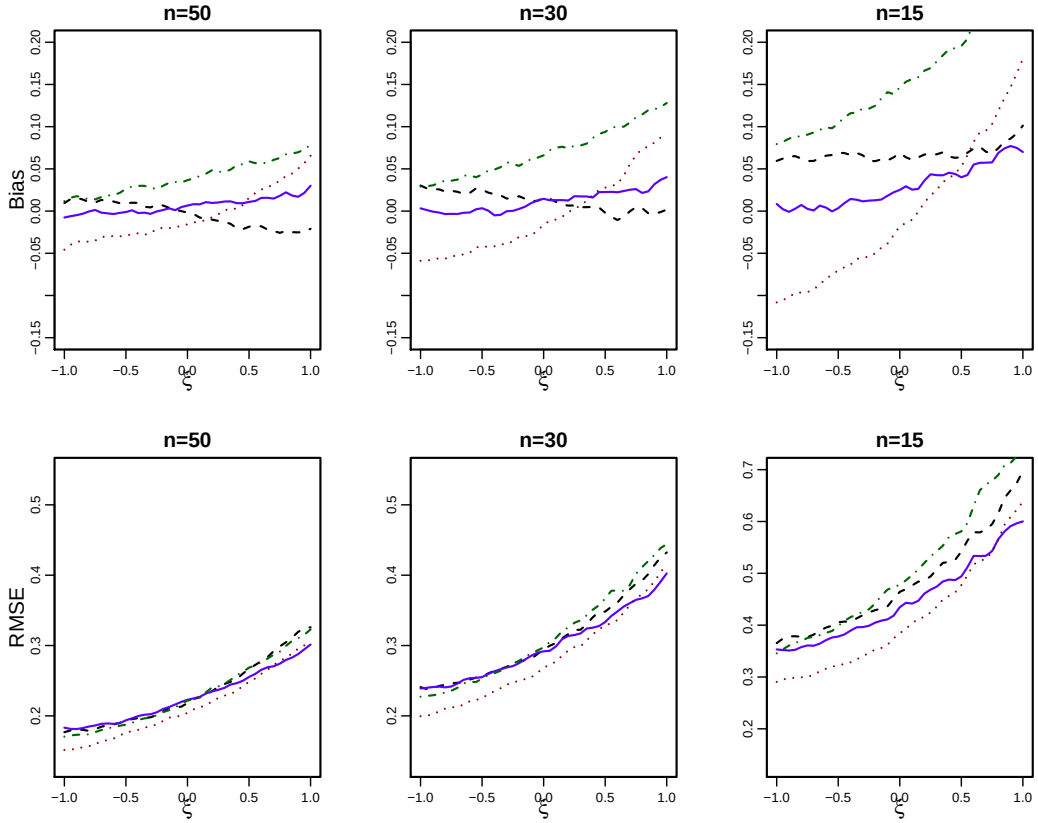
Table 3.1: Bias and RMSE of $\hat{\sigma}$

		Bias						RMSE					
n	ξ	MOM	PWM	EPM	LME	ZS	NEW2	MOM	PWM	EPM	LME	ZS	NEW2
50	-1	0.024	0.015	0.010	0.012	-0.045	-0.008	0.242	0.232	0.178	0.171	0.152	0.183
	-0.5	0.020	0.017	0.006	0.026	-0.031	0.002	0.217	0.226	0.194	0.187	0.178	0.194
	-0.25	0.030	0.027	0.008	0.024	-0.020	-0.007	0.205	0.222	0.201	0.200	0.187	0.205
	0	0.046	0.032	-0.001	0.033	-0.016	0.009	0.211	0.230	0.220	0.220	0.206	0.222
	0.25	0.121	0.040	-0.014	0.048	-0.006	0.006	0.249	0.235	0.240	0.242	0.226	0.233
	0.5	0.406	0.083	-0.018	0.056	0.016	0.009	0.812	0.264	0.268	0.267	0.248	0.255
	0.75	1.305	0.182	-0.021	0.064	0.041	0.017	4.200	0.356	0.303	0.292	0.282	0.274
	1	4.794	0.440	-0.021	0.079	0.065	0.029	25.755	0.920	0.326	0.323	0.304	0.301
30	-1	0.060	0.042	0.032	0.028	-0.058	0.003	0.346	0.318	0.242	0.227	0.200	0.239
	-0.5	0.050	0.045	0.029	0.042	-0.037	0.006	0.291	0.298	0.254	0.252	0.223	0.254
	-0.25	0.045	0.039	0.016	0.058	-0.035	0.002	0.277	0.295	0.272	0.275	0.246	0.270
	0	0.076	0.054	0.014	0.069	-0.017	0.014	0.285	0.306	0.296	0.301	0.270	0.292
	0.25	0.171	0.074	0.005	0.075	0.002	0.015	0.339	0.319	0.320	0.328	0.295	0.312
	0.5	0.462	0.124	-0.001	0.089	0.029	0.030	0.856	0.357	0.350	0.361	0.331	0.337
	0.75	1.471	0.261	0.003	0.108	0.066	0.032	8.929	0.614	0.384	0.394	0.364	0.365
	1	6.173	0.633	-0.001	0.129	0.091	0.042	65.418	3.127	0.431	0.442	0.415	0.402
15	-1	0.103	0.057	0.061	0.082	-0.107	0.006	0.543	0.451	0.366	0.348	0.290	0.352
	-0.5	0.098	0.079	0.070	0.101	-0.070	0.004	0.480	0.458	0.406	0.397	0.324	0.378
	-0.25	0.088	0.072	0.054	0.127	-0.059	0.014	0.429	0.440	0.420	0.431	0.346	0.395
	0	0.145	0.107	0.069	0.142	-0.017	0.023	0.456	0.473	0.467	0.479	0.386	0.429
	0.25	0.266	0.147	0.066	0.171	0.018	0.041	0.522	0.490	0.492	0.534	0.423	0.463
	0.5	0.567	0.223	0.071	0.192	0.058	0.035	1.129	0.564	0.545	0.570	0.479	0.489
	0.75	1.476	0.400	0.069	0.238	0.102	0.064	4.677	0.836	0.609	0.664	0.547	0.553
	1	5.270	0.948	0.104	0.267	0.181	0.070	46.190	4.448	0.708	0.745	0.646	0.602

Table 3.2: Bias and RMSE of $\hat{\xi}$

		Bias						RMSE					
n	ξ	MOM	PWM	EPM	LME	ZS	NEW2	MOM	PWM	EPM	LME	ZS	NEW2
50	-1	-0.031	-0.013	-0.007	-0.015	0.057	0.014	0.332	0.315	0.192	0.188	0.167	0.209
	-0.5	-0.019	-0.015	0.011	-0.029	0.033	0.005	0.201	0.222	0.153	0.150	0.144	0.166
	-0.25	-0.027	-0.022	0.024	-0.027	0.024	0.011	0.157	0.186	0.156	0.152	0.143	0.166
	0	-0.045	-0.031	0.052	-0.029	0.019	0.000	0.145	0.171	0.191	0.165	0.157	0.177
	0.25	-0.103	-0.046	0.083	-0.039	0.009	-0.003	0.168	0.173	0.255	0.189	0.181	0.196
	0.5	-0.223	-0.089	0.118	-0.044	-0.007	0.008	0.253	0.202	0.340	0.223	0.210	0.224
	0.75	-0.392	-0.173	0.153	-0.046	-0.027	0.003	0.404	0.247	0.420	0.255	0.241	0.254
	1	-0.594	-0.297	0.189	-0.057	-0.039	-0.004	0.599	0.336	0.503	0.295	0.277	0.290
30	-1	-0.077	-0.043	-0.028	-0.038	0.077	0.001	0.475	0.429	0.270	0.255	0.229	0.278
	-0.5	-0.047	-0.040	0.003	-0.048	0.044	0.002	0.271	0.290	0.211	0.211	0.191	0.226
	-0.25	-0.041	-0.034	0.036	-0.062	0.042	0.005	0.215	0.249	0.223	0.217	0.199	0.225
	0	-0.079	-0.057	0.053	-0.063	0.018	-0.003	0.203	0.233	0.261	0.228	0.217	0.239
	0.25	-0.144	-0.077	0.098	-0.066	0.007	0.001	0.224	0.235	0.328	0.259	0.241	0.258
	0.5	-0.266	-0.128	0.137	-0.068	-0.012	-0.005	0.303	0.250	0.406	0.298	0.271	0.295
	0.75	-0.434	-0.220	0.176	-0.081	-0.038	-0.011	0.450	0.303	0.498	0.334	0.305	0.336
	1	-0.634	-0.353	0.214	-0.090	-0.060	-0.001	0.642	0.404	0.609	0.383	0.350	0.377
15	-1	-0.133	-0.052	-0.053	-0.116	0.155	-0.001	0.742	0.601	0.436	0.408	0.370	0.431
	-0.5	-0.104	-0.078	-0.017	-0.119	0.084	0.007	0.461	0.448	0.374	0.357	0.317	0.372
	-0.25	-0.092	-0.071	0.034	-0.137	0.072	-0.001	0.349	0.375	0.374	0.355	0.316	0.361
	0	-0.144	-0.106	0.057	-0.138	0.028	0.000	0.321	0.350	0.415	0.371	0.328	0.377
	0.25	-0.230	-0.150	0.104	-0.144	-0.001	-0.012	0.342	0.354	0.483	0.412	0.361	0.402
	0.5	-0.362	-0.226	0.137	-0.148	-0.040	-0.003	0.424	0.381	0.562	0.436	0.391	0.433
	0.75	-0.519	-0.320	0.209	-0.182	-0.063	-0.019	0.551	0.430	0.678	0.500	0.435	0.498
	1	-0.715	-0.463	0.251	-0.177	-0.109	0.001	0.733	0.536	0.806	0.557	0.492	0.555

Figure 3.1: Bias and RMSE of the **scale estimator** $\hat{\sigma}$ plotted against ξ (from -1 to 1).

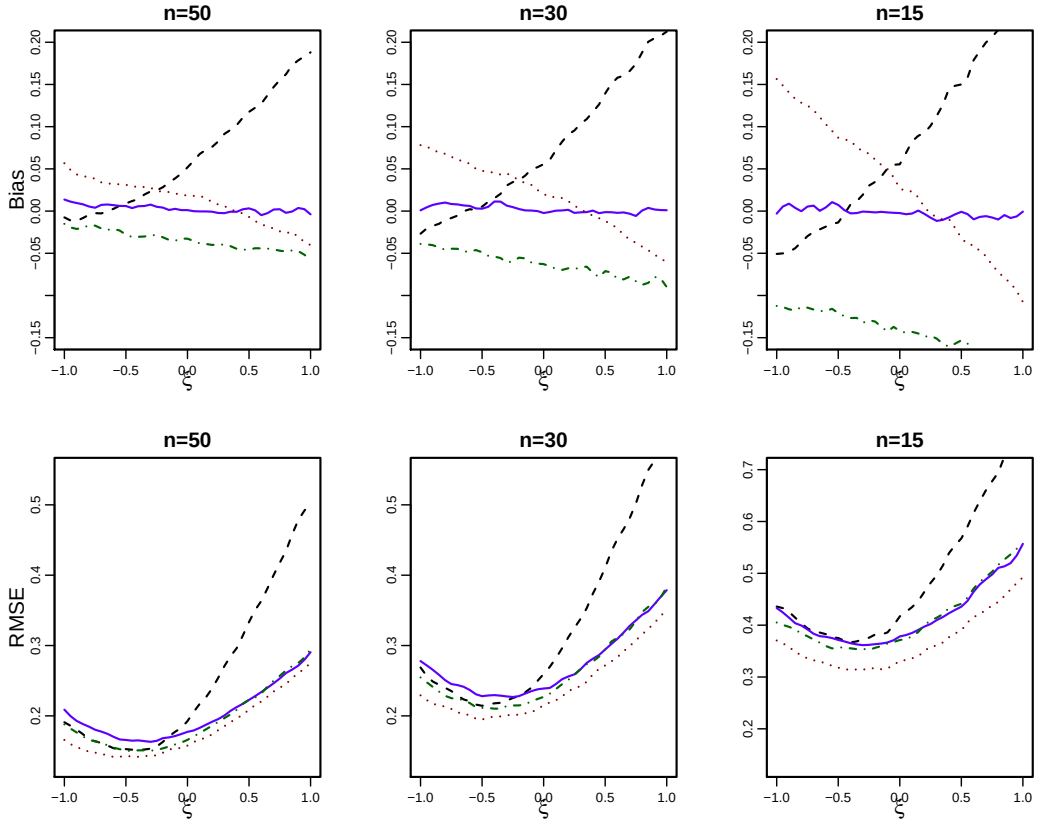


dashed: EPM, dot-and-dash: LME, dotted: ZS, solid: NEW2

3.3.2 Quantile estimation

In practice, it is of interest to assess quantiles of the GPD. Thus we proceed to analyse performances of different methods for quantile estimations using the same simulation data and results as in Section 3.3.1. We also notice that, in previous literature it is often assumed that, in general, parameter estimations and quantile estimations based on the same approach should perform consistently. However through simulation tests we show that this statement is arguably vague. For all tested approaches, performance of quantile estimations are poor for extreme upper quantiles of heavy-tailed GPD. In order to compare the results under various quantile level ($p = 0.5, 0.75, 0.9$) and shape parameter ($\xi = \{-1, -0.5, -0.25, 0.0.25, 0.5, 0.75, 1\}$), percentage-bias and percentage-RMSE for different methods are given in Table 3.3 to 3.5. The percentage-bias and percentage-RMSE are defined as the ratio of estimation

Figure 3.2: Bias and RMSE of the **shape estimator** $\hat{\xi}$ plotted against ξ (from -1 to 1).



dashed: EPM, dot-and-dash: LME, dotted: ZS, solid: NEW2

bias and RMSE over the theoretical quantile value, respectively:

$$\text{Percentage-Bias} = \frac{\text{Bias}[\hat{Q}(p;\hat{\sigma},\hat{\xi})]}{Q(p;\sigma,\xi)},$$
$$\text{Percentage-RMSE} = \frac{\text{RMSE}[\hat{Q}(p;\hat{\sigma},\hat{\xi})]}{Q(p;\sigma,\xi)}.$$

We also plot the percentage-bias and percentage-RMSE of $0.5, 0.75$ and 0.9 quantile for EPM, LME, ZS and NEW2 methods against the value of ξ in Figure 3.3 ($n = 30$). Figures for 95% quantile and other sample sizes are not included since similar conclusions can be drawn.

The simulation tests above show that the results are more sophisticated when it comes to quantile estimates. Nonetheless, tested estimators share some of the properties as they have displayed in parameter estimations. Firstly, the percentage-bias and percentage-RMSE for tested estimators increases when the value of ξ increases. Secondly, MOM and PWM estimators have large percentage-bias and percentage-

RMSE for large value of ξ such as 0.75 and 1. Thirdly, in general, bias and RMSE of all estimators increase when sample size decreases.

On the other hand, the properties in quantile estimation that are different from parameter estimation are summarised as follows:

1. For the three tested quantiles, all tested methods tend to overestimate for extreme heavy-tailed GPD, especially when $\xi \to 1$, despite different patterns of bias shown in parameter estimation.

2. As shown in Figure 3.3, generally the EPM estimator has significantly larger percentage-bias and percentage-RMSE compared to others, especially for $p = 0.75$ and 0.9. For some cases it is even worse than MOM and PWM estimates. Unfortunately, we are unable to compare our results with Castillo and Hadi (1997) since there is no similar data presented.

3. For 0.5 and 0.75 quantiles, he proposed NEW2 estimators generally have the best performance, in terms of consistently small bias and RMSE. However, its estimating bias and error are obviously higher than LME and ZS methods as $\xi \to 1$ under the 0.9 quantile case. On the other hand, performance of LME and ZS' 0.9 quantile estimators also dropped significantly, as their percentage-bias and percentage-RMSE almost doubled from the 0.75 quantile cases.

Table 3.3: Percentage-bias and Percentage-RMSE of estimating $Q(0.5; 1, \xi)$

		Percentage-bias						Percentage-RMSE					
n	ξ	MOM	PWM	EPM	LME	ZS	NEW2	MOM	PWM	EPM	LME	ZS	NEW2
50	-1	-0.001	-0.004	0.000	0.000	-0.034	-0.012	0.133	0.132	0.118	0.113	0.107	0.121
	-0.5	0.004	0.001	0.003	0.010	-0.027	-0.005	0.151	0.155	0.147	0.139	0.138	0.145
	-0.25	0.014	0.011	0.010	0.008	-0.018	-0.011	0.157	0.165	0.161	0.156	0.151	0.161
	0	0.025	0.013	0.009	0.016	-0.016	0.001	0.173	0.182	0.182	0.177	0.171	0.177
	0.25	0.078	0.017	0.007	0.027	-0.009	-0.002	0.209	0.196	0.200	0.198	0.191	0.192
	0.5	0.303	0.044	0.014	0.031	0.007	0.004	0.764	0.220	0.219	0.218	0.211	0.216
	0.75	1.020	0.107	0.022	0.038	0.023	0.010	3.771	0.303	0.246	0.241	0.240	0.234
	1	3.720	0.294	0.034	0.046	0.043	0.019	21.219	0.852	0.264	0.269	0.262	0.259
30	-1	0.007	0.003	0.009	0.004	-0.045	-0.011	0.171	0.170	0.153	0.146	0.140	0.154
	-0.5	0.017	0.013	0.018	0.014	-0.033	-0.007	0.192	0.196	0.187	0.183	0.171	0.186
	-0.25	0.018	0.012	0.015	0.024	-0.032	-0.010	0.205	0.213	0.211	0.207	0.196	0.208
	0	0.038	0.020	0.019	0.034	-0.021	-0.001	0.224	0.234	0.237	0.233	0.220	0.231
	0.25	0.109	0.035	0.025	0.038	-0.006	0.002	0.280	0.257	0.264	0.260	0.247	0.255
	0.5	0.335	0.066	0.032	0.049	0.012	0.014	0.785	0.297	0.295	0.287	0.281	0.279
	0.75	1.146	0.162	0.049	0.062	0.040	0.013	8.062	0.586	0.326	0.320	0.311	0.305
	1	4.823	0.452	0.055	0.076	0.056	0.026	53.844	3.069	0.367	0.358	0.359	0.343
15	-1	0.001	-0.007	0.012	0.018	-0.084	-0.026	0.230	0.230	0.215	0.206	0.204	0.219
	-0.5	0.022	0.013	0.032	0.032	-0.066	-0.025	0.275	0.278	0.274	0.266	0.246	0.264
	-0.25	0.029	0.016	0.034	0.050	-0.059	-0.017	0.290	0.299	0.304	0.301	0.271	0.289
	0	0.070	0.041	0.056	0.063	-0.030	-0.007	0.327	0.336	0.349	0.345	0.306	0.327
	0.25	0.161	0.068	0.069	0.084	-0.005	0.007	0.411	0.371	0.387	0.394	0.345	0.362
	0.5	0.385	0.114	0.085	0.104	0.019	0.003	1.014	0.443	0.440	0.440	0.395	0.392
	0.75	1.097	0.246	0.108	0.127	0.056	0.023	4.146	0.741	0.515	0.499	0.466	0.445
	1	4.008	0.685	0.157	0.154	0.114	0.035	37.769	4.311	0.601	0.574	0.549	0.490

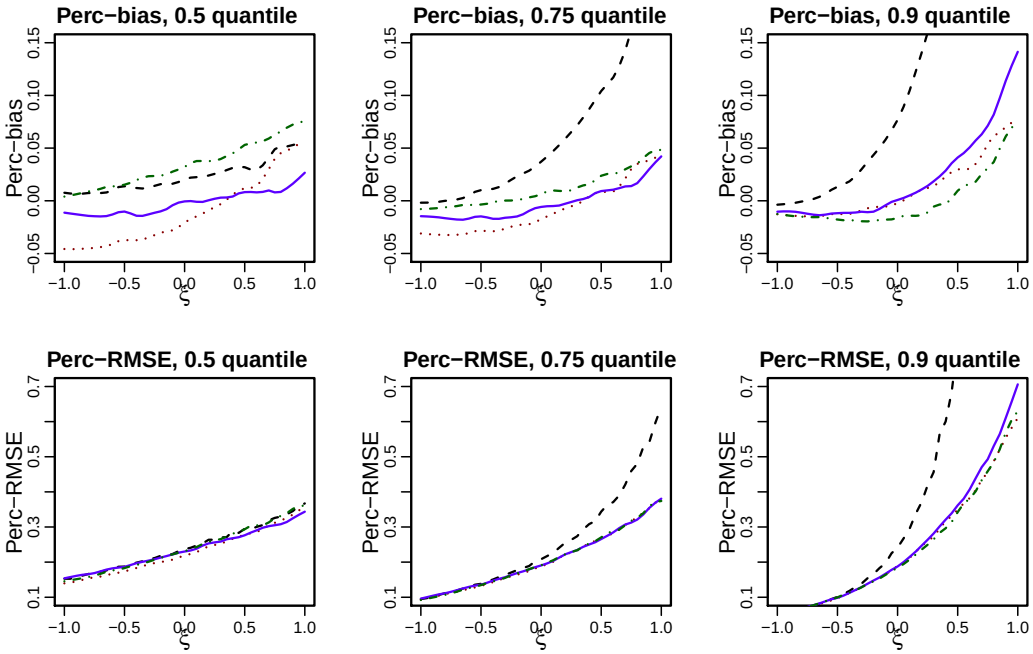
Table 3.4: Percentage-bias and Percentage-RMSE of estimating $Q(0.75; 1, \xi)$

		Percentage-bias						Percentage-RMSE					
n	ξ	MOM	PWM	EPM	LME	ZS	NEW2	MOM	PWM	EPM	LME	ZS	NEW2
50	-1	-0.009	-0.008	-0.004	-0.005	-0.023	-0.011	0.073	0.074	0.076	0.072	0.072	0.076
	-0.5	-0.005	-0.006	0.003	-0.001	-0.020	-0.006	0.107	0.108	0.111	0.104	0.106	0.107
	-0.25	0.003	0.000	0.015	-0.003	-0.013	-0.011	0.125	0.126	0.132	0.124	0.123	0.127
	0	0.007	0.000	0.026	0.003	-0.011	-0.002	0.150	0.152	0.164	0.150	0.149	0.147
	0.25	0.038	-0.001	0.041	0.011	-0.007	-0.005	0.190	0.176	0.200	0.177	0.174	0.172
	0.5	0.202	0.008	0.073	0.014	0.005	0.009	0.731	0.207	0.256	0.202	0.204	0.208
	0.75	0.744	0.033	0.112	0.021	0.015	0.016	3.345	0.289	0.340	0.238	0.241	0.241
	1	2.730	0.149	0.167	0.028	0.034	0.026	16.967	0.811	0.441	0.283	0.281	0.282
30	-1	-0.010	-0.008	-0.001	-0.008	-0.031	-0.015	0.093	0.094	0.095	0.093	0.093	0.096
	-0.5	-0.001	-0.003	0.013	-0.004	-0.025	-0.012	0.133	0.134	0.138	0.135	0.131	0.136
	-0.25	0.001	-0.004	0.022	0.000	-0.023	-0.014	0.161	0.163	0.170	0.163	0.159	0.163
	0	0.008	-0.003	0.035	0.008	-0.019	-0.006	0.190	0.193	0.208	0.192	0.188	0.191
	0.25	0.054	0.004	0.066	0.011	-0.005	0.000	0.253	0.229	0.263	0.227	0.227	0.229
	0.5	0.215	0.016	0.103	0.023	0.009	0.014	0.738	0.277	0.344	0.267	0.271	0.268
	0.75	0.837	0.069	0.162	0.034	0.031	0.015	7.188	0.606	0.459	0.316	0.317	0.314
	1	3.588	0.277	0.226	0.047	0.043	0.041	43.018	3.019	0.643	0.371	0.386	0.380
15	-1	-0.023	-0.019	-0.005	-0.013	-0.056	-0.033	0.130	0.134	0.129	0.128	0.136	0.138
	-0.5	-0.014	-0.017	0.017	-0.010	-0.052	-0.034	0.187	0.190	0.192	0.190	0.186	0.188
	-0.25	-0.007	-0.015	0.036	-0.002	-0.045	-0.027	0.223	0.227	0.236	0.226	0.220	0.221
	0	0.016	-0.003	0.071	0.007	-0.028	-0.016	0.268	0.269	0.299	0.275	0.261	0.267
	0.25	0.072	0.009	0.115	0.026	-0.009	-0.002	0.361	0.320	0.382	0.332	0.315	0.319
	0.5	0.223	0.025	0.168	0.047	0.006	0.004	0.944	0.402	0.515	0.401	0.382	0.373
	0.75	0.753	0.111	0.275	0.059	0.043	0.030	3.633	0.720	0.761	0.468	0.487	0.465
	1	2.881	0.445	0.433	0.098	0.095	0.067	29.957	4.182	1.172	0.586	0.615	0.562

Table 3.5: Percentage-bias and Percentage-RMSE of estimating $Q(0.9; 1, \xi)$

		Percentage-bias						Percentage-RMSE					
n	ξ	MOM	PWM	EPM	LME	ZS	NEW2	MOM	PWM	EPM	LME	ZS	NEW2
50	-1	-0.006	-0.003	-0.004	-0.007	-0.010	-0.006	0.053	0.052	0.040	0.039	0.040	0.039
	-0.5	-0.008	-0.007	0.007	-0.010	-0.010	-0.004	0.081	0.084	0.081	0.076	0.077	0.076
	-0.25	-0.006	-0.006	0.025	-0.013	-0.003	-0.005	0.107	0.109	0.118	0.105	0.105	0.106
	0	-0.011	-0.011	0.058	-0.007	0.000	0.002	0.144	0.144	0.187	0.144	0.146	0.143
	0.25	-0.009	-0.018	0.111	-0.001	0.005	0.003	0.198	0.188	0.307	0.196	0.197	0.196
	0.5	0.075	-0.031	0.220	0.005	0.018	0.033	0.711	0.241	0.600	0.253	0.261	0.278
	0.75	0.408	-0.057	0.379	0.022	0.025	0.052	2.809	0.334	1.069	0.337	0.337	0.360
	1	1.621	-0.033	0.626	0.038	0.055	0.077	12.110	0.800	1.856	0.447	0.444	0.465
30	-1	-0.009	-0.003	-0.004	-0.013	-0.013	-0.010	0.070	0.069	0.049	0.052	0.051	0.051
	-0.5	-0.011	-0.009	0.015	-0.018	-0.010	-0.010	0.103	0.106	0.101	0.100	0.097	0.099
	-0.25	-0.012	-0.012	0.040	-0.021	-0.005	-0.009	0.138	0.141	0.154	0.139	0.136	0.138
	0	-0.022	-0.023	0.074	-0.014	-0.005	-0.001	0.184	0.184	0.241	0.184	0.188	0.188
	0.25	-0.010	-0.024	0.163	-0.009	0.013	0.017	0.259	0.243	0.419	0.249	0.262	0.266
	0.5	0.067	-0.040	0.300	0.014	0.029	0.043	0.708	0.311	0.823	0.346	0.350	0.363
	0.75	0.465	-0.041	0.553	0.034	0.054	0.067	6.066	0.695	1.918	0.462	0.458	0.498
	1	2.210	0.066	1.019	0.071	0.078	0.137	30.676	2.963	4.642	0.628	0.609	0.702
15	-1	-0.016	-0.005	-0.005	-0.027	-0.017	-0.022	0.102	0.100	0.069	0.081	0.076	0.081
	-0.5	-0.030	-0.025	0.021	-0.040	-0.020	-0.023	0.152	0.155	0.146	0.146	0.144	0.142
	-0.25	-0.031	-0.030	0.064	-0.044	-0.010	-0.016	0.197	0.200	0.227	0.195	0.199	0.196
	0	-0.035	-0.036	0.136	-0.037	0.001	0.004	0.257	0.257	0.379	0.264	0.273	0.276
	0.25	-0.024	-0.044	0.272	-0.013	0.025	0.032	0.360	0.332	0.686	0.368	0.385	0.395
	0.5	0.037	-0.067	0.499	0.021	0.039	0.072	0.897	0.441	1.619	0.512	0.516	0.549
	0.75	0.357	-0.038	1.067	0.045	0.106	0.154	3.004	0.777	3.848	0.685	0.777	0.896
	1	1.657	0.172	2.388	0.160	0.192	0.301	21.138	4.019	14.784	1.141	1.191	1.328

Figure 3.3: Percentage-bias and percentage-RMSE of estimating $Q(p; 1, \xi)$, plotted against ξ, for different value of $p = 0.5, 0.75, 0.9$, respectively; sample size $n = 30$.



dashed: EPM, dot-and-dash: LME, dotted: ZS, solid: NEW2

To further explore the performance losses when estimating higher upper quantiles of the GPD, we consider following subsequent tests.

Hosking and Wallis (1987) show that, by Taylor expansion, $Q(p) = \sigma p(1 + \frac{1}{2}(1 + \xi)p + O(p^2))$, as $p \to 0$, thus the accuracy of $\hat{Q}(p)$ for small p is mainly affected by the accuracy of $\hat{\sigma}$. For large p, the mechanic is rather unclear. Hence, we proceed by examining upper quantile estimations for heavy-tailed GPD with EPM, LME, ZS and NEW2 methods. Models are tested for various values of p, ranging from 0.5 to 0.95 by 0.05, given sample size $n = 30$, and several fixed values of $\xi = \{0.25, 0.5, 0.75, 1\}$. The percentage-bias and percentage-RMSE are plotted in Figure 3.4.

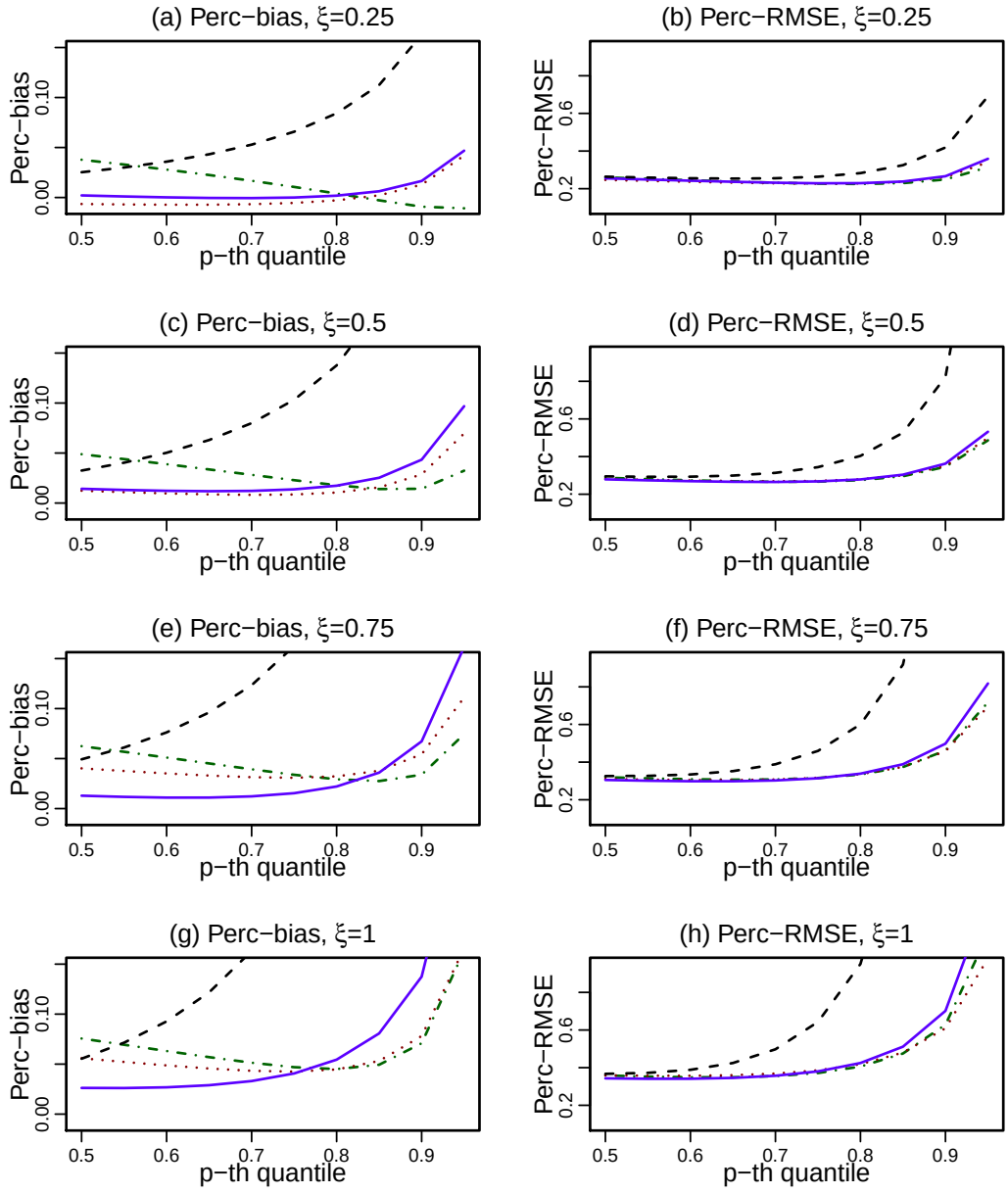
We would like to draw attentions to following facts from those figures:

1. The slope of estimation bias and error curves of all tested methods increases steeply when p value is close to 1. This implies performance of higher upper quantile estimates is dominated by the largest data points in the sample.

2. The performance loss of the NEW2 method is heavier than LME and ZS when estimating extreme quantiles of the GPD.

3. However, when estimating less extreme quantiles (as $p \leq 0.8$), the NEW2

Figure 3.4: Percentage-bias and percentage-RMSE of estimating $Q(p; 1, \xi)$, plotted against p ($0.5 \leq p \leq 0.95$), for heavy-tailed GPD with different values of $\xi = 0.25, 0.5, 0.75, 1$, respectively; sample size $n = 30$.



dashed: EPM, dot-and-dash: LME, dotted: ZS, solid: NEW2

estimator generally outperform LME and ZS estimators, as the percentage-bias and -RMSE are small and consistent.

Through the simulation tests we should note that, although the GPD is often used with the POT method to assess extreme quantiles of the original distribution and it

has been proved efficient, estimating extreme quantiles of the GPD *per se* should still be avoided when possible. Recall that $p = (\frac{N}{n}p^*)$, where p^* is the extreme probability or quantile level of the original distribution F_Y that need to be assessed, and N is the sample size of the original data. p^* and N are often fixed with given information, but value of p can still be adjusted, by changing the threshold value u (thus changing n), to be more 'optimal'. This is essentially the 'bias-variance-tradeoff' effect as in the threshold selection problem: when p^* and N are fixed, larger n leads to higher p, thus larger overall bias; smaller n leads to lower p, but larger estimation variance. Unfortunately, such problem is not the main scope of this chapter and should be explored in future. Nonetheless, it can be concluded that, choosing a threshold that leads to a moderate p value will benefit the overall performance of quantile estimation of the GPD, and the NEW2 method we proposed has displayed promising properties under this circumstance.

3.3.3 Confidence interval estimation

In practice, knowing the confidence interval of the estimators will also provide valuable information in modelling GPD. In this section we compare the performance of different confidence interval methods, including: profile log-likelihood method based on MLE estimates (Profile), asymptotic CI based on observed Fisher information of PWM and LME methods, respectively, and the generalised CI of NEW2 method. We also consider two bootstrap methods, the percentile bootstrap confidence intervals and the bias-corrected and accelerated confidence intervals. Details of both bootstrap methods can be found in Efron and Tibshirani (1986), and see Tajvidi (2003) for a more spesific study of bootstrap CI for heavy-tailed GPD.

We mainly focus on CI of quantile estimation of the GPD, which include 90% and 95% CI of $Q(p)$, where $p = \{0.75, 0.9\}$, respectively. CI for $\hat{Q}(0.5)$ is not reported because similar conclusions can be drawn from the results. Random samples of differenct sizes ($n = 50, 30$) were generated from $\sigma = 1$ and different values of shape parameter $\xi = \{-0.25, 0.25, 0.5, 0.75\}$, with 1000 replicates. For the NEW2 estimates, $\Theta_{(\tau)}$ was calculated based on 2000 Monte Carlo random samples; the bootstrap methods were calculated based on 1000 bootstrap samples. For each method, the average interval length and coverage probability was summarised in Table 3.6 and 3.7. It is worth to mention that CI estimations for sample size $n = 15$ are not reported, due to tested methods are all failed to provide reliable results in terms of coverage probability or average interval length.

Same rule applies for assessing interval estimations as in point estimations of the

GPD, namely, the CI should be consistently reliable for different tail behaviours (different ξ values). Also, ideally a good CI estimation should have coverage probability (CP) close to the nominal $1-\tau$ confidence level, but as little above $1-\tau$ as possible (Newcombe, 1998), combining with a reasonable interval length. In general, average length of CI of all tested method increase as sample size decreases.

Table 3.6 presents results for the less-extreme 75% quantile of the GPD. Overall, the generalised CIs provided by NEW2 method have the best performance in combination of CP and average length. The CP based on PWM estimates is always lower than the nominal level, decreasing as value of ξ increases. On the contrary, The CP based on LME estimates is obviously much higher than the nominal level, also decreasing as value of ξ increases. CIs based on profile log-likelihood and NEW2 method have CP close to the nominal level for different values of ξ, but the NEW2 estimates are slightly better. Furthermore, if the 'as little above $1-\tau$ as possible' rule is applied, the NEW2 method is preferred.

For $\hat{Q}(0.75)$, bootstrap methods (PERC and BC_a) have CP close to the nominal level in most situation, however, their interval length are very misleading when the GPD is extremely heavy-tailed, such as $\xi = 0.5, 0.75$. Extremely large average interval length ($> 1 \times 10^5$) have been observed for both bootstrap methods, as denoted with '***' signs in tables. For small sample sizes and extremely heavy-tailed distributions, estimations from naive bootstrap methods are dominated by the largest sample point (or several extremely large sample points), and may not converge. For more detailed analysis of naive bootstrap may fail in heavy-tailed cases, please refer to Athreya (1987) and Hall (1990). Such extremely wide confidence intervals have basically no practical uses.

For higher quantiles $\hat{Q}(0.9)$, CP of profile log-likelihood and PWM CIs drop significantly as the tail of the GPD goes heavier. The interval based on LME is obviously overestimated, with extreme cases that CP=1 when $\xi = -0.25$. The NEW2 method still has CP close to the nominal level within ± 0.03 differences for all cases. The bootstrap methods have the same convergence problem for heavy-tailed GPD, as shown in previous case, and their CPs are significantly below the nominal level. One the other hand, the average interval lengths of the NEW2 method are also large for extremely heavy-tailed situations (such as for $\xi = 0.75$ the average length of NEW2 CI is 26.35, while the true value of $Q(0.9; 1, 0.75) \approx 6.165$), but still much less extreme than bootstrap methods, and with coverage probability close to the nominal level. We should mention that, under extremely heay-tailed situations, for $n = 15$ (which is not reported here), the interval length problem of NEW2 CI estimations is more serious, which could render the estimation results useless. However, all other tested

results have either very poor CP or very poor interval length as well. Nonetheless, for $n = 50$ and 30, the confidence interval estimations based on the NEW2 method obviously outperform other tested methods, and it is contently reliable for different tail behaviours.

Table 3.6: Average length and coverage probability (in parentheses) of confidence interval of estimating $Q(0.75; 1, \xi)$

		90% CI						95% CI					
n	ξ	Profile	PWM	LME	NEW2	PERC	BC_a	Profile	PWM	LME	NEW2	PERC	BC_a
50	-0.25	0.474	0.482	1.066	0.505	0.521	0.527	0.566	0.574	1.263	0.603	0.623	0.631
		(0.893)	(0.897)	(0.997)	(0.904)	(0.886)	(0.894)	(0.922)	(0.926)	(0.997)	(0.95)	(0.947)	(0.951)
	0.25	0.961	0.926	1.703	1.016	0.963	0.982	1.170	1.110	2.038	1.235	1.165	1.193
		(0.881)	(0.868)	(0.989)	(0.89)	(0.894)	(0.891)	(0.937)	(0.924)	(0.996)	(0.947)	(0.945)	(0.953)
	0.5	1.373	1.212	2.228	1.451	1.371	1.414	1.712	1.457	2.700	1.804	***	***
		(0.883)	(0.859)	(0.98)	(0.897)	(0.896)	(0.894)	(0.93)	(0.905)	(0.986)	(0.938)	(0.943)	(0.948)
	1	2.030	1.644	3.054	2.154	***	***	2.490	1.864	3.621	2.633	***	***
		(0.885)	(0.788)	(0.97)	(0.896)	(0.908)	(0.901)	(0.942)	(0.869)	(0.987)	(0.943)	(0.929)	(0.931)
30	-0.25	0.612	0.611	1.365	0.660	0.697	0.702	0.730	0.729	1.632	0.793	0.836	0.844
		(0.868)	(0.874)	(0.998)	(0.913)	(0.909)	(0.897)	(0.935)	(0.933)	(0.999)	(0.957)	(0.948)	(0.943)
	0.25	1.245	1.169	2.184	1.371	1.303	1.348	1.569	1.410	2.644	1.732	1.576	1.640
		(0.889)	(0.884)	(0.984)	(0.918)	(0.898)	(0.896)	(0.933)	(0.919)	(0.99)	(0.948)	(0.949)	(0.950)
	0.5	1.849	1.538	2.901	2.031	1.816	1.908	2.292	1.756	3.405	2.531	2.171	2.310
		(0.900)	(0.861)	(0.980)	(0.917)	(0.890)	(0.886)	(0.947)	(0.885)	(0.986)	(0.958)	(0.941)	(0.932)
	0.75	2.796	1.842	3.989	3.107	***	***	3.400	2.292	4.718	3.911	***	***
		(0.901)	(0.774)	(0.967)	(0.908)	(0.86)	(0.853)	(0.953)	(0.836)	(0.982)	(0.955)	(0.935)	(0.930)

*** *denotes value larger than* 1×10^5

Table 3.7: Average length and coverage probability (in parentheses) of confidence intervals of estimating $Q(0.9; 1, \xi)$

		90% CI						95% CI					
n	ξ	Profile	PWM	LME	NEW2	PERC	BC_a	Profile	PWM	LME	NEW2	PERC	BC_a
50	-0.25	0.594 (0.864)	0.616 (0.866)	1.920 (1.000)	0.713 (0.897)	0.572 (0.858)	0.582 (0.878)	0.723 (0.913)	0.736 (0.913)	2.283 (1.000)	0.897 (0.952)	0.684 (0.916)	0.699 (0.923)
	0.25	2.243 (0.876)	1.976 (0.845)	3.869 (0.994)	2.539 (0.898)	1.841 (0.84)	2.011 (0.843)	2.810 (0.936)	2.368 (0.897)	4.567 (0.998)	3.174 (0.953)	2.190 (0.885)	2.466 (0.899)
	0.5	3.643 (0.902)	3.413 (0.787)	5.881 (0.973)	4.700 (0.877)	3.601 (0.846)	4.202 (0.863)	4.416 (0.938)	4.377 (0.849)	7.153 (0.986)	6.431 (0.942)	4.121 (0.886)	5.027 (0.917)
	0.75	4.587 (0.847)	4.898 (0.613)	9.720 (0.947)	9.525 (0.881)	*** (0.842)	*** (0.849)	5.011 (0.833)	6.011 (0.663)	11.350 (0.957)	12.355 (0.943)	*** (0.882)	*** (0.901)
30	-0.25	0.764 (0.849)	0.780 (0.842)	2.430 (0.998)	1.031 (0.900)	0.708 (0.823)	0.718 (0.841)	0.989 (0.905)	0.937 (0.897)	2.927 (1.000)	1.389 (0.946)	0.855 (0.881)	0.872 (0.889)
	0.25	3.011 (0.874)	2.489 (0.821)	5.003 (0.982)	3.950 (0.892)	2.330 (0.81)	2.642 (0.827)	3.662 (0.93)	2.917 (0.868)	5.848 (0.991)	5.271 (0.944)	2.701 (0.833)	3.140 (0.865)
	0.5	4.151 (0.862)	4.271 (0.743)	7.792 (0.957)	8.143 (0.903)	24.477 (0.82)	2352.063 (0.838)	4.730 (0.911)	4.706 (0.820)	9.244 (0.972)	11.586 (0.956)	5.660 (0.862)	7.650 (0.876)
	0.75	4.725 (0.710)	5.873 (0.610)	13.305 (0.919)	18.083 (0.869)	84.915 (0.772)	*** (0.811)	5.328 (0.784)	7.032 (0.700)	15.706 (0.954)	26.350 (0.947)	16.986 (0.851)	93.397 (0.88)

*** *denotes value larger than* 1×10^5

3.4 Example

We assess a dataset that is given by Castillo and Hadi (1997), Table 3, and has also been examined by Zhang and Stephens (2009). The dataset is zero-crossing hourly mean periods (in seconds) of the sea waves measured in a Bilbao buoy, in January 1997. Previous studies on this dataset suggested that, the fitted GPD tend to have very negative shape parameters, with estimated $\hat{\xi}$ range between $(-0.6, -1.8)$, depending on the method and threshold selected. Zhang and Stephens have already carried out a detailed analysis about this dataset, including estimated parameters, Q-Q plots and goodness-of-fit test, with consideration of different thresholds ranging from 7.0 to 9.5. Since small sample size is the major concern in this chapter, we only compare different methods with threshold $u = 9.0$ and 9.5, which have 41 and 17 samples, respectively. Parameter estimates of the NEW2 method for this data is:

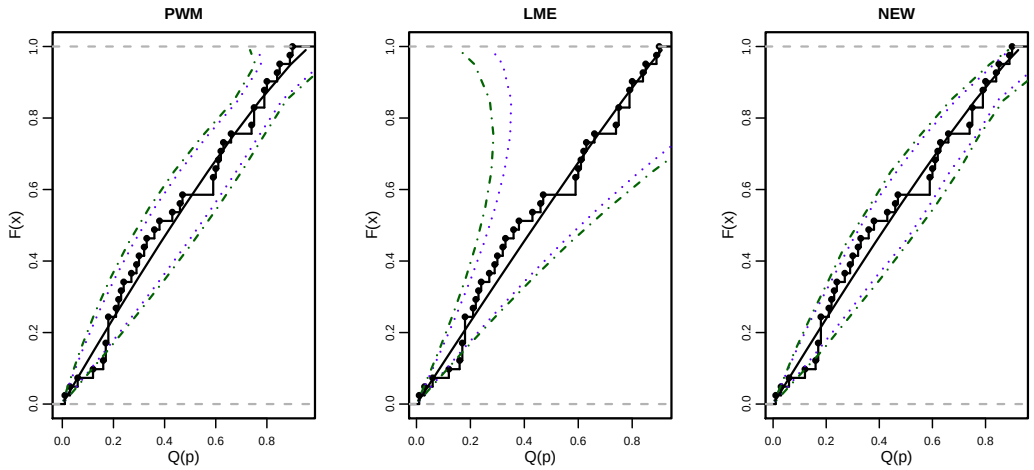
$$\begin{aligned} u = 9.0: &\quad \hat{\sigma} \approx 0.824, \quad \hat{\xi} \approx -0.877; \\ u = 9.5: &\quad \hat{\sigma} \approx 0.502, \quad \hat{\xi} \approx -1.243, \end{aligned}$$

which are close to results of ZS estimates, as shown in Zhang and Stephens, Table 3.

We also calculated the 90% and 95% confidence band of quantile estimations using threshold $u = 9.0$ and 9.5 with different methods. However, for both threshold values the MLE method does not converge, so the profile log-likelihood CI cannot be obtained. Results from the asymptotic method based on PWM estimates and LME estimates, and from the NEW2 inference for the two thresholds are plotted in Figure 3.5 and 3.6, respectively. The theoretical and empirical distribution function are also plotted accordingly.

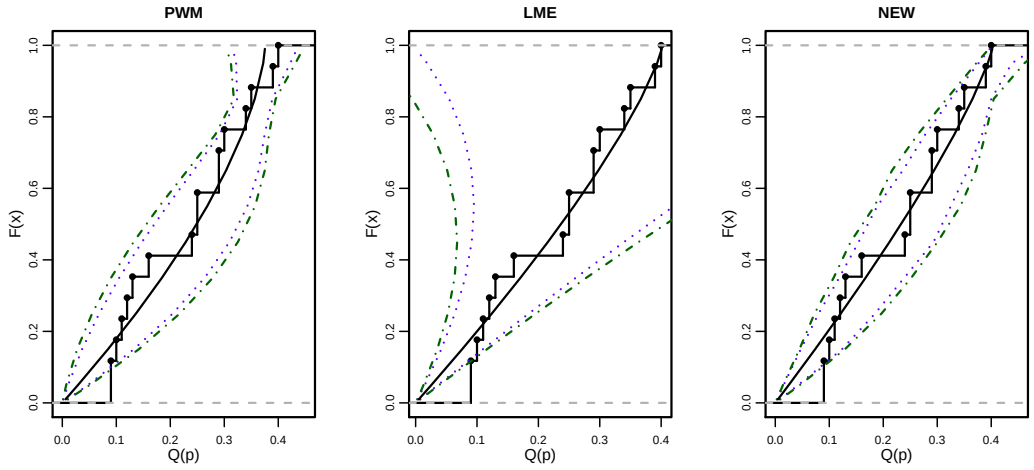
It is shown in the figures that, for the three methods (PWM, LME and NEW2), the model fitting is adequate, except that the PWM method has slightly worse fitting on the upper tail. The lines of CI estimation suggest that the asymptotic approach based on LME estimates has too wide intervals towards the upper tail. It can be observed on Figure 3.6 that, for $u = 9.5$, the lower bound of the 95% CI for $\hat{Q}(p)$ based on LME goes to negative value as p increases, which is theoretically invalid, as $Q(p)$ of the GPD is strictly larger than 0 for $p > 0$. This, however, does illustrate another drawback of the asymptotic CI approach which has not been shown in our simulation tests since we did not test for ξ smaller than -0.25: the CI estimation based on asymptotic normality assumptions may have its bound go beyond theoretical domain of the estimator. The confidence bands for the asymptotic method based on PWM and the exact inference of NEW2 have no significant difference, since for very

Figure 3.5: Confidence intervals of quantiles estimated by different methods plotted against different quantile level p, with threshold $u = 9.0$.



Connected-dots: empirical cumulative distribution; solid: estimated quantiles; inner-dash-line: 90% *CI; outer-dash-line:* 95% *CI*

Figure 3.6: Confidence intervals of quantiles estimated by different methods plotted against different quantile level p, with threshold $u = 9.5$.



Connected-dots: empirical cumulative distribution; solid: estimated quantiles; inner-dash-line: 90% *CI; outer-dash-line:* 95% *CI*

negative values of ξ both methods should have accurate quantile estimations.

3.5 Chapter Summary

In this chapter, we introduce a new method for point and interval estimations of the GPD, specifically focusing on small sample cases. The numerical evidence has shown that, the new method provides consistently reliable point estimates against the varying tail behaviour of the GPD. At the same time it provides consistently accurate confidence interval estimates in terms of coverage probability and with reasonable average interval length, for sample size as small as $n = 30$. However, it should be mentioned that the generalised confidence interval estimations described in Section 3.2.2 may be very computational extensive when the number of iterations of the Monte Carlo simulation algorithm is large.

Furthermore, because of some statistical properties and the tail behaviour of the GPD depend on the shape parameter ξ, some modelling methods may have better performance than others for certain interval of ξ. However, in reality, it is not always possible to choose the most 'appropriate' model, as one cannot always know for sure what range does the value ξ fall into or what tail behaviour does the underlying distribution have. Thus, an estimation method that maintains good (not necessarily the best) performance for most situations is certainly a very useful tool when modelling the GPD. It is true that the new approach we proposed is not uniformly better than any other methods, but it is certainly, for most circumstances that we have tested, the most reliable one.

Chapter 4

Volatility forecast using expected shortfall and expectile

This chapter studies a distribution free approach for forecasting financial volatility, which could be interpreted as, essentially, the variance of financial returns. Classical models of this approach use the interval between two symmetric extreme quantiles of the return distribution as a proxy of volatility. Two new models are proposed, which use intervals of expected shortfalls and expectiles, instead of interval of quantiles. Different models are compared with empirical stock indices data.

4.1 Introduction

In finance, volatility is the term to describe the variation of the return process. It is important because of its widely application in pricing, risk evaluating and decision making. In recent decades, tremendous efforts have been devoted to produce more thorough understanding and better forecasting of volatility. However, as volatility is not directly observable, the true behaviour and mechanism behind it still remained a mystery.

Empirical findings indicate that, in financial return series, volatility often shows a clustering effect. A number of methods have been developed to capture the movement of this time-varying variation. Popular methods such as the ARCH-GARCH (autoregressive conditional heteroskedasticity and generalized autoregressive conditional heteroskedasticity, Bollerslev, 1986, Bollerslev et al., 1992) class of models, stochastic volatility models and option implied volatility models (see, for example, Canina and Figlewski, 1993, Christensen and Prabhala, 1998, among others) often

have distributional assumptions such that the models are associated with certain distributions and its properties are usually fixed over time. This is a potential drawback as if these model specifications are not met, the performance and accuracy of volatility forecast will suffer.

In this chapter we propose a new class of volatility models based on a distribution-free standard deviation model originally introduced by Pearson and Tukey (1965). They showed that, the ratio between a random variable's standard deviation and the interval between its two symmetric extreme quantiles is remarkably fixed, even for random variables from various different distributions. Their model was not originally designed for time-series but some extensions have been made. Taylor (2005) proposes an alternative method that replaces the quantile estimates with the conditional autoregressive value at risk class of models introduced by Engle and Manganelli (2004). They also employ the least square regression framework to replace the fixed value of ratios in Pearson and Tuckey's original model. However, there is a question remains: is one single pair of quantiles enough to capture the behaviour of variation of the data? This was lately followed by Huang (2012), who questioned the sufficiency of information should be contained in one single pair of quantiles, and introduced a new class of models that involves not only one pair of quantiles, but a series of uniformly spaced quantiles that cover the entire distribution. Huang's model is computationally intensive and it is commonly known that quantiles away from the centre of distribution contain more information about variation. In this chapter we propose a new approach that employs not the interval between symmetric quantiles or multiple isolated quantiles, but the interval between symmetric expected shortfalls and expectiles, which are risk measures/statistics focusing on information on the tail of distribution, to forecast volatility. In the next section we briefly review previous research of quantile based volatility forecast, including the developments of Taylor and Huang's models. In section 4.3 and section 4.4 we introduce new approaches based on expected shortfall and expectile. In section 4.5 we show empirical findings of the performance of different models applied to financial stock indices data. The final section provides summaries and conclusion comments.

4.2 Literature review

In this chapter we focus on volatility estimates based on time-series methods. By definition, volatility of a financial asset is the variance of its return distribution, $\sigma_t^2 = \text{var}(r_t)$, where $r_t = ln(S_t/S_{t-1})$ is the log return at time t, and S_t is the price of the underlying asset at time t. In finance, the residual term, ε_t, is defined as

$\varepsilon_t = r_t - \bar{r}$, where $\bar{r}$ is the mean return, is described as an impact of some sort of 'shock' or 'news'. When assuming the average return is constant for a short period m and there is no autocorrelation between successive price 'shocks', the realised volatility (RV) is possibly the best estimation of variance, defined as

$$RV_t = \frac{1}{m-1}\sum_{t=1}^{m}\varepsilon_t^2. \tag{4.1}$$

The RV is especially effective for high-frequency data (see Barndorff-Nielsen, 2002, Andersen et al., 2003), for example, calculating the daily volatility using intra-day returns. In this chapter we use the RV as a benchmark in accessing performance of different volatility forecasting models.

The generalized autoregressive conditional heteroskedasticity (GARCH) class models (see Engle, 1982, Bollerslev, 1986 and Taylor, 2008b) are the most popular statistical volatility forecast models. The general GARCH model interprets volatility as a linear function of lagged time-varying components such as squared residual term and past conditional variance. For example, standard GARCH(1,1) model is showed as follow:

$$\sigma_t^2 = \alpha + \beta_1\varepsilon_t^2 + \beta_2\sigma_{t-1}^2,$$

where α, β_1 and β_2 are parameters. Empirical findings have shown that generally $\beta_1 \approx 1 - \beta_2$. If we force $\beta_1 = 1 - \beta_2$, then the GARCH(1,1) model transforms into the integrated GARCH (iGARCH) model (Nelson, 1990).

Furthermore,the price of financial assets often receive impact of different scale from negative and positive shocks, respectively. Glosten et al. (1993) introduce the gjr-GARCH model to cope with this asymmetric effects. The gjr-GARCH(1,1) model is given by the following expression:

$$\sigma_t^2 = \alpha + (1 - I[\varepsilon_{t-1} > 0])\beta_1\varepsilon_{t-1}^2 + I[\varepsilon_{t-1} > 0]\beta_2\varepsilon_{t-1}^2 + \beta_3\sigma_{t-1}^2,$$

where α, β_1, β_2 and β_3 are parameters. The indicator function $I[\varepsilon_{t-1} > 0]$ represents either a positive (value equals to 1) or a negative (value equals to 0) shock. The gjr-GARCH is good at capturing the so-called leverage effect, which is the empirically observed fact that negative 'shocks' usually have a stronger impact on the volatility.

However, simple volatility models such as exponentially weighted moving average mode sometimes out-perform GARCH models, especially in specific markets (see Tse, 1991 and Kuen and Hoong, 1992 as examples). For a review of comparison among different models, including EWMA, GARCH(1,1) and gjr-GARCH, please

refer to Brailsford and Faff (1996).

In the following two sections, we will focusing on introducing the quantile based model pioneered by Pearson and Tukey (1965) and later extended by Taylor (2005) and Huang (2012).

4.2.1 Volatility estimation using two symmetric extreme quantiles

Pearson and Tukey (1965) proposed a measure of the standard deviation σ of a distribution, which can be expressed directly as a proportion to the difference of the two symmetric extreme quantiles, or in other words, the inter-quantile range, of that distribution:

$$\tilde{\sigma} = \frac{\hat{Q}(1-p) - \hat{Q}(p)}{C(p)} \tag{4.2}$$

where $p \in (0, 1)$ and $\hat{Q}(p)$ is the p-th population quantile, and the value of the denominator $C(p)$ depends on p. They found that for $p = 0.01, 0.025, 0.05$ the corresponding values for $C(p)$ are given by 4.65, 3.92 and 3.25 respectively. These values are actually calculated based on $(\Phi^{-1}(1-p) - \Phi^{-1}(p))$ with some adaptation, where Φ^{-1} is the quantile function of the standard normal distribution. For a normal distribution, the denominator would be 4.653, 3.920 and 3.290, respectively. Obviously, $\sigma = (Q(1-p) - Q(p))/(\Phi^{-1}(1-p) - \Phi^{-1}(p))$ if the random variable is normally distributed. Nonetheless, Pearson and Tukey showed that the estimation in (4.2) holds for a number of different distributions. Thus it provides an approach of estimating standard deviation without making any distributional assumptions.

The original model of Pearson and Tukey (1965) can be improved in several ways. One way is to find the optimum value of the denominator $C(p)$, given different p. For example, Ally (2010) introduces a data-driven version of $C(\cdot)$, where it is estimated using the empirical cumulative distribution function of the standardised data. They indicated that accuracy of this model increases for large samples.

Also, the model in (4.2) can be extended to a regression type of model:

$$\hat{\sigma}^2 = \alpha + \beta \left(\hat{Q}(1-p) - \hat{Q}(p) \right)^2$$

Taylor (2005) has shown that, in many cases, parameters obtained from the quadratic regression above are not significantly different from the values in the original model of Pearson and Tukey (1965), *e.g.*, α is close to zero and β is approximately close to 4.65, 3.92 and 3.25 corresponding to $p = 0.01, 0.025, 0.05$, respectively.

4.2.2 Volatility forecasting based on quantile regression

Taylor (2005) proposes an alternative way of volatility estimation based on Pearson and Tukey's model. They replaced the estimated population quantile $\hat{Q}(p)$ with the value at risk (VaR). The $100p\%$ VaR is defined as the threshold value of loss at probability level p. In this chapter, we simply define VaR as the p-th quantile, $\text{VaR}(p) = Q(p)$. Engle and Manganelli (2004) introduce a type of conditional autoregressive VaR models that uses quantile regression, which also does not require assumptions of the distribution. Specifically, they present the following four CAViaR models:

- Indirect GARCH:

$$Q_t(p) = (\beta_1 + \beta_2 Q_{t-1}^2(p) + \beta_3 r_{t-1}^2)^{1/2},$$

- Adaptive:

$$Q_t(p) = Q_{t-1}(p) + \beta([1 + \exp(G[r_{t-1} - Q_{t-1}(p)])]^{-1} - p),$$

- Symmetric absolute value:

$$Q_t(p) = \beta_1 + \beta_2 Q_{t-1}(p) + \beta_3 |r_{t-1}|,$$

- Asymmetric slope:

$$Q_t(p) = \beta_1 + \beta_2 Q_{t-1}(p) + \beta_3 \max(r_{t-1}, 0) - \beta_4 \min(r_{t-1}, 0),$$

where $Q_t(p)$ is the p-th quantile at time t, βs are regression parameters, and r_t is the excess return at time t. G in the adaptive model is a constant, for example, $G = 10$ as suggested in Engle and Manganelli.

The first model corresponding to the situation where the underlying data is assumed from a GARCH(1,1) model and has an independent and i.i.d. error distribution. The second and third model treat both positive and negative past-return symmetrically, while the fourth model allows different response coefficients on positive and negative past-returns.

The parameters in the CAViaR models are obtained via quantile regression, which is originally introduced by Koenker and Bassett (1978). The population $100p\%$ quantile

of a random variable y can be defined as the parameter m that minimises the function

$$\mathbb{E}[|p - I[y < m]|\,|y - m|], \tag{4.3}$$

where $I[\cdot]$ is indicator function. Straightforward, parameters $\boldsymbol{\beta}$ in CAViaR models can be solved from the following minimisation problem:

$$\min_{\boldsymbol{\beta}} \sum_{t} (p - I[y_t < Q_t(p)])(y_t - Q_t(p)). \tag{4.4}$$

Then the 1-step ahead quantile forecast $\hat{Q}(p)$ can be obtained with CAViaR. Henceforth, the 1-step ahead volatility forecast can be produced by the following least square (LS) regression:

$$\hat{\sigma}^2_{t+1} = \alpha_1 + \beta_1 \left(\hat{Q}_{t+1}(1-p) - \hat{Q}_{t+1}(p)\right)^2. \tag{4.5}$$

For multi-period variance forecast, e.g., a k-period time window starting from $t+1$ to $t+k$, the following LS regression is suggested by Taylor (2005):

$$\hat{\sigma}^2_{t+1,k} = \alpha_k + \beta_k \left(\hat{Q}_{t+1}(1-p) - \hat{Q}_{t+1}(p)\right)^2, \tag{4.6}$$

and the k-period RV is used as a proxy of the true volatility, $\sigma^2_{t+1,k}$.

In the original model of Pearson and Tukey (1965), three different values of p were considered. However, the reason of why these specific values were used was not given. Pearson and Tuckey showed that, among these values, the 90% interval ($p = 0.05$) was most robust to different skewness and kurtosis of the data. But it is questionable that these values are optimal, for example, why not using 91%, 94%, or 99% intervals?

Huang (2012) argues that, instead of using one single pair of extreme quantiles, employing multiple quantiles that are uniformly spaced across the distribution (e.g., from 1% to 99%, by 1%) would reflect not only the tail behaviours but also the whole pattern of the entire distribution. Under the same LS regression framework, they propose a family of models:

$$\hat{\sigma}^2_{t+1} = \alpha_1 + \beta_1 \left(F(\hat{Q}_{t+1}(p))\right)^2, \tag{4.7}$$

where $\hat{Q}_{t+1}(p)$ are also obtained via CAViaR models, but the function $F(\hat{Q}_{t+1}(p))$

can be one of the following:

$$\begin{aligned}
\text{SD}: \qquad & F(\cdot) = \left(\frac{1}{m-1} \sum_{m=1}^{99} (Q(0.01m) - \bar{Q})^2 \right)^{1/2} \\
\text{Weighted SD}: \qquad & F(\cdot) = \left(\sum_{m=1}^{99} w(Q(0.01m) - \bar{Q})^2 \right)^{1/2} \\
\text{Median SD}: \qquad & F(\cdot) = \left(\frac{1}{m-2} \sum_{m=1}^{99} (Q(0.01m) - Q(0.5))^2 \right)^{1/2},
\end{aligned} \tag{4.8}$$

where $\bar{Q}$ is the mean of all these equally spaced quantiles, and $Q(0.5)$ is the central quantile, namely the median. w in the second model is a weight factor, w is set as $p/25$ for $p \leq 0.5$ and $(1-p)/25$ otherwise. Huang (2012) makes comments that for a return distribution, the behaviour of either tails as well as the section between them were driven by different forces, thus the approach considering multiple quantiles that cover the whole distribution would lead to a better performance.

However, it is well-known that quantiles away from the centre of the distribution contain more information about variation. In this chapter we propose new approaches of volatility forecasts that consider information on the tails of distribution in the forms of 'tail expectation'. Details are described in the next section.

4.3 Volatility forecast based on expected shortfall

In the previous section, the methods used for estimating and forecasting population quantiles were actually a risk measure, value at risk (VaR). Hence, it is institutive to think of another risk measure as the proxy of estimating variation: the expected shortfall (ES). The $100p\%$ expected shortfall is defined as the conditional tail expectation that value of the negative log return r_t exceeds $\text{VaR}(p)$

$$\text{ES}_t(p) = \mathbb{E}[r_t | r_t < \text{VaR}_t(p)].$$

Amongst recent research (e.g., Acerbi and Tasche (2002a), Acerbi and Tasche (2002b) and Tasche (2002)), the ES is considered as a better risk measure than the VaR. We considered using ES to replace quantiles in the interval quantile based volatility models for two reasons: first, the ES can be interpreted as the expectation of all more extreme quantiles that exceed probability level p. Thus, generally, the ES explains tail behaviours better than a single pair of quantiles. Secondly, risk measures are essentially measurements of variation, and that makes the ES a potential candidates

for volatility estimation as VaR is.

Note that similar to VaR, usually the ES is only defined on the 'loss' side of the return distribution. To avoid confusion with the conventional definition of ES, we introduce following alternative measure:

$$\upsilon_t(p) = \begin{cases} \mathbb{E}[r_t | r_t < Q_t(p)] & \text{if } p \text{ is close to 0} \\ \mathbb{E}[r_t | r_t > Q_t(p)] & \text{if } p \text{ is close to 1} \end{cases} \qquad 0 < p < 1. \tag{4.9}$$

When p is close to 0, $\upsilon_t(p)$ is identical with the conventional $\mathrm{ES}_t(p)$, represents the conditional expectation of r_t lower than $Q_t(p)$; when p is close to 1, $\upsilon_t(p)$ is defined as a 'mirror' version of ES, represents the conditional expectation of r_t higher than $Q_t(p)$.

Henceforth, the 1-step ahead volatility forecast can be obtained via LS regression:

$$\hat{\sigma}^2_{t+1} = \alpha_1 + \beta_1 \left(\hat{\upsilon}_{t+1}(1-p) - \hat{\upsilon}_{t+1}(p)\right)^2, \qquad p \leq 0.5 \tag{4.10}$$

and corresponding k-period volatility forecast:

$$\hat{\sigma}^2_{t+k} = \alpha_k + \beta_k \left(\hat{\upsilon}_{t+1}(1-p) - \hat{\upsilon}_{t+1}(p)\right)^2, \qquad p \leq 0.5 \tag{4.11}$$

However, estimating the ES is not a trivial work. In this chapter, we use a relatively simple way to estimate the ES (consequently the alternative measure $\upsilon_t(p)$). The expression in (4.9) is equivalent to the following integral form:

$$\upsilon_t(p) = \begin{cases} \dfrac{1}{p}\displaystyle\int_0^p Q_t(\delta)\, d\delta & \text{if } p \text{ is close to 0} \\ \dfrac{1}{1-p}\displaystyle\int_p^1 Q_t(\delta)\, d\delta & \text{if } p \text{ is close to 1,} \end{cases}$$

and this integral form can be approximated via numerical integration methods. For example, one can first produce a finite sets of integration points $\hat{Q}_t(\delta_j)$s with the CAViaR models described in previous sections, where $0 < \delta_1 < \delta_2 < ... < \delta_j < p$. Then the integral approximation is produced as a weighted sum of all integration points:

$$\int_0^p Q_t(\delta)\, d\delta \approx \sum_{i=1}^{j} \omega_i \hat{Q}_t(\delta_i), \tag{4.12}$$

where the value of weight ω_i depends on the algorithm used.

In the empirical study (Sec. 4.5), we also use CAViaR models to calculate $\hat{Q}_t(\tau_i)$ in

expression (4.12). And the midpoint rule was used for the numerical integration.

4.4 Volatility forecast based on expectile regression

In Section 4.2 we mentioned that the $100p\%$ quantile of random variable y was defined as the parameter m that minimises $E\left[|p - I[y < m]||y - m|\right]$ (as in expression 4.3). The expectile, as named by Newey and Powell (1987), is the solution of an asymmetric least squares (ALS) regression with the absolute deviations term $|y - m|$ replaced by squared deviations $(y - m)^2$. Thus the $100\tau\%$ expectile is defined as the parameter m that minimises

$$\mathbb{E}\left[|\tau - I[y < m]|(y - m)^2\right]. \tag{4.13}$$

The expectile regression specification is the same as the quantile regression, namely there is no assumption on the regression's error term except that the error is independent. On the other hand, the expectile regression has several different properties compared with the quantile regression, such as the expression in (4.13) is differentiable on the regression effects, while the quantile regression is not. Also, as for $p = 0.5$ the quantile regression becomes median regression, for $\tau = 0.5$ the expectile regression becomes the mean regression. But what interests us most in the context of estimating variance is that the calculation of population expectile uses more information from the whole distribution than the quantile does (Newey and Powell, 1987). As described by Sobotka and Kneib (2010), the expectile relies 'on the distance of observations from the regression predictor while quantiles only use the information on whether an observation is above or below the predictor'. However, this could be both an advantage and a drawback, as expectiles are then more sensitive to outliers and extreme values but at the same time capture more information with respect to deviation.

Analogue to the minimisation function of regression quantiles in (4.4), Taylor (2008a) introduce the asymmetric least square regression as in (4.14) to estimate the conditional model of expectile $\mu_t(\tau)$.

$$\min_{\mu_t(\tau)} \sum_t |\tau - I[y_t < \mu_t(\tau)]|(y_t - \mu_t(\tau))^2. \tag{4.14}$$

Taylor also introduce a set of conditional autoregressive expecile (CARE) models which are similar to the CAViaR models. For example, the symmetric CARE model is as follow:

$$\mu_t(\tau) = \beta_1 + \beta_2\mu_{t-1}(\tau) + \beta_3|r_{t-1}|,$$

where the regression parameters, βs, are solved via expression (4.14).

One important difference between conditional quantile and conditional expectile is that, the expectiles are determined by tail expectations, while the quantiles are determined by tail probabilities. Nonetheless, for variable y, its population expectile and population quantile have a one-to-one mapping relation, such as $Q(p) = \mu(\tau)$, for some p and τ (for instance, see Efron, 1991). However, it is not necessary that $p = \tau$. As a matter of fact, p and τ are not equal for most cases.

It is intuitive that the expectiles can be used as proxies for volatility estimating. One possible approach is that of using expression (4.5) but instead of using quantiles estimated by CAViaR, using quantiles estimated by CARE, such as $\hat{Q}(p) = \hat{\mu}(\tau)$. However, some major drawbacks of this approach are: firstly, accurate mapping from expectiles to quantiles throughout probability is required in order to produce accurate volatility forecast. Secondly, as symmetric quantiles are employed, for a matching pair of p and τ, there is no guarantee that $Q(1-p) = \mu(1-\tau)$, unless the distribution is symmetric. An alternative approach is that using expectiles directly as proxies for volatility estimation, without matching expectiles with quantiles. For extreme expectiles on lower and upper tails, the expression in (4.13) can be interpreted as minimising the sum of weighted squares of deviation, and the weights depend on whether the observations' value larger or smaller than the predictor m. Thus observations which are more extreme have higher impact on expectiles. By using two symmetric extreme expectiles, one can incorporate more distributional information into volatility estimating than using a single pair of quantiles, without losing focus on the data near the tails of the distribution. Specifically, the following model is employed:

$$\hat{\sigma}^2_{t+1} = \alpha_1 + \beta_1 \left(\hat{\mu}_{t+1}(1-\tau) - \hat{\mu}_{t+1}(\tau)\right)^2, \tag{4.15}$$

and the corresponding k periods forecast is given by

$$\hat{\sigma}^2_{t+1,k} = \alpha_k + \beta_k \left(\hat{\mu}_{t+1}(1-\tau) - \hat{\mu}_{t+1}(\tau)\right)^2.$$

4.5 Empirical study

4.5.1 The data

In this section, financial indices from different markets were used to study the performance of different volatility forecasting models. The empirical study employs daily returns from the following five stock indices: German DAX30, US's Standard & Poor

(S&P500) and NASDAQ, Japan's Nikkei 225 (N225) and UK's FTSE 100 (FTSE). For each index there are 4900 daily returns starting from 2 January 1992. Due to a different number of non-trading days in different countries, the dates of the final data point vary form 25 May 2010 (DAX30) to 06 December 2010 (N225). The first 4000 data points are used as in-sample data for estimating model parameters. The last 900 data points are used as post-sample data for 10- and 20-day period volatility forecasts.

Volatility forecasts are compared from the following models:

- GARCH family: GARCH(1,1) model, integrated GARCH model (iGARCH) and gjr-GARCH model (gjr-GARCH).
- VaR-based quantile interval approach: interval between two symmetric quantiles as described in Section 4.2, where the quantiles are estimated with CAViaR models. Two interval length are used (90% and 98%) which were suggested in Pearson and Tukey (1965) are having better performance. Symmetric absolute value (SAV), asymmetric slope (AS) and adaptive (ADPT) CAViaR models are used. Estimations from the indirect GARCH CAViaR model are omitted since its performance does not stand out in Taylor (2005).
- Multiple-quantile-based approach: as described in Section 4.2, introduced by Huang (2012). Models in this class all employ multiple uniformly spaced quantiles (USQ) ranged from 0.01 to 0.99, by 0.01, which are produced using CAViaR. For each model (SD, MSD and MSD), as in expression (4.8), quantile estimates from SAV, AS and APDT CAViaR are used.
- ES-based approach: the ES are calculated with numerical integration of quantiles as described in Section 4.3. Two different intervals of ES are tested as proxy of variance: the 90% interval $(\upsilon(.95) - \upsilon(.05))$ and the 80% interval $(\upsilon(.9) - \upsilon(.1))$. The ES' numerical integration are also based on CAViaR estimations with SAV, AS and ADPT specifications.
- Expectile-based approach: 98%, 95% and 90% expectile intervals calculated as stated in Sec. 4.4. SAV, AS, ADPT and indirect GARCH (INDI) CARE models are used.

4.5.2 In-sample estimations

For each stock indices, there are 4000 daily returns used as in-sample data. The first 3500 data points of each stock indices are used to estimate parameters in CAViaR

Table 4.1: OLS regression the coefficients of 10- and 20-day period realised volatility on intervals between symmetric interval of $ES(.95) - ES(.05)$ based on asymmetric slope CAViaR model (panel (a)) and symmetric interval of $CARE(.99) - CARE(.01)$ estimated by asymmetric slope CARE model (panel (b)). Standard errors in parentheses.

(a) ES.90.AS

	DAX30	SP500	N225	NASDAQ	FTSE
10-day					
$\alpha_{10} \times 10^5$	2.889	2.876	6.788	1.442	4.628
	(1.475)	(1.421)	(2.567)	(1.587)	(1.407)
β_{10}	46.594	51.601	42.064	48.080	46.752
	(1.592)	(1.331)	(1.831)	(1.356)	(1.702)
20-day					
$\alpha_{20} \times 10^5$	11.299	9.872	21.072	7.697	11.054
	(1.691)	(1.774)	(2.968)	(2.027)	(1.553)
β_{20}	40.894	46.752	32.558	43.015	41.332
	(1.819)	(1.657)	(2.106)	(1.728)	(1.873)

(b) CARE.98.AS

	DAX30	SP500	N225	NASDAQ	FTSE
10-day ahead					
$\alpha_{10} \times 10^5$	0.643	2.694	4.310	0.154	4.023
	(1.552)	(1.366)	(2.567)	(1.465)	(1.453)
β_{10}	0.106	0.098	0.081	0.085	0.094
	(0.004)	(0.002)	(0.003)	(0.002)	(0.004)
20-day ahead					
$\alpha_{20} \times 10^5$	8.138	9.780	19.105	9.154	10.570
	(1.779)	(1.735)	(2.992)	(1.916)	(1.601)
β_{20}	0.093	0.089	0.063	0.076	0.083
	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)

and CARE models, the next 500 data points are used for the least square regressions of realised volatility on interval between symmetric quantiles/ES/expectiles. In Table 4.1 we report coefficients of the LS regression and their standard errors of two models: the 90% ES interval based on numerical integration of asymmetric slope CAViaR quantiles and the 98% expectile interval estimated by asymmetric slope CARE model. These two models are the best performance models in their own classes in terms of R^2, which we will discuss in the next section. Both models have relatively very small intercept coefficients, α_{10} and α_{20}. The coefficients of independent variables, β_{10} and β_{20}, are very different from the Pearson and Tukey (1965) values of corresponding quantile intervals, which is expected, since the values of ES and expectiles usually depart from the value of quantile with the same probability level.

4.5.3 Post-sample estimations

The last 900 data points of each stock index are used as post-sample analysis. Following Taylor (2005) and Huang (2012), the R^2 coefficients of these least squares regressions are reported to see how much explanation power does each model have to the realized volatility. Table 4.2 and 4.3 summarise the results of 10- and 20-day periods volatility for R^2 coefficients, respectively. The R^2 are presented as percentage values, and the higher the value, the better explanatory power the model has. For each index, the highest R^2 value is highlighted with grey background.

For the 10-day periods volatility, Table 4.2 shows that, overall, estimations based on adaptive CAViaR or CARE models have poor performance compared with those based on other types of CAViaR and CARE models. This finding is consistent with Taylor (2005) but contradictory with Huang (2012), as in the later the ADPT models generally out-performed the others. Among the five stock indices, the 98% interval between symmetric expectiles estimated by AS CARE model (CARE.98.AS) has the highest R^2 value for two indices, the U.S. S&P500 ($R^2 = 65.41$) and NASDAQ ($R^2 = 62.45$). For the German DAX30, the method proposed by Taylor (2005) using 98% intervals and estimated by AS CAViaR (VaR.98.AS) has the highest value of $R^2 = 49.94$. Performance of benchmark GARCH family models do not stand out except for the Japanese Nikkie 225 index: the gjr-GARCH has the best measure of 42.57. Finally, for the UK FTSE100, the 90% interval between $ES(.95)$ and $ES(.05)$ that based on numerical integration of AS CAViaR quantiles , with value 45.93. The mean R^2 of each model across five indices were also calculated, for which the CARE.98.AS model has the highest value of $R^2 = 52.09$.

The multi-quantile-based methods, on the other hand, do not stand out for all five indices tested, although their R^2 values are not far behind the best ones. Amongst this class of models, the median SD model as in equation (4.8) with uniformly spaced quantiles estimated by AS CAViaR model (USQ.MSD.AS) has the best measure for three out of five indices, and highest mean R^2 value. To further compare multi-quantile-based methods with ES- and CARE-based methods, emcompassing tests are performed and results are discussed in the next section.

In addition, for ES- or CARE-based models with different interval range, there is no clear evidence that one interval range is better than others, as the preference differ according to the index. For example, the ES.80.AS model has better measure than the ES.90.AS for DAX30 and Nikkie 225, while opposite results are found for the other three indices. On the other hand, it seems the impact of choosing different interval ranges is relatively small, while the impact of the selection of different types

of CAViaR or CARE models are more significant.

Table 4.3 summarises the results for the 20-day periods volatility and it leads to conclusions similar to those in the 10-day periods case, except that for NASDAQ, the VaR.98.AS model has the highest value of R^2 (44.83) over the CARE.98.AS model, which is the best for this index in the previous case. Overall, the CARE.98.AS model still has the best mean measure amongst all tested models. Furthermore, by comparing values from Table 4.2 and Table 4.3, the explanatory power of different models are generally weaker when forecasting longer period of volatility, as the uncertainty in the farther future is more difficult to capture.

To sum up, expectile-based models with asymmetric slope CARE function consistently outperform other models in both 10- and 20-day period volatility forecast. Especially, the CARE.98.AS model with the 98% interval achieves the highest measure in both cases. For other interval models, those with asymmetric slope CAViaR outperform other types of regression models such as symmetric absolute value CAViaR and adaptive CAViaR. Interestingly, for the Nikkie 225 stock index, the gjr-GARCH model has the best performance over other tested models. Estimations based on a single pair of quantiles and expected shortfall also have remarkable performance for specific indices. Multi-quantile-based methods are competitive but do not stand out in any single cases.

4.5.4 J-tests and combining forecasts

We perform the Davidson and MacKinnon (1981)'s J-test as a criteria for comparing one model with another. Specifically, we want to investigate whether the proposed methods using ES intervals or CARE intervals is significantly better than methods using uniformly spaced quantiles. Two pairs of models were selected for comparison: the MSD model of uniformly spaced quantiles estimated by AS CAViaR function (USQ.MSD.AS) against the ES 90% interval model based on AS CAViaR function (ES.90.AS), and the USQ.MSD.AS against the AS CARE 98% interval model (CARE.98.AS). These models have the best performance in their own classes in term of R^2, as shown in previous section. Consider the following two regression models:

$$\begin{aligned} \text{LM1:} \quad & \sigma^2_{RV,t+k} = \alpha_1 + \beta_1 F^2_{\texttt{MSD},t} + \varepsilon_t, \\ \text{LM2:} \quad & \sigma^2_{RV,t+k} = \alpha_2 + \beta_2 F^2_{\texttt{NEW},t} + e_t, \end{aligned}$$

where $\sigma_{RV,t}$ is the realised volatility for k-period, ε_t and u_t are residual terms, $F_{\texttt{MSD}}$ is the regressor of USQ.MSD.AS model as claimed in expression (4.8), and $F_{\texttt{NEW}}$

represents the regressor of the ES interval model or the CARE interval model as specified in expression (4.10) and (4.15), respectively. The idea of the J-test is, if regression model LM1 is adequately fitted, then including fitted values from model LM2 as an additional set of regressor should not provide significant improvement, and *vice versa*. Let $\hat{\sigma}^2_{\texttt{MSD},t+k} = \alpha_1 + \beta_1 F^2_{\texttt{MSD},t}$ and $\hat{\sigma}^2_{\texttt{NEW},t+k} = \alpha_2 + \beta_2 F^2_{\texttt{NEW},t}$ be forecasts generated for the $t+k$ period. Then in the 'plug-in' models:

$$\sigma^2_{RV,t+1,k} = \alpha_1 + \beta_1 F^2_{\texttt{MSD},t} + \lambda_1 \hat{\sigma}^2_{\texttt{NEW},t+k} + \varepsilon_t,$$
$$\sigma^2_{RV,t+1,k} = \alpha_2 + \beta_2 F^2_{\texttt{NEW},t} + \lambda_2 \hat{\sigma}^2_{\texttt{MSD},t+k} + e_t,$$

the coefficients λ_1 and λ_2 are tested under null hypothesises $H_0^1 : \lambda_1 = 0$ and $H_0^2 : \lambda_2 = 0$, respectively.

If H_0^1 cannot be rejected, it implies that we cannot reject $\sigma^2_{RV,t+k}$ being better predicted by $\hat{\sigma}^2_{\texttt{MSD},t+k}$; if H_0^2 cannot be rejected, it implies that we cannot reject $\sigma^2_{RV,t+k}$ being better predicted by $\hat{\sigma}^2_{\texttt{NEW},t+k}$. Note that we also performed the encompassing test as in Section 5.3 of Taylor (2005). However there is no significant evidence showing one model encompasses another. For the J-test, we report the value of λ and corresponding p-values under three different significant levels: $a = 1\%, 5\%, 10\%$, in Table 4.4 and 4.5.

Table 4.4 shows the J-test results for comparing USQ.MSD.AS model against ES.90.AS model. We only consider one model is better than the other if results for λ_1 and λ_2 are significantly different (*e.g.*, one is zero and the other is not, at significant level a). For S&P500 and Nikkie 225 indices, both null hypothesises are rejected at 1% significant level. For NASDAQ index, $H_0^1 : \lambda_1 = 0$ is rejected for both 10- and 20-day period volatility forecast at $a = 1\%$, but $H_0^2 : \lambda_2 = 0$ can only be rejected at 10% significant level for both cases. For the 20-day period forecast for FTSE100, H_0^1 is rejected at $a = 10\%$, while unable to reject H_0^2. On the contrary, for DAX30 index, H_0^2 is rejected at $a = 1\%$ but H_0^1 can only be rejected at $a = 5\%$. To summarise, for five stock indices and two different periods, there are 3 cases indicate that ES.90.AS forecasts is better than USQ.MSD.AS forecasts, while there is only 1 case indicates the opposite.

Similar conclusions can be drawn from Table 4.5, which compares USQ.MSD.AS model against against CARE.98.AS model. There are 3 out of 10 cases (Nikkie 225 for both periods, and DAX30 for 10-day period) indicates that CARE.98.AS model is better, while no significant evidence shows the opposite. Combining the results of j-tests and results of R^2 measure in the previous section, conclusions can be made that, in general, for stock indices included in tests, the multi-quantile-based methods

do not have advantages over the proposed methods using ES intervals and CARE intervals. Overall, the interval between symmetric expectiles estimated by AS CARE model has appealing performance as a proxy for volatility forecasting.

We also consider combining forecasts from multiple models to see whether it improves the performance of volatility forecasting. Two groups of combined forecast models are tested: the VaR.98.AS model together with the ES.90.AS model forecasts, and the VaR.98.AS forecasts together with the CARE.98.AS forecasts. For example, the combined forecasting model of VaR.98.AS & CARE.98.AS is:

$$\hat{\sigma}^2_{comb,t+1,k} = \alpha_k + \beta_{k,1}\left(\hat{Q}(.99)_{t+1} - \hat{Q}(.01)_{t+1}\right)^2 + \beta_{k,2}\left(\hat{\mu}_{t+1}(.99) - \hat{\mu}_{t+1}(.01)\right)^2$$

We compare the R^2 of the LS regression of 10- and 20-day period volatility on combined forecasts for five stock indices, and the highest R^2 measures from individual forecast models as shown in Table 4.2 and 4.3 were also included as benchmarks. Results of the R^2 measure of combined forecasts are summarised in Table 4.6. Intuitively, by combining quantile interval model with ES interval model or expectile interval model, R^2 measure is improved significantly. Overall, the combined VaR.98.AS and CARE.98.AS forecasts have the highest mean measure across five stock indices. Interestingly, the improvement brought by combining forecast is more compelling for the 20-day period volatility. The values of R^2 were improved from 16.1% (S&P500) to 53.8% (Nikkie 225), for different stock indices.

4.6 Chapter summary

Based on works of Pearson and Tukey (1965), Taylor (2005) and Huang (2012), we propose two new volatility forecast methods that using intervals between symmetric expected shortfalls and conditional autoregressive expectiles under LS regression framework. The new methods emphasise on information contained on tails of underlying data, and there is no distributional assumption needed. Empirical studies using five different stock indices suggest that, for 10- and 20-day period realised volatility forecast, proposed new methods, especially the model based on 98% asymmetric slope CARE estimates, have overall better performance than other tested models, including the multiple uniformly spaced quantiles volatility forecast model introduced by Huang (2012). We also find that, combining the ES interval model or the CARE interval model with the CAViaR interval model improves the volatility forecasts, especially for the longer 20-day period volatility forecast, in terms of R^2, which represents the informational content or explanatory power of the LS regression.

We also would like to draw attention to the fact that methods we proposed are not perfect, and there are possible improvements worth look into. For example, both methods, especially models based on expectiles, are sensitive to extraordinary values and outliers. Further study of possible impacts caused by extreme values might be necessary, as future research topics.

Table 4.2: R^2 measure of informational content for 900 post-sample volatility forecasts for 10-day period for different stock indices

	DAX30	S&P500	N225	NASDAQ	FTSE	Mean
GARCH family						
GARCH(1,1)	43.04	54.20	37.78	49.08	41.80	45.18
iGARCH	42.27	52.62	36.37	46.96	41.08	43.86
gjr-GARCH	47.81	60.86	42.57	55.95	42.70	49.98
VaR-based intervals						
VaR.98.SAV	43.98	53.58	32.64	50.07	40.74	44.20
VaR.98.AS	49.94	63.67	37.09	61.37	44.40	51.29
VaR.98.ADPT	2.07	1.29	0.46	0.74	0.68	1.05
VaR.90.SAV	42.52	50.18	28.30	41.41	39.84	40.45
VaR.90.AS	48.94	53.09	37.88	55.23	40.99	47.23
VaR.90.ADPT	2.08	1.32	0.44	0.73	0.68	1.05
Multi-quantile-based methods						
USQ.SD.SAV	42.05	49.23	28.66	42.56	39.67	40.43
USQ.SD.AS	49.18	61.00	37.86	57.75	45.77	50.31
USQ.SD.ADPT	2.06	1.30	0.45	0.73	0.68	1.04
USQ.WSD.SAV	42.19	49.55	28.86	42.88	39.73	40.64
USQ.WSD.AS	49.23	61.13	37.78	57.78	45.72	50.33
USQ.WSD.ADPT	2.06	1.30	0.46	0.73	0.68	1.05
USQ.MSD.SAV	42.03	49.44	28.75	42.58	39.66	40.49
USQ.MSD.AS	49.26	61.05	37.99	57.75	45.80	50.37
USQ.MSD.ADPT	2.04	1.26	0.40	0.21	0.25	0.83
ES-based intervals						
ES.90.SAV	43.07	52.16	30.74	45.99	40.19	42.43
ES.90.AS	49.11	62.87	37.27	58.60	45.93	50.76
ES.90.ADPT	2.08	1.31	0.47	0.74	0.68	1.05
ES.80.SAV	42.77	51.02	29.33	43.00	39.87	41.20
ES.80.AS	49.39	61.18	37.77	57.54	45.37	50.25
ES.80.ADPT	2.06	1.31	0.47	0.71	0.68	1.05
Expectile-based intervals						
CARE.98.SAV	44.09	55.85	32.98	51.40	40.94	45.05
CARE.98.AS	49.22	65.41	39.17	62.45	44.20	52.09
CARE.98.ADPT	14.00	9.45	0.03	8.77	6.33	7.72
CARE.98.INDI	42.90	57.74	41.20	56.97	43.06	48.37
CARE.95.SAV	43.76	54.30	31.84	48.82	40.68	43.88
CARE.95.AS	49.48	62.18	38.23	60.40	44.79	51.02
CARE.95.ADPT	11.82	8.05	0.54	11.27	5.69	7.47
CARE.95.INDI	42.58	56.77	40.18	54.84	42.97	47.47
CARE.90.SAV	43.71	53.25	30.53	46.95	40.64	43.02
CARE.90.AS	49.45	61.13	38.22	59.38	45.03	50.64
CARE.90.ADPT	15.10	2.52	1.46	8.38	8.37	7.16
CARE.90.INDI	42.56	55.89	38.68	54.05	42.95	46.83

Table 4.3: R^2 measure of informational content for 900 post-sample volatility forecasts for 20-day period for different stock indices

	DAX30	S&P500	N225	NASDAQ	FTSE	Mean
GARCH family						
GARCH(1,1)	32.72	42.11	20.50	35.05	34.05	32.89
iGARCH	32.13	40.96	19.69	33.51	33.39	31.94
gjr-GARCH	35.16	48.13	23.27	40.03	32.06	35.73
VaR-based intervals						
VaR.98.SAV	33.59	40.89	18.55	35.28	33.08	32.28
VaR.98.AS	37.12	48.02	21.17	44.83	35.05	37.24
VaR.98.ADPT	2.24	1.51	0.59	0.84	0.96	1.23
VaR.90.SAV	32.36	38.34	16.05	29.37	32.05	29.63
VaR.90.AS	36.31	41.17	21.90	39.12	33.19	34.34
VaR.90.ADPT	2.25	1.55	0.59	0.83	0.97	1.24
Multi-quantile-based methods						
USQ.SD.SAV	32.03	37.72	16.27	30.19	31.96	29.63
USQ.SD.AS	36.54	46.24	21.88	40.72	35.28	36.13
USQ.SD.ADPT	2.23	1.53	0.60	0.83	0.96	1.23
USQ.WSD.SAV	32.13	37.94	16.38	30.41	32.01	29.77
USQ.WSD.AS	36.56	46.32	21.81	40.74	35.28	36.14
USQ.WSD.ADPT	2.22	1.53	0.61	0.83	0.96	1.23
USQ.MSD.SAV	32.02	37.91	16.33	30.21	31.95	29.69
USQ.MSD.AS	36.59	46.28	21.96	40.71	35.30	36.17
USQ.MSD.ADPT	2.02	1.48	0.14	0.29	0.42	0.87
ES-based intervals						
ES.90.SAV	32.76	39.82	17.43	32.43	32.45	30.98
ES.90.AS	36.53	47.57	21.39	41.36	35.67	36.51
ES.90.ADPT	2.24	1.53	0.61	0.84	0.96	1.24
ES.80.SAV	32.52	38.91	16.63	30.45	32.10	30.12
ES.80.AS	36.64	46.34	21.78	40.58	35.16	36.10
ES.80.ADPT	2.23	1.55	0.61	0.81	0.97	1.24
Expectile-based intervals						
CARE.98.SAV	33.67	42.66	18.80	36.07	33.29	32.90
CARE.98.AS	36.74	49.40	22.56	44.14	34.03	37.37
CARE.98.ADPT	13.78	7.30	0.03	5.96	4.58	6.33
CARE.98.INDI	32.61	45.12	22.45	40.70	35.56	35.29
CARE.95.SAV	33.35	41.45	18.13	34.33	32.98	32.05
CARE.95.AS	36.77	47.02	22.08	42.69	34.37	36.58
CARE.95.ADPT	11.10	6.26	0.08	8.11	4.85	6.08
CARE.95.INDI	32.38	44.43	21.85	39.17	35.44	34.66
CARE.90.SAV	33.27	40.67	17.37	33.12	32.94	31.47
CARE.90.AS	36.68	46.23	22.12	41.97	34.42	36.29
CARE.90.ADPT	13.74	2.21	1.92	6.23	8.25	6.47
CARE.90.INDI	32.36	43.82	21.00	38.64	35.42	34.25

Table 4.4: J-test for comparing USQ.MSD.AS forecasts ($\hat{\sigma}^2_{\texttt{MSD},t+k}$) and ES.90.AS forecasts ($\hat{\sigma}^2_{\texttt{ES},t+k}$). p-values are in parentheses. The star symbols denote significant levels: *** 1%, ** 5%, * 10%

	DAX30	SP500	N225	NASDAQ	FTSE
10-day period					
Fitted $\hat{\sigma}^2_{\texttt{ES},t+k}$, $\hat{\lambda}_1$:	-1.112 (0.044)**	4.659 (0.000)***	-2.455 (0.000)***	1.723 (0.000)***	0.257 (0.721)
Fitted $\hat{\sigma}^2_{\texttt{MSD},t+k}$, $\hat{\lambda}_2$:	2.104 (0.000)***	-3.717 (0.000)***	3.428 (0.000)***	-0.728 (0.098)*	0.743 (0.302)
20-day period					
Fitted $\hat{\sigma}^2_{\texttt{ES},t+k}$, $\hat{\lambda}_1$:	0.040 (0.948)	4.619 (0.000)***	-3.279 (0.000)***	2.068 (0.000)***	1.496 (0.068)*
Fitted $\hat{\sigma}^2_{\texttt{MSD},t+k}$, $\hat{\lambda}_2$:	0.960 (0.119)	-3.674 (0.000)***	4.234 (0.000)***	-1.076 (0.05)*	-0.497 (0.544)

Table 4.5: J-test for comparing USQ.MSD.AS forecasts ($\hat{\sigma}^2_{\texttt{MSD},t+k}$) and CARE.98.AS forecasts ($\hat{\sigma}^2_{\texttt{CARE},t+k}$). p-values are in parentheses. The star symbols denote significant levels: *** 1%, ** 5%, * 10%

	DAX30	SP500	N225	NASDAQ	FTSE
10-day period					
Fitted $\hat{\sigma}^2_{\texttt{CARE},t+k}$, $\hat{\lambda}_1$:	1.001 (0.064)*	1.852 (0.000)***	1.829 (0.000)***	2.427 (0.000)***	-1.297 (0.000)***
Fitted $\hat{\sigma}^2_{\texttt{MSD},t+k}$, $\hat{\lambda}_2$:	-0.001 (0.999)	-0.895 (0.000)***	-0.845 (0.028)**	-1.506 (0.000)***	2.264 (0.000)***
20-day period					
Fitted $\hat{\sigma}^2_{\texttt{CARE},t+k}$, $\hat{\lambda}_1$:	0.616 (0.309)	1.729 (0.000)***	1.792 (0.000)***	2.275 (0.000)***	-1.029 (0.005)***
Fitted $\hat{\sigma}^2_{\texttt{MSD},t+k}$, $\hat{\lambda}_2$:	0.385 (0.525)	-0.763 (0.000)***	-0.806 (0.102)	-1.340 (0.000)***	2.006 (0.000)***

Table 4.6: R^2 measure of combined forecast models for 900 post-sample volatility forecasts for 10- and 20-day period volatility. Values in the first row of each periods are the highest R^2 measures of individual-forecast models.

	DAX30	SP500	N225	NASDAQ	FTSE	Mean
10-day period						
Highest individual-forecast model	VaR.98.AS 49.94	CARE.98.AS 65.41	gjr-GARCH 42.57	CARE.98.AS 62.45	ES.90.AS 45.93	CARE.98.AS 52.09
Combination forecast						
ES.90.AS and VaR.98.AS	50.34	64.41	37.79	62.31	46.46	52.26
CARE.98.AS and VaR.98.AS	50.11	66.92	47.15	62.46	44.50	54.23
20-day period						
Highest individual forecast model	VaR.98.AS 37.12	CARE.98.AS 49.40	gjr-GARCH 23.27	VaR.98.AS 44.83	ES.90.AS 35.67	CARE.98.AS 38.06
Combination forecast						
ES.90.AS and VaR.98.AS	45.42	54.98	28.39	54.63	41.60	45.01
CARE.98.AS and VaR.98.AS	45.32	57.36	35.79	52.50	41.10	46.42

Chapter 5

Joint modelling of regression quantiles and heteroskedasticity with asymmetric Laplace distribution

The quantile regression has been a rewarding methodology of regression analysis and has received high attention in recent decades, since being introduced by Koenker and Bassett (1978). The original assumptions of quantile regression include homoskedasticity, however the heteroskedastic conditions of quantile regression have also been studied by many (for example, see Koenker and Bassett, 1982, Powell, 1984, Koenker and Zhao, 1994, among others). A common setting of the heteroskedastic quantile regression model is to introduce a scale component specified as a linear model of some explanatory variables. Estimators of the linear scale component model can be obtained with standard methods, but are usually obtained separately with the quantile regression estimators. In this chapter, we discuss a new model, which consists of a scale component that has a log linear form, and parameters of the heteroskedastic scale and parameters of the regression quantile can be estimated simultaneously, via the parametric link between quantile regression and the asymmetric Laplace distribution.

5.1 Introduction

The quantile regression is introduced by Koenker and Bassett (1978) (see also Koenker, 2005). Starting from the following standard linear model:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t, \qquad t = 1, ..., T, \tag{5.1}$$

where y_t is the dependent variable, $\mathbf{x}_t$ are the explanatory variables; $\boldsymbol{\beta}$ are unknown parameters, and ε_t is an error term which usually has no distributional assumption other than that the pth quantile of ε_t is zero and ε_t has finite variance.

The p-th $(0 < p < 1)$ conditional quantile of y_t given $\mathbf{x}_t$ is then in the expression as:

$$Q_{y_t|\mathbf{x}_t}(p) = \mathbf{x}_t'\boldsymbol{\beta}(p),$$

where $\boldsymbol{\beta}(p)$ is the coefficients vector depends on p.

Then the regression parameters $\boldsymbol{\beta}(p)$ for the p-th quantile can be defined as the solution to the following minimisation problem:

$$\min_{\boldsymbol{\beta}} \mathbb{E}\left[\rho_p(y_t - \mathbf{x}_t'\boldsymbol{\beta})\right], \tag{5.2}$$

where $\rho_p(u) = u(p - I[u \leq 0])$ is the so-called check function, and $I[\cdot]$ is an indicator function. The p-th regression quantile estimator $\hat{\boldsymbol{\beta}}(p)$ is obtained by solving the sample analog minimisation problem

$$\min_{\boldsymbol{\beta}} S(\boldsymbol{\beta}) \equiv \frac{1}{T}\sum_{t=1}^{T} \rho_p(y_t - \mathbf{x}_t'\boldsymbol{\beta}) \tag{5.3}$$

In the standard setting, ε_t is i.i.d., and has distribution function F and density function f. Under some regulatory conditions, the quantile regression is asymptotically normal (see, for example, Koenker, 2005):

$$\sqrt{T}\left(\hat{\boldsymbol{\beta}}(p) - \boldsymbol{\beta}(p)\right) \xrightarrow{d} \mathbf{N}(0, V),$$

with

$$V = \frac{p(1-p)}{f(0)^2}\mathbb{E}[\mathbf{x}\mathbf{x}']^{-1},$$

where $f(0)$ is the density of ε_t at zero. The asymptotic properties of the quantile regression under more complex conditions such as heteroskedasticity has been discussed by many, for example, Koenker and Bassett (1982), Powell (1984), among

others.

A special case of the regression quantiles with heteroskedasticity is the location-scale shift model (see, for example, Koenker and Xiao, 2002), which can be seen as induced from the following linear model:

$$y_t = \mathbf{x}_t'\boldsymbol{\alpha} + (\mathbf{x}_t'\boldsymbol{\gamma})\varepsilon_t, \qquad \mathbf{x}_t'\boldsymbol{\gamma} > 0 \tag{5.4}$$

where the error term ε_t is i.i.d., and $\mathbf{x}_t'\boldsymbol{\alpha}$ is the location shift component and $\mathbf{x}_t'\boldsymbol{\gamma}$ is the scale shift component, which is also in a linear form. The p-th conditional quantile of (5.4) is then given by

$$Q_{y_t|\mathbf{x}_t}(p) = \mathbf{x}_t'\boldsymbol{\alpha} + (\mathbf{x}_t'\boldsymbol{\gamma})F^{-1}(p),$$

where $F^{-1}(p)$ denotes the quantile function of the error term ε_t. Under this model specification, the quantile regression estimator

$$\hat{\boldsymbol{\beta}}(p) = \arg\min_{\boldsymbol{\beta}} \frac{1}{T}\sum_{t=1}^{T} \rho_p(y_t - \mathbf{x}_t'\boldsymbol{\beta}) \tag{5.5}$$

is actually an estimator of

$$\boldsymbol{\beta}(p) = \boldsymbol{\alpha} + \boldsymbol{\gamma}F^{-1}(p).$$

$\hat{\boldsymbol{\beta}}(p)$ converges to $\boldsymbol{\beta}(p)$ with probability one, and $\mathbf{x}_t'\hat{\boldsymbol{\beta}}(p)$ is an approximation of $q_{y_t|\mathbf{x}_t}(p)$. However, solving (5.5) only gives estimates of $\boldsymbol{\beta}(p)$. The parameters $\boldsymbol{\gamma}$ of the scale component are not assessed when using the classical regression estimator $\hat{\boldsymbol{\beta}}(p)$.

To estimate $\boldsymbol{\gamma}$ under the situation that a location shift component is included, Koenker and Zhao (1996) consider using a two-step procedure: first, remove the location shift component by 'de-meaning', and then regressing the quantile estimator on the residual. Let $\hat{\boldsymbol{\alpha}}$ be the ordinary least square estimators such that $\hat{y}_t = \mathbf{x}_t'\hat{\boldsymbol{\alpha}}$, and the residual $\hat{e}_t = y_t - \mathbf{x}_t'\hat{\boldsymbol{\alpha}}$. Then the quantile regression estimator $\check{\boldsymbol{\gamma}}$, given $\hat{\boldsymbol{\alpha}}$, is expressed as:

$$\check{\boldsymbol{\gamma}}(p, \hat{\boldsymbol{\alpha}}) = \arg\min_{\boldsymbol{\gamma}} \frac{1}{T}\sum_{t=1}^{T} \rho_p(\hat{e}_t - \mathbf{x}_t'\boldsymbol{\gamma}).$$

Another model that addresses heteroskedasticity is the weighted quantile regression model introduced by Koenker and Zhao (1994) and Koenker and Zhao (1996). For the linear model as shown in (5.4), if the estimation in (5.5) is referred as the 'unweighted' regression quantile estimator, then the p-th $(0 < p < 1)$ weighted regression quantile

is obtained by:

$$\check{\boldsymbol{\beta}}(p, \hat{\boldsymbol{\gamma}}) = \arg\min_{\boldsymbol{\beta}} \frac{1}{T} \sum_{t=1}^{T} \hat{w}_t^{-1} \rho(y_t - \mathbf{x}_t' \boldsymbol{\beta}), \tag{5.6}$$

given that $\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\| = \mathcal{O}(T^{-1/4}(\log T)^{1/2})$, and $\hat{w}_t = \mathbf{x}_t' \hat{\boldsymbol{\gamma}}$ denotes the 'weights'. Estimation methods for parameters and weights of the weighted regression quantiles are further discussed by Zhou and Portnoy (1998), among others.

In this chapter, we propose a new class of heteroskedastic quantile regression model by replacing the linear scale component in (5.4), $\mathbf{x}_t'\boldsymbol{\gamma}$, with a log linear form $\lambda_t = \exp(\mathbf{x}_t'\boldsymbol{\gamma})$. An obvious advantage of this change to the model is that, since $\exp(\mathbf{x}_t'\boldsymbol{\gamma}) > 0$ for any $\mathbf{x}_t'\boldsymbol{\gamma}$, the additional assumption on the conventional location-scale shift model ensuring $\mathbf{x}_t\boldsymbol{\gamma} > 0$ could be omitted. We illustrate that the new model can be estimated via the parametric link between the quantile regression and the asymmetric Laplace distribution, which is briefly introduced in Section 5.2. Details of the new quantile regression estimators are given in Section 5.3. In Section 5.4 we study properties of the new estimators when data departures from the assumption on the error distribution. In the last two sections, we make remarks and conclusions.

5.2 Asymmetric Laplace distribution

The asymmetric Laplace distribution (ALD), as in some literatures referred as the skewed-Laplace distribution, has the following density function:

$$f(u|\mu, \sigma, p) = \frac{p(1-p)}{\sigma} \exp\left(-\frac{u-\mu}{\sigma}(p - I[u \leq \mu])\right), \tag{5.7}$$

where μ, σ and p are location, scale and shape (or skew) parameters, respectively. If $p = 1/2$, (5.7) reduces to the density of the standard symmetric Laplace distribution; otherwise the density is asymmetric (skewed). One of the most important properties of the ALD is that, if a random variable u follows the ALD distribution, $u \sim \text{ALD}(\mu, \sigma, p)$, the p-th quantile of u equals to μ. Furthermore, the mean and the variance of this three-parameter ALD are given by

$$\mathbb{E}(u) = \mu + \frac{\sigma(1-2p)}{p(1-p)},$$
$$\text{var}(u) = \frac{\sigma^2(1-2p+2p^2)}{(1-p)^2 p^2},$$

respectively. For more details and properties of the three-parameter ALD, see Yu and Zhang (2005).

The link between the quantile regression and the asymmetric Laplace distribution is as follow: if we assume the error term in (5.1) $\varepsilon_t \overset{iid}{\sim} \text{ALD}(0, \sigma, p)$, and let $\mu(\mathbf{x}_t) = \mathbf{x}_t'\boldsymbol{\beta}$, then the estimation problem for $\boldsymbol{\beta}(p)$, the p-th conditional quantile $Q_p(y_t|\mathbf{x}_t)$ as in the minimisation function (5.3), is equivalent to parameter estimation problem for $y_t \sim \text{ALD}\left(\mu(\mathbf{x}_t), \sigma, p\right)$ using the MLE, with the log-likelihood function

$$\ell(\mu(\mathbf{x}_t), \sigma) = -T\log(\sigma) - \frac{1}{\sigma}\sum_{t=1}^{T}\rho_p(y_t - \mu(\mathbf{x}_t)). \tag{5.8}$$

Yu and Zhang (2005) show that, solutions to the MLE of the three-parameter ALD is as follows:

$$\hat{\mu}(\mathbf{x}_t) = \arg\min_{\mu(\mathbf{x}_t)} \frac{1}{T}\sum_{t=1}^{T}\rho_p(y_t - \mu(\mathbf{x}_t)),$$
$$\hat{\sigma} = \frac{1}{T}\sum_{t=1}^{T}\rho_p(y_t - \mu(\mathbf{x}_t)),$$

for p known.

Interestingly, if $\boldsymbol{\beta}(p)$ denotes the true p-th quantile regression parameter, and $\mathbf{x}_t'\boldsymbol{\beta}(p)$ denotes the true value of the p-th conditional quantile, then $\sigma \equiv E[\rho_p(y_t - \mathbf{x}_t'\boldsymbol{\beta}(p))]$, namely, σ is the expectation of the check function at quantile level p. Then $\hat{\sigma}$ is the sample analog estimator of the expectation of the check function. Consider the ordinary least square counterpart, *i.e.*, $\boldsymbol{\alpha}^*_{OLS}$ is the true value of minimiser to the minimisation problem

$$\text{OLS:} \qquad \min_{\boldsymbol{\alpha}} \mathbb{E}[(y_t - \mathbf{x}_t'\boldsymbol{\alpha})^2],$$

then $\mathbb{E}[(y_t - \mathbf{x}_t'\boldsymbol{\alpha}^*_{OLS})^2]$ is actually the sample variance of y. Furthermore, let $\boldsymbol{\alpha}^*_{LAD}$ be the the true value of minimiser to the least absolute deviation minimisation problem

$$\text{LAD:} \qquad \min_{\boldsymbol{\alpha}} \mathbb{E}[|y_t - \mathbf{x}_t'\boldsymbol{\alpha}|],$$

then $\mathbb{E}[|y_t - \mathbf{x}_t'\boldsymbol{\alpha}^*_{LAD}|]$ actually yields the sample's mean absolute deviation. Both variance and MAD are well-known measure of variability. Similarly, if seen as a function of p, the scale parameter of the ALD , $\sigma(p) = \mathbb{E}\left[\rho_p(y_t - \mathbf{x}_t'\boldsymbol{\beta})\right]$, when considered in the framework of quantile regression, can be seen as a measure of statistical dispersion of the variability of dependent variable y around its p-th conditional quantile $\mathbf{x}_t'\boldsymbol{\beta}(p)$, and $\hat{\sigma}(p)$ is its estimator arising from the sample analog. We name $\sigma(p)$ the *p-th quantile deviation*, to be distinguished from the conventional standard deviation and MAD. Then, if a random variable y has an ALD, $y \sim \text{ALD}(\mu, \sigma, p)$,

$0 < p < 1$, then its standard deviation, $\mathrm{SD}(y)$, and its p-th quantile deviation, σ, has the following relationship:

$$\mathrm{SD}(y) = \sigma \frac{\sqrt{(1-2p+2p^2)}}{(1-p)p}.$$

Furthermore, let G and g be the c.d.f. and p.d.f. of y, respectively, and $\eta = \mathbb{E}(y)$. Then, given that the check function can be equivalently written as $\rho_p(y-\mu) = p(y-\mu) - (y-\mu)I[y \leq \mu]$, a relatively general representation for $QD(p)$ of y is as follow

$$\begin{aligned}
\mathrm{QD}(p) \equiv & \mathbb{E}[\rho_p(y - Q(p)] \\
= & \mathbb{E}\big[p(y-Q(p)) - (y-Q(p))I[y \leq Q(p)]\big] \\
= & p(\eta - Q(p)) - \mathbb{E}\big[(y-\eta)I[y \leq Q(p)]\big] - \mathbb{E}\big[(\eta - Q(p))I[y \leq Q(p)]\big] \\
= & p(\eta - Q(p)) - \int_{-\infty}^{Q(p)} (y-\eta)g(y)\,\mathrm{d}y - p(\eta - Q(p)) \\
= & -\int_{-\infty}^{Q(p)} (y-\eta)g(y)\,\mathrm{d}y.
\end{aligned}$$

For some distributions, the unconditional population QD can be derived explicitly using the expression above. For example, if random variable y has a normal distribution with variance υ^2, it is straightforward to obtain its p-th QD

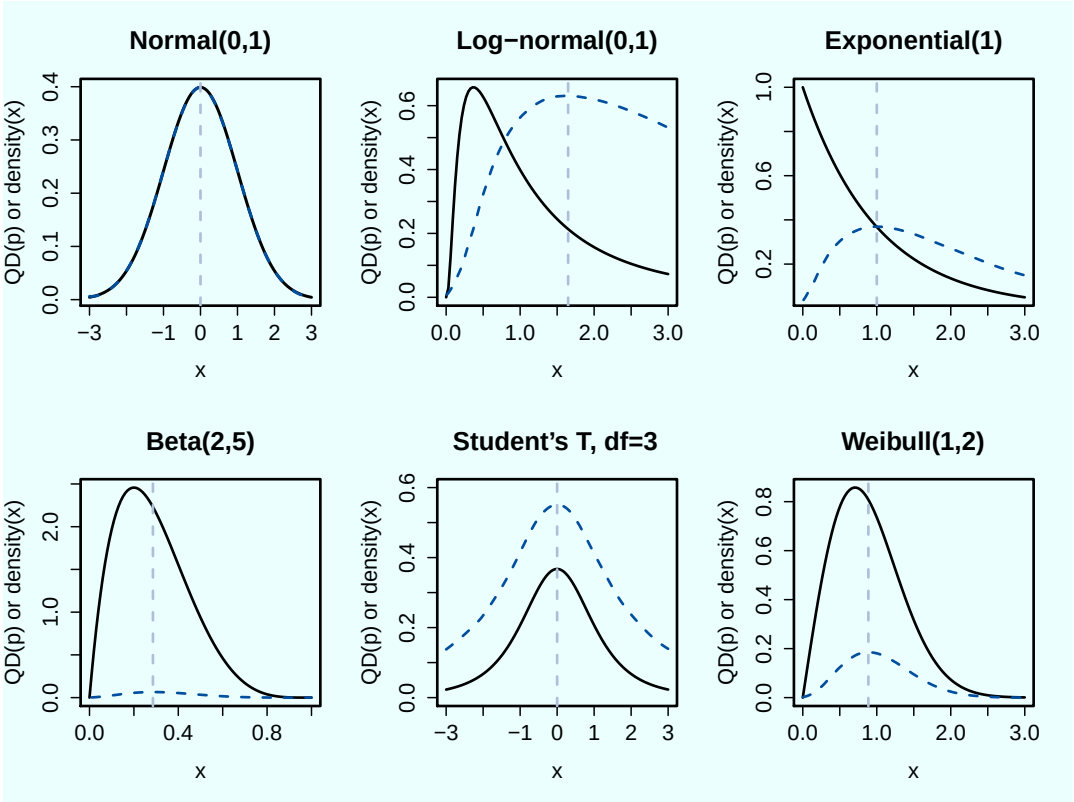
$$\mathrm{QD}(p) = \upsilon\phi(\psi^{-1}(p)),$$

where $\phi(\cdot)$ is the p.d.f. of standard normal, and $\psi^{-1}(p)$ is the inverse of the c.d.f. of standard normal.

Hence, the normal distribution's QD is identical with the normal density function, which is considered as a very special case. In Figure 5.1, we present graphs of unconditional population QD functions of several common distributions against their density functions $g(x)$, where the QD is displayed as a function with respect to x, namely $\mathrm{QD}(G(x))$.

The parametric link between the quantile regression and the ALD has been extensively used in the Bayesian inference of regression quantiles. Yu and Moyeed (2001) and Tsionas (2003), among others, have illustrated that, using the link between quantile regression and the ALD allows accurate Bayesian estimation and inference to be achieved, through computational algorithm such as the Markov Chain Monte Carlo. In following sections we study heteroskedastic quantile regression with assumption

Figure 5.1: The QD function graphs for some common distributions. Solid line: density function $f(x)$; dashed line: quantile deviation $\mathrm{QD}(p)$, where $p = F(x)$; vertical dashed line: distribution mean



that the error term ε_t has ALD, thus parameters of the quantile regression can be obtained via MLE method.

5.3 Proposed quantile regression with dynamic variance

For classic linear regression model, it is typical that the error term ε_t in (5.1) is normal. However, authors such as Koenker and Bassett (1982), Powell (1984), and Chamberlain (1994) have considered the quantile regression in under more general conditions, leading to the conclusion that, the quantile regression is robust when the normality assumption of ε_t is not met. Namely, under certain regulatory conditions and without assuming ε_t is normal, the regression coefficient estimator $\hat{\boldsymbol{\beta}}(p)$ is asymptotically normal. We further investigate and propose a new inference that allows assessing the joint modelling problem of quantile regression and the heteroskedastic variance of conditional quantile simultaneously.

The model considered is under the following assumptions

A 5.1 Given that y_t, $\mathbf{x}_t$, $\boldsymbol{\gamma}$ and ε_t are defined as in expression 5.1, let $\lambda_t > 0$ be the heteroskedasticity scale component, then y_t and $\mathbf{x}_t$ have the following model

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \lambda_t\varepsilon_t, \qquad \lambda_t = \exp(\mathbf{x}_t'\boldsymbol{\gamma}), \quad t = 1, ..., T \tag{5.9}$$

A 5.2 The ε_t in A 5.1 has an i.i.d. asymmetric Laplace distribution, $\varepsilon_t \sim \text{ALD}(0, 1, p)$.

Then it is straightforward to show that the combination of $\lambda_t\varepsilon_t$ has an asymmetric Laplace $\text{ALD}(0, \lambda_t, p)$ distribution. Then the p-th conditional quantile and its variance are given by:

$$q_y(p) = \mathbf{x}_t'\boldsymbol{\beta}(p), \tag{5.10}$$

with

$$\text{var}(\lambda_t\varepsilon_t) = \frac{\exp(\mathbf{x}_t'\boldsymbol{\gamma})(1 - 2p + 2p^2)}{(1-p)^2p^2}.$$

The estimations for parameters in (5.10) can also be obtained via maximum likelihood estimation (MLE) for the parameters in the corresponding ALD. Replacing the σ in (5.8) with $\lambda_t = \mathbf{x}_t'\boldsymbol{\gamma}$, the likelihood function of parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ is then

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\gamma}|\boldsymbol{X}) &\propto \prod_{t=1}^{T} \exp(-\mathbf{x}_t'\boldsymbol{\gamma}) \exp\left(-\frac{\rho_p(y_t - \mathbf{x}_t'\boldsymbol{\beta})}{\exp(\mathbf{x}_t'\boldsymbol{\gamma})}\right) \\ &= \exp\left(\sum_{t=1}^{T} -\frac{\rho_p(y_t - \mathbf{x}_t'\boldsymbol{\beta})}{\exp(\mathbf{x}_t'\boldsymbol{\gamma})} - \sum_{t=1}^{T} \mathbf{x}_t'\boldsymbol{\gamma}\right). \end{aligned}$$

The log-likelihood function is given by:

$$\ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{t=1}^{T} -\frac{\rho_p(y_t - \mathbf{x}_t'\boldsymbol{\beta})}{\exp(\mathbf{x}_t'\boldsymbol{\gamma})} - \sum_{t=1}^{T} \mathbf{x}_t'\boldsymbol{\gamma}. \tag{5.11}$$

Denote $\tilde{\boldsymbol{\beta}}(p, \tilde{\boldsymbol{\gamma}})$ and $\tilde{\boldsymbol{\gamma}}(p)$ as the MLEs of parameters $\boldsymbol{\beta}(p)$ and $\boldsymbol{\gamma}(p)$. Then we have the following likelihood equations:

$$\tilde{\boldsymbol{\beta}}(p, \tilde{\boldsymbol{\gamma}}) = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} \frac{\rho_p(y_t - \mathbf{x}_t'\boldsymbol{\beta})}{\exp(\mathbf{x}_t'\tilde{\boldsymbol{\gamma}}(p))}, \tag{5.12}$$

$$\sum_{t=1}^{T} \frac{\rho_p(y_t - \mathbf{x}_t'\tilde{\boldsymbol{\beta}}(p, \tilde{\boldsymbol{\gamma}}))}{\exp(\mathbf{x}_t'\tilde{\boldsymbol{\gamma}}(p))}\mathbf{x}_t = \sum_{t=1}^{T} \mathbf{x}_t. \tag{5.13}$$

So the $\tilde{\boldsymbol{\beta}}(p, \tilde{\boldsymbol{\gamma}})$ is essentially the weighted quantile regression estimator as shown in (5.6). Nonetheless, by using the parametric link between the ALD and the quantile

regression, the weighted estimators, $\tilde{\boldsymbol{\beta}}(p, \tilde{\boldsymbol{\gamma}})$ and the weight estimator, $\tilde{\lambda}_t = \mathbf{x}_t'\tilde{\boldsymbol{\gamma}}(p)$ can be obtained simultaneously.

Furthermore, the maximum likelihood function with respect to λ requires that

$$\begin{aligned}\tilde{\lambda}_t &= \exp(\mathbf{x}_t'\tilde{\boldsymbol{\gamma}}(p)) \\ &\equiv \frac{1}{T}\sum_{t=1}^{T}\rho_p(y_t - \mathbf{x}_t'\tilde{\boldsymbol{\beta}}(p, \tilde{\boldsymbol{\gamma}})).\end{aligned}$$

As discussed in the previous section, the estimator $\tilde{\lambda}_t$ is also an estimator of the p-th quantile deviation, $\mathbb{E}[\rho_p(y_t - \mathbf{x}_t'\boldsymbol{\beta}(p))]$, given $\mathbf{x}_t$.

Unfortunately, it seems difficult to obtain the asymptotic properties of $(\tilde{\boldsymbol{\beta}}(p, \tilde{\boldsymbol{\gamma}}), \tilde{\boldsymbol{\gamma}}(p))$. It is not reasonable to use the information matrix to give asymptotic variance of the MLE, as under usual regular conditions the likelihood function is not differentiable with respect to $\mathbf{x}_t'\boldsymbol{\beta}$. Koenker and Zhao (1994) proved that the weighted regression quantile estimator $\check{\boldsymbol{\beta}}(p, \hat{\boldsymbol{\gamma}})$ defined in (5.6) with weights $\hat{w}_t$ is asymptotically normal , under certain regulatory conditions (conditions C1 - C5 as in their paper). The estimator $\tilde{\lambda}_t$ we proposed can be seen as an estimator of weights. However, under the assumption A 5.2 in this chapter, C5 from Koenker and Zhao (1994) is not satisfied since the c.d.f. of ε_t, $F(z)$, is not twice differentiable at the point $z = F^{-1}(p) = 0$. Thus, the Bahadur representation they derived is not available for the estimators we introduced. The asymptotic properties of regression parameters $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\gamma}}$ is a topic worth investigating for future research.

The computational estimation of parameters are simply implemented in the R language (R Core Team, 2013), by employing the `optim` function with the quasi-Newton algorithm to search for solutions to the minimisation problems above. We recommend using $\tilde{\boldsymbol{\beta}}(p, \tilde{\boldsymbol{\gamma}}) = \hat{\boldsymbol{\beta}}(p)$ and $\tilde{\boldsymbol{\gamma}}_p = \mathbf{0}$ as initial values for the algorithm, where $\hat{\boldsymbol{\beta}}(p)$ are parameters estimated as in (5.3), and $\mathbf{0}$ is a vector of zeros.

One may argue that the distribution of the error term, ε_t, usually departs from the ALD in real data. In the next section, we illustrate that even the true distribution of ε_t is not ALD, namely the model in (5.7) is misspecified, the results of the quantile regression are still accurate, via simulation tests.

5.4 Simulation study

We perform simulation tests to investigate the performance of the proposed model when the underlying distribution actually depart from ALD. We also consider the

situation that the model is misspecified, since model misspecification is most likely in reality. Random samples are generated from the following models.

Model 1. Simple linear model, homoskedastic:

$$Y^{(1)} = -2X_1 + \varepsilon_{1,t}, \quad X_1 \sim \mathbf{N}(0, 2), \quad \varepsilon_{1,t} \sim \mathbf{N}(0, 0.0625),$$

where $\mathbf{N}(\cdot)$ denotes the normal distribution. The true value of p-th conditional quantile is $q_Y(p, X) = -2X + \mathbf{N}^{-1}(p|0, 0.0625)$, where $\mathbf{N}^{-1}(p|\cdot)$ denotes the p-th quantile of the normal distribution with given parameters.

Model 2. Purely heteroskedastic model, with linear scale shift component, as described in Section 1:

$$Y^{(2)} = 2X_1 + (.25 + .75X_1)\varepsilon_{2,t}, \quad X_1 \sim \mathbf{U}(0, 5), \quad \varepsilon_{2,t} \sim \mathbf{N}(0, 1),$$

where $\mathbf{U}(\cdot)$ denotes the uniform distribution. Note that for data generated from this model, the heteroskedastic quantile regression model with ALD error term we proposed is actually misspecified. The true value of p-th conditional quantile is given by $q_Y(p, X) = 2X(.25 + .75X)\mathbf{N}_p^{-1}(0, 1)$.

Model 3. Heteroskedastic model with exponential scale component and normally distributed error term

$$Y^{(3)} = 1 + X_1 \exp(-0.5X_1)\varepsilon_{3,t}, \quad X_1 \sim \mathbf{U}(0, 5), \quad \varepsilon_{3,t} \sim \mathbf{N}(0, 1).$$

The true value of p-th conditional quantile is given by $q_Y(p, X) = 1+X\exp(-.5X)\mathbf{N}_p^{-1}(0, 1)$.

Model 4. Heteroskedastic model with exponential scale component and Student's t error term:

$$Y^{(4)} = 1 + X_1 + \exp(-0.5X_1)\varepsilon_{4,t}, \quad X_1 \sim \mathbf{U}(0, 5), \quad \varepsilon_{4,t} \sim \mathbf{t}(6),$$

where $\mathbf{t}(6)$ denotes the Student's t distribution with 6 degrees of freedom. The true values of the p-th quantile is $q_Y(p, X) = 1+X+\sqrt{4/6}\exp(-.5X)\mathbf{t}_p^{-1}(6)$, where $\mathbf{t}_p^{-1}(\cdot)$ denotes the p-th quantile of the Student's t distribution.

For each model, 500 replicates were generated with two different sample sizes, $n = 500$ and $n = 100$. Based on these samples, we compute the $p = 0.5, 0.75, 0.9$ conditional quantile of Y, using the standard quantile regression model and the heteroskedastic model proposed in this chapter. Results of the standard quantile regression estimators $\hat{\boldsymbol{\beta}}$ and the heteroskedastic weighted estimators $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\gamma}}$ of the 0.5 and 0.75 conditional

quantiles are summarised in Table 5.1 and 5.2, respectively, in terms of means and standard errors of the estimation.

Table 5.1: Mean and standard estimation error of the 0.5 conditional quantile estimators from the the standard model ($\hat{\beta}$) and the proposed heteroskedastic model ($\tilde{\beta}$ and $\tilde{\gamma}$), for four different models.

	Model 1		Model 2		Model 3		Model 4	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
n=100								
$\hat{\beta}_1$	-0.001	0.008	0.000	0.214	0.007	0.123	0.990	0.127
$\hat{\beta}_2$	-2.000	0.004	2.005	0.165	-0.002	0.031	1.002	0.032
$\tilde{\beta}_1$	0.000	0.008	-0.001	0.166	0.007	0.107	0.990	0.110
$\tilde{\beta}_2$	-2.000	0.004	2.004	0.146	-0.002	0.026	1.002	0.026
$\tilde{\gamma}_1$	-3.713	0.093	-1.434	0.178	-0.917	0.157	-0.788	0.181
$\tilde{\gamma}_2$	0.004	0.074	0.437	0.062	-0.511	0.057	-0.506	0.062
n=500								
$\hat{\beta}_1$	0.000	0.003	-0.002	0.088	0.001	0.059	0.998	0.058
$\hat{\beta}_2$	-2.000	0.002	2.002	0.074	0.000	0.015	1.001	0.014
$\tilde{\beta}_1$	0.000	0.003	-0.002	0.073	0.002	0.051	1.000	0.049
$\tilde{\beta}_2$	-2.000	0.002	2.003	0.062	0.000	0.012	1.000	0.012
$\tilde{\gamma}_1$	-3.697	0.034	-1.412	0.076	-0.916	0.069	-0.778	0.081
$\tilde{\gamma}_2$	0.000	0.018	0.433	0.027	-0.502	0.024	-0.501	0.028

Note that for Model 1, the true values of $\boldsymbol{\beta}(p)$ are given by $(\beta_1, \beta_2) = (0, 2)$ for $q(0.5)$, and $(\beta_1, \beta_2) = (0.042, 2)$ for $q(0.75)$; and for Model 2, $(\beta_1, \beta_2) = (0, 2)$ for $q(0.5)$, and $(\beta_1, \beta_2) = (0.168, 2.506)$ for $q(0.75)$. Furthermore, for Model 1, $\gamma_2 = 0$ regardless of the value of p, since the model is homoskedastic. The true values of parameters for other cases are difficult to obtain explicitly. Table 5.1 and 5.2 show that, both classic estimators and the proposed estimators for Model 1 are almost identical and very close to the true values. For other three models, the differences between $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ are obvious, but not largely depart from each other. On the other hand, the standard errors of $\tilde{\boldsymbol{\beta}}$ are smaller than of $\hat{\boldsymbol{\beta}}$ in almost every cases.

To further investigate the performance of the two quantile estimators, we compared estimated quantiles from the standard quantile regression model as in (5.3) and the proposed model as in (5.10), denoted by $\hat{q}$ and $\tilde{q}$, respectively. Integrated squared error, as shown in the equation (5.14) below, were calculated over the range of $-5 \leq x \leq 5$ for Model 1, and $0 \leq x \leq 5$ for Model 2, 3, and 4, and averaged over replicates. Results were summarised in Table 5.3.

$$\text{ISE} = \int_a^b (\hat{q}_y(p, x) - q_y(p, x))^2 \, dx \tag{5.14}$$

Table 5.3 shows that, unsurprisingly, for Model 1, which is homoskedastic, the mean ISE of $\tilde{q}$ are higher than the mean ISE of $\hat{q}$ for most cases, although the differences

Table 5.2: Mean and standard estimation error of the 0.75 conditional quantile estimators from the the standard model ($\hat{\beta}$) and the proposed heteroskedastic model ($\tilde{\beta}$ and $\tilde{\gamma}$), for four different models.

	Model 1		Model 2		Model 3		Model 4	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
n=100								
$\hat{\beta}_1$	0.042	0.008	0.192	0.230	0.467	0.135	1.492	0.157
$\hat{\beta}_2$	-2.000	0.004	2.502	0.171	-0.092	0.033	0.904	0.038
$\tilde{\beta}_1$	0.042	0.008	0.181	0.184	0.399	0.119	1.413	0.134
$\tilde{\beta}_2$	-2.000	0.004	2.506	0.157	-0.074	0.029	0.924	0.032
$\tilde{\gamma}_1$	-3.934	0.128	-1.666	0.178	-1.128	0.180	-0.982	0.212
$\tilde{\gamma}_2$	0.002	0.113	0.437	0.062	-0.514	0.064	-0.509	0.073
n=500								
$\hat{\beta}_1$	0.042	0.004	0.172	0.098	0.470	0.062	1.496	0.069
$\hat{\beta}_2$	-2.000	0.002	2.508	0.079	-0.092	0.015	0.904	0.017
$\tilde{\beta}_1$	0.042	0.004	0.170	0.079	0.398	0.053	1.422	0.056
$\tilde{\beta}_2$	-2.000	0.002	2.508	0.068	-0.073	0.012	0.923	0.013
$\tilde{\gamma}_1$	-3.922	0.056	-1.643	0.077	-1.124	0.078	-0.971	0.095
$\tilde{\gamma}_2$	0.002	0.041	0.434	0.027	-0.506	0.027	-0.504	0.032

Table 5.3: Results of simulation tests using the four data generating models. Mean ISE of standard quantile regression ($\hat{q}$) and proposed quantile regression ($\tilde{q}$), multiplied by 1000, are summarised in this table, for sample size $n = 100, 500$, and $p = 0.1, 0.5, 0.9$. 500 replicates.

		n=100			n=500		
		$p = 0.5$	$p = 0.75$	$p = 0.9$	$p = 0.5$	$p = 0.75$	$p = 0.9$
1	$\tilde{q}$	0.136	0.156	0.244	0.025	0.027	0.046
	$\hat{q}$	0.130	0.153	0.242	0.025	0.027	0.047
2	$\tilde{q}$	70.992	71.813	122.598	12.600	15.504	27.418
	$\hat{q}$	73.188	76.750	147.327	16.392	16.904	29.932
3	$\tilde{q}$	1.705	8.358	26.360	0.339	7.268	24.006
	$\hat{q}$	2.218	6.179	17.336	0.478	4.086	13.512
4	$\tilde{q}$	2.001	5.944	18.500	0.336	3.730	14.071
	$\hat{q}$	2.618	6.258	20.066	0.429	3.808	16.657

are only marginal. Both methods have significantly higher mean ISE for Model 2, which has a linear scale component. Note that the model of $\tilde{q}$ is actually misspecified for data generated from Model 2. Interestingly, the ISE of $\tilde{q}$ is smaller than of $\hat{q}$ under this case. Model 3 is a heteroskedastic model with an exponential scale component and error term follows the standard normal distribution, as specified in (5.9). The standard estimator $\hat{q}$ has smaller ISE, except for $p = 0.5$, which implies the median regression. Model 4 follows a more general form of (5.9), under this model the proposed estimator $\tilde{q}$ is correctly specified. The error term in Model 4 ε is assumed to have a Student's t distribution with 6 degree of freedom, and it seems has little impact on the performance of both estimators. Compared with results from Model 3, the introduction of location component improves the performance of $\tilde{q}$ but compromise the performance of $\hat{q}$.

To sum up, in simulation study we tested the standard quantile regression model and the proposed heteroskedastic quantile regression model using data generated from distribution depart from the assumed ALD, and data generated from different model specification, including homoskedasticity and linear heteroskedasticity (other than exponential). In general, the proposed heteroskedastic quantile regression model outperform the standard model, especially when heteroskedasticity present.

5.5 Remarks for future research

It should be not difficult to extend the heteroskedastic quantile regression model proposed in previous sections to auto-regressive models, and thus it can be applied with time-series data. We are particularly interest in estimating and forecasting of financial time-series. Koenker and Zhao (1996) has already considered inferences of quantile regression applied to autoregressive conditional heteroskedasticity (ARCH) models, and part of their research has been introduced in Section 5.1. We suggest two possible future research paths. The first is to look at the conditional regression quantiles of financial returns used as a risk measure, which is commonly known as the value-at-risk (VaR). Engle and Manganelli (2004) introduce a class of forecast models called conditional autoregressive VaR (CAViaR), which has been mentioned in previous chapters. The CAViaR models do not treat heteroskedasticity explicitly, but it has already received considerable popularity due to its performance in practice. If heteroskedasticity is introduced into the CAViaR model through the approach proposed in this chapter, an instantly encountered problem, which may eventually lead to the second research path, would be: what model should one choose for the exponential heteroskedastic component, $\exp(\mathbf{x}_t'\boldsymbol{\gamma})$? The second research path is to

address this issue. A natural choice would seem to be to search in the GARCH family of volatility forecast models. For example, one can consider the quantile analog version of the EGARCH model, which is introduced by Nelson (1991) and also assumes exponential heteroskedasticity. However, if applied to financial time-series, one must note that a key difference between assumptions for financial time-series and assumptions we use in this chapter. That is, for financial returns, it is typical to assume that the residual term has a zero-mean distribution. Thus, it is necessary to change the error term in (5.9) $e_t = \varepsilon_t - \mu$, where $\mu = \frac{1-2p}{p(1-p)}$ is the mean of $\text{ALD}(0, 1, p)$, as pointed out by Chen et al. (2012). Then the new error term e_t has mean 0 as standard. The goal of future research will focus on two aspects: first, a new approach for the confidence interval of VaR estimates and forecasts; secondly and more importantly, a potential new type of 'volatility', which is defined as the variance of quantile of return, instead of variance of the mean return.

5.6 Chapter summary

In this chapter, we reviewed classical literature on parametric heteroskedastic quantile regression models. Then we introduced a new method, which assumes the model's error term follows an ALD, thereby estimators of the regression quantiles and estimators of the weights, or the quantile deviation, can be obtained simultaneously via MLE method for parameters of the ALD. Simulation tests show that, even for models generated using distributions that depart from the assumed ALD, the proposed heteroskedastic quantile regression model outperform the standard model in general.

Chapter 6

Concluding remarks

The thesis studied several theoretical and applied order-statistics-based methods relating to distribution estimation, measurement and forecasting of some risk measures. In final chapter, the main contributions from Chapter 2 to 5 are summarised, along with a brief discussion for possible future research topics.

6.1 Main contributions

In Chapter 2, a new method for constant stress ALT model with Weibull distribution when the data are progressively censoring is considered. We have derived the unbiased estimators of unknown parameters, exact confidence interval of shape parameter and generalized confidence intervals of other parameters. The method and theoretic results are new, totally different from MLE-based inference. The numerical analysis and comparison are promising, even for small sample and different censoring rate or schemes.

An alternative approach is proposed in Chapter 3 for fitting the GPD, especially under small sample sizes. For parameter estimation, the proposed estimators are unbiased and stable for extreme shape parameters and small sample sizes. For quantile estimations the situation is more complicated. But for non-extreme quantiles, estimates based on the alternative approach also has overall small bias. Amongst all tested estimators, we find that the method proposed in Zhang and Stephens (2009) also has outstanding performance in both parameter and quantile estimations. The alternative approach also provides exact and generalised confidence interval estimations for parameters and quantile of the GPD, and it clearly provides more accurate results, in terms of coverage probabilities with reasonable interval width.

In Chapter 4, we propose two new volatility forecast methods that use intervals between symmetric expected shortfalls and conditional autoregressive expectiles under least square regression framework. The new methods elaborate on information contained on tails of underlying data, and there is no distributional assumption needed. Empirical studies using five different stock indices for short period RV forecasts suggest that, the proposed new methods, especially the model based on 98% asymmetric slope CARE estimates, have overall better performance than other tested models, including the multiple uniformly spaced quantiles volatility forecast model introduced by Huang (2012). We also found that, combining the ES interval model or the CARE interval model with the CAViaR interval model improves the volatility forecasts, in terms of R^2.

Chapter 5 elaborates the joint modelling of quantile regression and heterosckedasticity. A new method has been discussed, which assumes the model's error term follows an ALD, thereby estimators of the regression quantiles and estimators of the weights, or the quantile deviation as defined in this chapter, can be obtained simultaneously via MLE method. Simulation tests show that, even for models generated using distributions that depart from the assumed ALD, the proposed heteroskedastic quantile regression model outperform the standard model in general.

6.2 Recommendations for Future Research

- The model analysed in Chapter 2 is for constant-stress accelerated life-tests under certain data censoring. A natural progression is the study of step-stress ALT, with progressive Type II censoring. In step-stress tests, the values of stress on units are not constant, but increased by pre-planned 'stress patterns'. *i.e.*, if a unit survived a stress level, it is then moved to a higher stress level, and continues until the end of the test. The cumulative effects of each stress level that a unit has survived have to be taken into consideration for the model. The simple step-stress ALT (Miller and Nelson, 1983), which has only two stress levels, could be considered as a starting point.

- In Chapter 3 we do not consider more realistic circumstances or empirical examples for financial data. That is because financial data is well-known not Gaussian, which is one of the most important assumptions in the POT approach. Applying EVT for financial data often requires the heteroskedasticity been 'filtered', see McNeil (1999) and Gilli and Këllezi (2006) as examples.

- In Chapter 4, the accuracy of forecasts of RV is affected by the accuracy of

the estimation of extreme tail quantiles, ES, or expectiles. One could explore possible ways of improving the forecast performance by considering improved models for, say, extreme quantiles or ES. Furthermore, it may worth consideration that, when volatility jumps occur, the possible impacts on the proposed methods, as they are more sensitive to extreme values.

- The work presented in Chapter 5 raises a number of interesting questions for further consideration.

    1. It should be not difficult to extend the heteroskedastic quantile regression model proposed to auto-regressive models, and thus it can be applied with time-series data, especially financial time-series. Koenker and Zhao (1996) has already considered inferences of quantile regression applied to autoregressive conditional heteroskedasticity (ARCH) models.
    2. It would be of interest to investigate if the joint modelling approach proposed can be considered together with the CAViaR models. The CAViaR models do not treat heteroskedasticity explicitly, but it has already received considerable popularity due to its performance for real problems.
    3. In Chapter 5 we considered a simple log linear relationship for the scale component. If considered for financial time-series, a natural progression would be replacing the simple log linear scale model with the EGARCH model, which is introduced by Nelson (1991) and also assumes log heteroskedasticity.
    4. The definition and properties of the quantile deviation could be further investigated.

Bibliography

C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503, 2002a.

C. Acerbi and D. Tasche. Expected shortfall: a natural coherent alternative to value at risk. *Economic Notes*, 31(2):379–388, 2002b.

S. Alizadeh, M. W. Brandt, and F. X. Diebold. Range-based estimation of stochastic volatility models. *The Journal of Finance*, 57(3):1047–1091, 2002.

A. Ally. Quantile-based methods for prediction, risk measurement and inference. *School of Information Systems, Computing and Mathematics*, 2010.

T. Andersen, T. Bollerslev, F. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, 2003.

K. Athreya. Bootstrap of the mean in the infinite variance case. *The Annals of Statistics*, 15(2):724–731, 1987.

N. Balakrishnan. Progressive censoring methodology: an appraisal. *Test*, 16(2): 211–259, 2007.

N. Balakrishnan and R. Aggarwala. *Progressive censoring: theory, methods, and applications*. Boston: Birkhäuser, 2000.

N. Balakrishnan and D. Han. Optimal step-stress testing for progressively type-i censored data from exponential distribution. *Journal of Statistical Planning and Inference*, 139(5):1782–1798, 2009.

N. Balakrishnan and Q. Xie. Exact inference for a simple step-stress model with type-ii hybrid censored data from the exponential distribution. *Journal of statistical planning and inference*, 137(8):2543–2563, 2007.

N. Balakrishnan, N. Kannan, C. Lin, and S. Wu. Inference for the extreme value distribution under progressive type-ii censoring. *Journal of Statistical Computation and Simulation*, 74(1):25–45, 2004.

N. Balakrishnan, D. Kundu, H. K. Tony Ng, and N. Kannan. Point and interval estimation for a simple step-stress model with type-ii censoring. *Journal of Quality Technology*, 39(1):35–47, 2007.

A. Balkema and L. De Haan. Residual life time at great age. *The Annals of Probability*, 2(5):792–804, 1974.

O. Barndorff-Nielsen. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B*, 64(2):253–280, 2002.

G. Bates. Joint distributions of time intervals for the occurrence of successive accidents in a generalized polya scheme. *The Annals of Mathematical Statistics*, 26(4): 705–720, 1955.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

T. Bollerslev, R. Y. Chou, and K. F. Kroner. Arch modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics*, 52(1):5–59, 1992.

T. Brailsford and R. Faff. An evaluation of volatility forecasting techniques. *Journal of Banking & Finance*, 20(3):419–438, 1996.

L. Canina and S. Figlewski. The informational content of implied volatility. *Review of Financial studies*, 6(3):659–681, 1993.

E. Castillo and A. S. Hadi. Fitting the Generalized Pareto Distribution to Data. *Journal of the American Statistical Association*, 92(440), 1997. ISSN 01621459.

G. Chamberlain. Quantile regression, censoring, and the structure of wages. In *Advances in Econometrics: Sixth World Congress*, volume 2, pages 171–209, 1994.

Q. Chen, R. Gerlach, and Z. Lu. Bayesian value-at-risk and expected shortfall forecasting via the asymmetric laplace distribution. *Computational Statistics & Data Analysis*, 56(11):3498–3516, 2012.

B. J. Christensen and N. R. Prabhala. The relation between implied and realized volatility. *Journal of Financial Economics*, 50(2):125–150, 1998.

A. C. Cohen. Progressively censored samples in life testing. *Technometrics*, 5(3): 327–339, 1963.

R. Davidson and J. G. MacKinnon. Several tests for model specification in the presence of alternative hypotheses. *Econometrica*, pages 781–793, 1981.

A. Davison. Modelling excesses over high thresholds, with an application. *Statistical extremes and applications*, 131:461–482, 1984.

A. Davison and R. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B*, pages 393–442, 1990.

L. De Haan and A. Ferreira. *Extreme value theory*. Springer Science+ Business Media, 2006.

F. Diebold, T. Schuermann, and J. Stroughair. Pitfalls and opportunities in the use of extreme value theory in risk management. *Journal of Risk Finance*, 1(2):30–35, 2000.

D. Dupuis and M. Tsao. A hybrid estimator for generalized pareto and extreme-value distributions. *Communications in Statistics-Theory and Methods*, 27(4):925–941, 1998.

B. Efron. Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, 1(93):125, 1991.

B. Efron and R. Tibshirani. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1(1), 1986.

R. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, pages 987–1007, 1982.

R. Engle and S. Manganelli. Caviar. *Journal of Business and Economic Statistics*, 22(4):367–381, 2004.

R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Univ Press, 1928.

M. B. Garman and M. J. Klass. On the estimation of security price volatilities from historical data. *Journal of business*, pages 67–78, 1980.

M. Gilli and E. Këllezi. An application of extreme value theory for measuring financial risk. *Computational Economics*, 27(2):207–228, 2006.

L. Glosten, R. Jagannathan, and D. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of finance*, pages 1779–1801, 1993.

B. V. Gnedenko. On a local limit theorem of the theory of probability. *Uspekhi Matematicheskikh Nauk*, 3(3):187–194, 1948.

E. Gouno, A. Sen, and N. Balakrishnan. Optimal step-stress test under progressive type-i censoring. *IEEE Transactions on Reliability*, 53(3):388–393, 2004.

S. Grimshaw. Computing maximum likelihood estimates for the generalized pareto distribution. *Technometrics*, pages 185–191, 1993.

A. J. Gross and V. Clark. *Survival distributions: reliability applications in the biomedical sciences*, volume 11. Wiley New York, 1975.

R. D. Gupta and D. Kundu. Exponentiated exponential family: An alternative to gamma and weibull distributions. *Biometrical journal*, 43(1):117–130, 2001.

P. Hall. Asymptotic properties of the bootstrap for heavy-tailed distributions. *The Annals of Probability*, 18(3):1342–1360, 1990.

J. R. Hosking. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B*, pages 105–124, 1990.

J. R. M. Hosking and J. R. Wallis. Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, 29(3):1609–1620, 1987.

A. Huang. Volatility forecasting by quantile regression. *Applied Economics*, 44(4): 423–433, 2012.

S. Huang. Statistical inference in accelerated life testing with geometric process model. *a Master's thesis*, 2011.

J. Hüsler, D. Li, and M. Raschke. Estimation for the generalized pareto distribution using maximum likelihood and goodness of fit. *Communications in StatisticsTheory and Methods*, 40(14):2500–2510, 2011.

A. F. Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348):158–171, 1955.

T. J. Kielpinski and W. Nelson. Optimum censored accelerated life tests for normal and lognormal life distributions. *IEEE Transactions on Reliability*, 24(5):310–320, 1975.

R. Koenker. *Quantile regression*, volume 38. Cambridge university press, 2005.

R. Koenker and G. Bassett. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

R. Koenker and G. Bassett. Robust tests for heteroscedasticity based on regression quantiles. *Econometrica: Journal of the Econometric Society*, pages 43–61, 1982.

R. Koenker and Z. Xiao. Inference on the quantile regression process. *Econometrica*, 70(4):1583–1612, 2002.

R. Koenker and Q. Zhao. L-estimation for linear heteroscedastic models. *Journal of Nonparametric Statistics*, 3(3-4):223–235, 1994.

R. Koenker and Q. Zhao. Conditional quantile estimation and inference for arch models. *Econometric Theory*, 12:793–813, 1996.

T. Kuen and T. Hoong. Forecasting volatility in the singapore stock market. *Asia Pacific Journal of Management*, 9(1):1–13, 1992.

J. F. Lawless. *Statistical models and methods for lifetime data*, volume 54. Wiley-Interscience, 1982.

R. V. Leon, R. Ramachandran, A. J. Ashby, and J. Thyagarajan. Bayesian modeling of accelerated life tests with random effects. *Journal of quality technology*, 39(1): 3–16, 2007.

X. Liu. Planning of accelerated life tests with dependent failure modes based on a gamma frailty model. *Technometrics*, 54(4):398–409, 2012.

X. Liu and L. C. Tang. Accelerated life test plans for repairable systems with multiple independent risks. *IEEE Transactions on Reliability*, 59(1):115–127, 2010.

H. Ma and W. Q. Meeker. Strategy for planning accelerated life tests with small sample sizes. *IEEE Transactions on Reliability*, 59(4):610–619, 2010.

A. J. McNeil. Extreme value theory for risk managers. *Departement Mathematik ETH Zentrum*, 1999.

W. Q. Meeker and W. Nelson. Optimum accelerated life-tests for the weibull and extreme value distributions. *IEEE Transactions on Reliability*, 24(5):321–332, 1975.

R. Miller and W. Nelson. Optimum simple step-stress plans for accelerated life testing. *IEEE Transactions on Reliability*, 32(1):59–65, 1983.

R. G. Miller, G. Gong, and A. Muñoz. Survival analysis, 1981.

D. Nelson. Stationarity and persistence in the garch (l, 1) model. *Econometric theory*, 6(3):318–334, 1990.

D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370, 1991.

W. Nelson. Graphical analysis of accelerated life test data with the inverse power law model. *IEEE Transactions on Reliability*, 21(1):2–11, 1972.

W. Nelson. Analysis of accelerated life test data-least squares methods for the inverse power law model. *IEEE Transactions on Reliability*, 24(2):103–107, 1975.

W. Nelson. Accelerated life testing-step-stress models and data analyses. *IEEE Transactions on Reliability*, 29(2):103–108, 1980.

W. B. Nelson. *Accelerated testing: statistical models, test plans, and data analysis*, volume 344. Wiley-Interscience, 2009.

R. Newcombe. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in medicine*, 17(8):857–872, 1998.

W. Newey and J. Powell. Asymmetric least squares estimation and testing. *Econometrica*, pages 819–847, 1987.

H. K. T. Ng, P. S. Chan, and N. Balakrishnan. Estimation of parameters from progressively censored data using em algorithm. *Computational Statistics & Data Analysis*, 39(4):371–386, 2002.

H. K. T. Ng, P. S. Chan, and N. Balakrishnan. Optimal progressive censoring plans for the weibull distribution. *Technometrics*, 46(4):470–481, 2004.

M. Parkinson. The extreme value method for estimating the variance of the rate of return. *Journal of Business*, pages 61–65, 1980.

F. Pascual. Accelerated life test planning with independent weibull competing risks. *IEEE Transactions on Reliability*, 57(3):435–444, 2008.

E. Pearson and J. Tukey. Approximate means and standard deviations based on distances between percentage points of frequency curves. *Biometrika*, 52(3/4): 533–546, 1965.

J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, pages 119–131, 1975.

J. L. Powell. Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, 25(3):303–325, 1984.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/.

C. R. Rao. *Linear statistical inference and its applications*, volume 22. Wiley-Interscience, 1973.

J. René Van Dorp and T. A. Mazzuchi. A general bayes weibull inference model for accelerated life testing. *Reliability Engineering & System Safety*, 90(2):140–147, 2005.

M. Ribatet. *A Users Guide to the POT Package (Version 1.4)*, 2011. URL http://cran.r-project.org/web/packages/POT/vignettes/POT.pdf. [Accessed 9 December 2011].

L. Rogers and S. Satchell. Estimating variance from high, low and closing prices. *The Annals of Applied Probability*, 1(4):504–512, 1991.

R. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90, 1985.

R. Smith. Estimating tails of probability distributions. *The Annals of Statistics*, 15 (3):1174–1207, 1987.

F. Sobotka and T. Kneib. Geoadditive expectile regression. *Computational Statistics & Data Analysis*, 2010.

N. Tajvidi. Confidence intervals and accuracy estimation for heavy-tailed generalized pareto distributions. *Extremes*, 6(2):111–123, 2003.

L. C. Tang and X. Liu. Planning and inference for a sequential accelerated life test. *Journal of quality technology*, 42(1):103–118, 2010.

L. C. Tang, T. N. Goh, Y. S. Sun, and H. L. Ong. Planning accelerated life tests for censored two-parameter exponential distributions. *Naval Research Logistics (NRL)*, 46(2):169–186, 1999.

D. Tasche. Expected shortfall and beyond. *Journal of Banking & Finance*, 26(7): 1519–1533, 2002.

J. Taylor. Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, 6(2):231–252, 2008a.

J. W. Taylor. Generating volatility forecasts from value at risk estimates. *Management Science*, 51(5):712–725, 2005.

S. Taylor. *Modelling financial time series*. World Scientific Publishing, 2008b.

Y. Tse. Stock returns volatility in the tokyo stock exchange. *Japan and the World Economy*, 3(3):285–298, 1991.

E. G. Tsionas. Bayesian quantile inference. *Journal of statistical computation and simulation*, 73(9):659–674, 2003.

S. A. Vander Wiel and W. Q. Meeker. Accuracy of approx confidence bounds using censored weibull regression data from accelerated life tests. *IEEE Transactions on Reliability*, 39(3):346–351, 1990.

D. VENZON and S. MOOLGAVKAR. A method for computing profile-likelihood-based confidence intervals. *Applied statistics*, 37(1):87–94, 1988.

R. Viveros and N. Balakrishnan. Interval estimation of parameters of life from progressively censored data. *Technometrics*, pages 84–91, 1994.

B. Wang, K. Yu, and M. Jones. Inference under progressively type ii right-censored sampling for certain lifetime distributions. *Technometrics*, 52(4):453–460, 2010.

B. X. Wang. Interval estimation for exponential progressive type-ii censored step-stress accelerated life-testing. *Journal of Statistical Planning and Inference*, 140 (9):2706–2718, 2010.

B. X. Wang and K. Yu. Optimum plan for step-stress model with progressive type-ii censoring. *Test*, 18(1):115–135, 2009.

W. Wang and D. B. Kececioglu. Fitting the weibull log-linear model to accelerated life-test data. *IEEE Transactions on Reliability*, 49(2):217–223, 2000.

A. J. Watkins. Review: Likelihood method for fitting weibull log-linear models to accelerated life-test data. *IEEE Transactions on Reliability*, 43(3):361–365, 1994.

A. J. Watkins and A. M. John. On constant stress accelerated life tests terminated by type ii censoring at one of the stress levels. *Journal of Statistical Planning and Inference*, 138(3):768–786, 2008.

S. Weerahandi. Generalized confidence intervals. *Journal of the American Statistical Association*, 88(423):899–905, 1993.

S. Weerahandi. *Generalized inference in repeated measures: Exact methods in MANOVA and mixed models*, volume 500. Wiley-Interscience, 2004.

S. S. Wilks. Order statistics. *Bulletin of the American Mathematical Society*, 54(1): 6–50, 1948.

D. Yang and Q. Zhang. Drift-independent volatility estimation based on high, low, open, and close prices. *The Journal of Business*, 73(3):477–492, 2000.

G. B. Yang. Optimum constant-stress accelerated life-test plans. *IEEE Transactions on Reliability*, 43(4):575–581, 1994.

I. T. Yu and C. L. Chang. Applying bayesian model averaging for quantile estimation in accelerated life tests. *IEEE Transactions on Reliability*, 61(1):74–83, 2012.

K. Yu and R. A. Moyeed. Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447, 2001.

K. Yu and J. Zhang. A three-parameter asymmetric laplace distribution and its extension. *Communications in StatisticsTheory and Methods*, 34(9-10):1867–1879, 2005.

J. Zhang. Likelihood moment estimation for the generalized Pareto distribution. *Australian & New Zealand Journal of Statistics*, 49(1):69–77, Mar. 2007.

J. Zhang and M. Stephens. A new and efficient estimation method for the generalized pareto distribution. *Technometrics*, 51(3):316–325, 2009.

K. Q. Zhou and S. L. Portnoy. Statistical inference on heteroscedastic models based on regression quantiles. *Journal of Nonparametric Statistics*, 9(3):239–260, 1998.