Quantile-based Methods for Prediction, Risk Measurement and Inference

A thesis submitted for the degree of

Doctor of Philosophy

by

Abdallah K. Ally

Supervised by

Dr. Keming Yu

and

Dr. Veronica Vinciotti

Department Mathematical Sciences
School of Information Systems, Computing and Mathematics
Brunel University, West London

December 2010

Abstract

The focus of this thesis is on the employment of theoretical and practical quantile methods in addressing prediction, risk measurement and inference problems. From a prediction perspective, a problem of creating model-free prediction intervals for a future unobserved value of a random variable drawn from a sample distribution is considered. With the objective of reducing prediction coverage error, two common distribution transformation methods based on the normal and exponential distributions are presented and they are theoretically demonstrated to attain exact and error-free prediction intervals respectively.

The second problem studied is that of estimation of expected shortfall via kernel smoothing. The goal here is to introduce methods that will reduce the estimation bias of expected shortfall. To this end, several one-step bias correction expected shortfall estimators are presented and investigated via simulation studies and compared with one-step estimators.

The third problem is that of constructing simultaneous confidence bands for quantile regression functions when the predictor variables are constrained within a region is considered. In this context, a method is introduced that makes use of the asymmetric Laplace errors in conjunction with a simulation based algorithm to create confidence bands for quantile and interquantile regression functions. Furthermore, the simulation approach is extended to an ordinary least square framework to build simultaneous bands for quantiles functions of the classical regression model when the model errors are normally distributed and when this assumption is not fulfilled.

Finally, attention is directed towards the construction of prediction intervals for realised volatility exploiting an alternative volatility estimator based on the difference of two extreme quantiles. The proposed approach makes use of AR-GARCH procedure in order to model time series of intraday quantiles and forecast intraday returns predictive distribution. Moreover, two simple adaptations of an existing model are also presented.

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".

.....

Abdallah K. Ally

Acknowledgements

The completion of the this thesis owes heartfelt gratitude to a number of individuals who have been instrumental towards its completion in one way or another.

I am heavily indebted to my supervisors Dr. Keming Yu and Dr. Veronica Vinciotti for their invaluable guidance, support and advise throughout the three years of research. I am particularly grateful for their patience and humility in their dealings with me, especially in their effort to unscramble and decipher my explanations during times in which I experienced difficulties in expressing myself as eloquently I as would have liked. I would also like to extend sincere gratitude to Dr. David Cappitt for introducing me to the world of statistics and for his unwavering support, assistance and kindness not only during my PhD but throughout my undergraduate studies.

Special thanks goes to my dear friends in Brunel who have had a profound impact on my life and without whom it would have been very difficult to achieve whatever little I have managed to achieve. Specifically, I would like to thank (alphabetically), Ali Al-Kenani, Dr. Afandi Ahmad, Fahad Alnaim, Fatmir Qirezi, Hakim Mezali, His excellency Dr. Nafiu Hussaini and Rahim Al-Hamzawi.

Last, but by no means least, sincere and heartfelt gratitude goes out to the head of two households Dr. Fahmi Said and his family for their unwavering and perpetual support during times of ease and hardships.

Author's Publication

1. Yu, K. and **A. Ally** (2009). Improving Prediction Intervals: Some Elementary Methods, *The American Statistician*, 63, 17-19.

2. Yu, K., **Ally, A**., Yang, S. and Hand, D. J. (2010) Kernel Quantile based Estimation of Expected Shortfall, *The Journal of Risk*, 12, 15-32. .

3. Yu, K., Liu, W. and **Ally, A** (2010). Simultaneous confidence bands for quantile regression, *Submitted to the Journal of Computational Statistics and Data Analysis*.

4. **Ally, A**., Vinciotti, V. and Yu, K. (2010) Confidence and prediction intervals for integrated volatility, *Submitted to the Journal of Econometrics*.

5. **Ally, A**., Yu, K. and Vinciotti, V (2010). GARCH induced prediction intervals for realised volatility, *Working paper*.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In statistics, distribution functions play a fundamental role in characterising and describing the attributes of population of interest and thus serve as an invaluable tool towards decision making. Traditionally, the mean and standard deviation, the respective measures of location and dispersion, have been extensively employed as parameters that can depict the entire distribution. However, the information content of the mean and standard deviation is limited and there is therefore a need to explore other measures with more attractive properties that may provide profound insight into a distribution. To address such deficiencies, quantiles which divides a distribution into two parts, have recently received great interest as alternative robust statistical methods that transcend beyond the mean framework to capture different features of an entire distribution.

The usage of quantiles have a history stretching over a century when Galton (1889) used a sample of almost a thousand subjects to calculate conditional quartiles of height of sons given the height of their fathers. Realising the need to explore different aspect of a distributional function Galton calls on statisticians not to restrict their investigations to averages but should adjust to finding 'more comprehensive views'. Since then there has been a revolution towards a more holistic outlook of representing and drawing inference from distribution functions, and due to the

fact that quantiles are the components making up any distribution academia has directed more effort towards identifying quantiles features that can assist in problem solving. As a result, quantiles have received immense interest theoretically and found numerous applications in different fields such as medicine, social science and finance.

Although the potential of quantiles in solving statistical problems through modelling and inferencing has been identified in the fall of the nineteenth century, the fruits of this promising new thinking did not materialise until mid to late twentieth century through the work of Tukey (1965) and the publication of books on order statistic by Sarhan and Greenberg (1962) and David (1970) which have a natural implications on the calculation of quantiles. Similarly, since his groundbreaking paper (Parzen (1979)) Parzen has played an instrumental role in the advancement of quantile statistical thinking towards a unified approach that address both frequentist and Bayesian perspectives as well as parametric and non-parametric approaches, see Parzen (1993, 2004).

The introduction of quantile regression by Koenker and Bassett (1978) as an alternative and more rewarding methodology of modelling the relationship between response and predictor variables has proven to be one of the most successful application of quantiles. The original idea has been extended to a non-parametric setting through the work of Yu (1997) amongst others and has found application in financial modelling as substitute to standard techniques, see for example Engle and Manganelli (2004), Koenker and Zhao (1996), Xiao and Koenker (2009).

It is often of economic interest in areas such as finance, insurance seismic analysis and hydrology to calculate extreme quantiles to quantify the effect of some event or outcome under investigation. Hosking and Wallis (1987) introduced an extreme value theory viewpoint in dealing with such problems through studying the estimation of quantiles and parameters of generalised Pareto distribution (GPD).

In assessing the goodness-of-fit of some distribution on specific dataset or in examining similarities between two distributions the work of Hazen (1914) on

quantile-quantile $(q-q)$ plots as well as that of Parzen (1979) on probability-probability $(p-p)$ plots have found numerous applications in visually detection of parameters of interest such as skewness and heavy tails of one distribution in comparison to another.

This conclude a brief historical analysis of some revolutionary and important contributions in the literature that have enhanced the advancement of quantiles methods. The areas of research presented are by no means exhaustive; for a general history on statistics and modelling via quantile functions see Hald (1998) and Gilchrist (2000) respectively.

The focal point of the thesis is on the theoretical and practical application of quantile methods for prediction and risk analysis in the context of bias reduction in estimating expected shortfall and confidence and prediction intervals for integrated volatility. To this end, in the following discussion a formal introduction of quantiles together with some common properties are given.

1.1 Quantiles and Their Properties

Let $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ denote independent and identically distributed random variables from a distribution function F, then for $p \in [0,1]$ the inverse of cumulative distribution function, $F^{-1}(p)$, is the quantile function and is defined as

$$F^{-1}(p) = \inf\{x : F(x) \geq p\}. \tag{1.1}$$

The $100p\%$ quantile ξ_p is obtained as the unique solution to $F^{-1}(p) = \xi_p$. In parallel with the definition of equation (1.1) one can make use of order statistics to define sample quantiles. This is achieved by introducing order statistics $\tilde{X}_1 < \tilde{X}_2 < \ldots, < \tilde{X}_n$ obtained from sorting the original sample in ascending order and defining the $100p\%$ sample quantile as $\hat{\xi}_p = \tilde{X}_{[p(n+1)]}$, where $[a]$ is the integer part of a.

From a different angle quantiles can be represented as solution of an optimisation

problem as follows. Given a random variable $\mathbf{X}$ from a distribution function F with a corresponding density f then the objective is to find a quantity ξ_p by minimising the expected loss

$$E\rho_p(X-\xi_p) = (p-1)\int_{-\infty}^{\xi_p}(x-\xi_p)f(x)dx + p\int_{\xi_p}^{\infty}(x-\xi_p)f(x)dx, \tag{1.2}$$

where the loss function

$$\rho_p(\gamma) = pI(\gamma \geq 0) + (1-p)I(\gamma < 0).$$

To obtain a minimum of equation (1.2) one differentiate with respect to ξ_p resulting in

$$\frac{\partial}{\partial \xi_p}E\rho_p(X-\xi_p) = (1-p)\int_{-\infty}^{\xi_p}f(x)dx - p\int_{\xi_p}^{\infty}f(x)dx$$

and on equating to zero, one finds the solution as $\xi_p = F^{-1}(p)$. In the event that there exist a number of solutions, in consistence with the definition of equation (1.1) the convention is to choose the lowest value, Koenker (2005). In a more computational friendly format the same problem can be presented as

$$\min_{\xi_p}\frac{1}{n}\left(p\sum_{x_i>\xi_p}|x_i-\xi_p| + (1-p)\sum_{x_i\leq\xi_p}|x_i-\xi_p|\right).$$

When F is continuous with density function f the equivalent relations linking density, distribution and quantile functions

$$f(x) = \frac{d}{dx}F(x); \qquad x = F^{-1}(p).$$

Similarly, Tukey (1965) first noted that the relationship between distribution and density functions can be extended to define a new analogous quantile density measure

which he called the sparsity function

$$q_p = \frac{d(F^{-1}(p))}{dp}.$$

Parzen (1979) present an almost identical quantity which he termed density quantile function defined as $f_p = f(F^{-1}(p))$. These quantities play an important role when calculating confidence intervals for regression quantile coefficients as well as modelling with quantile functions as illustrated by Koenker (2005) and Gilchrist (2000).

The advantage, attraction and practical usefulness of quantiles spring from their properties. To this end, in what follows some common properties of quantiles are briefly explained.

Robustness against outliers is a key inherent property of quantiles. This attribute is of significant importance in many application and thus consequently elevate the usage of quantiles and quantile-based techniques particularly in those fields in which outliers pose the risk of distorting the final outcome of some analysis of interest.

Another attractive property of quantile functions is their equivariance to monotone transformation. This can be illustrated by considering a monotonic function g that is used to create a transformed random variable $\mathbf{Y} = g(\mathbf{X})$. The quantiles of the new random variable $\mathbf{Y}$ are obtained by transforming the quantile function of $\mathbf{X}$ such that $\xi_p^Y = g(F_X^{-1}(p)) = g(\xi_p^X)$.

Furthermore, a well known usage of quantiles is in their ready interpretability when calculating confidence and prediction intervals. In this respect, for a pre-specified confidence level, quantiles naturally form the upper and lower bounds enclosing some unknown parameter or future observation.

1.2 Quantile-based Measures of Location, Scale, Skewness and Kurtosis

A distribution function can be characterised by infinite number of quantiles spanning its support. This view lends itself useful in re-defining moment-based estimators in terms of quantiles.

The mean has long been employed as measure of location. However, there are occasions when it is more informative to observe and analyse different parts of a distribution and the flexibility of quantiles allows one do just that by shedding light on and examining any location of interest. This idea form the building block of quantile regression which unlike its ordinary least square counterpart, lies in its ability to capture heterogeneity and characterising the entire conditional distribution of the variable of interest given its covariates.

Quantiles, although implicitly, have always been used in estimating the the second moment of a distribution. Specifically, the range which correspond to the difference between the extreme quantiles

$$R = F^{-1}(1) - F^{-1}(0)$$

has for quite some time been used as a measure of variability, Tippett (1925). Other well known quantile measures of scale are the interquartile range (IQR) and twice the interquartile range (IQR2),

$$IQR = \xi_{0.75} - \xi_{0.25}, \qquad IQR2 = 2(\xi_{0.75} - \xi_{0.25}).$$

Using the difference of two extreme quantiles, Pearson and Tukey (1965) proposed quantile-based measure of standard deviation defined as

$$\tilde{\sigma} = \frac{\xi_{1-p} - \xi_p}{C(p)},$$

where the order of quantile p are chosen to be 0.01, 0.025, and 0.05 with corresponding correction constant $C(p)$ as 4.65, 3.92 and 3.25 respectively. Moment-based measure of skewness is the standard procedure used to quantify the asymmetry of distributions. However, quantiles due to their robustness against outliers have the potential to accurately and simply capture deviations from normality. One such measure is that of Hinkley (1975)

$$QSK_3 = \frac{\xi_{1-p} - \xi_p - 2\xi_{0.5}}{\xi_{1-p} - \xi_p} \qquad 0 < p < 0.5$$

which is a generalised version of coefficient of skewness presented earlier by Bowley (1920). Similarly, Crow and Siddiqui (1967) uses extreme quantiles to define quantile-based measure of kurtosis as

$$QKU = \frac{\xi_{1-p} + \xi_p}{\xi_{1-\eta} - \xi_\eta}$$

where $0 < p, \eta < 1$.

1.3 Moment and Quantile-based Risk Measures

The need to quantify and manage risk is of fundamental importance in finance. Great effort has been invested in developing risk measurement techniques that can bridge the gap between financial institutions, practitioners and regulators, with the goal of obtaining accurate and easy to implement methodologies for measuring risk.

Financial markets have always been risky but over the last three to four decades this phenomenon has become more apparent. One possible reason is an increase in volatility across a spectrum of financial products such as equity returns, interest and foreign exchange rates, see Dowd (2002). Similarly, another contributor to an upsurge in volatility emanates from the expansion in the magnitude of derivative products such as options whose returns are generally riskier than those of the underline stocks, Ruppert (2006). For these reasons together with infamous cases that have led to the

collapse of institutions such as that Baring Banks, Orange County and Enron greatly accentuate the need for companies to not only understand their exposure to risk but also monitor and understand the risks incurred by by individual employees as a result of their dealings.

From a finance point of view risk is defined as being made up of uncertainty and exposure components, see Holton (2004). The author agues that due to its multi-facet it is very difficult to define perceived risk while it is possible to define components of risk using risk metrics measures such as the standard deviations of returns of an asset or maximum likely credit exposure.

In the proceeding subsection brief overviews of popular risk measures are given with the aim of creating a link between these measures with quantiles.

1.3.1 Standard Deviation

In finance, standard deviation (also known as volatility) as popularised by the modern portfolio theory of Markowitz (1952, 1959) has become one of the most widely used risk measure. The set up of the modern portfolio theory hinged on the assumption of the normality of asset returns and from this Markowitz (1952, 1959) concluded that a natural risk measure is the standard deviation. In this respect, the problem of efficient portfolio selection can be viewed as that of maximising expected returns coupled with the condition of minimising risk (standard deviation), which for a random variable R representing financial asset returns the quantity is defined as

$$\sigma = \sqrt{E(R - E(R))^2}.$$

It is by now well documented that the distribution of financial are not normally distributed, see for instance Jondeau et al. (2006). When the distribution under investigation is Gaussian standard deviation can be employed as a measure of dispersion, however, when such an assumption is violated the second moment may give a distorted picture of dispersion, see Poon and Granger (2003). To circumvent

this drawback other measures such as interquartile range and mean absolute return have been proposed.

Over the last two decades, the availability of high frequency data has resulted in the adaptation of the model free continuous measure of variance. The new popular estimator of true unobserved volatility is known as realised variance, see McAleer and Medeiros (2008) for a review. Realised variance obtained as the summation of squared high frequency returns over a given day and thus serve as a model free estimator of the true unobserved variance called integrated variance.

From a risk management perspective the employment of quantiles have been instrumental in the introduction of the so called quantile-based risk measures. Two such measures that have been extensively used by regulators as well as financial institutions are value at risk (VaR) and expected shortfall (ES).

1.3.2 Value at Risk

In the early 90's a new risk measuring technique known as value at risk has since revolutionised the risk management world has emerged from the RiskMetrics methodology developed by JP Morgan. Value at risk of a portfolio is defined as an upper bound such that for a specified time horizon and confidence level $100(1-p)\%$ the market value of the portfolio will depreciate beyond the bound with probability $p \in (0,1)$. In most industry application the value of p is chosen to be either 1% or 5%, which can be interpreted to indicate the the risk-averseness of the decision-maker.

For a given a vector $\{r_t\}_{t=1}^{T}$ denoting the evolution of returns of a portfolio over time, the calculation of VaR can be considered as that of forecasting the conditional quantiles of the distribution of portfolio returns. Mathematically, the next period VaR calculated at time $t-1$ can be expressed as

$$\Pr(r_t \leq \xi_p | \mathcal{F}_{t-1}) = p$$

such that

$$VaR_t(p) = -\xi_p,$$

where $\mathcal{F}_{t-1}$ denotes the information set available up until time $t-1$.

A major breakthrough in the acceptance of VaR by regulators followed as result of the Basle Committee on Banking Supervision (BCBS) requesting banks to use their internal models, subject to calibration, to estimate probable losses and set aside capital to cover for market risk based on the outcome of their estimations. Furthermore, the recognition and acceptance of VaR as a risk measure amongst financial institutions is due to its ability to summarise the market risk of a portfolio to just a single monetary value together with the ease with which this information can be communicated to the concerned parties.

1.3.3 Expected Shortfall

Despite it popularity and wide acceptance by regulators and financial institutions, VaR as a risk measure has received a number of criticisms. The major of these criticisms center on the inability of VaR to satisfy all the conditions of a coherent risk measure as proposed by Artzner et al. (1999). Specifically, for the four properties defining a coherent risk measure; monotonicity, homogeneity, translational invariance and subadditivity VaR satisfy the first three and falls short of the sub-additive property. Another criticism of VaR come forth from the fact that this risk measure sets an upper bound on the maximum loss, and thus its inability to account for losses beyond the confidence level and may therefore underestimate the true risk associated with a portfolio. To this end, Artzner et al. (1999) introduced expected shortfall (ES) as an alternative coherent risk measure directly related to VaR.

Expected shortfall is defined as the expected loss of a portfolio provided the

bound set by VaR is violated. In mathematical terms, this can be expressed as

$$ES_t(p) = -\mathbb{E}(r_t | r_{t-1} \leq \xi_p).$$

1.4 Thesis Outline

The outline of the thesis is as follows. Chapter 2 addresses the problem of constructing prediction intervals for future values of a random variable drawn from a sample distribution. Two simple methods based on the Gaussian and exponential distributions transformation are proposed with the focus on improving the coverage accuracy of prediction intervals. The Gaussian distribution method is constructed using the well known Box-Cox transformation and is theoretically shown to produce exact prediction intervals. The exponential distribution method is demonstrated to attain error free prediction intervals and is shown to have a natural extension from a parametric to a nonparametric setting. Furthermore, the practical accuracy of the two proposed methods are studied through simulation experiments. The application of the proposed methods encompasses a wide range of problems ranging from the well known usage of creating prediction intervals for regression models to quality control and manufacturing, where, for instance, one is interested in creating intervals for future defective products. The types of data employed by the proposed methods are independent and identically distributed (*iid*).

The focus of Chapter 3 is on a number of kernel quantile based estimators including jacknife based bias-correction estimators which have theoretically been proven to reduce bias. Bias-reduction is particularly effective in reducing the tail estimation bias as well as the consequential bias that arises in kernel smoothing and finite sampling fitting, and thus serves as a natural approach to the estimation of extreme quantiles of assets prices distribution. These estimators are studied numerically through simulation and real data example to assess their competitiveness in estimating expected shortfall. Although, by construction, the proposed methods

are aimed to be used in conjunction with (*iid*) data. Through filtering, as demonstrated at the end of Chapter 3, the application of these estimators can be readily extended beyond (*iid*) and thus account for heteroscedasticity observed within real financial data.

Chapter 4 address the problem of constructing simultaneous confidence bands for quantile regression functions when the covariates are constrained within a region. Through exploiting the relationship between the asymmetric Laplace distribution and quantile regression a simulation-based method employing the active set algorithm of (Liu et al., 2004, 2005) is presented. It is also shown that the simulation based algorithm can be readily employed to create confidence bands for the difference of two regression functions. Furthermore, attention is also focused on the construction of confidence bands for quantile functions of classical ordinary least square regression when the model errors are both normally distributed and when they are not. Through a different outlook of the problem it is demonstrated that the problem can be solved by alternative optimisation techniques, namely simulated annealing/acceptance threshold and in some special cases, depending on the regression model, as one dimension constrained optimisation. The practicality of the methods are examine through empirical studies using the coverage probabilities as a distinguishing criterion. From a data and application point of view, the methods presented in this chapter can be employed in numerous fields, such as medicine, economics and social science, where the aim is to model the relationship between a dependent variable and a set of covariates for different quantiles, and to attach a measure of uncertainty on the true quantile regression function. Crucially, the difference between pointwise and simultaneous bands considered in this chapter is that, for the latter the bands are constructed, given the covariates, simultaneously on the whole range or a subset of the regression function. For application, consider a situation where for a given regression model one has a number of possible candidates for a particular quantile, then simultaneous bands can be used to distinguish between the possible candidates based on the coverage.

In Chapter 5 we take a practical approach to creating prediction intervals for RV. In particular we exploit the quantile-based estimator of volatility originally proposed by Pearson and Tukey (1965) to create prediction intervals for realised volatility. Using the presentation of the difference of two extreme quantiles we establish a link with realised volatility calculated using intraday returns. An AR-GARCH model is employed to model intraday time series of quantiles in order to forecast next period density of intraday returns and from which prediction intervals are calculated using the distribution of the difference of two quantiles. Furthermore, we present a bootstrap approach of this idea as well as introduce two adaptations of the Heterogeneous AutoRegressive (HAR) model of Corsi (2009) with a view to predict end points of a prediction interval for realised volatility. Thereafter, we apply the proposed methods to two equities and assess their coverage accuracy. This chapter calls on high frequency financial data, that is data in which the price of an asset is recorded in equally spaced time periods over the course of a trading day.

Finally, Chapter 6 summarises the main results of the research as well as proposing recommendations for possible future research directions.

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

Improving the Accuracy of Prediction Intervals

In this chapter the problem of constructing prediction intervals for future values of a random variable drawn from a sample distribution is considered. Two simple calibration methods based on distribution transformation are proposed and are shown both theoretically and via simulation study to improve the coverage accuracy of the prediction intervals.

2.1 Introduction

Prediction is one of the fundamental problems tackled by statistics. Consequently, there exist numerous statistical techniques such as time series and regression models designed to address different aspects of forecasting.

Traditionally, forecasters have relied on point forecasts as a helpful tool in decision making. While the production of a point forecast serves as an important function, this approach overlooks a vital aspect of dealing with uncertainty associated with such prediction. In failing to consider the likely range of outcomes, the decision maker foregoes key information which can assist in reaching more instructive and

economical conclusions.

Two related concepts that are widely used to account for point prediction uncertainties are prediction intervals and density forecasts, as lucidly discussed by Chatfield (1993) and Tay et al. (2000) respectively. Density forecasts of a random variable involve the estimation of the entire distribution function depicting the spectrum of possible future values of the random valuable under study.

Prediction intervals, also known as interval forecasts is a notion that is of a middle path between the two extremes of a point and a density forecast and can be viewed as a special case of the latter. Prediction intervals of a random variable are constructed by estimating upper and lower bound quantiles such that for a pre-specified probability the future values of random variable drawn independently from the original sample will be captured within the interval. The idea of prediction intervals is closely related to that of confidence intervals; the latter attach probability statements regarding the likely values of a fixed but unobservable population parameter whereas the former provide bounds on future observations. A summary of applications of prediction intervals including a predictive application of the Bayesian approach are given by Geisser (1993), Lawless and Fredette (2005), Hamada et al. (2004) and the references therein, and for a review of time series prediction intervals see Chatfield (2001).

The problem can be formally posited as follows: Let $\mathbf{X} = \{X_1, ..., X_n\}$ denote a random sample of size n from a distribution with distribution function F. Furthermore, for $p \in (0, 1)$ let $\xi_p = F^{-1}(p)$ denote a distribution's $100p\%$ quantile assumed to be uniquely defined. Additionally, let $\hat{\xi}_p$ be the sample estimator of ξ_p, and suppose that X denotes a random variable independent of the original sample but drawn from the same population as that of $\mathbf{X}$. By defining two end point statistics, $\hat{\xi}^{\mathbf{X}}_{p/2}$ and $\hat{\xi}^{\mathbf{X}}_{1-p/2}$, obtained from the original sample $\mathbf{X}$, then $100(1-p)\%$ symmetric prediction interval for X can be written as

$$\Pr[\hat{\xi}^{\mathbf{X}}_{p/2} < X < \hat{\xi}^{\mathbf{X}}_{1-p/2}] = 1 - p, \tag{2.1}$$

where the quantity $1-p$ is known as the nominal level. Analogously, a level p one sided prediction interval for X can be defined as $\Pr[X < \hat{\xi}_p^{\mathbf{X}}] = p$.

The objective is to investigate if there exist a sample estimator $\hat{\xi}_p$ of the distribution $100p\%$ quantile, ξ_p, such that equations

$$\Pr[X < \hat{\xi}_p] = p \tag{2.2}$$

or

$$\mathbb{E}_F\{\Pr[X < \hat{\xi}_p]\} = p \tag{2.3}$$

are satisfied.

Equation (2.2) defines an exact prediction interval for X while equation (2.3) which is equivalent to $\mathbb{E}\{F(\hat{\xi}_p)\} = p$ defines a prediction interval for X with zero coverage error.

The main ingredient in achieving the objective rests on the estimation of the quantile. In this respect, it follows from the edgeworth expansion of the distributions of order statistics (see, Reiss (1989) for example) that by substituting the $100p\%$ sample quantile in place of $\hat{\xi}_p$ the relation given by equation (2.3) is the most accurate attainable. That is, by defining the corresponding order statistics of the original sample as $Y_1 < Y_2 < \ldots, Y_n$ and taking the estimator of $100p\%$ quantile ξ_p as $\tilde{\xi}_p = Y_{[(n+1)p]}$, where $[a]$ is the integer part of a, then coverage probability of the interval $(-\infty, \tilde{\xi}_p)$ has a coverage error of order n^{-1} as a prediction interval for future data values:

$$P[X \leq \tilde{\xi}_p] = p + \mathcal{O}(n^{-1}), \tag{2.4}$$

in the sense of $\mathbb{E}\{F(\tilde{\xi}_p)\} - p = \mathcal{O}(n^{-1})$.

It can be deduced that the accuracy of equation (2.4) depend crucially on the estimation of the quantile ξ_p. To this end, a number of nonparametric methods have been proposed to improve the coverage accuracy. For instance, Hall and Rieck (2001) have put forward a calibration method that is based on the extrapolation of

adjacent order statistics. Specifically, the authors show that if $\tilde{\xi}_p$ is an appropriate interpolation among three and five order statistics it follows that the coverage error can be reduced from $\mathcal{O}(n^{-1})$ to orders $\mathcal{O}(n^{-3})$ and $\mathcal{O}(n^{-4})$ respectively. Furthermore, Hall et al. (1999) and Hall and Rieck (2001) respectively demonstrate the validity of the bootstrap and smoothed bootstrap techniques in reducing the coverage error of prediction intervals. The smoothed bootstrap is shown, under specific conditions on the choice of the smoothing parameter, to reduce the coverage error from the inverse of sample size to $\mathcal{O}(n^{-2})$ and $\mathcal{O}(n^{-3})$. From a Bayesian school of thought, Sweeting (1999) among others, presents an approach of obtaining approximate zero coverage probability bias associated with equation (2.4).

The focus of the chapter is on the traditional frequentist stance with the motivation centering on proposing methods that can go beyond reducing the order of the coverage error to those with exact or zero coverage error. The following subsection presents exact prediction interval based on the Box-Cox transformation and in section 2.3 a zero coverage error exponential distribution based prediction interval is proposed.

2.2 Normal Distribution-based Method

Let $\mathbf{X} = \{X_1, ..., X_n\}$ denote a random sample from the normal population $N(\mu, \sigma^2)$, with parameters μ and σ^2 unknown, then the $100p\%$ quantile of the normal distribution is given by $\xi_p = \mu + \sigma\Phi^{-1}(p)$, where $\Phi(.)$ is the standard normal cumulative distribution. Instead of the popular order statistic based pth sample quantile estimate, $\tilde{\xi}_p$, an alternative estimate for ξ_p built from the estimated parameters is given by

$$\widehat{\xi}_p = \overline{X} + cS_n,$$

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean, $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is the sample variance and c is a known quantity which depends only on the sample size n and centile p.

When c is chosen as

$$c = \Phi^{-1}(p)\sqrt{\frac{n-1}{2}}\frac{\Gamma((n-1)/2)}{\Gamma(n/2)},$$

where $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1}\exp(-u)du$ denotes the gamma function, then it can be shown that (see, Rao (1973)) $\tilde{\xi_p}$ gives an unbiased and even minimum variance estimate of ξ_p. However, similar to the $100p\%$ sample quantile estimator, this estimate has a predictive accuracy up to order n^{-1}. That is, for a random variable X independent of sample $\mathbf{X}$, $P[X < \hat{\xi}_p] = p + \mathcal{O}(n^{-1})$.

It can be noted that the parameter set $(\overline{X}, S_n^2)$ is the minimum sufficient statistic for the normal distribution, and thus should provide the same information as any interpolation of order statistic for improving coverage accuracy of prediction interval.

By selecting the constant c as

$$c = \sqrt{\frac{n+1}{n}}\ t_{n-1}^p, \tag{2.5}$$

where t_{n-1}^p denotes the $100p\%$ quantile from a student-t distribution with $(n-1)$ degree of freedom, then one can construct exact prediction intervals as stated by the following theorem.

Theorem 1: If X is a random variable independent of a normally distributed random sample $\mathbf{X}$, estimate ξ_p by

$$\widehat{\xi}_p = \overline{X} + cS_n,$$

with constant c given by (2.5), then $\Pr[X < \hat{\xi}_p] = p$.

The derivation of the exact prediction interval, $\Pr[X < \hat{\xi}_p] = p$, directly follows from the fact that $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ and thus $X - \overline{X} \sim N\left(0, (1+\frac{1}{n})\sigma^2\right)$, then $\frac{X-\overline{X}}{S_n\sqrt{1+\frac{1}{n}}} \sim t_{n-1}$ and

$$\Pr[X < \hat{\xi}_p] = \Pr[X < \overline{X} + cS_n] = \Pr\left[\frac{X-\overline{X}}{S_n} < c\right] = p.$$

This conclusion shows that, when the original sample $\mathbf{X}$ is from a normal distribution an exact prediction interval for a new random variable X can be readily obtained.

The derivation of the exact prediction interval rest on the assumption that the original sample is independent and identically drawn from a normal distribution. Although, the normal distribution is extensively applied in both theoretical and practical statistics, it is often the case that the normality assumption fail to hold. If a random sample $\mathbf{X} = \{X_1, ..., X_n\}$ is from an unknown distribution F, then there exist different approaches to transform it to be approximately normal with the most employed of these known as the Box-Cox transformation, (see Sakia (1992) for a review).

To transform a random sample $\mathbf{X} = \{X_1, ..., X_n\}$ to one that is approximately normally distributed, say $\mathbf{Y} = \{Y_1, ..., Y_n\}$, Box and Cox (1964) propose a transformation function

$$\mathbf{Y} = g_\lambda(\mathbf{X}) = \begin{cases} \frac{\mathbf{X}^\beta - \mathbf{1}}{\beta}, & \text{when } \beta \neq 0 \\ \log(\mathbf{X}) & \text{when } \beta = 0, \end{cases} \tag{2.6}$$

where β is a transformation parameter to be estimated. Without loss of generality, the Box-Cox transformation techniques (2.6) only admits strictly positive random variables, however, there are variants of other similar transformations (see Sakia (1992)) that can accommodate any real values. Generally, consider a random variable $X \sim F(x)$ and a random variable $Y = G_\beta(X) \sim N(\mu_\beta, \sigma_\beta^2)$ where $G(.)$ is a known monotonic function and the inverse $G^{-1}(.)$ exist. Let ξ_p be the $100p\%$ quantile of X, and ξ_p^Y be the pth quantile of Y, then

$$\xi_p^Y = \mu_\beta + \sigma_\beta \Phi^{-1}(p),$$

and

$$\xi_p = G_\beta^{-1}(\xi_p^Y).$$

From the preceding discussion under normal distribution and according to Theorem 1, it follows that if ξ_p^Y is estimated by

$$\hat{\xi}_p^Y = \bar{Y} + cS_{Y,n},$$

with the constant c given by equation (2.5) then for any random variable Y independent of sample $\mathbf{Y} = \{Y_1, ..., Y_n\}$, $P[Y < \hat{\xi}_p^Y] = p$. Now define the estimate of ξ_p by

$$\hat{\xi}_{p,\beta} = G_\beta^{-1}(\hat{\xi}_p^Y), \tag{2.7}$$

thus, for any a random variable X independent of sample $\mathbf{X}$,

$$P[X \leq \hat{\xi}_{p,\beta}] = P[G_\beta(X) \leq \hat{\xi}_p^Y] = P[Y < \hat{\xi}_p^Y] = p.$$

In order to examine the small sample performance of the proposed method and to facilitate numerical comparison simulations are conducted based on four different populations: D1, standard normal; D2, standard exponential; D3, two-parameter Weibull with shape and scale parameters equal to 2 and 1 respectively; D4, standard log-normal. For each population model 5000 replications are used.

Let the coverage error be the difference between the coverage probability and the nominal coverage; $P[X < \hat{\xi}_p] - p$. Table 2.1 lists approximate coverage errors obtained from simulations for $p = 0.90$ where the transformation parameter λ is calculated using the MATLAB function *boxcox* which makes use of unconstrained linear optimisation where the likelihood function, derived from assuming that the transformed data is normally distributed, is maximised. The results indicate that the method produce very accurate coverage with increased accuracy for large sample sizes. However, as it can be noted from Table 2.1 for $p = 0.90$ the proposed method appear to overestimate.

Table 2.1: Coverage errors (multiplied by 1000) with $p = 0.9$ based on Box-Cox normal transformation.

n	$D1$ (Norm)	$D2$ (Exp)	$D3$ (Weib)	$D4$ (Lognorm)
12	11.0	12.0	13.0	15.0
14	9.1	9.7	11.0	11.0
16	7.7	8.2	8.8	9.4
18	6.7	7.1	7.5	8.0
50	3.7	4.4	4.8	4.8
100	2.1	2.3	2.4	2.4

2.3 Exponential Distribution-based Methods

Due to its attractive features, the most celebrated and noteworthy of which is the memoryless property, the exponential distribution has become one of the most extensively employed in both theoretical and applied statistics.

Let $\mathbf{X} = \{X_1, ..., X_n\}$ be a positive exponential distributed random sample with respective density and distribution functions $f(x) = \frac{1}{\lambda}\exp(-\frac{1}{\lambda}x)$ and $F(x) = 1 - \exp(-\frac{1}{\lambda}x)$. Then the $100p\%$ quantile of the distribution is given by

$$\xi_p = -\lambda \log(1-p). \tag{2.8}$$

A natural unbiased estimate of ξ_p can be achieved by replacing the unknown parameter λ by its maximum likelihood estimator $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

However, estimating the quantile ξ_p with $\hat{\xi_p} = -\hat{\lambda}\log(1-p)$ does not improve the predictive accuracy, in the sense that for a new random variable X independent of sample $\mathbf{X}$, $\mathbb{E}\{\Pr[X < \hat{\xi_p}]\} < p$. The proof of this arises from noting that the sum of independent and identically exponential random variates follows a Gamma distribution, that is, $Y \equiv \sum_{i=1}^{n} X_i \sim \Gamma(n, \lambda)$, where

$$\Gamma(n, \lambda) = \frac{y^{n-1}}{\Gamma(n)\lambda^n}\exp\left(-\frac{1}{\lambda}y\right). \tag{2.9}$$

The proportion p of the area below $\hat{\xi_p}$ is

$$\begin{aligned} E\{P[X < \hat{\xi_p}]\} &= 1 - \int_0^\infty \exp\left(-\frac{1}{\lambda}\log(1-p)^{-\frac{1}{n}}y\right) \frac{y^{n-1}}{\Gamma(n)}\frac{1}{\lambda^n}\exp(-\frac{1}{\lambda}y)dy \\ &= 1 - \int_0^\infty \frac{y^{n-1}}{\Gamma(n)}\frac{1}{\lambda^n}\exp\left(-\frac{1}{\lambda}[1+\log(1-p)^{-\frac{1}{n}}]y\right)dy \\ &= 1 - [1+\log(1-p)^{-\frac{1}{n}}]^{-n} \\ &\times \int_0^\infty \frac{y^{n-1}}{\Gamma(n)}\frac{[(1+\log(1-p)^{-\frac{1}{n}})]^n}{\lambda^n}\exp\left(-\frac{1}{\lambda}[1+\log(1-p)^{-\frac{1}{n}}]y\right)dy \\ &= 1 - [1+\log(1-p)^{-\frac{1}{n}}]^{-n}.1 \\ &= 1 - [1 - \frac{1}{n}\log(1-p)]^{-n}. \end{aligned} \tag{2.10}$$

Given that $\log\left(1-\frac{1}{n}\log(1-p)\right)^n = n\log\left(1-\frac{1}{n}\log(1-p)\right)$ then using the logarithmic series, $-\log(1-z) = \sum_{k=1}^\infty \frac{z^k}{k}$ for $|z| < 1$, one can express the last component of equation (2.10) as

$$-n\log\left(1-\frac{1}{n}\log(1-p)\right) = -n\sum_{k=1}^\infty \frac{(\log(1-p))^k}{kn^k}.$$

Subsequently, it follows that

$$-\log(1-p) - n\sum_{k=2}^\infty \frac{(\log(1-p))^k}{kn^k} < -\log(1-p),$$

and thus $1 - [1 - \frac{1}{n}\log(1-p)]^{-n} < p$.

To circumvent the deficiency in the predictive accuracy, ξ_p can be estimated by an alternative calibrator

$$\hat{\xi_p} = \left(\frac{1}{(1-p)^{1/n}} - 1\right)\sum_{i=1}^n X_i, \tag{2.11}$$

then for a new random variable X independent of sample $\mathbf{X}$, a zero coverage prediction interval

$$\mathbb{E}\{P[X < \hat{\xi}_p]\} = p,$$

is attained.

The estimator $\hat{\xi}_p$ defined in equation (2.11) is not an unbiased estimator of ξ_p, however it provides a prediction interval with zero coverage error. As a matter of fact, let $c_n(p) = \frac{1}{(1-p)^{1/n}} - 1$, and $Y = \sum_{i=1}^{n} X_i$, then $Y \sim \Gamma(n, \lambda)$,

$$E_F\{F(\hat{\xi}_p)\} = 1 - E\{\exp(-\frac{1}{\lambda}c_n(p)Y)\}$$

$$= 1 - \int_0^\infty \exp(-\frac{1}{\lambda}c_n(p)y)\ \frac{y^{n-1}}{\Gamma(n)}\frac{1}{\lambda^n}\exp(-\frac{1}{\lambda}y)dy$$

$$= 1 - (1-p)\ \int_0^\infty \frac{y^{n-1}}{\Gamma(n)\{\lambda(1-p)^{1/n}\}^n}\exp(-\frac{1}{\lambda(1-p)^{1/n}}y)dy = 1 - (1-p) = p.$$

Due to

$$\int_0^\infty \exp(-\frac{1}{\lambda}c_n(p)y)\frac{y^{n-1}}{\Gamma(n)}\frac{1}{\lambda^n}\exp(-\frac{1}{\lambda}y)dy = \int_0^\infty \exp(-c_n(p)y)\frac{y^{n-1}}{\Gamma(n)}\exp(-y)dy,$$

the conclusion of zero prediction coverage error, $\mathbb{E}\{P[X < \hat{\xi}_p]\} = p$, is independent of the selection of parameter λ and holds true for all $\lambda > 0$. When $\lambda = 1$ which corresponds to the simple exponential distribution $Exp(1)$, $\hat{\xi}_p = c_n(p)\sum_{i=1}^{n} X_i$ provides a predictive interval end point with zero coverage error.

In the following discussion, the the idea of the exponential distribution method is extended to general distributions.

Consider a sample $\mathbf{X} = \{X_1, ..., X_n\}$ from a general but known distribution F with an existing and unique inverse function F^{-1}. Let Y denote an exponentially distributed random variable such that $Y = -\log F(X)$ and $Y_i = -\log F(X_i)$. Further, let $d_n(p) = \frac{1}{p^{1/n}} - 1$ and

$$\hat{Y}_p = d_n(p)\sum_{i=1}^{n} Y_i; \tag{2.12}$$

if the $100p\%$ quantile $\xi_p = F^{-1}(p)$ is estimated by

$$\hat{\xi}_{p,F} = F^{-1}(\exp(-\hat{Y}_p)), \tag{2.13}$$

then it is still the case that, for any random variable X independent of sample $\mathbf{X}$,

$$E_F P[X < \hat{\xi}_{p,F}] = p.$$

By noting that,

$$P[X < \hat{\xi}_{p,F}] = P[-\log F(X) > -\log F(\hat{\xi}_{p,F})] = P[Y > \hat{Y}_p],$$

together with the fact that $Y \sim Exp(1)$, then the $\hat{Y}_p$ given by equation (2.12) equals a constant times a Gamma distribution $\Gamma(n, 1)$. Consequently, it follows that $EP[X < \hat{\xi}_{p,F}] = E\exp(-\hat{Y}_p) = p$.

On substituting $Y_i = -\log F(X_i)$ into equation (2.12) it can be observed that an alternative expression for $\hat{\xi}_{p,F}$ is given by

$$\hat{\xi}_{p,F} = F^{-1}\left((\prod_{i=1}^{n} F(X_i))^{\frac{1}{p^{1/n}}-1}\right). \tag{2.14}$$

Theorem 2: If X is a random variable independent of sample $\mathbf{X}$ from distribution F, then $\hat{\xi}_{p,F}$ given by equation (2.14) provides a prediction interval with zero coverage error for X.

Table 2.2: Coverage errors (multiplied by 1000) with $p = 0.9$ based on exponential distribution method.

n	$D1$ (Norm)	$D2$ (Exp)	$D3$ (Weib)	$D4$ (Lognorm)
12	0.18	-0.55	0.43	-0.30
14	-0.48	0.16	-0.22	0.32
16	-0.37	0.14	0.31	0.32
18	0.085	-0.16	0.015	0.30
50	0.014	-0.070	-0.064	-0.010
100	0.0040	0.020	-0.054	0.0071

To illustrate the accuracy of prediction interval $(-\infty, \hat{\xi}_{p,F})$, Table 2.2 display approximate coverage errors obtained from 5000 simulations for the same four

distributions as employed for Table 2.1. Unlike Table 2.1, Table 2.2 contains both positive and negative coverage errors. It is also worth noting that the proposed prediction interval is comparable with $I5$ of Hall and Rieck (2001) which in their simulations is shown to have the lowest values of coverage error.

The zero coverage quantile estimator of equation (2.14) is not restricted to a known functional form distribution function, F. The idea readily extends to a nonparametric data driven methods for estimating F. For instance, the cumulative distribution function F can be replaced by an empirical function, F_n, or a smoothed distribution function, $F_{n,h}$, with a smoothing parameter h.

2.4 Chapter Summary

The main idea and findings of this chapter can be summarised as follows.

- Two novel and easy-to-implement methods are proposed for improving the coverage accuracy of prediction interval. The methods are constructed based on well known distribution transformations and are theoretically shown to provide exact and zero coverage error prediction intervals.
- The exponential distribution-based method is demonstrated to have a natural extension to a nonparametric setting, which maybe useful when one is not prepared to assume the distributional functional form of a random variable under study.
- The numerical study, encompassing the analysis of small sample performance, indicate that both the normal and exponential transformation methods produce very accurate coverage for prediction intervals.

Chapter 3

Kernel Quantile-based Estimation of Expected Shortfall

The focus of this chapter is on the introduction of kernel-based expected shortfall estimators with the objective of reducing bias. Bias-reduction technique is particularly effective in reducing the tail estimation bias as well as the induced bias arising from kernel smoothing and finite sampling fitting, and thus serve as a natural approach to the estimation of extreme quantiles of return distribution. By taking advantage of the integral representation of expected shortfall, a new type of expected shortfall estimator is proposed. The performance of the proposed estimators are investigated through simulation studies and the methods are applied to real data.

3.1 Introduction

The ability to accurately and meaningfully measure the risk associated with a portfolio plays an important role in market risk management. The most common and widely used risk measure is value at risk (VaR), which under normal market conditions is defined as the maximum potential loss of a portfolio over a prescribed holding period and for a given confidence level. Since its emergence from JP

Morgan and subsequent acceptance and endorsement by the Basel Committee in 1996 (Basle Committee Banking Supervision (1996)) and in the latest proposed Basel II norms (Basle Committee on Banking Supervision (2003)) the estimation of VaR has been the subject of abundant research. Following the Basle guidelines, financial regulators have adopted VaR for designing capital adequacy standard for banks and financial institutions. In addition, financial firms have adopted VaR for internal risk management and allocation of resources, see (Danelsson et al. (2005)).

Shortly after the introduction of VaR as a benchmark for calculating the market risk of a portfolio research was underway to determine the meaning of economically meaningful risk measure. To this end, Artzner et al. (1997) introduced the concept of coherent risk measures (see also Artzner (1999), Artzner et al. (1999) and references therein). They argued that a risk measure should satisfy a set of four desirable properties: monotonicity, sub-additivity, positive homogeneity and translation invariance. In this context Artzner (1999) demonstrate that VaR is not a coherent risk measure due the fact that it fails to satisfy the sub-additivity property. This implies that the risk of a portfolio can be larger than the sum of stand-alone risks of its components when measured by VaR, the consequence of which goes against the modern portfolio theory which states that diversification leads to reduction in risk.

In order to construct a risk measure that is both coherent as well as easy to compute and estimate, the expected shortfall (ES), defined as the expectation of losses above VaR for a given time horizon, was proposed and discussed by Acerbi et al. (2001). Furthermore, Acerbi and Tasche (2002) then provided an integral presentation of ES while Acerbi and Tasche (2001) showed that for $\alpha \in (0, 1)$ ES arises in a natural way from the estimation of the "average of the $100\alpha\%$ worst losses" in a sample of returns of a portfolio.

Estimating and forecasting VaR has received a considerable amount of attention in the literature (see Jorion (2001), Manganelli and Engle (2001)) and there has recently been an increase interest in nonparametric estimation of ES. One of the simplest nonparametric approach for computing ES is the sample average estimator

in which first VaR is calculated as the quantile from the empirical distribution of returns and thereafter ES is computed as the average of realizations falling below the VaR estimate.

With the intention on improving the accuracy of expected shortfall estimate, Scaillet (2004) proposed an alternative estimator to the sample average that is based on kernel distribution smoothing. Chen (2008) studied the sample average and kernel estimator proposed by Scaillet (2004) and contrary to expected result, the author conclude that in comparison with the simple sample weighted average estimator, kernel smoothing does not produce more accurate calculation of ES.

The truncated mean estimator together with the kernel smoothing method proposed by Scaillet (2004) can all be classified to fall under the umbrella of two-step nonparametric methods: first nonparametric estimation of VaR, then nonparametric estimation of the expectation of a variable truncated at the VaR estimate. Fermanian and Scaillet (2005) also explore some interesting applications of these methods in credit risk environment.

While the estimator of Scaillet (2004) is interesting, it may have a substantial bias arising from the boundary effects of kernel estimation for small probability levels. In fact Chen (2008) demonstrate that kernel smoothing induces bias and since the variance is not reduced this consequently leads to an increase in mean square error.

In this chapter an integral representation which provides mathematical tractability for studying the analytic properties of expected shortfall is exploited. Using the integral expression, one-step nonparametric expected shortfall estimation method is presented and a number of kernel based estimators including bias-corrected ones are proposed.

3.2 Expected Shortfall

Let X be the random variable describing the future value of the profit or loss of a portfolio at some fixed time horizon T from today, and $\alpha \in (0,1)$ be a probability level. Usually, α is taken as a small percentage of the profit and loss distribution. The quantile with level α is defined as

$$q(\alpha) = \sup\{x|P(X \leq x) \leq \alpha\},$$

and, in keeping with the convention of reporting losses as positive, VaR is defined by

$$VaR(\alpha) = -q(\alpha). \tag{3.1}$$

If the truncated mean exists, that is $E[X^-] < \infty$, where X^-=max$(-X, 0)$, then the tail conditional expectation and the expected shortfall are respectively defined as

$$TCE(\alpha) = -E\{X|X \leq q(\alpha)\},$$

and

$$ES(\alpha) = -\frac{1}{\alpha}\{E[XI(X \leq q(\alpha))] - q(\alpha)[P(X \leq q(\alpha)) - \alpha]\} \tag{3.2}$$

where $I(A)$ is an indicator function taking a value of unity when A is satisfied and zero otherwise. Delbaen et al. (2000) reported that the TCE does not in general satisfy sub-additivity. Acerbi and Tasche (2001) proved that while $ES(\alpha)$ satisfies sub-additivity, in general TCE(α) does not. However, if X is a continuous random variable, then

$$ES(\alpha) = TCE(\alpha) = -\frac{1}{\alpha}E[XI(X \leq q(\alpha))]. \tag{3.3}$$

Based on the structure of truncated expectation, Scaillet (2004) presented a nonparametric kernel estimator of ES associated with a portfolio, and derived the

asymptotic properties of the kernel estimator and its first-order derivative with respect to portfolio allocation in the context of a stationary process satisfying strong mixing conditions. Its asymptotic performance and optimal bandwidth were considered by Chen (2008).

The proceeding discussion draws on the fact that ES can be expressed as an integral of a quantile function via a simple integral transformation. That is

$$ES(\alpha) = -\frac{1}{\alpha}\int_0^\alpha q(p)dp, \tag{3.4}$$

see Pflug and Römisch (2007), for example.

This expression provides mathematical tractability for studying the analytical properties of ES. For instance, it is clear from (3.4) that ES(α) is continuous in α while this is not obvious from (3.2). Moreover, from (3.4) it can be observed that given the quantile function, $q(p)$, is specified, ES(α) can be estimated via an explicit computation.

3.3 Kernel Density and Distribution Estimation

This section reviews the concept of kernel density and distribution function estimation which form the foundation of the expected shortfall estimation method to be introduced afterwards. The material of this section relies heavily on the work of Wand and Jones (1995) and Silverman (1986).

3.3.1 Kernel Density Estimation

Given a sequence of n independent and identically distributed observations $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ from an unknown univariate probability density function f with corresponding distribution function F, the kernel density estimation of f is given

by

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right),$$

where $K(u)$ is a kernel function and h is a strictly positive constant called the bandwidth.

In most applications, the kernel function is chosen to be a positive zero-centred symmetric density function satisfying

$$\int_{-\infty}^{\infty} K(u)du = 1, \qquad \int_{-\infty}^{\infty} uK(u)du = 0$$

$$\int_{-\infty}^{\infty} K^2(u)du < \infty, \qquad \int_{-\infty}^{\infty} u^2K(u)du = \sigma_K^2$$

where $0 < \sigma_K^2 < \infty$. Despite the existence of numerous kernel functions, it is by now well documented (see Wand and Jones (1995)) that $K(u)$ has minimal influence on the accuracy of the density estimator and as result a natural choice that has gained great popularity is the Gaussian density $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$. Furthermore, the usage of density functions such as the Gaussian, as kernels, ensures that the density estimator $\hat{f}_h(x)$ inherits all the continuity and differentiability properties of $K(u)$.

The selection mechanism of the scale parameter h is of crucial importance in kernel density estimation, and consequently this area has been the subject of on going research focusing on identifying data-driven procedures of selecting the parameter and more importantly reduce the associated bias, see Jones et al. (1996) for a discussion. In order to choose h and differentiate between alternatives it is necessary to establish a measure of closeness between the estimated $\hat{f}_h(x)$ and target density function $f(x)$. A widely used criterion for finding the optimal bandwidth h is to minimise the Mean Integrated Square Error (MISE) . Let $Bias(\hat{f}_h(x)) = E(\hat{f}_h(x)) - f(x)$, MISE can be

expressed as

$$\int MSE(\hat{f}_h(x))dx = \int \left[Var\left(\hat{f}_h(x)\right) + \left(Bias^2\hat{f}_h(x)\right)\right]dx$$

where $MSE(\hat{f}_h(x)) = E\left[(\hat{f}_h(x) - f(x))^2\right]$.

Under the assumption of continuity of the second order derivatives of $f(x)$ it has been shown (see Silverman (1986)) that the respective variance and bias of a kernel estimator are

$$Var\left(\hat{f}_h(x)\right) = \frac{f(x)}{nh}\int K^2(u) + o\left(\frac{1}{nh}\right)$$

and

$$Bias(\hat{f}_h(x)) = \frac{h^2}{2}\sigma_K^2 f''(x) + o(h^2).$$

The last two equations illustrate the trade off between the bias and variance of kernel estimator in the minimisation of MISE. That is, as the scale parameter h increases, variance decreases while bias increases and thus resulting in oversmoothed density estimate that obscure possible important features. As h decreases the opposite effect is observed.

3.3.2 Kernel Distribution Function Estimation

The simplest and most commonly used used estimator of the cumulative distribution function is the empirical distribution function, defined as

$$\hat{F}_n(x) = n^{-1}\sum_{i=1}^{n} I(x_i \leq x),$$

where I is an indication function. By calling upon the standard result $F' = f$, then a natural smooth alternative estimator to distribution function is the kernel distribution

function estimator (Nadaraya, 1964) defined as

$$\hat{F}_h(x) = \int_{-\infty}^{x} \hat{f}_h(u)du = \frac{1}{n}\sum_{i=1}^{n} A_h(x - x_i),$$

where $A_h(u) = A(u/h)$ and

$$A(x) = \int_{-\infty}^{x} K(u)du. \tag{3.5}$$

By considering second-order properties of the estimator $\hat{F}_h(x)$ and the assumption of continuity of $f(x)$ and existence of $f^{'}(x)$, Azzalini (1981) has shown that as $n \to \infty$ MSE of kernel distribution function can be approximated by

$$ah^4 - ch/n + F(x)\{1 - F(x)\}, \tag{3.6}$$

where

$$\sqrt{a} = \frac{f^{'}(x)}{2}\int_{-b}^{b} \sigma_K^2 du, \qquad c = f(x)\left(b - \int_{-b}^{b} A^2(u)du\right).$$

3.4 Nonparametric Estimation of Expected Shortfall

Let $q(p)$ be a $100p\%$ $(0 < p < 1)$ quantile of F. Then there are three basic kernel quantile function estimators of q_p given by:

(1) distribution function based estimation from the inverse solution of

$$F_h(x) = p; \tag{3.7}$$

(2) density function based estimation from the integral solution of

$$\int_{-\infty}^{q(x)} f_h(x)dx = p; \tag{3.8}$$

(3) kernel weighted sum of order statistics (Parzen, 1979):

$$\hat{q}(p) = \sum_{i=1}^{n} \left[\int_{\frac{i-1}{n}}^{\frac{i}{n}} K_h(t-p)dt \right] x_{(i)}, \tag{3.9}$$

where $x_{(1)}$, $x_{(2)}$, ..., $x_{(n)}$ are the sample order statistics.

The above three kernel quantile estimators (3.7), (3.8), and (3.9) are shown (see Sheather and Marron (1990) and Cheng and Sun (2006)) to have equivalent asymptotic performance under the asymptotic mean square error (AMSE) criterion:

$$AMSE = \frac{p(1-p)}{n} v^2(p) + \frac{h^4}{4}(v'(p)\sigma_K)^2 - \frac{h}{n} v^2(p)\psi(K),$$

where $v = q(p)'$, $v' = q(p)''$, $\sigma_K^2 = \int_{-\infty}^{\infty} u^2 K(u)du$ and $\psi(K) = 2\int uK(u)A(u)du$.

Let $\{x_i\}_{i=1}^n$ denote the returns of a profit and loss distribution. The sample average ES estimator is defined as

$$ES^{\star}(\alpha) = -\frac{\sum_{i=1}^{n} x_i I(x_i \leq \hat{q}^{\star}(\alpha))}{\sum_{i=1}^{n} I(x_i \leq \hat{q}^{\star}(\alpha))} \tag{3.10}$$

where $I(.)$ is an indication function and $\hat{q}^{\star}(\alpha) = x_{([(n+1)\alpha])}$ is the value at risk (100α sample quantile estimator) obtained from the $[(n+1)\alpha]$-th order statistic.

The idea of kernel estimator proposed by Scaillet (2004) is based on replacing the indicator of equation (3.10) by a smooth kernel distribution function such that a two-step kernel estimation of ES from the basic definition (3.3) is given by

$$\widetilde{ES}_h(\alpha) = -\frac{1}{n\alpha} \sum_{i=1}^{n} x_i A_h(\hat{q}(\alpha) - x_i),$$

where $\hat{q}(\alpha)$ is the solution of $F_h(x) = \alpha$ and $A(x)$ is the integrated kernel (3.5) with $A_h(u) = A(u/h)$.

Alternatively, a one-step kernel estimator of ES from equation (3.4) is given by

$$\widehat{ES}_h(\alpha) = -\frac{1}{\alpha} \int_0^{\alpha} \widehat{q}(p)dp.$$

Clearly, $\widetilde{ES}_h(\alpha)$ is a sum-type estimator, whereas $\widehat{ES}_h(\alpha)$ is an integral-type estimator.

From asymptotic results on kernel estimation (Parzen, 1979), the bias of all these estimators is $\mathcal{O}(h^2)$ as $h \to 0$. Theoretically, it is difficult to compare the performance of a one-step kernel estimator to a two-step one. However, in the proceeding discussion some analytical properties of these estimators including asymptotic bias and mean-square errors are considered. Under some conditions, to be discussed in Section (3.5), these asymptotic properties do support $\widehat{ES}_h(\alpha)$ over $\widetilde{ES}_h(\alpha)$.

3.5 Analytic Properties of $\widetilde{ES}_h(\alpha)$ and $\widehat{ES}_h(\alpha)$

In comparing the properties of the two estimators the following basic assumptions from Chen and Tang (2005) are employed.

A1 The process $\{X_i : 1 \leq i \leq n\}$ is strictly stationary and α-mixing, and there exists a $\rho \in (0,1)$ such that the mixing coefficient $\alpha(k) \leq C\rho^k$ for all $k \geq 1$. X_1 is continuously distributed with f and F as its density and distribution functions respectively.

A2 $f(q(p)) > 0$ where $p \in (0,1)$ and f has continuous second derivative in a neighborhood $B(q(p))$ of $q(p)$. The second partial derivatives of F_k, which is the joint distribution function of (X_1, X_{k+1}) for $k \geq 1$, are all bounded in $B(q(p))$ uniformly with respect to k.

A3 Let the kernel function $A(t) = \int_{-\infty}^{t} K(u)du$, where K is a univariate probability density function with continuous bounded second derivative and satisfies the following moment conditions:

$$\int_{-\infty}^{\infty} uK(u)du = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} u^2K(u)du = \sigma_K^2 < \infty.$$

A4 The smoothing bandwidth h satisfies $h \to 0, nh^{3-\tau} \to \infty$ for any $\tau > 0$ and $nh^4 \log^2(n) \to 0$ as $n \to \infty$.

Assumptions **A3** and **A4** are the most commonly used in kernel smoothing literature with the latter imposing the range of values admissible by the bandwidth, h. Similarly, **A2** capture standard conditions for kernel quantile estimation and is required when evaluating bias and mean square error of $f(\hat{q}(p))$. Finally, the stationarity and α-mixing condition **A1** are satisfied by a number of common stochastic models such as the GARCH, continuous-time diffusion and stochastic volatility models.

Under the assumptions **A1**–**A4**, Chen and Tang (2005) gave the mean square error of $\widehat{q}(p)$ as follows

$$\begin{aligned} MSE(\widehat{q}(p)) = \quad & n^{-1}\sigma^2(p;n)f^{-2}(q(p)) - 2n^{-1}h\,b_K\,f^{-1}(q(p)) \\ & + \frac{1}{4}h^4\,\sigma_K^4\{f'(q(p))f^{-1}(q(p))\}^2 + o(\frac{h}{n} + h^4), \end{aligned} \tag{3.11}$$

where $b_K = \int_{-\infty}^{\infty} uK(u)G(u)du$, $\sigma^2(p;n) = \{p(1-p) + 2\sum_{k=1}^{n-1}(1-k/n)\gamma(k)\}$ and $\gamma(k) = cov\{I(X_1 < q(p)), I(X_{k+1} < q(p))\}$ for positive integers k. Clearly, $\sigma^2(p;n) = p(1-p)$ for independent process.

From (3.11) an upper bound for the mean square error of one-step estimator $\widehat{ES}_h(\alpha)$ can be expressed as:

$$\begin{aligned} MSE(\widehat{ES}_h(\alpha)) \quad \leq \quad & \frac{2}{n\alpha^2}\left(\int_\beta^\alpha \sigma^2(p;n)f^{-2}(q(p))dp\right) \\ + \quad & \frac{h^4\sigma_K^4}{4\alpha^2}\left[\int_\beta^\alpha f'(q(p))f^{-1}(q(p))dp\right]^2 - \frac{4hb_K}{n\alpha^2}\left(\int_\beta^\alpha f^{-1}(q(p))dp\right) \\ + \quad & o(h/n + h^4). \end{aligned}$$

Owing to mathematical complexity, it is difficult to compare the two type of estimators by their mean square errors. However, one may be able to compare their biases. In fact, from Chen and Tang (2005) and Chen (2008) the bias of a quantile estimator is:

$$\text{Bias}(\widehat{q}(p)) = -\frac{1}{2}h^2\sigma_K^2 f'(q(p))f^{-1}(q(p)) + o(h^2) \tag{3.12}$$

and the bias of a two step kernel ES estimator is

$$\text{Bias}(\widetilde{ES}_h(\alpha)) = -\frac{1}{2\alpha}h^2\sigma_K^2 f(q(\alpha)) + o(h^2). \tag{3.13}$$

Using the fact that $q'(p) = f^{-1}(q(p))$ and equation (3.12), the bias of a one step ES estimator can be expressed as:

$$\text{Bias}(\widehat{ES}_h(\alpha)) \quad = \quad \frac{1}{2\alpha}h^2\sigma_K^2\left[f(q(\alpha)) - \lim_{\beta\to 0} f(q(\beta))\right] + o(h^2). \tag{3.14}$$

Remark 1: Asymptotically, equation (3.13) shows that the two-step kernel estimator always underestimates ES, but integral-type estimator may not underestimate it, depending on the sign of $1 - \lim_{\beta\to 0} f(q(\beta))/f(q(\alpha))$ from equation (3.14).

Remark 2: The condition $|1 - \lim_{\beta\to 0} f(q(\beta))/f(q(\alpha))| \leq 1$ holds for some (but not all) distributions. In fact, the following observations can be made:

1. Because $f(q(\beta))$ tends to the left end of the support of $f(x)$ where $f(x)$ is usually increasing in the left tail, when β tends to zero. This implies $f'(q(\beta)) > 0$ and $1 - f(q(\beta))/f(q(\alpha)) < 1$ provided $f'(q(\beta)) > 0$;

2. It can be seen that $f(q(\beta))/f(q(\alpha)) \geq 0$ when β is very small. For example, under normal distribution with $f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$, we may see that $0 < \frac{f(q(\beta))}{f(q(\alpha))} < 1$.

3.5.1 Bias Reduction of Kernel Estimators

According to the kernel based jackknife rule (see Jones and Signorini (1997) among others) if $\hat{\eta}_h$ is the kernel estimator of η_h with bias $\mathcal{O}(h^2)$, then as $h \to 0$,

$$\tilde{\eta} = 2\hat{\eta}_h - \hat{\eta}_{\sqrt{2}h}$$

theoretically improves bias from $\mathcal{O}(h^2)$ to $\mathcal{O}(h^4)$. There is usually data scarcity in the tail of a distribution, especially in the far tail. Consequently, for small p, proper kernel estimation of the quantile $q(p)$ is difficult. The simulation study in Section 3.6 shows that the bias reduction technique is particularly effective for the estimation of $ES(\alpha)$ when α is small.

3.5.2 One and Two Steps Kernel Estimators

In this section a number of two-step kernel ES estimators are introduced. Furthermore, for comparison purpose a two-step estimator together with its bias reduction version is also presented.

The two-step sum type kernel estimator and its bias reduced form are given by:

Est1(2S): $\widetilde{ES}_h(\alpha) = \frac{1}{n\alpha}\sum_{i=1}^{n} x_i A_h(\hat{q}(\alpha) - x_i)$, where $\hat{q}(\alpha)$ is a kernel estimator of $q(\alpha)$.

Est2(2S)$_{br(1)}$: the bias-reduction version $\widetilde{ES}_r(\alpha) = 2\widetilde{ES}_h(\alpha) - \widetilde{ES}_{\sqrt{2}h}(\alpha)$.

Two-step kernel estimation requires selection of the smoothing parameter h twice; initially to estimate value at risk and the second time to smooth the excessive losses. The selection of h is of crucial importance and is well understood to be a difficult task especially for smoothing the tails of underlying distributions with possible data scarcity. Hence, the fewer times h is selected the simpler the estimator. To this end, the following three one-step kernel estimators are considered.

Est3(1S)$_{dens}$:$\widehat{E}S_f(\alpha) = -\frac{1}{\alpha}\int_0^{\alpha} \hat{q}(p)dp$, with $\hat{q}(p)$ being estimated by kernel

density function;

$Est4(1S)_{dist}$: $\widehat{E}S_F(\alpha) = -\frac{1}{\alpha}\int_0^\alpha \hat{q}(p)dp$ with $\hat{q}(p)$ being estimated by kernel distribution function; and

$Est5(1S)_{order}$: $\widehat{E}S_{Kq}(\alpha) = -\frac{1}{\alpha}\int_0^\alpha \hat{q}(p)dp$ with $\hat{q}(p)$ being estimated by kernel weighted order statistics.

Note that $Est5(1S)_{order}$ can be expressed explicitly as $\widehat{E}S_{Kq,h}(\alpha) \equiv \widehat{E}S_{Kq}(\alpha) = -\frac{1}{\alpha}\sum_{i=1}^{n} x_{(i)} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \int_0^\alpha K_h(t-p)dp\,dt$. Furthermore, a bias reduced version of this estimator,

$$Est6(1S)_{br(5)} : \widehat{ES}_{Kqr}(\alpha) = 2\widehat{ES}_{Kq,h}(\alpha) - \widehat{ES}_{Kq,\sqrt{2}h}(\alpha)$$

is considered. Additionally, *Remark 1* of section 3.5 shows that the two-step kernel estimator almost always underestimates ES but the one-step kernel estimator may not.

Henceforth, where it is convenient the above estimators shall be referred to as estimators 1 to 6 respectively.

3.6 Monte Carlo Study

In this section simulation experiments are conducted to assess the performance of the six estimators discussed in the previous section. In all simulations the standard normal density is fixed as the kernel function and the bandwidth selection rules proposed by Sheather and Jones (1991) and Bowman et al. (1998) are employed for kernel density and distribution estimation respectively.

Model 1: Normal Distribution

Suppose that the return of a financial asset X is normally distributed such that

$$X \sim N\left(\mu, \sigma^2\right), \tag{3.15}$$

then

$$E[XI(X \leq q(\alpha))] = \mu\Phi\left(\frac{q(\alpha)-\mu}{\sigma}\right) - \frac{\sigma}{\sqrt{2\pi}}e^{-\frac{[q(\alpha)-\mu]^2}{2\sigma^2}}.$$

Hence the true value of ES is

$$ES(\alpha) = \frac{\sigma}{\alpha}\phi\left(\frac{q(\alpha)-\mu}{\sigma}\right) - \mu,$$

where and $\phi(.)$ and $\Phi(.)$ respectively denote the standard normal density and distribution functions.

Although a normal distribution is widely applied, it is by now well documented that financial asset returns have a distinct feature of leptokurtosis and are therefore non-Gaussian. A possible remedy to this problem is to consider mixture distributions, see (Zhang and Cheng, 2005; Giacomini et al., 2008) and the references therein.

Model 2: Mixture Normal Distribution

Assume that X is the return of an asset with a mixture normal distribution

$$f(x) = \tau f_1(x) + (1-\tau)f_2(x), \tag{3.16}$$

where $f_1(x)$ and $f_2(x)$ are the density functions of $N(a_1, \sigma_1^2)$ and $N(a_2, \sigma_2^2)$ respectively. Then under (3.16) the true value of ES is

$$\begin{aligned} ES(\alpha) \quad = \quad & \frac{\tau\sigma_1}{\alpha}\phi\left(\frac{q(\alpha)-a_1}{\sigma_1}\right) + \frac{(1-\tau)\sigma_2}{\alpha}\phi\left(\frac{q(\alpha)-a_2}{\sigma_2}\right) \\ & -\tau a_1 - (1-\tau)a_2. \end{aligned}$$

Model 3: Mixture t Distribution

Assume that X is the return of an asset with a mixture Student-t distribution

$$f(x) = \tau f_1(x) + (1-\tau)f_2(x), \tag{3.17}$$

where $f_1(x)$ and $f_2(x)$ are respectively the density functions of Student-t distribution with degrees of freedom 3 and 5. Under (3.17) the true value of ES is given by

$$ES(\alpha) = \frac{1}{\alpha}\left(\frac{9\tau}{2\sqrt{3\pi}\Gamma(3/2)(3+q(\alpha)^2)} + \frac{125(1-\tau)}{2\sqrt{5\pi}\Gamma(5/2)(5+q(\alpha)^2)^2}\right).$$

Before proceeding to the calculation of ES, the finite sample performance of the three kernel quantile estimators described in Section 3.4 are assessed empirically. Table 3.1 shows the average bias induced by the three quantile estimators for a sample size of 100 and 1000 replications. From the table it can be observed that there are no significant differences in results obtained from the three estimators. The table further confirms the fact that the accuracy of the kernel quantile estimators diminish as one attempts to estimate extreme quantiles.

Table 3.1: Mean bias comparison of the kernel quantile estimators based on cumulative distribution, density and order statistics for a sample size of 100 with standard error in brackets multiplied by 100.

	Distribution		Density		Order statistics	
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$
Model 1	-0.0173	-0.0127	-0.0171	-0.0127	0.0056	0.0092
	(0.2883)	(0.1818)	(0.2881)	(0.1810)	(0.3433)	(0.2077)
Model 2	0.0210	0.0148	0.0210	0.0148	0.0256	0.0186
	(0.0762)	(0.0492)	(0.0762)	(0.1809)	(0.0904)	(0.0554)
Model 3	0.7270	0.4226	0.8316	0.4226	0.7526	0.5242
	(7.6468)	(0.0492)	(9.3140)	(0.1809)	(12.8185)	(2.6595)

To investigate the accuracy of the six estimators, now attention is directed towards the estimation of ES for all three models discussed in Section 3.5.1 at loss probability levels 0.01 and 0.05 and sample sizes n=100 and 300, where the mixing proportion $\tau = 0.5$ and mean and variance in (3.15) are chosen to be 0.05 and 0.01 respectively.

Furthermore, since the extreme value theory (EVT) approach has gained great popularity in estimating ES (see McNeil (1999), and the references therein) this method will be included in the comparison study by fitting a generalised Pareto distribution (GPD) to the the lower tail of the distribution as explained in McNeil and Frey (2000).

Table 3.2 reports the bias of the ES estimators and 95 percent bootstrapped confidence bands of biases with sample size 100 based on 1000 replications. The results for sample size 300 are presented in Table 3.3.

Numerical results show that, in most cases, all kernel-based methods tend to underestimate the theoretical ES and this fact is consistent with the theoretical result explored in remark 1 of Section 3.5. However, these estimators have different performances. First, despite having similar asymptotic results and with reference to remark 1 of Section 3.5, the four non-bias reduced estimators (1, 3, 4 and 5) have different bias performances under finite samples $n = 100$ and $n = 300$. In particular, at both 1% and 5% levels the two step estimator, estimator 1, performs better than the two one-step non-bias reduced estimators (3, 4) but more or less the same as another one-step non-bias reduced estimator 5. Secondly, on comparing the performance of estimator 2 to estimator 1 and that of estimator 6 to estimator 5, it can be observed that, in agreement with asymptotic results, bias-reduction does improve the estimation accuracy with finite sample. Additionally, it is worth noting that, the biases from estimator 1 to 4 maybe partially caused by the kernel estimation of $q(p)$ in the first place, whereas estimators 5 and 6 do not rely on any initial estimation of $q(p)$. Finally, it can be observed that the one-step bias reduced estimator 6, derived from estimator 5, performs best with the smallest bias and narrower confidence bands.

Table 3.2 empirically demonstrate that the EVT method also underestimate the true expected shortfall.

In order to gain a better understanding of the bias as a function of α, Figure 3.1

Table 3.2: Bias of kernel estimation of ES with sample size $n = 100$ where the bias is multiplied by 1000. 95% bootstrapped confidence interval are listed in brackets.

$\alpha = 0.01$	*Est1(2S)*	$Est2(2S)_{br(1)}$	$Est3(1S)_{dens}$	$Est4(1S)_{dist}$	$Est5(1S)_{order}$	$Est6(1S)_{br(5)}$	EVT
Model 1	-18.0	-15.5	-31.9	-19.8	-9.7	-2.7	-15.7
	(-64.6,2.7)	(-57.7,2.4)	(-63.3,-4.2)	(-39.3,-3.0)	(-51.4,12.8)	(-37.4, 10.6)	(-17.4,-13.5)
Model 2	-8.7	-2.4	-25.0	-10.8	22.3	-6.3	-5.3
	(-29.2,10.2)	(-21.7,14.2)	(-41.4,-8.6)	(-17.3,-3.7)	(5.4, 39.1)	(-16.8,2.0)	(-9.4, 0.8)
Model 3	-2.7	-1.9	-5.0	12.1	-2.1	-1.3	-8.2
	(-12.4,5.9)	(-8.3,3.3)	(-10.1,-1.1)	(2.3,14.4)	(-7.1,-4.2)	(-4.3,7.8)	(-20.9, 0.30)
$\alpha = 0.05$							
Model 1	-4.9	-1.5	-13.2	-9.3	-1.4	-1.4	-13.4
	(-16.1,3.7)	(-6.4,5.4)	(-20.7,-5.3)	(-14.8,-3.5)	(-12.0,7.2)	(-11.3,6.9)	(-16.0, -9.2)
Model 2	-34.6	-5.0	-60.1	-16.6	-11.1	1.8	-4.8
	(-131.2,3.4)	(-75.7,22.1)	(-123.6, -8.1)	(-30.8,-2.4)	(-52.3,4.5)	(-83.4,5.1)	(-8.8, -1.2)
Model 3	-3.1	-1.7	-3.3	-9.4	12.0	-1.1	-7.4
	(-18.5,1.1)	(-10.4,3.5)	(-21.3,1.4)	(-10.7,0.6)	(-8.9,22.1)	(-7.6,8.8)	(-14.1, -3.0)

Table 3.3: Bias of kernel estimation of ES with sample size $n = 300$ where the bias is multiplied by 1000.
(95% bootstrapped confidence interval are listed in brackets).

$\alpha = 0.01$	*Est1(2S)*	$Est2(2S)_{br(1)}$	$Est3(1S)_{dens}$	$Est4(1S)_{dist}$	$Est5(1S)_{order}$	$Est6(1S)_{br(5)}$	EVT
Model 1	-5.1	-4.9	-8.1	-2.2	6.3	-1.3	-16.3
	(-19.2,2.6)	(-10.1, 6.7)	(-16.3,-2.1)	(-3.8, -1.2)	(-8.5,13.6)	(-7.3, 2.1)	(-37.3,-5.3)
Model 2	-17.7	-18.0	-27.7	-7.4	21.2	-4.8	-4.27
	(-67.7,13.2)	(-62.9,15.3)	(-52.5,-7.1)	(-12.6,-3.4)	(-16.5,51.0)	(-25.2,7.1)	(-10.4,7.2)
Model 3	-149.4	-144.5	-260.4	-83.5	124.5	-45.4	-126.4
	(-912.2,315.0)	(-942.6,338.1)	(-696.8,29.3)	(-182.7,-5.3)	(26.6,217.1)	(-360.4,119.8)	(-136.8,-117.2)
$\alpha = 0.05$							
Model 1	-1.4	-2.4	-3.5	-1.5	7.5	-0.9	-11.3
	(-4.7,2.2)	(-3.1,1.8)	(-5.2,-2.4)	(-1.9,-0.7)	(5.2,10.4)	(-1.5,1.3)	(-30.3,-0.34)
Model 2	-4.4	-4.4	-12.1	-5.2	2.6	-2.9	-2.7
	(-16.3,5.8)	(-17.2,6.0)	(-18.5,-6.4)	(-8.2,-2.9)	(17.0,35.8)	(-8.1,2.2)	(-17.7,6.4)
Model 3	-70.8	-65.4	-181.2	-77.2	71.1	-41.4	-22.5
	(-304.2,151.9)	(-290.1,153.0)	(-307.6,-0.67)	(-122.2,-40.4)	(48.1,96.8)	(-141.8,44.5)	(-68.6,21.3)

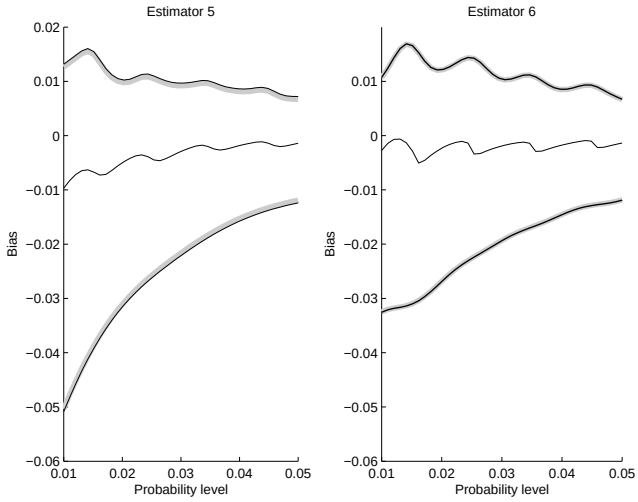

Figure 3.1: Based on model 1, the estimated bias with 95% bootstrapped confidence bands for $Est5(1S)_{order}$(left) and $Est6(1S)_{br(5)}$ (right) with $n = 100$.

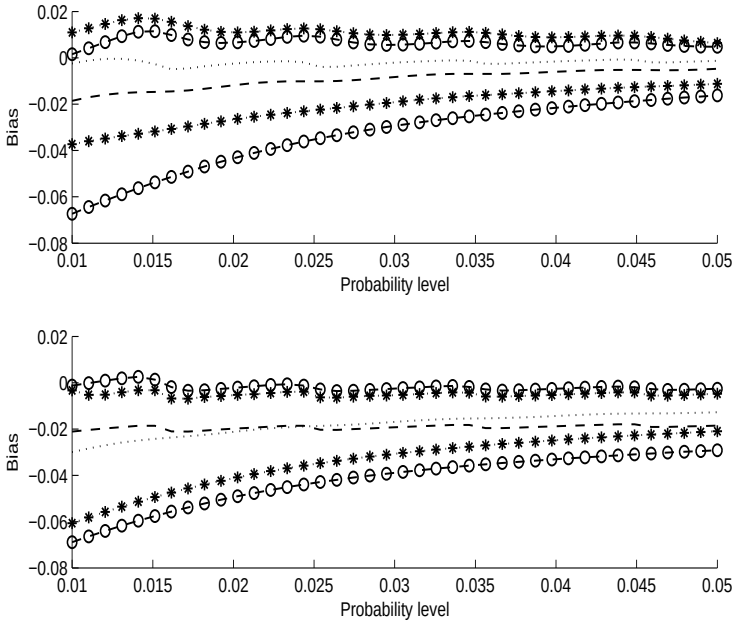

Figure 3.2: Model 1 estimated bias, $n = 100$ and α taking equally spaced values between 0.01 and 0.05. The top graph displays the bias and 95 percent confidence bands for *Est1(2S)* (dash) and $Est2(2S)_{br(1)}$ (dot); the bottom graph show $Est3(1S)_{dens}$ (dot) and $Est4(1S)_{dist}$ (dash). The confidence bands for the dotted and dashed lines are respectively given by solid starred and solid circled lines.

displays smoothed bias estimates and their 95 percent bootstrapped confidence bands for model 1 with $n = 100$ and 40 equally spaced levels of α ranging from 0.01 to 0.05. Graphs for estimators 1 to 4 are shown in Figure 3.2. Plots of all estimators show an increase in bias as α decreases; however, such a slight increase does not make estimation accuracy weak. The wider confidence bands for the lower quantiles indicate greater uncertainty in estimating ES, primarily due to scarcity of data.

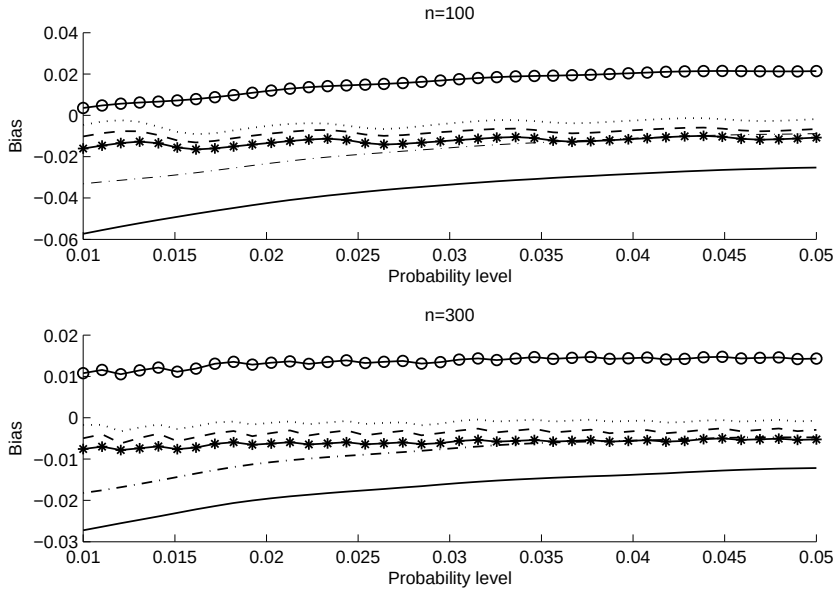

Figure 3.3: Based on Model 2, the estimated bias for $n = 100$ (top) and $n = 300$ (bottom). The dash-dotted, dashed, circled, solid, starred and dotted lines denote estimators 1 to 6 respectively.

Comparison of the biases of the estimators for $n = 100$ and $n = 300$ reveal that bias does not always decrease with large sample size. This observation is further illustrated by Figure 3.3 which shows the calculated bias for all six estimators under the normal mixture model for $n = 100$ and $n = 300$. The results for the remaining three models are roughly the same and are therefore omitted. Furthermore, the bias reduction version of estimators 3 and 4 were also considered but it was concluded that these performed no better than estimator 6.

3.7 Empirical Study

In this section the proposed kernel estimators are applied to estimate the ES of two financial series. These two financial series are the daily returns of the Dow Jones and S&P500 indexes for the period from 01/01/2002 to 31/12/2004, comprising 750 observations. Figure 3.4 displays the log-returns series for the two indexes.

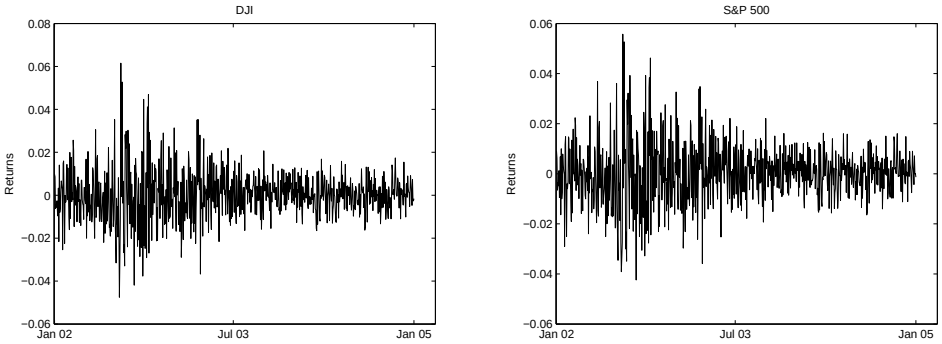

Figure 3.4: Log-returns of the DJI and S&P500 indexes.

All six ES estimators investigated in this study are based on nonparametric methods and thus making no distribution assumptions. It is by now well established that one of the stylised fact of real financial time series is the presence of heteroscedasticity as can be observed from Figure 3.4. In order to account for the effect of changing volatility and forecast next period expected shortfall, the proposed six estimators could be applied after the removal of heteroscedasticity through filtering methods such as the one proposed by McNeil and Frey (2000). In their approach the authors proposed a semi-parametric method (GARCH-EVT) of calculating VaR and ES in which a GARCH model is used to model the conditional volatility which are then used to filter the dependence in the return series by computing standardised residuals. Since the filtered residuals are independent and identically distributed, this allows the application of EVT in the form of a GPD to model the extreme left tail of the residual distribution. To obtain an estimate of next period value at risk and expected shortfall the left tail of the residual distribution is used to calculate VaR and ES which are then multiplied by forecasted standard deviation from the GARCH

model. However, it is noted that, the application of filtering approach proposed by McNeil and Frey (2000) would render the six estimators semiparametric as this method requires one to choose a specific model for the conditional mean and volatility dynamics. Assuming that the conditional mean is zero and the distribution of the return process is constant, as is commonly the case in many financial applications, to circumvent the need of following a specific volatility model here Parkinson volatility estimator (Parkinson, 1980) is used in conjunction with the six proposed estimators to forecast one period ahead expected shortfall and thus maintain the nonparametric nature of the six estimators.

Let H_t and L_t denote the respective highest and lowest prices on day t, then the Parkinson volatility estimator is defined as

$$\hat{\sigma}_{P,t} = \frac{ln(H_t) - ln(L_t)}{\sqrt{4ln(2)}}. \tag{3.18}$$

Given the historical evolution of the log transformed prices $\{r_t\}_{t=1}^{T} = ln(P_t) - ln(P_{t-1})$ of a financial asset together with the highest $\{H\}_{t=1}^{T}$ and lowest $\{L\}_{t=1}^{T}$ prices, our approach similar to that of McNeil and Frey (2000) can be partitioned into two stages, where P_t denotes the closing price on day t.

1. Estimate $\{\hat{\sigma}_{P,t}\}_{t=1}^{T}$ then standardise the returns such that $\{z_t\}_{t=1}^{T} = \{r_t\}_{t=1}^{T}/\{\hat{\sigma}_{P,t}\}_{t=1}^{T}$ and use the proposed estimators to estimate the expected shortfall of the standardised returns, $ES_z(\alpha)$.

2. To obtain an estimate of next day expected shortfall multiply the standardised expected shortfall with $\hat{\sigma}_{P,T}$, that is, $ES(\alpha) = \hat{\sigma}_{P,T} ES_z(\alpha)$. Implicitly this model free approach uses the current volatility as an estimator of next period volatility and is different from the method proposed by McNeil and Frey (2000) which by construction uses a model to forecast volatility. Furthermore, McNeil and Frey (2000) employed the Generalised Pareto distribution to estimate the expected shortfall of the standardised residuals.

Table 3.4: One period ahead Expected Shortfall estimates for loss probability levels $\alpha = 0.01$ and $\alpha = 0.05$ for kernel-based and parametric estimators, where the original value is multiplied by 100.

	DJI		S&P500	
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$
Est1(2S)	0.7262	0.4177	0.9842	0.4477
$Est2(2S)_{br(1)}$	0.5427	0.4087	0.7146	0.4258
$Est3(1S)_{dens}$	0.6396	0.5091	0.6784	0.5780
$Est4(1S)_{dist}$	0.5638	0.4710	0.5937	0.5345
$Est5(1S)_{order}$	0.5581	0.4805	0.8151	0.5952
$Est6(1S)_{br(5)}$	0.5317	0.3854	0.7091	0.5654
EVT	0.5329	0.4441	0.6109	0.4132

Table 3.4 shows one step ahead ES estimates from the six estimators together with the EVT approach for both DJI and S&P500 indexes using the procedure outlined above.

3.8 Chapter Summary

Kernel smoothing is a useful nonparametric method for estimating expected shortfall (ES). The main results of the chapter can be summarised as follows.

- One step kernel estimators of ES are proposed and are investigated through Monte Carlo simulation studies with an emphasis on bias. Through numerical experiments it is demonstrated that these estimators have appealing numerical performance in comparison to existing two step kernel estimators.

- Several estimators proposed in this chapter, such as estimators *3, 4, 5*, are all shown to be fast, efficient and valid for estimation of ES. The failure of an estimator of ES is often due not to substandard methodology but rather to the inaccurate and difficult estimation in the tail of a distribution.

- In comparison to their non-bias reduced estimators, it was noted that bias reduction technique with estimators 2 and 6 improve the accuracy of ES estimates. In particular, estimator 6, a computationally efficient new bias reduction estimator with an explicit expression is shown to be unrivalled in reducing ES estimation bias.

Chapter 4

Simultaneous Confidence Bands for Linear Quantile Regression

Simultaneous confidence bands are frequently used in regression analysis to quantify the plausible ranges of a regression function or the differences between regression functions. While this technique has been studied extensively in an ordinary least square regression setting, the procedures have yet to be extended to a linear quantile regression framework. In this chapter we fill this gap by proposing a method for constructing confidence bands for quantile regression function. In particular, we exploit the asymmetric Laplace distribution together with a modification of an existing simulation-based method employing the active set method in order to easily construct simultaneous confidence band for a linear quantile regression functions over a pre-specified region of the independent variables.

Furthermore, through the introduction of an alternative simulation algorithm we extend the simulation idea to a classical linear regression and create simultaneous confidence bands for regression quantile functions when the error terms are normal and when this assumption is not met. In the latter case the Box-Cox transformation is utilised to transform the response variable and thus making the procedure non-linear.

Finally, it is demonstrated that compared to the active set algorithm the new

procedure calling upon the threshold acceptance algorithm can improve the coverage accuracy of confidence bands.

4.1 Introduction

The standard ordinary least square regression method has long been the workhorse in modelling the mean relationship between independent and response variables. However, it is well known by now that the information obtained from this attractive technique is restricted to the conditional mean and fail to offer key insight on different regions of the conditional distribution.

A natural extension that can account for this deficiency is the quantile regression technique introduced by Koenker and Bassett (1978). This technique has received immense interest theoretically and found numerous applications in different fields such as medicine, social science and finance, see Buchinsky (1998) and Yu et al. (2003).

Let $q_\theta(y|\boldsymbol{x})$ be the $100\theta\%$ $(0 < \theta < 1)$ quantile of a response variable y conditional on the independent variables $\boldsymbol{x} = (1, x_1, \cdots, x_p)'$. In linear quantile regression models, it is assumed that the relationship between $q_\theta(y|\boldsymbol{x})$ and $\boldsymbol{x}$ can be described by a linear model

$$q_\theta(y|\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta}_\theta \tag{4.1}$$

where $\boldsymbol{\beta}_\theta = (\beta_{0,\theta}, \beta_{1,\theta} \cdots, \beta_{p,\theta})'$ is the vector of unknown regression coefficients as in classical mean regression. The unknown parameters $\boldsymbol{\beta}_\theta$ are estimated by minimising the sum of residuals

$$\sum_{i=1}^{n} \rho_\theta(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_\theta), \tag{4.2}$$

where

$$\rho_\theta(z) = \begin{cases} \theta z, & \text{if } z > 0 \\ (\theta - 1)z, & \text{otherwise} \end{cases}$$

and for $i = 1, 2, \ldots, n$, y_i is the observed response corresponding to independent variables $\boldsymbol{x}_i = (1, x_{i1}, \cdots, x_{ip})'$.

The central point of this chapter is on the construction of simultaneous confidence bands for the quantile function $\boldsymbol{x}'\boldsymbol{\beta}_\theta$ over a given region of the independent variables $\boldsymbol{x} \in \mathcal{X}$. Specifically, we aim to find lower and upper bound functions $l(\boldsymbol{x})$ and $u(\boldsymbol{x})$ such that

$$\Pr\{l(\boldsymbol{x}) < \boldsymbol{x}'\boldsymbol{\beta}_\theta < u(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathcal{X}\} = 1 - \alpha,$$

where $\alpha \in (0, 1)$, $1 - \alpha$ is the confidence level and the region of interest is given by

$$\mathcal{X} = \{\boldsymbol{x} : -\infty \leq a_j < x_j < b_j \leq \infty, \quad j = 1, 2, \ldots, p\}. \tag{4.3}$$

The flexibility on the region of interest allows an analyst to easily provide bounds a_j and b_j that are of some practical or economical importance.

The construction of confidence bands in the manner described provide a useful measure of uncertainty of the true unknown regression quantile model and assess the plausibility of different regression functions. In particular, a linear quantile regression function is a credible candidate on the condition that the entire function is enveloped within the calculated bands.

Recently, Hardle et al. (2010) considered a bootstrap approach for constructing uniform confidence bands for nonparametric quantile estimates of regression functions. Similarly, Hardle and Song (2010) proposed a smoothing method for building uniform confidence bands for quantile curves by exploiting extreme value theory together with the asymptotic maximal absolute deviation between the localised quantile smoothers

and true unknown quantile curve.

On a topic closely related to this chapter, Chernozhukov et al. (2009) introduce a finite sample distribution free inference method for quantile regression that exploits the conditional pivotal nature of the quantile regression objective function. The authors demonstrate the application of the method in constructing confidence regions as well as bounds for quantile regression parameters that makes use of simulation techniques for calculating the critical constant. For a generic treatment of simulation inference on parametric models consult Hothorn et al. (2008).

The original idea of creating simultaneous confidence bands for regression functions can be traced as far back as eight decades ago through the work of Working and Hotelling (1929). For a detailed chronological treatment on the evolvement of the field consult Liu et al. (2005).

Simultaneous confidence bands have a number of practical applications. For instance Al-Saidy et al. (2003) construct simultaneous bands for dosage risk estimation, Liu et al. (2004) apply the confidence bands in medicine by considering the simultaneous comparison of several linear regression models and Sun et al. (2000) create bands for growth and response curves. As an alternative application consider a regression quantile model, $\boldsymbol{x}'\boldsymbol{\beta}_\theta$, used to model the relationship between customer satisfaction (dependent variable) and a number of explanatory variables such as the quality of service, age of customers, annual income, social status and availability of products. In this context, simultaneous bands provide the plausible range of the true but unknown regression quantile model. For instance, a marketing manager may wish to investigate the key drivers of satisfaction by fitting a 90% regression quantile and thus by creating simultaneous bands this offers a criterion of deciding whether any other model (say, fitted with a subset of the original covariates) is a plausible candidate for $\boldsymbol{x}'\boldsymbol{\beta}_{0.90}$. Similarly, as an alternative interpretation of the confidence bands one can also test the hypothesis regarding the true parameters simultaneously by testing $H_0 : \boldsymbol{\beta}_\theta = \hat{\boldsymbol{\beta}}_\theta$ against $H_0 : \boldsymbol{\beta}_\theta \neq \hat{\boldsymbol{\beta}}_\theta$. Additionally, in inequality studies social scientists may be interested in investigating the relationship between income

and a number of explanatory variables such as age, ethnicity, level of education, occupation and marital status. Furthermore, the researchers may wish to investigate the disparity between higher and lower earners by taking the difference between, say, the upper and lower regression quartiles. Through building simultaneous bands for the difference of two regression quantiles this provide a criterion for differentiating between regression functions employed to directly model the difference. For a detailed theoretical treatment and application simultaneous confidence bands consult Liu (2010).

4.2 Simultaneous Confidence Bands

To construct regression quantiles confidence bands we make use of the simulation algorithm Liu et al. (2005) who investigated the problem in the classical ordinary least square setting. To this end, we begin by discussing the problem tackled by Liu et al. (2005) and demonstrate a natural progression towards the regression quantile framework by exploiting the Asymmetric Laplace Distribution (ALD) of Yu and Zhang (2005).

4.2.1 Confidence Bands for Multiple Linear Regression

Given a vector of response variable $\mathbf{Y}_{n\times 1}$ and a set of covariates $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p\}$, then assuming the relationship between the response and independent variables is linear with an intercept term the model can be expressed as

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{4.4}$$

In the classical presentation of the ordinary least square regression (4.4) $\boldsymbol{\beta}_{(p+1)\times 1} = (\beta_0, \beta_1, \ldots, \beta_p)'$ is a vector of coefficients, $\boldsymbol{\epsilon}_{n\times 1}$ is vector of error terms assumed to be normally distributed, $N(\mathbf{0}, \sigma^2\boldsymbol{I})$, with unknown variance σ^2. The i-th row of the design matrix X is given by $(1, x_{i,1}, x_{i,2}, \ldots, x_{i,p})$ and assuming the inverse of $(X'X)^{-1}$

exist the least square estimates of $\boldsymbol{\beta}$ are given by $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{Y}$ and unbiased estimator of the variance $\hat{\sigma}^2 = \sum_{i=1}^{n} \hat{\epsilon}_i^2/\nu \sim \sigma^2\chi_\nu^2/\nu$ where $\nu = n - p - 1$ is the degrees of freedom. The objective of the work of Liu et al. (2005) was to provide a $100(1-\alpha)\%$ simultaneous confidence bands of the form

$$\Pr\left\{-c < \frac{\boldsymbol{x}'\boldsymbol{\beta} - \boldsymbol{x}'\hat{\boldsymbol{\beta}}}{\hat{\sigma}\sqrt{\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}}} < c, \quad \forall \boldsymbol{x} \in \mathcal{X}\right\} = 1 - \alpha \tag{4.5}$$

for the true regression line

$$\boldsymbol{x}'\boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \ldots \beta_p x_p,$$

where the rectangular region is as defined in equation (4.3), c is the critical constant to be calculated and the term in the denominator of (4.5) correspond to the square root of variance of $\boldsymbol{x}'\hat{\boldsymbol{\beta}}$, $var(\boldsymbol{x}'\hat{\boldsymbol{\beta}})$. The focal point of the work of Liu et al. (2005) is to determine the value of the critical constant c, however since the distribution of

$$T = \max_{\boldsymbol{x}\in\mathcal{X}} \frac{|\boldsymbol{x}'\boldsymbol{\beta} - \boldsymbol{x}'\hat{\boldsymbol{\beta}}|}{\hat{\sigma}\sqrt{\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}}} \tag{4.6}$$

is very difficult to derive analytically, the authors propose a simulation approach of obtaining the distribution of T and therefore determining the value of c. To achieve this objective Liu et al. (2005) present two algorithms for simulating T; the first is based on the branching method while the other is called the active set method that solves a quadratic programme with inequality constraints. Using real data the authors show that through simulating a large number of T's one is able to determine the critical constant c as accurate as required.

Throughout the chapter we shall call on the active set method to create confidence bands for regression quantiles. To this end, it is therefore instructive at this point to begin with a brief explanation of quadratic programming which is the main ingredient of the active set algorithm and then produce an explicitly step by step account of the active set algorithm.

4.2.2 Quadratic Programme

Quadratic programming is an optimisation technique that can be considered as a generalisation of the linear progamming method capable of dealing with quadratic functions. The problem can be described as that of optimising (maximising or minimising) a quadratic function of p variables subject to m linear constraints over some region of interest. For a positive semi-definite matrix $H_{p\times p}$ and a column vector $\boldsymbol{b}^*_{m\times 1}$ the problem can be formally presented as

$$\min_{\boldsymbol{x}} \frac{1}{2}\boldsymbol{x}'H\boldsymbol{x} + f'\boldsymbol{x} \tag{4.7}$$

subject to

$$A\boldsymbol{x} \underset{\geqslant}{\leqslant} \boldsymbol{b}^*, \qquad \boldsymbol{a} \leq \boldsymbol{x} \leq \boldsymbol{b},$$

where $\boldsymbol{a}$ and $\boldsymbol{b}$ are the respective lower and upper bound constraints on the variable $\boldsymbol{x}$, A is a $m \times p$ and $\boldsymbol{f}$ is a $p \times 1$ vector. In the special case when $H = 0$ the problem reduces to that of solving a linear programme. In the event that all constraints of equation (4.7) are satisfied such that there exist a solution $\boldsymbol{x}^* = (x_1^*, x_2^*, \ldots, x_p^*)'$ the problem is termed as feasible, otherwise it is unfeasible.

4.2.3 Active Set Algorithm

Since T as given by equation (4.6) involves maximisation over p-variate function over the specified region an explicit solution is not easy to obtain. To circumvent this difficulty, Liu et al. (2005) propose an alternative representation of the quantity (4.6) that can facilitate simulation of the distribution of T. Through the introduction of a unique symmetric positive definite matrix Q such that

$$Q = (X'X)^{-1/2} \tag{4.8}$$

and by noting that $\mathbf{N} = \sigma^{-1}(Q')^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim N(\mathbf{0}, \boldsymbol{I}_{p+1})$, equation (4.6) can be written as

$$T = \max_{\boldsymbol{x}\in\mathcal{X}} \left(\frac{\|\mathbf{N}\|}{(\hat{\sigma}/\sigma)} \right) \frac{|(Q\boldsymbol{x})'\mathbf{N}|}{\|Q\boldsymbol{x}\|\|\mathbf{N}\|}. \tag{4.9}$$

Using the standard result $\nu\hat{\sigma}^2/\sigma^2 \sim \chi^2_\nu$ the vector $\boldsymbol{f} = \mathbf{N}/(\hat{\sigma}/\sigma)$ can be simulated as a quotient of an independent standard normal variable and the square root of an independent chi-square distribution divided by the number of degrees of freedom, $\hat{\sigma}/\sigma \sim \sqrt{\chi^2_\nu/\nu}$.

A direct consequence of this presentation is that the optimisation of (4.9) takes the form of minimising a quadratic programming problem subject to the constraint:

$$(Q^{-1}_{(j+1,)} - b_j Q^{-1}_{(1,)})\boldsymbol{x} \leq \mathbf{0}, \qquad (a_j Q^{-1}_{(1,)} - Q^{-1}_{(j+1,)})\boldsymbol{x} \leq \mathbf{0}$$

where for $j = 1, 2, \ldots, p+1$, $Q^{-1}_{(j,)}$ denotes the j-th row of the matrix Q^{-1}. A detailed algorithm is as follows:

1. Given the design matrix X calculate Q from equation (4.8)

2. Create a matrix $A_{(2p+1)\times(p+1)}$ such that for an index $j = 1, 2, \ldots, p$ the odd rows are composed of

$$A_{(2j-1,)} = Q_{(j+1,)} - b_j Q_{(1,)},$$

even rows are given by

$$A_{(2j,)} = a_j Q_{(1,)} - Q_{(j+1,)}$$

and the last row of A is $A_{(2p+1,)} = Q_{(1,)}$

3. Create an identity matrix $H_{(p+1)\times(p+1)}$ and a column vector of zeros $\boldsymbol{b}^* = \mathbf{0}_{(2p+1)}$

4. Generate the random vector, $\boldsymbol{f} = \left(\frac{z_1}{\sqrt{\chi^2_\nu/\nu}}, \frac{z_2}{\sqrt{\chi^2_\nu/\nu}}, \ldots, \frac{z_{p+1}}{\sqrt{\chi^2_\nu/\nu}} \right)'$ where z_j are independent standard normal random variables

5. Use the inputs A, $\boldsymbol{b}^*$, $\boldsymbol{f}$ and H to solve a quadratic programming problem (4.7) giving $\boldsymbol{x}^{*+} = (x_1^{*+}, x_2^{*+}, \ldots, x_p^{*+})'$ and find the norm of the solution, $\boldsymbol{v}^+ = \|\boldsymbol{x}^{*+}\|$

6. Replace $\boldsymbol{f}$ with $-\boldsymbol{f}$ and solve the quadratic programme to get $\boldsymbol{x}^{*-} = (x_1^{*-}, x_2^{*-}, \ldots, x_p^{*-})'$ and find the norm of the solution, $\boldsymbol{v}^- = \|\boldsymbol{x}^{*-}\|$

7. Find the maximum of the two norms, $C = \max(\boldsymbol{v}^+, \boldsymbol{v}^-)$

8. Repeat steps 4 - 7 a large number of times $\mathcal{N}$ and use the $[\mathcal{N}(1-\alpha)]$-th largest simulated value as an approximation of $c = \boldsymbol{C}^*_{[\mathcal{N}(1-\alpha)]}$, where is $\boldsymbol{C}^*$ is a vector obtained from sorting the elements of $\boldsymbol{C}$ in ascending order

4.2.4 Confidence Bands for Regression Quantiles

In the proceeding discussions we shall make use of the asymmetric Laplace distribution of Yu and Zhang (2005) as a building block towards the creation of confidence bands for linear regression quantiles.

For a random variable Z with support on the real line the density function of the asymmetric Laplace distribution, $\text{ALD}(\mu, \sigma, \theta)$, is given by

$$f(z; \mu, \sigma, \theta) = \frac{\theta(1-\theta)}{\sigma} \begin{cases} \exp\left(\frac{(z-\mu)(1-\theta)}{\sigma}\right), & \text{if } z \leq \mu \\ \exp\left(\frac{-(z-\mu)\theta}{\sigma}\right), & \text{if } z > \mu, \end{cases}$$

where $0 < \theta < 1$ is the skewness parameter and $\sigma > 0$ and $\mu \in (-\infty, \infty)$ are the respective scale and location parameters.

(Yu et al., 2003; Yu and Zhang, 2005) observed a direct connection between the ALD and quantile regression and the authors demonstrate that the minimisation of the check function

$$\min_{\boldsymbol{\beta}_\theta \in \Re^{p+1}} \sum_{i=1}^{n} \rho_\theta(y_i - \boldsymbol{x}_i' \boldsymbol{\beta}_\theta) \tag{4.10}$$

can alternatively be viewed as the maximisation of

$$\max_{\beta_\theta \in \Re^{p+1}} \left(\frac{\theta(1-\theta)}{\sigma}\right)^n \exp\left\{-\sum_{i=1}^{n} \rho_\theta\left(\frac{y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_\theta}{\sigma}\right)\right\}. \tag{4.11}$$

Viewed from this perspective, the estimator $\hat{\boldsymbol{\beta}}_\theta$ of the true $100\theta\%$ regression quantile coefficient $\boldsymbol{\beta}_\theta$ can be construed as the maximum likelihood estimator of $\boldsymbol{\beta}_\theta$ based on the regression model

$$y_i = \boldsymbol{x}'\boldsymbol{\beta}_\theta + U_i, \quad i = 1, \cdots, n \tag{4.12}$$

where $\{U_1/\sigma, U_2/\sigma \cdots, U_n/\sigma\}$ are independently and identically distributed (iid) random variables each with a standard asymmetric Laplace distribution, ALD$(0, 1, \theta)$. In the classical regression models, $\{U_1/\sigma, U_2/\sigma \cdots, U_n/\sigma\}$ are assumed to be iid standard normal random variables, $N(0, 1)$.

Justification of the ALD model errors has been established by a number of authors in the literature. Specifically, Komunjer (2005) introduce the tick exponential family distribution and illustrate that maximum likelihood (MLE) based on this family is the correct specification for quantile regression inference in order to achieve consistency and other attractive properties associated with MLE. The author further points out that the asymmetric Laplace distribution is a well known member of the tick-exponential family.

Additionally, Lee (2009) noted that for parametric models the tick-exponential family of distributions introduced by (Komunjer, 2005) provide consistent estimators even when the likelihood function is misspecified. Finally, the employment of ALD-based likelihood for quantile regression inference was studied more than a decade ago by Koenker and Machado (1999).

The k-th central moment of a random variable $Z \sim$ALD(μ, σ, θ) is given by

$$\mathbb{E}(Z-\mu)^k = k!\sigma^k\theta(1-\theta)\left(\frac{1}{\theta^{k+1}} + \frac{(-1)^k}{(1-\theta)^{k+1}}\right),$$

and thus for $\mathbf{U} = (U_1, U_2 \cdots, U_n)'$ it directly follows that

$$\mathbb{E}(U_i) = \sigma \frac{1-2\theta}{\theta(1-\theta)} = \sigma e_\theta, \quad Var(U_i) = \sigma^2 \frac{2\theta^2 - 2\theta + 1}{\theta^2(1-\theta)^2} = \sigma^2 v_\theta. \tag{4.13}$$

Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ and without loss of generality the non-singular design matrix $X = [\mathbf{1}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_p]$ such that i-th row is given by $(1, x_{i,1}, x_{i,2}, \ldots, x_{i,p})$, then it follows that

$$\mathbb{E}(\tilde{\boldsymbol{\beta}}_\theta) = \mathbb{E}\{(X'X)^{-1}X'\boldsymbol{y}\} = \boldsymbol{\beta}_\theta + \sigma e_\theta (X'X)^{-1}X'\mathbf{1}$$

where $\mathbf{1}$ is an n by 1 vector of whose elements are 1's. Hence, when σ is a known constant, an unbiased estimator of $\boldsymbol{\beta}_\theta$ is given by

$$\begin{aligned} \tilde{\boldsymbol{\beta}}_\theta &= (X'X)^{-1}X'\boldsymbol{y} - \sigma e_\theta (X'X)^{-1}X'\mathbf{1} \\ &= \boldsymbol{\beta}_\theta + (X'X)^{-1}X'\mathbf{U} - \sigma e_\theta (X'X)^{-1}X'\mathbf{1} \end{aligned} \tag{4.14}$$

and its variance is

$$Var(\tilde{\boldsymbol{\beta}}_\theta) = \sigma^2 v_\theta (X'X)^{-1}.$$

In what follows we shall discuss the creation of regression quantiles simultaneous confidence bands under two different cases; when the scale parameter σ is known and later when this quantity is not known.

4.2.5 When $\sigma > 0$ is a Known Constant

When $\sigma > 0$ is a known constant we construct the following simultaneous confidence band based on

$$\boldsymbol{x}'\boldsymbol{\beta}_\theta \in \boldsymbol{x}'\tilde{\boldsymbol{\beta}}_\theta \pm c_\theta \sigma \sqrt{v_\theta \boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}} \qquad \forall\ \boldsymbol{x} \in \mathcal{X} \tag{4.15}$$

where c_θ is the critical constant chosen so that the simultaneous confidence level of the band is $1-\alpha$. Note that the expression inside the square-root sign in equation (4.15) is simply $Var(\boldsymbol{x}'\tilde{\boldsymbol{\beta}}_\theta)/\sigma^2$. In order to determine c_θ for a specified confidence level $1-\alpha$, we need to evaluate the simultaneous confidence level of the band (4.15) which is given by

$$\begin{aligned}
& \Pr\left\{\boldsymbol{x}'\boldsymbol{\beta}_\theta \in \boldsymbol{x}'\tilde{\boldsymbol{\beta}}_\theta \pm c_\theta\sigma\sqrt{v_\theta \boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}} \quad \forall\ \boldsymbol{x}\in\mathcal{X}\right\} \\
= & \Pr\left\{\max_{\boldsymbol{x}\in\mathcal{X}} \frac{|\boldsymbol{x}'(\tilde{\boldsymbol{\beta}}_\theta-\boldsymbol{\beta}_\theta)|}{\sigma\sqrt{v_\theta \boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}}} < c_\theta\right\} \\
= & \Pr\left\{\max_{\boldsymbol{x}\in\mathcal{X}} \frac{|\boldsymbol{x}'(X'X)^{-1}X'(\mathbf{U}/\sigma - e_\theta\mathbf{1})|}{\sqrt{v_\theta \boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}}} < c_\theta\right\} \qquad (4.16) \\
= & \Pr\left\{\max_{\boldsymbol{x}\in\mathcal{X}} \frac{|(Q\boldsymbol{x})'QX'(\mathbf{U}/\sigma - e_\theta\mathbf{1})/\sqrt{v_\theta}|}{\sqrt{(Q\boldsymbol{x})'(Q\boldsymbol{x})}} < c_\theta\right\} \\
= & \Pr\left\{\max_{\boldsymbol{x}\in\mathcal{X}} \frac{|(Q\boldsymbol{x})'\boldsymbol{R}|}{\|(Q\boldsymbol{x})\|} < c_\theta\right\}, \qquad (4.17)
\end{aligned}$$

where

$$\boldsymbol{R} = QX'(\mathbf{U}/\sigma - e_\theta\mathbf{1})/\sqrt{v_\theta} \qquad (4.18)$$

is a random vector.

It can be observed from equation (4.16) that the calculation of the critical value c_θ for simultaneous confidence bands depend on the design matrix X, the rectangular region $\mathcal{X}$ and the constant θ but not on the unknown coefficients $\boldsymbol{\beta}_\theta$. Furthermore, as will be clarified in the discussion to follow, equation (4.16) is also independent of σ and thus serve as a major advantage of the simultaneous confidence band (4.15).

To obtain the distribution of $|(Q\boldsymbol{x})'R|/\|(Q\boldsymbol{x})\|$ we employ a similar simulation method as proposed by Liu et al. (2005). In particular, we apply the same algorithm as described in subsection 4.2.3 with the exception that the vector $\boldsymbol{f}$ is now defined by $\boldsymbol{R}$ as given by equation (4.18).

The simulation of the asymmetric laplace distribution random variable U can be generated by the inverse transformation method. That is, given standard uniformly

distributed random variable $U^* \in [0,1]$ then a standard ALD$(0,1,\theta)$ random variable U/σ can be generated using the ALD quantile function as

$$\frac{U}{\sigma} = \frac{1}{1-\theta}\log\left(\frac{U^*}{\theta}\right)_{\boldsymbol{I}(U^* \le \theta)} - \frac{1}{\theta}\log\left(\frac{1-U^*}{1-\theta}\right)_{\boldsymbol{I}(U^* > \theta)}$$

Similarly, the simulation of U/σ can easily be accomplished using the presentation of Yu and Moyeed (2001)

$$\frac{U}{\sigma} = \frac{\eta_1}{\theta} + \frac{\eta_2}{\theta - 1} \tag{4.19}$$

where η_1 and η_2 are independent exponential random variables with unit mean.

It is worth noting that for the special case of $a_j = -\infty$ and $b_j = \infty$ we have, from the Cauchy-Schwarz inequality, $\max_{\boldsymbol{x}\in\mathcal{X}} |(Q\boldsymbol{x})'\boldsymbol{R}|/\|(Q\boldsymbol{x})\| = \|\boldsymbol{R}\|$. However, even in this simple case, it is difficult to determine the distribution of $\|\boldsymbol{R}\|$ analytically and thus a simulation method as given above is a practical way to determine c_θ, especially for a general $\mathcal{X}$. Furthermore, as mentioned earlier, since the number of replications $\mathcal{N}$ can be chosen to be as large as one requires, the simulation method presented above can be regarded to provide an exact estimate of c_θ.

4.2.6 When $\sigma > 0$ is an Unknown Constant

In most practical applications the scale parameter $\sigma > 0$ is unknown, and in these situations the maximum likelihood estimator of σ can be calculated by

$$\hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n} \rho_\theta(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\theta), \tag{4.20}$$

which is obtained from maximizing the expression in equation (4.11) with the vector of coefficients $\boldsymbol{\beta}_\theta$ replaced by $\hat{\boldsymbol{\beta}}_\theta$.

On replacing the unknown σ in the confidence band (4.15) with its estimate $\hat{\sigma}$ we construct $100(1-\alpha)\%$ simultaneous confidence band as

$$\boldsymbol{x}'\boldsymbol{\beta}_\theta \in \boldsymbol{x}'\tilde{\boldsymbol{\beta}}^*_\theta \pm c_\theta\hat{\sigma}\sqrt{v_\theta \boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}} \qquad \forall\ \boldsymbol{x} \in \mathcal{X} \tag{4.21}$$

where

$$\tilde{\boldsymbol{\beta}}_{\theta}^{*} = (X'X)^{-1}X'\boldsymbol{y} - \hat{\sigma}e_{\theta}(X'X)^{-1}X'\mathbf{1}, \tag{4.22}$$

and c_θ is a critical constant as in (4.15).

When the sample size n is large, one would expect $\hat{\sigma}$ to be close to σ and thus the confidence level of band (4.21) would be close to that of band (4.15).

4.2.7 Confidence Bands for the Difference of two Symmetric Regression Quantiles

One of the main advantage of the quantile regression technique is on the flexility of characterising the entire conditional distribution of some variable given its associated covariates. This advantage allows one to analyse and compare the behaviour of different parts, such as extreme lower and upper regression quantiles, of the conditional distribution of response variable. For instance, interquartile range is useful in quality control. Similarly, the difference of two symmetric extreme quantiles is a useful measure of variation, (see (Taylor, 2005; Pearson and Tukey, 1965)).

Suppose that the objective is to build simultaneous confidence bands for the difference of two symmetric quantile regression functions $\boldsymbol{x}'\boldsymbol{\beta}_{\theta_1}$ and $\boldsymbol{x}'\boldsymbol{\beta}_{\theta_2}$ corresponding to regression model (4.12) with scale parameters σ_{θ_1} and σ_{θ_2}, respectively. Let $\boldsymbol{\beta}_{\theta,d} = \boldsymbol{\beta}_{\theta_1} - \boldsymbol{\beta}_{\theta_2}$ and $\tilde{\boldsymbol{\beta}}_{\theta,d} = \tilde{\boldsymbol{\beta}}_{\theta_1} - \tilde{\boldsymbol{\beta}}_{\theta_2}$ then it follows from the presentation of equation (4.14) the simultaneous confidence bands of $\boldsymbol{x}'\boldsymbol{\beta}_{\theta,d}$ is

$$\boldsymbol{x}'\boldsymbol{\beta}_{\theta,d} \in \boldsymbol{x}'\tilde{\boldsymbol{\beta}}_{\theta,d} \pm c_{\theta,d}\sigma_d\sqrt{v_{\theta_1}\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}} \qquad \forall\ \boldsymbol{x} \in \mathcal{X} \tag{4.23}$$

where, for $\sigma_d^2 = \sigma_{\theta_1}^2 + \sigma_{\theta_2}^2$ and assuming the errors for two quantiles are independent, the variance of the difference of two regression functions is given by

$$Var(\boldsymbol{x}'(\tilde{\boldsymbol{\beta}}_{\theta_1} - \tilde{\boldsymbol{\beta}}_{\theta_2})) = v_{\theta_1}\sigma_d^2\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}.$$

Note that v_{θ_1} appears in the last equation due to the fact that for symmetric quantiles θ_1 and θ_2 such that $\theta_1 = 1-\theta_2$ it follows from (4.13) that $v_{\theta_1} = v_{\theta_2}$. The simultaneous confidence band (4.23) can be expressed as

$$\begin{aligned}
& \Pr\left\{\boldsymbol{x}'\boldsymbol{\beta}_{\theta,d} \in \boldsymbol{x}'\tilde{\boldsymbol{\beta}}_{\theta,d} \pm c_{\theta,d}\sigma_d\sqrt{v_{\theta_1}\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}} \quad \forall\ \boldsymbol{x} \in \mathcal{X}\right\} \\
= & \Pr\left\{\max_{\boldsymbol{x}\in\mathcal{X}} \frac{|\boldsymbol{x}'(\tilde{\boldsymbol{\beta}}_{\theta,d} - \boldsymbol{\beta}_{\theta,d})|}{\sigma_d\sqrt{v_{\theta_1}\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}}} < c_{\theta,d}\right\} \\
= & \Pr\left\{\max_{\boldsymbol{x}\in\mathcal{X}} \frac{|\boldsymbol{x}'(X'X)^{-1}X'((\mathbf{U}_{\theta_1} - \mathbf{U}_{\theta_2}) + (\sigma_{\theta_2}e_{\theta_2} - \sigma_{\theta_1}e_{\theta_1})\mathbf{1})|}{\sigma_d\sqrt{v_{\theta_1}\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}}} < c_{\theta,d}\right\} \\
= & \Pr\left\{\max_{\boldsymbol{x}\in\mathcal{X}} \frac{|(Q\boldsymbol{x})'QX'((\mathbf{U}_{\theta_1} - \mathbf{U}_{\theta_2}) + (\sigma_{\theta_2}e_{\theta_2} - \sigma_{\theta_1}e_{\theta_1})\mathbf{1})|}{\sigma_d\sqrt{v_{\theta_1}}\sqrt{(Q\boldsymbol{x})'(Q\boldsymbol{x})}} < c_{\theta,d}\right\} \\
= & \Pr\left\{\max_{\boldsymbol{x}\in\mathcal{X}} \frac{|(Q\boldsymbol{x})'\boldsymbol{R}_d|}{\|(Q\boldsymbol{x})\|} < c_{\theta,d}\right\} \qquad (4.24)
\end{aligned}$$

where

$$\boldsymbol{R}_d = \frac{QX'((\mathbf{U}_{\theta_1} - \mathbf{U}_{\theta_2}) + (\sigma_{\theta_2}e_{\theta_2} - \sigma_{\theta_1}e_{\theta_1})\mathbf{1})}{\sigma_d\sqrt{v_{\theta_1}}}. \qquad (4.25)$$

In order to apply the active set algorithm and calculate the critical constant $c_{\theta,d}$, the unknown parameters σ_{θ_1} and σ_{θ_2} are replaced by their respective maximum likelihood estimators given by (4.20), where θ is appropriately replaced by θ_1 and θ_2.

4.3 Normal Transformation-based Simultaneous Bands for Regression Quantiles

In this section we employ a different simulation-based method to construct simultaneous confidence bands for quantiles of the normal regression model when the residuals are normally distributed and when the assumption of normality is not tenable.

4.3.1 Normally Distributed Residuals

For a given vector of response variable $\boldsymbol{y}$ and a set of covariates $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p\}$ the general classical regression setting of equation (4.4) stipulates that the error terms be normally distributed, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I})$. The problem of constructing simultaneous confidence bands for quantile functions

$$q_{\theta,N}(y|\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta} + \sigma\Phi^{-1}(\theta) \tag{4.26}$$

in the general linear regression model has been considered by a number of authors. For instance, Steinhorst and Bowden (1971) considered simultaneous confidence bands for the true regression quantile (4.26) simultaneous in $\boldsymbol{x}$ for both θ fixed and when θ is variable. Turner and Bowden (1977) presented a generalisation of the work of Steinhorst and Bowden (1971) to a general linear model framework in the case when θ is fixed. Using the width of the bands as a measuring criterion the authors noted some improvement over the Scheffé-type bands of Steinhorst and Bowden (1971) but they conclude that these gains are not conclusive and consequently state that the "best" procedure has yet to be uncovered. Similarly, Thomas and Thomas (1986) proposed two different methods for creating simultaneous confidence bands for quantiles of a standard regression model which can be considered as an alternative extension to the bands of Steinhorst and Bowden (1971) and the modified version of the Kanofsky (1968) bands to a regression model.

In the implementation of all of the aforementioned methods the explanatory variables are not constrained in a some finite intervals like the one given by equation (4.3) and none are based on simulation technique. In keeping with the simulation theme of the chapter we put forward a simulation-based method to form simultaneous confidence bands for quantiles of a linear regression model in which the variables are constrained.

The objective here is to construct $100(1-\alpha)\%$ confidence bands of the form

$$q_{\theta,N}(y|\boldsymbol{x}) \in \hat{q}_{\theta,N}(y|\boldsymbol{x}) + c_{\theta,N}\sigma\sqrt{\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x} + (\Phi^{-1}(\theta))^2\kappa} \quad \forall\ \boldsymbol{x} \in \mathcal{X} \tag{4.27}$$

such that $\kappa > 0$ the covariates are constrained within the region (4.3) and where $\hat{q}_{\theta,N}(y|\boldsymbol{x}) = \boldsymbol{x}'\hat{\boldsymbol{\beta}} + \hat{\sigma}\Phi^{-1}(\theta)$. As before, the formation of the confidence band (4.27) relies on accurate estimation of the critical constant $c_{\theta,N}$ in order to ensure $\Pr(\mathfrak{T} \leq c_{\theta,N}) = 1-\alpha$ where

$$\mathfrak{T} = \max_{\boldsymbol{x}\in\mathcal{X}} \frac{|\boldsymbol{x}'\hat{\boldsymbol{\beta}} + \hat{\sigma}\Phi^{-1}(\theta) - (\boldsymbol{x}'\boldsymbol{\beta} + \sigma\Phi^{-1}(\theta))|}{\sigma\sqrt{\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x} + (\Phi^{-1}(\theta))^2\kappa}}, \tag{4.28}$$

with the subscript N indicating Gaussian errors.

Note that the denominator is the square root of a linear combination of $Var(\boldsymbol{x}'\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}$ and $Var(\Phi^{-1}(\theta)\hat{\sigma}) = (\Phi^{-1}(\theta))^2\sigma^2\kappa$. In order for $\mathfrak{T}$ to be feasible, in a sense that will become apparent in the proceeding explanation, the unknown constant κ must be determined. To achieve this one needs to calculate $Var(\hat{\sigma})$ which can be accomplished through employing the standard result $Z_1 = \nu\hat{\sigma}^2/\sigma^2 \sim \chi^2_\nu$ and note that $Z_2 = \sqrt{Z_1} \sim \chi_\nu$, with χ_ν denoting a Chi distribution with ν degrees of freedom. From the properties of a Chi distribution it follows that $Var(Z_2) = \nu - \omega^2$, where $\omega = \sqrt{2}\Gamma((\nu+1)/2)/\Gamma(\nu/2)$ and $\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}dt$ is the gamma function. Expressing the variance as

$$Var(Z_2) = \nu\sigma^{-2}Var(\hat{\sigma}) = \nu - \omega^2$$

we have $Var(\hat{\sigma}) = (\nu-\omega^2)\sigma^2\nu^{-1}$ and which straightforwardly follows that

$$\kappa = 1 - \omega^2\nu^{-1}.$$

Similar to the mean regression bands given by (4.5) it is possible to express $\mathfrak{T}$ as function that depends on a combination of standard normal and Chi-square

random variables, and thus becomes readily set for simulation approach to obtain its distribution. To observe this, note that equation (4.28) can be written as

$$\mathfrak{T} = \max_{\boldsymbol{x}\in\mathcal{X}} \frac{\left| \frac{\boldsymbol{x}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{\sigma\sqrt{\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}}} + \frac{\Phi^{-1}(\theta)}{\sqrt{\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}}}\left(\frac{\hat{\sigma}}{\sigma}-1\right)\right|}{(\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x})^{-1/2}\sqrt{\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}+(\Phi^{-1}(\theta))^2\kappa}} \tag{4.29}$$

From this presentation it can easily be deduced that $\boldsymbol{x}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\sigma^2\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x})^{-1/2} \sim N(\mathbf{0},1)$ and $\hat{\sigma}/\sigma \sim \sqrt{\chi^2_\nu/\nu}$. However, in this formulation the maximisation of $\mathfrak{T}$ can no longer be accomplished via the active set algorithm of (Liu et al., 2004, 2005) as described in section (4.2.3). Fortunately, through defining $\mathfrak{T}$ as a function of p variates one can nevertheless solve it via different optimisation techniques as will be illustrated in subsection (4.4).

4.3.2 Non-Normally Distributed Residuals

Although taken for granted, it is by now well established that for numerous practical applications the assumption of normality is not always tenable and thus there is a need to account for this lack of fit. In dealing with this issue, one approach that has gained great popularity is the Box-Cox power transformation Box and Cox (1964) as given by equation (2.6) in chapter 2. In this technique, after the response variable or equivalently the errors have been identified to deviate from normality, an assumption that is particularly sensitive when calculating confidence intervals and hypothesis testing for regression parameters, an analyst can transform the response variable via the Box-Cox procedure and then fit a regression function. Finally, the fitted response variable can be transformed back to the original scale.

Suppose the response variable $\boldsymbol{y}$ is non-Gaussian and let $\boldsymbol{y}_\lambda = g_\lambda(\boldsymbol{y})$ denote a vector of transformed response variable $\boldsymbol{y}$ obtained via the Box-Cox transformation $y_\lambda = ((y^\lambda - 1)/\lambda)_{I(\lambda\neq 0)} + \log(y)_{I(\lambda=0)}$ such that $\boldsymbol{y}_\lambda$ is approximately normally distributed. Given this setup our proposed approach can be regarded as a two step one, with an ultimate goal of creating simultaneous bands for quantiles of

the conditional distribution of y given $\boldsymbol{x}$, that is, $q_\theta(y|\boldsymbol{x}) = g_\lambda^{-1}(q_{\theta,N}(y_\lambda|\boldsymbol{x}))$, where $q_{\theta,N}(y_\lambda|\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta}_\lambda + \sigma_\lambda\Phi^{-1}(\theta)$. The execution of this method can be outlined as follows; first, build confidence bands of the transformed normal regression quantiles

$$q_{\theta,N}(y_\lambda|\boldsymbol{x}) \in \hat{q}_{\theta,N}(y_\lambda|\boldsymbol{x}) \pm c_{\theta,N}\sigma_\lambda\sqrt{\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x} + (\Phi^{-1}(\theta))^2\kappa} \quad \forall\ \boldsymbol{x} \in \mathcal{X}$$

by maximising $\mathfrak{T}$ from equation (4.28) and finally calculate the simultaneous bands for the original quantile function via the inverse transform

$$g_\lambda^{-1}(q_{\theta,N}(y_\lambda|\boldsymbol{x})) \in g_\lambda^{-1}\left(\hat{q}_{\theta,N}(y_\lambda|\boldsymbol{x}) \pm c_{\theta,N}\sigma_\lambda\sqrt{\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x} + (\Phi^{-1}(\theta))^2\kappa}\right) \quad \forall\ \boldsymbol{x} \in \mathcal{X}. \tag{4.30}$$

While the construction of simultaneous bands for regression quantiles of the linear regression has been studied by (Steinhorst and Bowden, 1971; Turner and Bowden, 1977, 1979; Thomas and Thomas, 1986) the simple idea of a Box-Cox transformation in conjunction with a simulation based technique has not been exploited. The strategy of transforming the response variable is particularly useful in the sense that the bounds constraining the explanatory variables are unaffected and remain valid when transforming back the fitted response variable to its original scale via the monotonic function, g_λ^{-1}. Additionally, the Box-Cox transformation technique utilises the equivariance property of quantile regression which guarantee the ordering of observations is preserved. For extensive theoretical and implementation treatment of Box-Cox quantile regression refer to (Chamberlain, 1994; Buchinsky, 1995; Powell, 1999; Machado and Mata, 2000).

4.4 Calculation of the Critical Constant $c_{\theta,N}$

The maximisation of (4.29) can be achieved by representing the function $\mathfrak{T}$ as a quotient of two functions $\mathcal{F}_1$ and $\mathcal{F}_2$ for the numerator and denominator respectively.

Let

$$\begin{aligned}\mathcal{V} &= \sqrt{\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}} \\ &= \Big(|(a_{1,1}+a_{2,1}x_1+\ldots+a_{j,1}x_p)+x_1(a_{1,2}+a_{2,2}x_1+\ldots+a_{j,2}x_p)+\ldots \\ &+ x_p(a_{1,p+1}+a_{2,p+1}x_1+\ldots+a_{j,p+1}x_p)|\Big)^{1/2}, \end{aligned} \tag{4.31}$$

where the constant $a_{j,k}$ correspond to the j-th row and k-th column elements of $(X'X)^{-1}$. Further, let z_1 denote a standard normal random variate and $\gamma \sim \sqrt{\chi^2_\nu/\nu}$, then the numerator of (4.29) can be expressed as

$$\mathcal{F}_1 = |z_1 + \mathcal{V}^{-1}\Phi^{-1}(\theta)(\gamma - 1)|. \tag{4.32}$$

Similarly, the non-random function corresponding to the denominator of (4.29) can also be expressed in terms of the unknown variable $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)'$ as

$$\mathcal{F}_2 = \mathcal{V}^{-1}(|\mathcal{V}^2 + (\Phi^{-1}(\theta))^2\kappa|)^{1/2}. \tag{4.33}$$

To obtain the critical constant $c_{\theta,N}$ the procedures are as follows.

1. Generate z_1 and γ
2. Maximise $\mathfrak{T} = \mathcal{F}_1/\mathcal{F}_2$ using one of the algorithms to be described below (4.4.1)
3. Repeat the last two steps a large number of times $\mathcal{N}$ and use the $[(1-\alpha)\mathcal{N}]$-th largest simulated value as an approximation for $c_{\theta,N}$

By the same argument as before, for sufficiently large number of simulations $\mathcal{N}$ the critical constant $c_{\theta,N}$ can be considered to be exact.

Note that, it is also possible to express the ALD-based confidence bands (4.16) as a quotient of two functions. One can easily observe this by defining the numerator of (4.16) as

$$\frac{(X'X)^{-1}X'}{\sqrt{v_\theta}}\left(\frac{\mathbf{U}}{\sigma} - e_\theta \mathbf{1}\right) = \boldsymbol{w} = (w_0, w_1, \ldots, w_p)' \tag{4.34}$$

where, without loss of generality, the regression model (4.1) has an intercept such that the first column of the design matrix X consist of a vector of ones. Then we can write (4.16) as

$$\mathcal{F}_{ALD} = \max_{\boldsymbol{x}\in\mathcal{X}} \frac{|w_0 + \sum_{j=1}^{p} x_j w_j|}{\mathcal{V}}. \tag{4.35}$$

The multivariate function (4.35) can be solved via simulation by; (i) generate a vector of standard ALD random variable $\boldsymbol{U}/\sigma$ and calculate the vector $\boldsymbol{w}$, (ii) Optimise (4.35) and (iii) repeat the last two steps a large number of times $\mathcal{N}$ and estimate the critical constant c_θ for $100(1-\alpha)\%$ confidence band as $[(1-\alpha)\mathcal{N}]$-th largest simulated value.

In the next subsection we proceed by outlining two algorithms that can accommodate the maximisation of bounded functions.

4.4.1 Simulated Annealing and Threshold Acceptance Algorithms

Simulated annealing (SA) optimisation technique introduced by (Kirkpatrick et al., 1983; Černý, 1985) was motivated by the adaptation of statistical mechanic techniques towards solving optimisation problems. This method has attracted wide application due to it ability to handle multivariate objective functions of different degrees of complexity and with or without constraints on both the cost function and the variables of interest, see Ingber (1993).

The mechanics of this generic algorithm, as the name suggest, operate by closely mimicking the microstructure behaviour of atoms when heating and slowly cooling a metal such that the objective function is slowly reduced from an initial high (temperature) state to one with low energy, and thus giving an approximate solution of the global optimum. The reduction of the objective function is obtained through an iterative process that at each state substitute a current solution with a better randomly generated solution in the neighbourhood until a final solution minimising

the objective function is reached. During the transition period from one state to another it is also possible for the algorithm to branch towards a worse solution. However, this seemingly backward step can sometimes be helpful in the sense of identifying new regions which could lead to a better minimum.

With the goal of creating a faster and more efficient alternative to SA a modified version called *threshold acceptance* algorithm was introduced. This algorithm function in the same way as the original SA with an added condition on selecting a solution from one state to another. Specifically, transitional solutions are chosen to be below some pre-specified threshold which is systematically lowered as the algorithm progress.

For implementation purpose both the SA and acceptance threshold algorithms can be carried out using the software MATLAB using the functions *simulannealbnd* and *threshacceptbnd* respectively. Finally, although SA and acceptance threshold algorithms are iterative procedures that provide an approximate solution, when the sample size is large enough this bootstrap-like method can accurately recover the distribution of $\mathfrak{T}$.

4.4.2 One Dimension Constrained Optimisation

In the special cases when there is one predictor variable and when the covariates have a polynomial functional relationship the optimisation of $\mathcal{T}$ reduces from that of maximising p-variate function to one of maximising single variate. That is, the problem can be written as

$$\max_{x_1} \mathcal{T}(x), \qquad a_1 \leq x_1 \leq b_1$$

and the MATLAB function *fminbnd* which is an algorithm based on the golden section search and parabolic interpolation can be utilised to obtain a solution.

4.5 Assessing the Coverage Accuracy of the Confidence Bands

In order to quantify the accuracy of confidence bands and distinguish between alternative methods of constructing confidence bands two techniques corresponding to the validity and optimality properties are widely employed, (see Canay (2010)). The primary of these properties, validity, centres on assessing the coverage levels of confidence bands with more accurate rule of constructing confidence bands producing coverage probabilities closest to the nominal confidence level. The optimality property is usually assessed by calculating the width of confidence bands with a construction method classified superior if it results in shorter lengths compared to others.

For the empirical work to be carried in section 4.6 we shall consider the validity criterion for assessing confidence bands built under both the ALD and normal model errors assumptions described in the last two sections. It can be observed that the exact coverage probability of bands (4.21), (4.23) and (4.27) depends on the unknown parameters, $\boldsymbol{\beta}_\theta$, σ, $\boldsymbol{\beta}_{\theta,d}$, σ_d and $\boldsymbol{\beta}$. For given values of the unknown parameters the coverage probability of each of the bands (4.21), (4.23) and (4.27) can be assessed via simulation. In order to clarify this approach, we shall demonstrate it for ALD-based band (4.21), and the parallel will be applicable for the difference of two regression quantiles (4.23) and normal error-based band (4.27).

Note that the coverage probability of band (4.21) is given by $\Pr\{V < c_\theta\} = 1-\alpha$ where

$$V = \max_{\boldsymbol{x}\in\mathcal{X}} \frac{|\boldsymbol{x}'(\tilde{\boldsymbol{\beta}}^*_\theta - \boldsymbol{\beta}_\theta)|}{\hat{\sigma}\sqrt{v_\theta \boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}}}.$$

For given values of $\boldsymbol{\beta}_\theta$ and σ the algorithm for assessing the coverage accuracy is as follows:

1. Use the specified $\boldsymbol{\beta}_\theta$ and σ to simulate $\boldsymbol{y}$ according to model (4.12)

2. Compute the MLEs $\hat{\boldsymbol{\beta}}_\theta$ and $\hat{\sigma}$ followed by $\tilde{\boldsymbol{\beta}}^*_\theta$ as given by (4.22)

3. Compute the random variable V

4. Repeat the last three steps a large number of times $\mathcal{N}$ and approximate coverage probability by $\mathcal{N}^{-1}\sum_{l=1}^{\mathcal{N}}(V_l \leq c_\theta)$

The calculation of V can be carried out by maximising the function $\max_{\boldsymbol{x}\in\mathcal{X}} V = \max_{\boldsymbol{x}\in\mathcal{X}} \mathcal{F}_1^*/\mathcal{F}_2^*$ using acceptance threshold or SA algorithms where the functions $\mathcal{F}_1^*$ and $\mathcal{F}_2^*$ are respectively given by

$$\mathcal{F}_1^* = |\tilde{\beta}_{0,\theta}^* - \beta_{0,\theta} + x_1(\tilde{\beta}_{1,\theta}^* - \beta_{1,\theta}) + \ldots + x_p(\tilde{\beta}_{p,\theta}^* - \beta_{p,\theta})|$$

and

$$\begin{aligned}\mathcal{F}_2^* &= \hat{\sigma}\sqrt{v_\theta}\Big(|(a_{1,1} + a_{2,1}x_1 + \ldots + a_{j,1}x_p) + x_1(a_{1,2} + a_{2,2}x_1 + \ldots + a_{j,2}x_p) + \ldots \\ &+ x_p(a_{1,p+1} + a_{2,p+1}x_1 + \ldots + a_{j,p+1}x_p)|\Big)^{1/2}.\end{aligned}$$

On a similar note, for the difference of two regression quantiles one can also construct $\mathcal{F}_1^*$ and $\mathcal{F}_2^*$ by replacing $\boldsymbol{\beta}_{\theta_1}$, $\boldsymbol{\beta}_{\theta_2}$ and σ_d with $\tilde{\boldsymbol{\beta}}_{\theta_1}$, $\tilde{\boldsymbol{\beta}}_{\theta_2}$ and $\hat{\sigma}_d$ respectively.

For confidence bands of normally distributed model errors, (4.27), the above algorithm can be slightly modified to reflect the Gaussian distribution assumption. That is, for given values of σ and $\boldsymbol{\beta}$ simulate $\boldsymbol{y}$ according to model (4.4) and estimate the parameters $\hat{\sigma}$ and $\hat{\boldsymbol{\beta}}$. The remaining two steps follows in an analogous manner.

4.6 Empirical Study

To illustrate the practicality of the different methods discussed in this chapter we employ two data sets and construct confidence bands for quantile regression functions. The first data set with the response variable approximately normal is the classical dataset of (Galton, 1889; Pearson and Lee, 1896) which was originally used to study the relationship between the heights of children and their fathers. The dataset

consists of 1078 observations of fathers and sons heights at maturity, and since the objective is to investigate the relationship between the two we use father's heights as an explanatory variable x and employ a simple regression model (**model 1**)

$$q_\theta(y|x) = \beta_{0,\theta} + \beta_{1,\theta}x,$$

to capture different quantiles of son's heights. The data set is available from R Documentation (2010) and for the purpose of keeping the sample size moderately large we only use the first 250 cases.

Secondly, we employ the immunoglobulin G data discussed by Isaacs et al. (1983). The dataset consists of serium concentration (grams per litre) of immunoglobulin G of 298 children between the ages of six months to 6 years. For this dataset we adapt a quadratic model (**model 2**)

$$q_\theta(y|x) = \beta_{0,\theta} + \beta_{1,\theta}x + \beta_{2,\theta}x^2,$$

in which the explanatory variable age is used to predict quantiles of the immunoglobulin G (IgG). Descriptive statistics together with Jarque-Bera p-values are presented in Table 4.1 for the response variables; son's heights, IgG and the Box-Cox transformation of IgG.

Table 4.1: Descriptive statistics of the response variables.

	Skewness	Kurtosis	p-value
Sons heights	0.1546	2.9423	> 0.5000
IgG	0.7685	4.0294	0.0000
IgG (Box-Cox)	0.0019	3.0988	> 0.5000

The p-value indicate that at 5% significance level the assumption of normality of the son's heights cannot be rejected while for the variable IgG this does not hold. However, after the Box-Cox transformation of the variable IgG the test does not detect deviation from the normality assumption and thus confirm the validity of the

power transformation in producing an approximate normal distribution, a fact that is further confirmed by the skewness and kurtosis values.

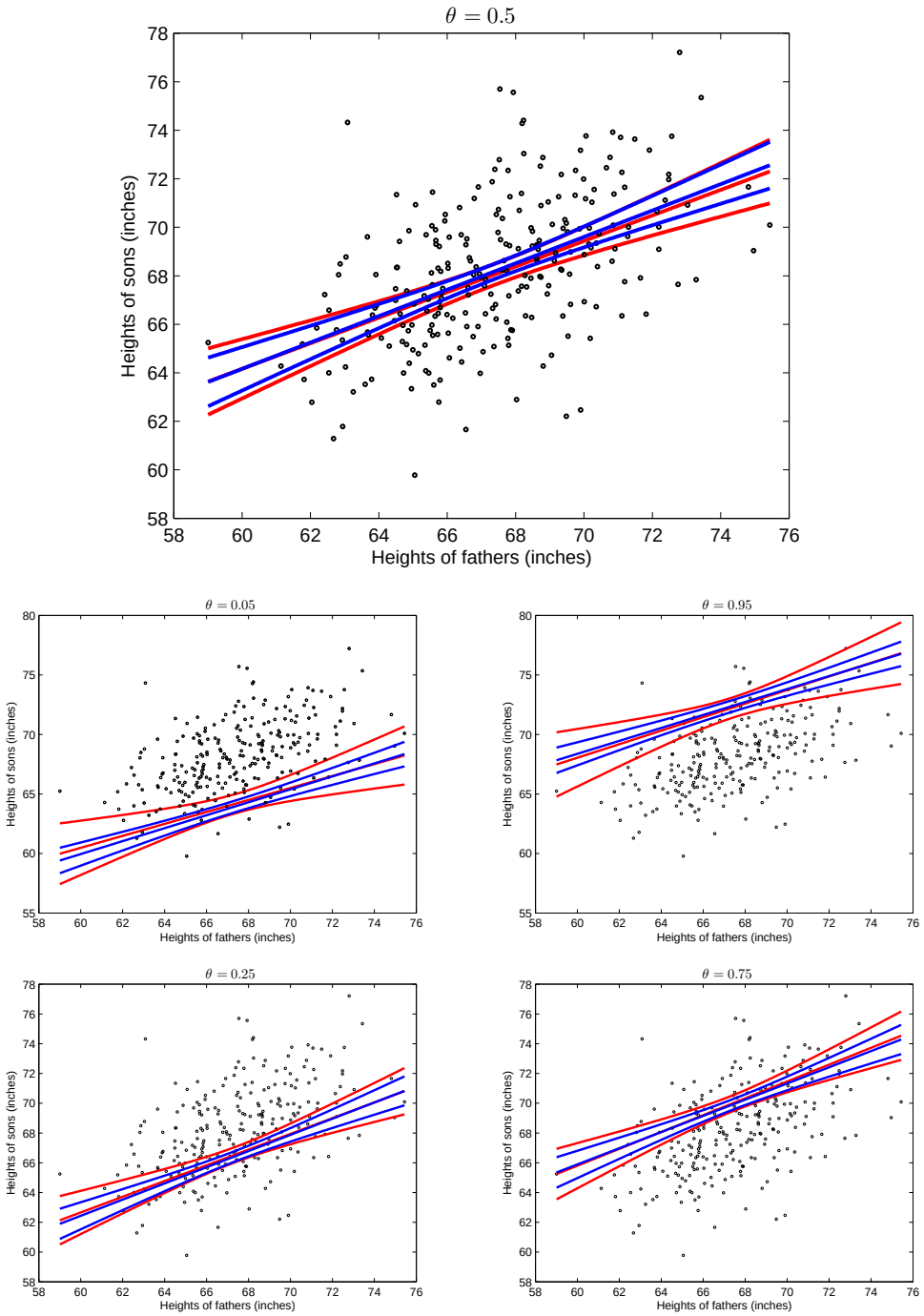

Figure 4.1: 95% simultaneous confidence bands for the 5%, 25%, 50%, 75% and 95% regression quantiles based on the ALD errors (red) and normal errors (blue). These bands correspond to equations (4.21) and (4.27).

Table 4.2 reports the estimated parameters of the two models. It should be

Table 4.2: Estimated parameters for the classical linear and quantile regression models with (standard errors) in parenthesis.

	Quantile regression			OLS regression			
	$\hat{\beta}_{0,\theta}$	$\hat{\beta}_{1,\theta}$	$\hat{\beta}_{2,\theta}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}$
Model 1							
$\theta = 0.05$	32.6868	0.4696		31.5262	0.5439		2.5599
	(7.4392)	(0.1107)		(3.9085)	(0.0579)		(0.4157)
$\theta = 0.25$	31.2148	0.5247					
	(5.7217)	(0.0868)					
$\theta = 0.50$	33.0797	0.5192					
	(4.3069)	(0.0640)					
$\theta = 0.75$	31.8656	0.5656					
	(6.4446)	(0.0954)					
$\theta = 0.95$	33.9047	0.5690					
	(11.2469)	(0.1658)					
Model 2							
$\theta = 0.05$	0.7651	1.1914	-0.1337	1.3398	0.5284	-0.0411	0.7892
	(0.4605)	(0.5078)	(0.0906)	(0.1539)	(0.1252)	(0.0204)	(0.0362)
$\theta = 0.25$	1.6021	1.2587	-0.1262				
	(0.4944)	(0.3928)	(0.0643)				
$\theta = 0.50$	2.8011	1.1587	-0.0752				
	(0.6409)	(0.5105)	(0.0890)				
$\theta = 0.75$	4.3425	0.7055	0.0191				
	(0.5180)	(0.4532)	(0.0752)				
$\theta = 0.95$	7.23563	-0.4817	0.2460				
	(1.3357)	(1.1560)	(0.1944)				

noted that for both models 1 and 2 there are two distinct cases corresponding to quantile regression and ordinary least square regression . For the quantile regression case parameters are obtained using the R software *rq* function available in the *quantreg* package. Additionally, for model 2 the classical regression parameters are obtained from regressing Box-Cox transformed IgG on age whereas the quantile regression parameters are obtained from fitting the model to the original data.

We proceed by way of graphical illustration of $95\%(\alpha = 0.05)$ simultaneous confidence bands for $5\%, 25\%, 50\%, 75\%$ and 95% regression functions of model 1 using the ALD and normal errors procedures as respectively described in subsections (4.2.4) and (4.3.1). Figure 4.1 depicts the confidence bands with the constraint, $x \in [59.01, 75.43]$, corresponding to the minimum and maximum values of fathers heights. The bounds used here are only for illustrative purposes, in practical applications one has the flexibility of choosing these such that inference of the model in the chosen region is of particular importance.

Estimates of the critical constants c_θ and $c_{\theta,N}$ for models 1 and 2 are presented in Table 4.3 using programs written in MATLAB. For both the ALD errors approach and the classical regression model employing normal errors, the critical bounds are obtained based on 50,000 simulations. The number of simulations is set at 50,000 since this was found by Liu et al. (2005) to be the minimum number that ensures reasonable accuracy of the critical constants to within two decimal places. The simulations were run on Toshiba PC Intel Pentium dual-core with processor speed of 1.73GHz. For model 1, the respective average times for running 50,000 simulations using active set and threshold acceptance algorithms were 218.94 and 605.24 seconds. Similarly, the average times for model 2 using the active set algorithm and *threshacceptbnd* MATLAB function were 345.74 seconds and 608.31 seconds respectively. Following Liu et al. (2005) the standard errors (se) are calculated as a consequence of the central limit theorem which under some regularity conditions states that $\hat{c}_\theta \sim N(c_\theta, \alpha(1 -$

$\alpha)/(f^2(c_\theta)\mathcal{N})$), where $\mathcal{N}$ is the number simulations used. Thus, standard errors is

$$se = \left(\frac{\alpha(1-\alpha)}{f^2(c_\theta)\mathcal{N}}\right)^{1/2},$$

with $f(\hat{c}_\theta)$ evaluated from the kernel density estimate of f and the bandwidth h is chosen by $\hat{h} = \min(1.06\hat{\sigma}\mathcal{N}^{-1/5}, 1.06(\hat{\xi}_{0.75} - \hat{\xi}_{0.25})\mathcal{N}^{-1/5}/1.348)$, where $\hat{\xi}_\theta$ denotes an estimator of the $100(\theta)\%$ quantile of the simulated values.

Table 4.3: Critical constants for ALD errors $\hat{c}_\theta$ and Gaussian errors $\hat{c}_{\theta,N}$ with corresponding (standard errors) in parenthesis.

θ	0.05	0.25	0.50	0.75	0.95
Model 1					
$\hat{c}_\theta$	2.4327	2.4394	2.4320	2.4403	2.4417
	(0.0036)	(0.0035)	(0.0036)	(0.0036)	(0.0034)
$\hat{c}_{\theta,N}$	1.9561	1.9527	1.9704	1.9701	1.9507
	(0.0084)	(0.0089)	(0.0083)	(0.0085)	(0.0083)
Model 2					
$\hat{c}_\theta$	2.5383	2.5301	2.5284	2.5310	2.5443
	(0.0078)	(0.0082)	(0.0080)	(0.0077)	(0.0080)
$\hat{c}_{\theta,N}$	2.1288	1.9920	1.9712	1.9777	2.1332
	(0.0199)	(0.0257)	(0.0296)	(0.0285)	(0.0192)

From figure 4.1 it can be observed that for all five quantiles the regression lines calculated from optimising the check function (4.2) and an estimate of the classical linear regression quantiles (4.26) are almost coinciding. However, it is also apparent that for all quantiles the bands created with the assumption of normal model errors are clearly narrower compared with those constructed using the ALD errors in conjunction with the active set algorithm. Having stated that, it is noteworthy to re-emphasise that for this dataset the assumption of normality of the response variable cannot be rejected, as found by the Jarque-Bera test.

Figure 4.2 shows (model 2) 95% confidence bands for $5\%, 25\%, 75\%$ and 95% regression functions for both ALD-procedure and for Box-Cox transformation approach (4.30). For the latter case, the Box-Cox transformation parameter $\hat{\lambda} =$

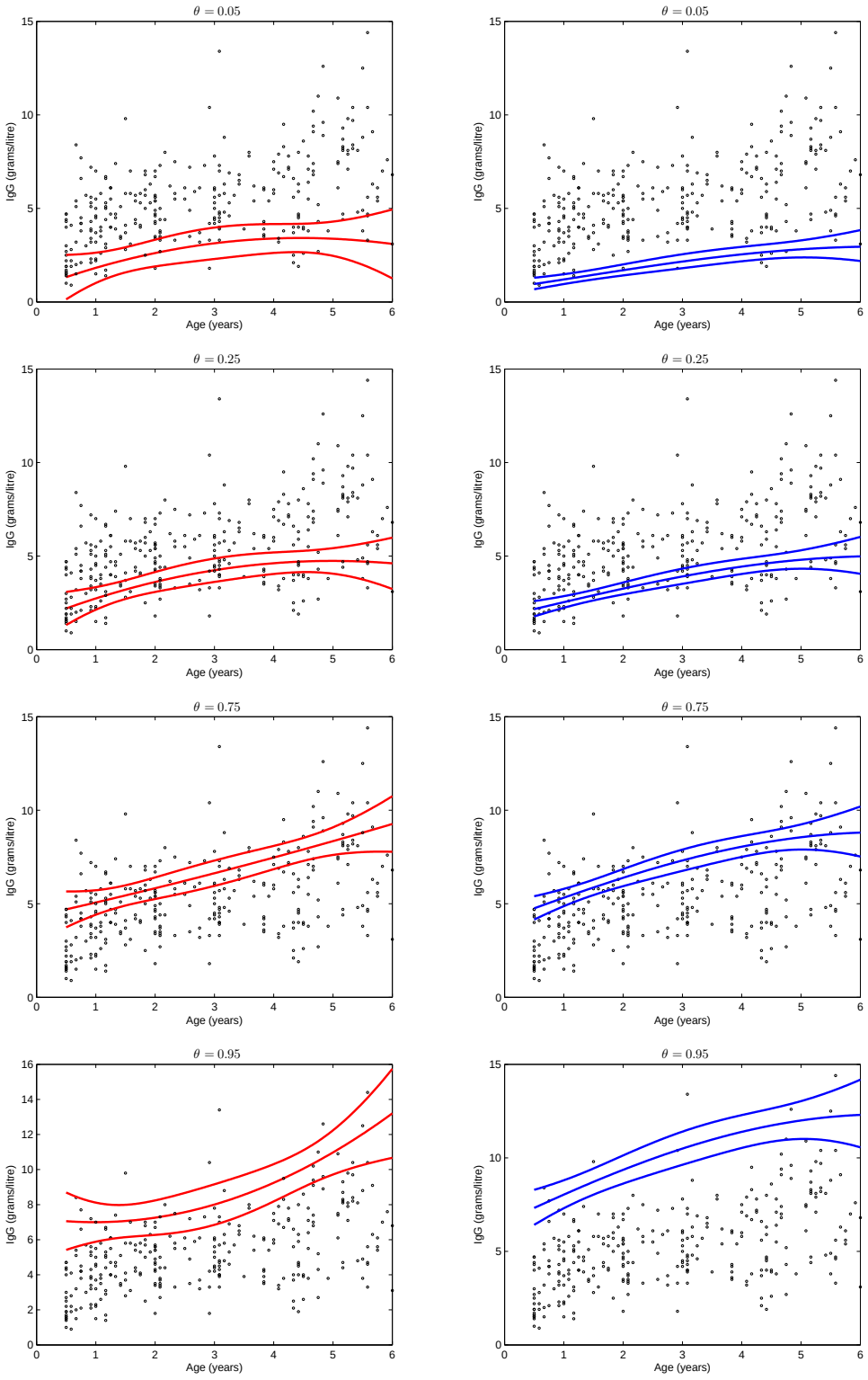

Figure 4.2: Model 2 - 95% simultaneous confidence bands for 5%, 25%, 75% and 95% regression quantiles based on the ALD errors (red) and normal errors (blue). These bands correspond to equations (4.21) and (4.30).

0.4518 is calculated from the *boxcox* MATLAB function with 95% confidence interval given by [0.2450, 0.6200]. The covariates are restricted within the interval, [0.5, 6], corresponding to the minimum and maximum ages. Furthermore, to illustrate the creation of simultaneous bands for the difference of two regression quantiles figures 4.3 and 4.4 depicts 95% confidence bands for the functions $\boldsymbol{x}'(\boldsymbol{\beta}_{0.75} - \boldsymbol{\beta}_{0.25})$ and $\boldsymbol{x}'(\boldsymbol{\beta}_{0.95} - \boldsymbol{\beta}_{0.05})$ respectively.

Careful inspection of figure 4.2 reveals that for 5% and 25% regression quantiles the bands obtained under the normal model errors are once gain narrower than those calculated with the assumption of ALD errors. However, for 75% and 95% quantiles deviations from the bands between the ALD and normal model errors are less pronounced. At this point it is important to highlight that these bands are build with two fundamental differences. First, under the ALD errors assumption the object is to construct bands for model 2 as presented by (4.1) where the parameters $\hat{\beta}_{0,\theta}, \hat{\beta}_{1,\theta}$ and $\hat{\beta}_{2,\theta}$ are calculated by minimising the 'check' function (4.2), whereas for the normal errors the parameter set $\hat{\boldsymbol{\beta}} = \{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\}$ is obtained by maximising the normal likelihood function. For the latter case, quantile regression functions are derived through scaling of the estimated mean function $\boldsymbol{x}'\hat{\boldsymbol{\beta}}$ by adding $\hat{\sigma}\Phi^{-1}(\theta)$. Secondly, while for the ALD errors the bands are directly build for the regression function $\boldsymbol{x}'\boldsymbol{\beta}_\theta$ using the active set algorithm in the case of normal errors the non-normal response variable is first transformed to be approximately normal and then bands are build for the regression functions $q_{\theta,N}(y_\lambda|\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta}_\lambda + \sigma\Phi^{-1}(\theta)$ using threshold acceptance algorithm calling on the MATLAB function *threshacceptbnd*. Once the bands of the transformed quantile are acquired the final step is to form bands for the original quantile function by transforming back to the original scale and in so doing making the whole procedure non-linear.

From figure 4.2 it can be observed that since the estimated parameters $\hat{\boldsymbol{\beta}}_\lambda$, $\hat{\sigma}$ and $\hat{\lambda}$ remain fixed the quantile regression functions calculated with an assumption of Gaussian model errors do not have the flexibility of adapting for different quantiles. On the other hand, since the parameters of quantile regression are re-estimated for

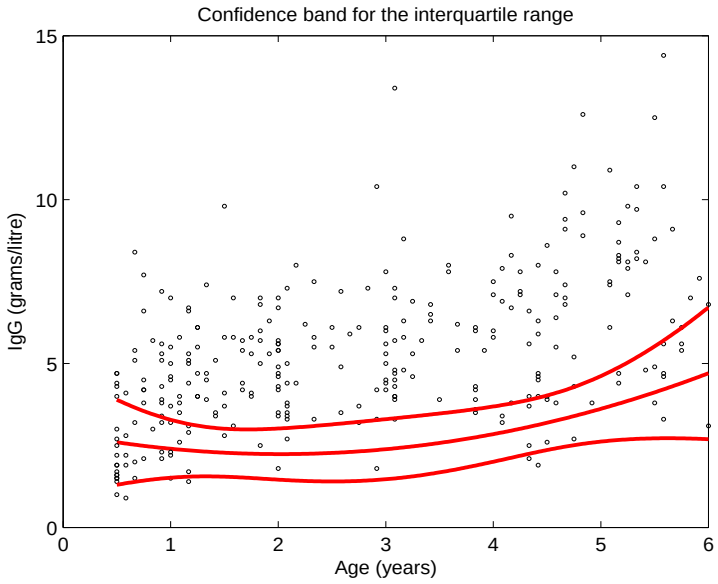

Figure 4.3: Model 2 - 95% simultaneous confidence band (4.23) for $\boldsymbol{x}'(\boldsymbol{\beta}_{0.75} - \boldsymbol{\beta}_{0.25})$. An estimate of the critical constant is $\hat{c}_\theta$=2.7558 with corresponding standard error (0.0075).

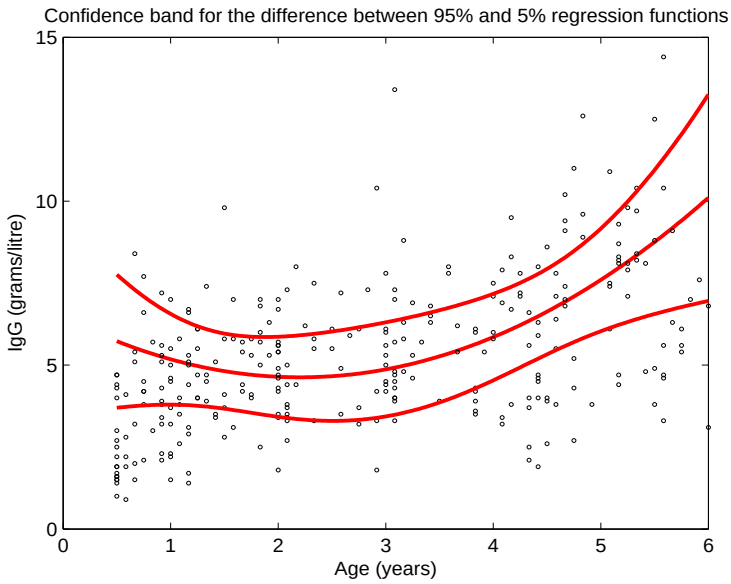

Figure 4.4: Model 2 - 95% simultaneous confidence band (4.23) for $\boldsymbol{x}'(\boldsymbol{\beta}_{0.95} - \boldsymbol{\beta}_{0.05})$. An estimate of the critical constant is $\hat{c}_\theta$=2.7661 with corresponding standard error (0.0082).

each quantile this results in evolving quantile functions as illustrated by figure 4.2 where the functional form of quantile regressions evolve from a negative quadratic for 5% and 25%, to approximately a polynomial of degree one for $\theta = 0.75$ and finally turning to a positive quadratic for 95% quantile.

We now turn attention to the calculation of coverage probabilities for the confidence bands as described in section 4.5 and we explain the set up using band (4.21) with an analogously extension to the other bands.

The number of replications $\mathcal{N}$ is set at 1000 and the unknown parameters $\boldsymbol{\beta}_\theta$ and σ are replaced by their estimators $\hat{\boldsymbol{\beta}}_\theta$ and $\hat{\sigma}$. To assess the coverage accuracy we keep the parameter set $\hat{\boldsymbol{\beta}}_\theta$ fixed and use the algorithm described in section 4.5 to calculate coverage probabilities under four different cases, namely, $\hat{\sigma} \pm se_{\hat{\sigma}}$, and $\hat{\sigma} \pm 2se_{\hat{\sigma}}$, with $se_{\hat{\sigma}}$ denoting the standard deviation of $\hat{\sigma}$. Thereafter, we repeat the same experiment by keeping $\hat{\sigma}$ fixed and allowing the vector of parameters $\hat{\boldsymbol{\beta}}_\theta$ to vary by $\hat{\boldsymbol{\beta}}_\theta \pm \Delta_\theta$ and $\hat{\boldsymbol{\beta}}_\theta \pm 2\Delta_\theta$, where Δ_θ is a vector of standard errors of the quantile regression parameters. Finally, we allow $\hat{\sigma}$ and each parameter in the set $\hat{\boldsymbol{\beta}}_\theta$ to randomly vary within two standard errors such that for each of the 1000 replications one generates different regression parameters $\{\hat{\boldsymbol{\beta}}^{(k)}\}_{k=1}^{\mathcal{N}}$ and standard deviation $\{\hat{\sigma}^{(k)}\}_{k=1}^{\mathcal{N}}$ which are then substituted for the real unknown parameters $\boldsymbol{\beta}_\theta$ and σ in step 1 of the algorithm outlined in section 4.5.

Tables 4.4, 4.5 and 4.6 presents the respective coverage probabilities for the two models for the cases when $\hat{\sigma}$ is fixed, the vector of parameters is fixed and random combination of both $\hat{\sigma}$ and the parameter vector within two standard errors. From the three tables it can be observed that for model 1 confidence bands built under Gaussian model errors produce coverage probabilities closer to nominal level compare to bands depending on the ALD errors. Furthermore, the tables reveal the conservative nature of the ALD-based bands for extreme quantiles with improved accuracy for the median and lower and upper quartiles regression functions.

For model 2 the coverage accuracy of the Box-Cox transformation procedure are

Table 4.4: 95% coverage probabilities for fixed $\hat{\sigma}$.

	Quantile regression				OLS regression			
	$\boldsymbol{\beta}_\theta - \Delta_\theta$	$\boldsymbol{\beta}_\theta + \Delta_\theta$	$\boldsymbol{\beta}_\theta - 2\Delta_\theta$	$\boldsymbol{\beta}_\theta + 2\Delta_\theta$	$\boldsymbol{\beta} - \Delta$	$\boldsymbol{\beta} + \Delta$	$\boldsymbol{\beta} - 2\Delta$	$\boldsymbol{\beta} + 2\Delta$
Model 1								
$\theta = 0.05$	0.9880	0.9860	0.9890	0.9840	0.9650	0.9460	0.9520	0.9410
$\theta = 0.25$	0.9780	0.9700	0.9740	0.9730	0.9580	0.9490	0.9430	0.9400
$\theta = 0.50$	0.9410	0.9420	0.9320	0.9510	0.9600	0.9310	0.9320	0.9550
$\theta = 0.75$	0.9630	0.9720	0.9610	0.9710	0.9590	0.9440	0.9580	0.9490
$\theta = 0.95$	0.9930	0.9880	0.9830	0.9870	0.9500	0.9630	0.9520	0.9610
Model 2								
$\theta = 0.05$	1	0.9990	0.9980	1	0.9690	0.9620	0.9630	0.9670
$\theta = 0.25$	0.9870	0.9960	0.9940	0.9890	0.9540	0.9460	0.9580	0.9460
$\theta = 0.50$	0.9880	0.9870	0.9890	0.9920	0.9400	0.9520	0.9510	0.9430
$\theta = 0.75$	0.9980	0.9950	0.9940	0.9890	0.9390	0.9480	0.9620	0.9430
$\theta = 0.95$	0.9990	0.9990	0.9960	0.9980	0.9660	0.9610	0.9750	0.9680

Table 4.5: 95% coverage probabilities for fixed $\hat{\boldsymbol{\beta}}_\theta$ and $\hat{\boldsymbol{\beta}}$.

	Quantile regression				OLS regression			
	$\hat{\sigma} - se_{\hat{\sigma}_\theta}$	$\hat{\sigma} + se_{\hat{\sigma}_\theta}$	$\hat{\sigma} - 2se_{\hat{\sigma}_\theta}$	$\hat{\sigma} + 2se_{\hat{\sigma}_\theta}$	$\hat{\sigma} - se_{\hat{\sigma}}$	$\hat{\sigma} + se_{\hat{\sigma}}$	$\hat{\sigma} - 2se_{\hat{\sigma}}$	$\hat{\sigma} + 2se_{\hat{\sigma}}$
Model 1								
$\theta = 0.05$	0.9920	0.9860	0.9840	0.9890	0.9440	0.9390	0.9480	0.9550
$\theta = 0.25$	0.9710	0.9790	0.9660	0.9750	0.9490	0.9560	0.9590	0.9560
$\theta = 0.50$	0.9540	0.9390	0.9560	0.9510	0.9610	0.9550	0.9550	0.9450
$\theta = 0.75$	0.9710	0.9690	0.9700	0.9790	0.9650	0.9510	0.9470	0.9500
$\theta = 0.95$	0.9900	0.9910	0.9880	0.9860	0.9580	0.9500	0.9630	0.9440
Model 2								
$\theta = 0.05$	0.9980	0.9850	0.9970	0.9980	0.9590	0.9720	0.9660	0.9580
$\theta = 0.25$	0.9960	0.9970	0.9940	0.9920	0.9530	0.9460	0.9490	0.9480
$\theta = 0.50$	0.9870	0.9810	0.9890	0.9880	0.9430	0.9430	0.9440	0.9500
$\theta = 0.75$	0.9940	0.9960	0.9930	0.9950	0.9430	0.9530	0.9470	0.9420
$\theta = 0.95$	0.9990	0.9990	0.9990	0.9990	0.9650	0.9660	0.9630	0.9720

Table 4.6: Coverage probabilities where the parameters are selected randomly within two standard errors.

θ	0.05	0.25	0.50	0.75	0.95
Model 1					
ALD errors	0.9870	0.9570	0.9390	0.972	0.9910
Normal errors	0.9510	0.9450	0.9590	0.9500	0.9470
Model 2					
ALD errors	0.9990	0.9950	0.9900	0.9950	0.9980
Normal errors	0.9650	0.9550	0.9550	0.9610	0.9660

Table 4.7: Critical constants for ALD errors $\hat{c}_\theta$ with corresponding (standard errors) in parenthesis using the *threshacceptbnd* algorithm. The parameters are randomly selected within two standard errors.

θ	0.05	0.25	0.50	0.75	0.95
Model 1					
$\hat{c}_\theta$	1.9719	1.9530	1.9607	1.9666	1.9402
	(0.0084)	(0.0083)	(0.0083)	(0.0084)	(0.0079)
	0.9530	0.9490	0.9500	0.9510	0.9470
Model 2					
$\hat{c}_\theta$	2.1914	2.3560	2.0480	2.1425	2.2862
	(0.0250)	(0.0322)	(0.0225)	(0.0260)	(0.0277)
	0.9570	0.9530	0.9470	0.9380	0.9570

very close to the nominal level. However, across all five quantiles the coverage results show that the ALD-based approach produce very conservative confidence bands as indicated by coverage probabilities very close to unity. These results are indicative of the deficiency of the active set algorithm in producing accurate confidence bands when the covariates have a polynomial relationship as in model 2. To this end, it will be instructive to compare the performance of the ALD-based confidence bands in terms of the critical constants and coverage probabilities by substituting the active set algorithm with the threshold acceptance one.

Table 4.7 reproduce ALD-based critical constants and 95% coverage probabilities for models 1 and 2 using the threshold acceptance algorithm. From the table it can be noticed that the critical constants for both models are smaller compare to those calculated by the active set algorithm. Consequently, the coverage probabilities of the bands are no longer conservative with coverage closer to nominal level. This observation is a testament of robustness of ALD maximum likelihood approach for quantile regression even when the error distribution is misspecified, as noted by Lee (2009). The improved accuracy in the coverage probabilities however is at the expense of longer computational time, with the threshold acceptance algorithm approximately running for twice the time of the active set algorithm.

4.7 Chapter Summary

The focal point of the chapter is on the construction of simultaneous confidence bands for regression quantile functions when the covariates as bounded in a region. The proposed techniques make use of simulation techniques to generate distributions from which the critical constants are calculated. The main theoretical and practical contribution of the chapter can be summarised as follows:

- Simultaneous confidence bands are created for quantile and interquantile regression functions exploiting the properties of the asymmetric Laplace

distribution.

- With a slight modification, the simulation approach is shown to readily extend in building bands for quantile functions calculated from classical ordinary least square regression technique when the model errors are assumed to normal or otherwise. In the latter case, a two step procedure that makes use of the Box-Cox transformation technique is utilised.

- Through a different presentation of the problem it is demonstrated that one is able to simulate distributions from which critical constants can be calculated using either simulated annealing, threshold acceptance and in some special cases by solving a one dimension constrained optimisation problem.

- Empirical studies demonstrate that the proposed approaches provide accurate coverage. Notedly, it is shown that in some scenarios, such as when the covariates have a functional relationship as in model 2, the active set algorithm provide conservative coverages, an observation that is shown to be remedied by the acceptance threshold algorithm.

Chapter 5

GARCH Induced Quantile-based Prediction Intervals for Realised Volatility

In this chapter we take a practical approach to calculating prediction intervals for realised volatility using quantile based procedures in conjunction with the GARCH model. We make use of the intraday returns to model time series of quantiles with the view of forecasting next period density followed by an application of the distribution of the difference of two extreme quantiles to create prediction intervals for RV. We also propose a bootstrap version of this idea together with two adapted extensions of the HAR model that employs the intraday quantiles of an approximate realised volatility measure and quantile regression. The proposed procedures are applied to two data sets and their coverage performances of realised volatility are assessed.

5.1 Introduction

The increasing availability of high frequency intraday financial data has facilitated for a rich understanding of asset dynamics as well as attracted a great deal of research in

the estimation and forecasting of volatility, a quantity that is of great importance in financial fields such as risk management. Since the introduction of the AutoRegressive Conditional Heteroscedastic (ARCH) model and its generalised version (GARCH) by Engle (1982) and Bollerslev (1986), numerous GARCH-type models have been proposed with the aim of capturing different characteristics of financial time series, see Poon and Granger (2003) for an extensive overview. However, despite their popularity, GARCH-type models have been criticised for relying heavily on model specification for volatility dynamics and distribution assumptions. Furthermore, due to the latent nature of volatility and therefore the absence of a benchmark with which to compare the forecasting performance, the task of models comparison becomes very difficult.

Arising from the need to have a benchmark from which various volatility forecasting models and estimators could be compared, Andersen and Bollerslev (1998) introduced the concept of realised variance for high frequency data, that is data where a number of observations are recorded for each time point during the course of the trading day. The realised variance is defined as the summation of intraday squared returns and its square root is referred to as the Realised Volatility (RV). In evaluating the performance of the GARCH models the authors treat RV as an observable benchmark and they demonstrate, through the use of Mincer-Zarnowitz regression technique, that the GARCH models provide accurate forecasts.

Treating volatility as observable rather than latent has given rise to abundant volatility models, and the interest has transcended beyond the need to only measure and model volatility, with numerous contributions exploiting the increase availability of intraday data for forecasting. For instance, Ghysels et al. (2006) introduced the Mixed Data Sampling (MIDAS) method for time series data sampled at different frequencies: here the intraday returns are used to forecast realised volatilities via the use of lag polynomials. Similarly, with the aim of capturing the long memory property of volatility in a parsimonious manner, Corsi (2009) proposed the Heterogeneous AutoRegressive (HAR) model making use of realised volatilities obtained from daily,

weekly and monthly periods in both modelling and forecasting realised volatilities. Using daily returns, Blair et al. (2001) found great improvements in the forecasting performance of GARCH models by incorporating realised variance as an explanatory variable in the conditional variance specification.

Despite having been criticised, GARCH models have been the subject of numerous research in the field of high frequency data. One area that has received a number of contributions is on the estimation of parameters of GARCH models using high frequency data. For instance, Visser (2008) utilises high frequency data and employs different volatility proxies in the form of realised volatility and range-based volatility and demonstrate that the estimation of the classical close-to-close GARCH model can be significantly improved and thus resulting in accurate volatility forecasts. On the same note, Drost and Nijman (1993) proposed a method for aggregating GARCH parameters obtained from low frequency data, say monthly, in order to derive high frequency parameters, such as weekly or daily. Galbraith and Zinde-Walsh (2000) proposed two methods that uses high frequency intra-day returns to estimate the GARCH parameters of daily returns in order to estimate conditional volatility. One of these methods uses the aggregation procedure proposed by Drost and Nijman (1993) and the other makes use of the method proposed by Galbraith and Zinde-Walsh (1994) and Galbraith and Zinde-Walsh (1997) where they introduce measurement error to a realised volatility measure and perform a two stage process whereby first the parameters of ARCH are approximated and thereafter the GARCH parameters are derived via a link function.

With regard to estimation and forecasting of realised volatility using GARCH models numerous modifications and extensions have been proposed in the literature. For instance, Giot and Laurent (2004) employs realised volatility in conjunction with ARCH type-models to model value at risk. Similarly, with the objective of forecasting value at risk Shao et al. (2009) uses high frequency data to forecast realised volatility via the Conditional AutoRegressive Range (CARR) of Chou (2005). Recently, Hansen et al. (2010) introduced the so-called realised GARCH model which simultaneously

models returns and realised volatility measures, such as, realised volatility, bipower variation and intraday range via a link function relating the realised measure to the conditional variance of returns. Through application, the authors demonstrate that their model provide an improvement over GARCH models based solely on daily returns. Similarly, Christoffersen et al. (2010) propose Generalised realised volatility model (GRV) with GARCH component that incorporates both the realised volatility and daily returns.

While the modelling of realised volatility and other realised measures has attracted a great deal of attention in the literature primarily due to its important link with risk measures such as value at risk and expected shortfall, the construction of prediction intervals has taken second stage. From a low frequency (close-to-close) perspective, this deficiency has been addressed, amongst others, by Pascual et al. (2006) who proposed a bootstrap approach for predicting the densities of future volatilities and returns (close-to-close) from the GARCH model. Through simulation the authors demonstrate quite accurate coverage probabilities for both returns and volatilities forecasts, a result that is attributed to the ability of the proposed procedures in dealing with parameter uncertainty. From a high frequency angle, Galbraith et al. (2001) propose a quantile regression approach for calculating the conditional quantiles of GARCH models conditioning on past realised volatilities and past squared returns. While this approach can be used to forecast conditional quantiles and create prediction intervals, the authors do not conduct such an evaluation restricting attention only on the analysis of patterns for different within sample conditional quantiles.

In this chapter we make use of high frequency data and propose a method for constructing prediction intervals for realised volatility that employs AR-GARCH model. Specifically, we exploit an alternative approximation of volatility based on the difference of two extreme quantiles as proposed by Pearson and Tukey (1965) to directly model and forecast intraday quantiles with the help of AR-GARCH model. Once the intraday quantiles with their corresponding volatilities are respectively

forecasted from the AR-GARCH, we make use of known result on order statistics to obtain a distribution of the difference of two quantiles and from which the prediction intervals of realised volatility are calculated. Furthermore, we also propose two adapted extensions of the HAR model of Corsi (2009) with the view to extend the functionality to construct prediction intervals for RV. One of these adaptations is based on modelling intraday quantiles of volatility in conjunction with a GARCH model while the other directly employs the HAR model using quantile regression.

It should be acknowledge that, the idea of using the difference of two extreme quantiles to forecast volatility has already been presented in the literature. Specifically, Taylor (2005) exploited the difference of two extreme quantiles to model and forecast time varying quantiles of (close-to-close) returns using the Conditional Autoregressive Value at Risk (CAViaR) models of Engle and Manganelli (2004) and then forecast volatility using the difference of predicted quantiles. On a similar note, Chou (2005) introduced a GARCH-type volatility model based on the range of price dynamics.

While there is a close resemblance of the work presented here with that of Taylor (2005), as it will become apparent in our discussion to follow, we differ in three fundamental points. First, our primary objective is to create prediction intervals for realised volatility with a point forecast as a by product; secondly, we make use of high frequency data. Finally to achieve our goal of creating prediction intervals for RV we employ an AR-GARCH model to model intraday quantiles of returns to project the quantiles corresponding to different time points during the day and from which intervals are calculated using the distribution of the difference of two order statistics.

5.2 Data

For the empirical illustration to follow in Section 5.8 and in order to demonstrate the proposed ideas during the course of discussion, we use two different equities (AXA and France Telecom (FT)) obtained from Dow Jones Euro Stoxx 50 covering the

three and half years period between 25th January 2005 and 24th June 2008. The prices are recorded every minute over a trading period 9:00 to 17:30 and thus giving 510 minute-by-minute observations. After the exclusion of holidays, weekends and dormant trading days we are left with 873 days. In the very few events when the price is missing we employ an imputation procedure whereby the last recorded price of an earlier time is imputed, thus assuming the price is constant and the return is zero. Without calculating the optimal sampling frequency, we adopt the approach of Andersen and Bollerslev (1998) by choosing five minutes sampling frequency which leaves us with 102 intradaily returns.

For the purpose of forecast evaluation we partition the data into two periods. The first period, January 2005 to September 2006 (473 observations) will be used as the in-sample and the remaining 400 days, will serve as the out of sample period.

5.3 Volatility Measures

The central point of the chapter is on the exploitation of the difference of two symmetric extreme quantiles to create prediction intervals for realised volatility and thus in this section a brief explanation of this idea and an introduction to the notion of realised volatility are given calling on the work of McAleer and Medeiros (2008). We also highlight the relationship between the realised volatility, standard deviation and quantile measure of volatility.

Let p_t denote the logarithm price of an asset observed at time t. A general and widely employed continuous-time (semimartingale) model for the evolution of the log price is given by

$$p_t = p_0 + \int_0^t \mu_u du + \int_0^t \sigma_u dW_u,$$

where μ and σ denote the drift and volatility terms respectively, W is a standard Brownian motion assumed to be independent of σ. Without loss of generality, let the

drift term μ be zero and let $\tau_1, \tau_2, \ldots, \tau_{M+1}$ denote M+1 equally spaced time periods over the course a trading day at which the prices of a financial security are observed. The central parameter of interest is the integrated volatility over a day t

$$IV_t = \left(\int_0^1 \sigma_u^2 du \right)^{1/2}, \tag{5.1}$$

where the day interval is normalised to be in the region [0,1], $0 = \tau_1 < \tau_2 < \ldots < \tau_{M+1} = 1$. Given a continuous model for the evolution of prices, equation 5.1 shows that IV is obtained as the integral of instantaneous volatilities over the trading day.

As IV is unobserved, a key quantity in estimating IV is the realised volatility. Building on the work of Merton (1980), Andersen and Bollerslev (1998) and Barndorff-Nielsen and Shephard (2002) demonstrate that the daily model-free realised variance can be constructed through the summation of equidistant intraday returns computed over very short time periods. That is, for a given day t, the daily realised volatility is defined as

$$RV_t = \left(\sum_{l=1}^{M} r_{t,l}^2 \right)^{1/2}, \tag{5.2}$$

where

$$r_{t,l} = p_{t,\tau_l} - p_{t,\tau_{l-1}}, \qquad l = 2, \ldots, M+1,$$

is the continuously compounded return, and p_{t,τ_l} is the logarithmic price observed at time τ_l on day t. As the sampling frequency approaches zero, $\sup_l \{\tau_{l+1} - \tau_l\} \to 0$, the realised variance converges uniformly in probability to the integrated volatility:

$$RV_t \to IV_t,$$

and thus under the assumption of no leverage effect it gives a consistent estimator of integrated volatility, McAleer and Medeiros (2008).

5.3.1 The Difference of two Quantiles as a Measure of Volatility

The realised volatility has a natural link with the sample standard deviation, which is used to estimate the overall variability σ_t on day t. In particular, the sample standard deviation for zero mean returns on day t is given by

$$\hat{\sigma}_t = \left(\frac{1}{M-1} \sum_{l=1}^{M} r_{t,l}^2 \right)^{1/2}. \tag{5.3}$$

On comparing equations (5.2) and (5.3) it can easily be observed that $\sqrt{(M-1)}\hat{\sigma}_t = RV_t$. In this chapter, we suggest an approximation of RV_t which is based on a different estimation of the standard deviation σ_t.

In particular, using the idea that the standard deviation σ of a distribution is directly proportional to the difference of the two symmetric extreme quantiles of that distribution, Pearson and Tukey (1965) propose a quantile-based measure of σ as

$$\tilde{\sigma} = \frac{\xi_{1-p} - \xi_p}{C(p)}, \tag{5.4}$$

where $p \in (0, 1)$, ξ_p is the $100p\%$ population quantile and $C(p)$ is a constant dependent on p. An estimate of σ is then obtained by using the sample quantile $\hat{\xi}_p$. Pearson and Tukey (1965) 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 obtained from the normal distribution, with a slight adaptation for the 5% quantile, and are found to be suited to a number of other standard distributions. Pearson and Tukey (1965) also showed that the accuracy of the proposed estimators are directly influenced by the skewness and kurtosis of the underlining distribution and consequently recommend the usage of $p = 0.05$ as it is robust against different skewness and kurtosis values.

The focus of this chapter is on using the formula of Pearson and Tukey (1965) to find an estimate of the standard deviation σ_t, but we adapt the constant $C(p)$ to a more flexible data-driven one. The constant $C(p)$ used by Pearson and Tukey

(1965) is essentially given by $\Phi^{-1}(1-p)-\Phi^{-1}(p)$ where $\Phi(.)$ is the standard Gaussian cumulative distribution function. Equivalent to that, when the distribution $F(.)$ of the data $\mathbf{X} = \{X_1, X_2, \ldots, X_M\}$ is not known, an estimator of $C(p)$ can be defined as

$$\hat{C}(p) = \hat{G}^{-1}_{M,Y}(1-p) - \hat{G}^{-1}_{M,Y}(p),$$

where $\hat{G}_{M,Y}(y)$ is the empirical cumulative distribution function of the standardised $\mathbf{X}$. Using the result of the ratio between standard deviation and mean absolute deviation of zero mean normally distributed variables X the standardised variable Y is calculated as

$$Y_l = \frac{(X_l - \overline{X})}{\hat{\sigma}_{MAD}},$$

where $\hat{\sigma}_{MAD} = \sqrt{\pi/2}\mathbb{E}|X|$ and $\overline{X}$ is the mean of $\mathbf{X}$. We note that (Andersen et al., 2000, 2003) demonstrate empirically that intraday returns standardised by RV (5.2) are approximately Gaussian distributed. In the proceeding empirical studies we shall use this method to estimate the constant $C_t(p)$ for each day t using the daily compounded returns. In particular, using this new estimate for σ_t, we propose to approximate the commonly used RV_t by the following measure;

$$\widehat{RV}_t = \sqrt{(M-1)}\frac{\hat{\xi}_{1-p,t} - \hat{\xi}_{p,t}}{\hat{C}_t(p)}, \tag{5.5}$$

where $\hat{\xi}_{p,t}$ is $100p\%$ sample quantile calculated from the empirical distribution of intraday returns on day t.

To illustrate the close proximity between the traditional measure of realised volatility (5.2) and the one based on the difference of two extreme quantiles (5.5), Figure 5.1 displays the realised volatility RV_t against the difference of quantiles measure of volatility $\sqrt{M-1}(\hat{\xi}_{1-p,t} - \hat{\xi}_{p,t})/\hat{C}_t(p)$ for AXA and France Telecom. From this figure it can be observed that, in consistence with findings of Pearson and

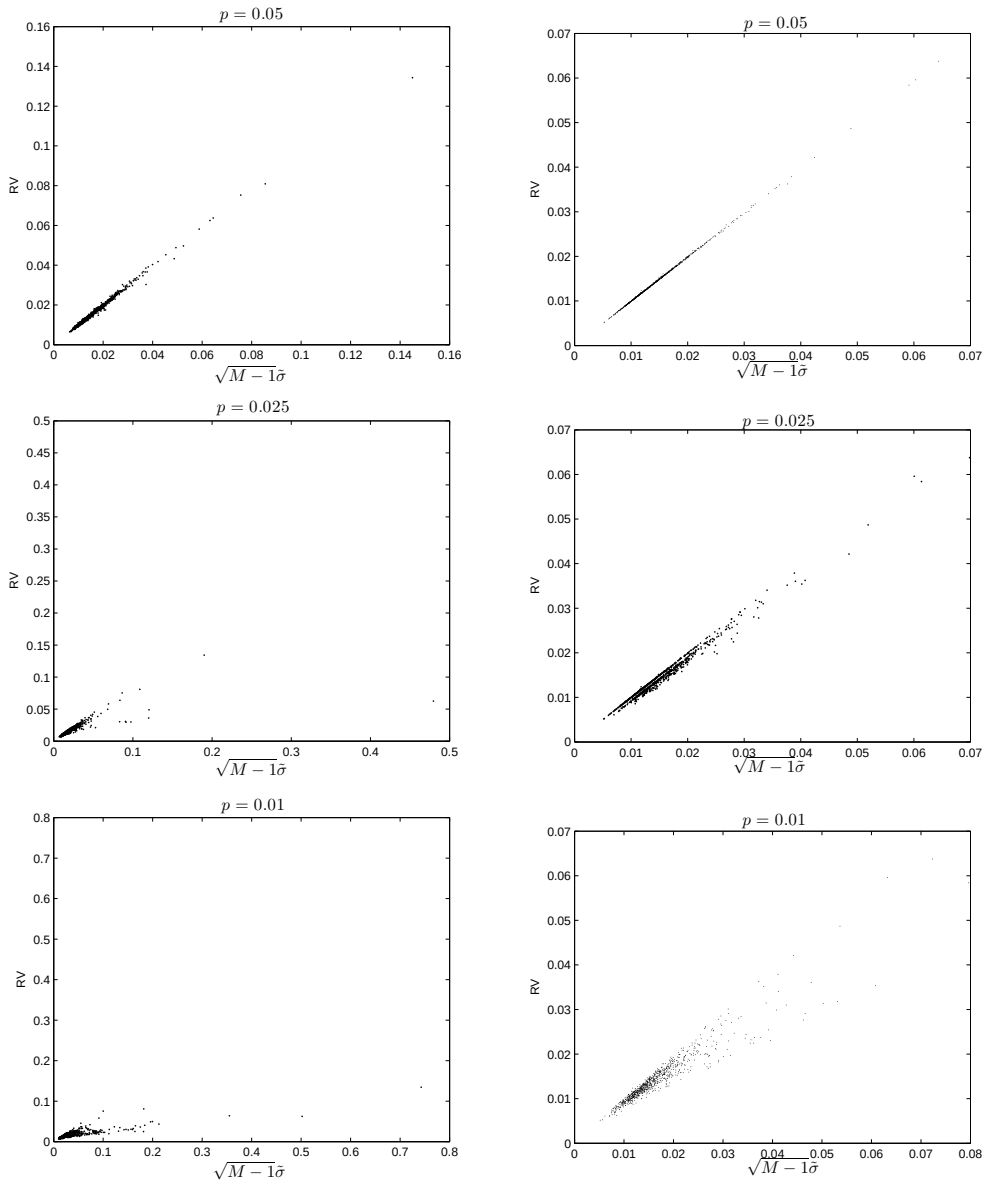

Figure 5.1: Realised volatilities RV_t (5.2) for AXA (left) and France Telecoms (right) against the estimator $\widehat{RV}_t$ (5.5) for p=0.05, 0.025 and 0.01 over a period of 873 days.

Tukey (1965), for $p = 0.05$ the difference of quantile measure gives an almost perfect estimation of realised volatility. Moreover, for p =0.025 and 0.01 realised volatility for AXA is always overestimated with increased discrepancies in more volatile periods. Similarly, for France Telecom it can be observed that $p = 0.05$ produce the best fit, with increase divergence from the RV as the quantile is reduced from $p = 0.025$ to 0.01.

Table 5.1: Mean absolute errors between the realised volatility and the difference of quantile measure, multiplied by 100.

	$p = 0.05$	$p = 0.025$	$p = 0.01$
AXA	0.054	0.472	1.820
FT	0.012	0.075	0.230

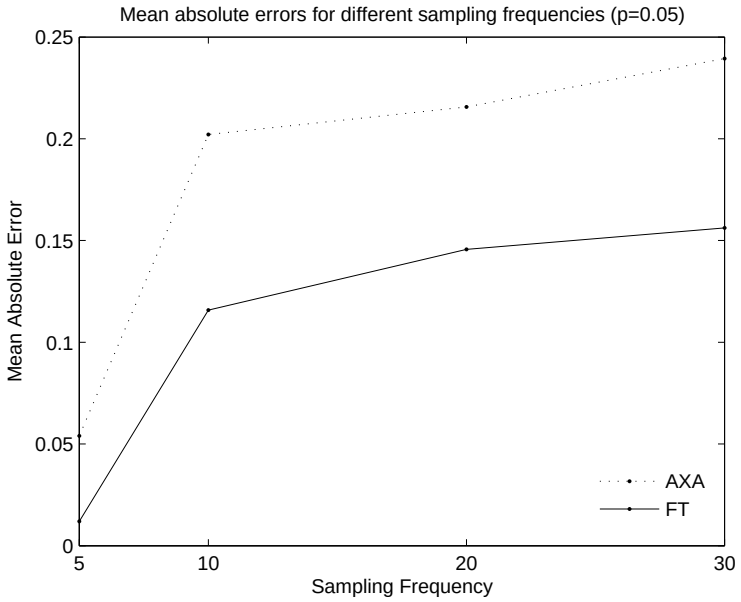

Figure 5.2: Mean absolute errors for sample frequencies 5, 10, 20 and 20 minutes.

This observation is further demonstrated by higher mean absolute error for $p = 0.025$ and 0.01 compared with p =0.05, as shown in Table 5.1. Henceforth, we shall use $p = 0.05$ to estimate the RV given by equation (5.5). Furthermore, to assess the accuracy of the quantile-based measure of RV for different sample sizes in Figure 5.2 we fix $p = 0.05$ and plot the mean absolute errors for sampling frequencies 5, 10, 20 and 30 minutes which correspond to 102, 51, 25 and 17 intraday returns observations. As expected the figure demonstrate that, for the two stocks under consideration, the accuracy of the quantile-based approximation of RV increases with more frequent sampled data (large sample size). However, since high frequency prices are recorded at much shorter time intervals such as every 15 and 30 seconds and the fact that some markets such as the currency exchange are operating for longer than eight hours considered here, the availability of a large sample size is not a restricting

factor.

5.4 The Probability Distribution of $D_t = \hat{\xi}_{1-p,t} - \hat{\xi}_{p,t}$

Our proposed method of constructing prediction intervals for realised volatility requires the knowledge of the distribution of RV_t or equivalently $D_t = \hat{\xi}_{1-p,t} - \hat{\xi}_{p,t}$. Starting with the estimation of quantiles, ξ_p, in the following discussion this concept is presented using known results from order statistics.

Assuming the sequence of random sample of returns $\{R_{t,l}\}_{l=1}^{M}$ on day t are from an unknown continuous distribution $F_t(r)$ with a density function $f_t(r)$, then the corresponding order statistics associated with this sample are defined as $U_{t,(1)}, U_{t,(2)}, ..., U_{t,(M)}$ with $U_{t,(1)} \leq U_{t,(2)} \leq ... \leq U_{t,(M)}$. Dropping the time subscript, an estimator of a quantile of order $p \in (0,1)$, ξ_p, can be defined through the following argument. Let $l = [p(M+1)]$, where $[A]$ is the largest integer less than or equal to A, then the area under probability density function less than $u_{(l)}$ is $P(R \leq u_{(l)}) = F(u_{(l)})$. The expected value of the area is given by

$$E(F(U_{(l)})) = \int_a^b F(u_l) d_l(u_l) du_l \tag{5.6}$$

where

$$d_l(u_l) = \frac{1}{B(l, M-l+1)} (F(u_l))^{l-1} (1 - F(u_l))^{M-l} f(u_l)$$

is the pdf of the l-th order statistic and a and b are its support (Hogg et al., 2005). The quantity $B(.,.)$ is a beta function, and for strictly positive constants ϕ and φ it is defined as

$$B(\phi, \varphi) = \int_0^1 x^{\phi-1} (1-x)^{\varphi-1} dx = \frac{\Gamma(\phi)!\Gamma(\varphi)!}{\Gamma(\varphi+\phi)!}.$$

Introducing a change of variables $z = F(u_l)$ then equation (5.6) can be written as

$$E(F(U_{(l)})) = \int_0^1 \frac{M!}{(l-1)!(M-l)!} z^l (1-z)^{M-l} dz,$$

and on comparing the right hand side of the last equation to the integral of a beta density function

$$f_{beta}(u;\phi,\varphi) = \frac{(\varphi+\phi-1)!}{(\phi-1)!(\varphi-1)!} u^{\phi-1}(1-u)^{\varphi-1}$$

with support $u \in (0,1)$, it follows that

$$E(F(U_{(l)})) = \frac{l}{M+1} \int_0^1 f_{beta}(u;\phi,\varphi)du = \frac{l}{M+1}.$$

From the last equation, the expected area to the left of $u_{(l)}$ is given by $E(P(R \leq u_{(l)})) = l(M+1)^{-1}$ and since $p \approx l(M+1)^{-1}$ then as a direct consequence of the above argument one can estimate the quantile ξ_p by $u_{(l)}$. There exist a number of alternative methods in the estimation of sample quantiles all of which give approximately the same results when the sample size is large, see for example (Parrish, 1990).

By defining $l_1 = [p(M+1)]$ and $l_2 = [(1-p)(M+1)]$ we can find the pdf of $d_{l_1 l_2} = U_{(l_2)} - U_{(l_1)} = \hat{\xi}_{1-p} - \hat{\xi}_p$ (David, 1981) as

$$f_{d_{l_1 l_2}}(d_{l_1 l_2}) = K \int_{-\infty}^{\infty} (F(u))^{l_1} f(u) \{F(u+d_{l_1 l_2}) - F(u)\}^{l_2-l_1-1} \times f(u+d_{l_1 l_2})(1-F(u+d_{l_1 l_2}))^{M-l_2} du, \quad (5.7)$$

where the constant K is defined as

$$K = \frac{M!}{(l_1-1)!(l_2-l_1-1)!(M-l_2)!}.$$

It is widely accepted that the distribution of returns evolves over time and thus in estimating the pdf $f_t(r)$ we take a kernel density estimation approach. That is, for

day t an estimate of the density function of returns is given by

$$\hat{f}_{t,h}(r) = \frac{1}{Mh} \sum_{i=1}^{M} K\left(\frac{r - r_i}{h}\right),$$

where $K(.)$ is a kernel function and h is the bandwidth. Throughout the proceeding discussion we make use of the standard normal density as the kernel function and employ the bandwidth selection procedure proposed by Sheather and Jones (1991). The execution of calculating the density (5.7) as well as the corresponding distribution function requires the usage of quadrature. In this respect, we employ the MATLAB function `trapz` which uses the trapezoidal method.

For ease of referencing, in the course of discussion we will make mention of the distribution of $\widehat{RV}$ as (5.7) where the correction constant $\sqrt{M-1}/\hat{C}(p)$ is implicitly factored in the calculation.

5.5 GARCH (p, q) Model

Due to its immense importance in such fields as risk management, portfolio allocation and derivatives pricing the modelling of volatility dynamics has received a great deal of attention since the introduction of ARCH models by Engle (1982) and its generalisation by Bollerslev (1986). Our proposed forecasting procedures will make use of the GARCH model and thus in what follows we give a brief overview of the model.

Let $\boldsymbol{r}_t = \{r_1, r_2, ..., r_T\}$ be a sequence of logarithmic returns obtained from taking the logarithmic of (close-to-close) prices $P_t = \{P_0, P_1, ..., P_T\}$ between time periods $t-1$ and t such that $r_t = \log(P_t) - \log(P_{t-1})$. A general ARCH$(p)$ can be written as a combination of the mean and conditional variance equations as

$$\sigma_t^2 = c + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2,$$

$$r_t = f(., \varphi) + \varepsilon_t, \qquad \varepsilon_t = \sigma_t z_t \tag{5.8}$$

where z_t is independent and identically distributed (i.i.d) with zero mean and unit variance. The quantity ε_t represent the innovations and the constants $\{c, \alpha_1, \alpha_2, ..., \alpha_p\}$ are parameters to be estimated which by construction of the model must be nonnegative in order to ensure the volatility σ_t is always positive. Furthermore, the conditional mean function $f(., \varphi)$ given by (5.8) is of a general form and where $\varphi = \{\varphi_0, \varphi_1, .., \varphi_m\}$ is a set of parameters associated with the mean equation. The mean equation can be presented in a number of ways such as autoregressive of order m;

$$r_t = \varphi_0 + \sum_{k=1}^{m} \varphi_k r_{t-k} + \varepsilon_t,$$

and depending on the modeller's perspective in some other application this is set to be zero or a constant, see Lunde and Hansen (2005).

As an improvement of the ARCH(p) model, Bollerslev (1986) introduced a generalisation of the ARCH model by adding q autoregressive in the conditional variance equation. The result is a GARCH(p, q) with the following specification,

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

where for $i = 1, 2, ..., p$ and $j = 1, 2, ..., q$ the constraints $\omega > 0, \alpha_i, \beta_j \geq 0$ ensures that the volatility is always positive with corresponding mean equation as in equation (5.8). The GARCH(p, q) has gained immense popularity due to its parsimonious representation of volatility, and its accuracy in capturing volatility clustering. Furthermore, in most application small p and q values are found to be sufficient in modelling a variety of financial time series such as equities and currency exchange rates. A GARCH(p, q) model is known to be stationary with finite variance

if its satisfies

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1,$$

and in the special case the previous equation sum to unity the GARCH(p, q) is referred to as integrated I-GARCH(p, q) process.

The estimation of parameters of the GARCH model are achieved through a number of techniques, with the maximum likelihood approach by far the most popular. In addition to the conditional mean and variance equations, an essential ingredient prior to the application of ML method is the specification of the distribution of z_t terms. Although the normal distribution is widely used due to its good properties, there exist abundant empirical evidences (see for example, Jondeau et al. (2006)) indicating that the distribution of the residuals ε_t deviates from normality. However, it has been shown by Weiss (1986) that the maximum likelihood estimator produce consistent estimates of parameters with a limiting Gaussian distribution.

For empirical application we shall use a student-t distribution which, in comparison to the Gaussian, provide a better capture of kurtosis effect observed in financial time series.

5.6 Prediction Intervals for Realised Volatility

In this section we describe our proposed method for constructing prediction intervals for realised volatility through modelling intraday return quantiles using the AR-GARCH model. For comparison purposes we also outline a bootstrap procedure based on the same idea. Furthermore, we introduce two adapted extensions of the HAR of Corsi (2009).

5.6.1 AR(1)-GARCH(1,1) Prediction Intervals

Let $\boldsymbol{P}_{\widetilde{t}} = \{P_0, P_1, \ldots, P_M\}$ denote the evolution of intraday prices observed, without loss of generality, at equidistant time points $\widetilde{t} = 0, 1, \ldots, M$. The dash superscript is used to distinguish between intraday and daily time periods with the latter denoted by t. Furthermore, let $\boldsymbol{r}_{\widetilde{t}} = \{r_1, r_2, \ldots, r_M\}$ be a vector of observed logarithmic returns obtained from the corresponding prices.

For each day t of an in-sample period $t = 1, 2, \ldots, T$ we employ the richness of high frequency data to calculate empirical quantiles corresponding to the number (M) of logarithmic return observations available during the day and thus create time series of quantile dynamics. To elucidate the idea consider $\boldsymbol{r}_{\widetilde{t}} = \{r_1, r_2, \ldots, r_M\}$ obtained on a specific day, then the procedures are as follows:

1. Create a vector $\boldsymbol{\phi} = \{(M+1)^{-1}, 2(M+1)^{-1}, \ldots, M(M+1)^{-1}\}$ and for each $\{\phi_{\widetilde{t}}\}_{\widetilde{t}=1}^{M}$ calculate the corresponding quantile as $u_{([\phi_{\widetilde{t}}(M+1)])}$ where u's are the order statistics of $\boldsymbol{r}_{\widetilde{t}}$;

2. Repeat the last step for all time periods $t = 1, 2, \ldots, T$ such that for each $100\phi_{\widetilde{t}}\%$ quantile we construct a univariate time series of the evolution of returns as $\hat{\boldsymbol{r}}_{\phi_{\widetilde{t}},t} = \{\hat{r}_{\phi_{\widetilde{t}},1}, \hat{r}_{\phi_{\widetilde{t}},2}, \ldots, \hat{r}_{\phi_{\widetilde{t}},T}\}$ with a total of M time series corresponding to different quantiles;

3. For each quantile time series fit an AR(1)-GARCH(1,1)

$$\hat{r}_{\phi_{\widetilde{t}},t} = \varphi_0 + \varphi_1 \hat{r}_{\phi_{\widetilde{t}},t-1} + \varepsilon_{\phi_{\widetilde{t}},t}, \qquad \sigma^2_{\phi_{\widetilde{t}},t} = \omega + \alpha \varepsilon^2_{\phi_{\widetilde{t}},t-1} + \beta \sigma^2_{\phi_{\widetilde{t}},t-1} \tag{5.9}$$

   and extract one step ahead mean and volatility forecasts from the corresponding AR and GARCH specifications as $\{\hat{r}_{\phi_{\widetilde{t}},T+1}\}_{\widetilde{t}=1}^{M}$ and $\{\hat{\sigma}_{\phi_{\widetilde{t}},T+1}\}_{\widetilde{t}=1}^{M}$;

4. De-mean the forecasted quantiles as, $\hat{\boldsymbol{r}}^{D}_{T+1} = \{\hat{r}_{\phi_1,T+1} - \overline{r}, \hat{r}_{\phi_2,T+1} - \overline{r}, \ldots, \hat{r}_{\phi_M,T+1} - \overline{r}\}$ where $\overline{r} = M^{-1} \sum_{\widetilde{t}=1}^{M} \hat{r}_{\phi_{\widetilde{t}},T+1}$

5. Use the de-meaned mean forecasts $\hat{\boldsymbol{r}}^{D}_{T+1}$ to calculate $\hat{\xi}_{1-p}$ and $\hat{\xi}_{p}$ quantiles from order statistics and compute $\hat{C}(p) = \hat{\xi}^{*,Y}_{1-p} - \hat{\xi}^{*,Y}_{p}$ where $\hat{\xi}^{*,Y}_{1-p}$ and $\hat{\xi}^{*,Y}_{p}$ are quantiles calculated from the empirical distribution of standardised mean forecasts, $\hat{\boldsymbol{Y}} = \left\{ \frac{\hat{r}_{\phi_1,T+1}}{\hat{\sigma}_{\phi_1,T+1}}, \frac{\hat{r}_{\phi_2,T+1}}{\hat{\sigma}_{\phi_2,T+1}}, \ldots, \frac{\hat{r}_{\phi_M,T+1}}{\hat{\sigma}_{\phi_M,T+1}} \right\}$

6. Finally, calculate an estimate of RV from equation (5.5) and employ numerical integration to recover the density function (5.7) from which the end points $RV^{*}_{\gamma/2}$ and $RV^{*}_{1-\gamma/2}$ of $100(1-\gamma)\%$ prediction interval are obtained

The calculation of the normalising constant $\hat{C}(p)$ in step 5 can be approached from a different perspective. Specifically, for each quantile of time series $\hat{r}_{\phi_{\tilde{t}},t}$ the in-sample conditional volatilities $\{\hat{\sigma}_{\phi_{\tilde{t}},t}\}^{T}_{t=1}$ can be used to obtain standardise returns, $\hat{\boldsymbol{z}}_{t,\phi_{\tilde{t}}} = \hat{\boldsymbol{r}}_{\phi_{\tilde{t}},t}/\hat{\sigma}_{\phi_{\tilde{t}},t}$, and thereafter the end quantiles $\hat{\xi}^{*,Z}_{1-p,\phi_{\tilde{t}}}$ and $\hat{\xi}^{*,Z}_{p,\phi_{\tilde{t}}}$ are evaluated from the empirical distribution. Finally, the correction constant is calculated by $\hat{C}(p) = \overline{\xi}_{1-p} - \overline{\xi}_{p}$ where $\overline{\xi}_{p}$ and $\overline{\xi}_{1-p}$ are the averages of the end quantiles $\{\hat{\xi}^{*,Z}_{p,\phi_{\tilde{t}}}\}^{M}_{\phi_{\tilde{t}}=1}$ and $\{\hat{\xi}^{*,Z}_{1-p,\phi_{\tilde{t}}}\}^{M}_{\phi_{\tilde{t}}=1}$ respectively.

From the procedures described above, a number of key assumptions and remarks are needed.

- In calculating intraday quantiles of returns we implicitly assume that the returns are independent and identically distributed. Justification of such an assumption is supported by the work of Gonçalves and Meddahi (2009) who proposed an *iid* and wild bootstrap methods for calculating confidence intervals for integrated volatility. The premise of the *iid* method is based on the empirical finding of high persistence of, and almost constant, volatility over the course of a trading day.

- In comparison to volatility, the forecasting of next period returns is a difficult task. This is due to the fact that empirical evidences indicate that the mean of financial time series returns is very close to zero. However, quantile modelling and forecasting offer a distinct advantage in the sense that the sign of time series

remain constant over time (especially for quantiles away from the medians) and thus improves on the forecasting accuracy as demonstrated by (Engle and Manganelli, 2004; Xiao and Koenker, 2009) in the context of predicting value at risk.

- As will be demonstrated in the empirical study of Section 5.8, the use of AR-GARCH accounts for the heteroscedaticity observed within quantile time series.

5.6.2 AR(1)-GARCH(1,1) Bootstrapped Prediction Intervals

Following Pascual et al. (2006) who used a bootstrap procedure to create prediction intervals for volatilities of close-to-close returns obtained from GARCH model, here we employ their idea to construct prediction intervals for realised volatility using the difference of extreme quantiles.

We begin by calculating $100p\%$ and $100(1-p)\%$ intraday quantiles of returns and form respective time series of the quantile dynamics over the in-sample period as, $\hat{\boldsymbol{r}}_{p,t} = \{\hat{r}_{p,1}, \hat{r}_{p,2}, \ldots, \hat{r}_{p,T}\}$ and $\hat{\boldsymbol{r}}_{1-p,t} = \{\hat{r}_{1-p,1}, \hat{r}_{1-p,2}, \ldots, \hat{r}_{1-p,T}\}$. We then proceed by fitting AR(1)-GARCH(1,1) model to $\hat{\boldsymbol{r}}_{p,t}$ and $\hat{\boldsymbol{r}}_{1-p,t}$ and estimate the parameter sets, $\hat{\boldsymbol{\theta}}_p = \{\hat{\varphi}_{0,p}, \hat{\varphi}_{1,p}, \hat{\omega}_p, \hat{\alpha}_p, \hat{\beta}_p\}$ and $\hat{\boldsymbol{\theta}}_{1-p} = \{\hat{\varphi}_{0,1-p}, \hat{\varphi}_{1,1-p}, \hat{\omega}_{1-p}, \hat{\alpha}_{1-p}, \hat{\beta}_{1-p}\}$.

Next, for each of the two time series we obtain one-step ahead quantile forecasts, $\hat{r}_{p,T+1}$ and $\hat{r}_{1-p,T+1}$, as well as the associated volatility forecasts, $\hat{\sigma}_{p,T+1}$ and $\hat{\sigma}_{1-p,T+1}$. Next period forecast of realised volatility is calculated as,

$$\widehat{RV}_{T+1} = \frac{\sqrt{M-1}(\hat{r}_{1-p,T+1} - \hat{r}_{p,T+1})}{\widehat{C}_p}, \tag{5.10}$$

where $\widehat{C}_p = \hat{r}_{1-p,T+1}\hat{\sigma}^{-1}_{1-p,T+1} - \hat{r}_{p,T+1}\hat{\sigma}^{-1}_{p,T+1}$. Note that, the correction factor of $\sqrt{M-1}$ is used based on our observation in section 5.3.1 equation (5.5).

In order to apply the bootstrap procedure for each quantile we employ the estimated parameter sets to reproduce the original series. That is, for $100p\%$ a

replication $\hat{\boldsymbol{r}}^*_{p,t} = \{\hat{r}^*_{p,1}, \hat{r}^*_{p,2}, \ldots, \hat{r}^*_{p,T}\}$ of the original series is obtained by

$$\hat{r}^*_{p,t} = \hat{\varphi}_{p,0} + \hat{\varphi}_1 \hat{r}^*_{p,t-1} + z^*_t \hat{\sigma}^*_{p,t}, \qquad \sigma^{*2}_{p,t} = \hat{\omega}_p + \hat{\alpha}_p \varepsilon^{*2}_{t-1} + \hat{\beta}_p \sigma^{*2}_{t-1} \tag{5.11}$$

where the initial value $\hat{r}^*_{p,1}$ is calculated as $100p\%$ quantile from re-sampling with replacement intraday returns of the first day, $\hat{\sigma}^{*2}_{p,1} = \hat{\omega}_p(1 - \hat{\alpha}_p - \hat{\beta}_p)^{-1}$ is the unconditional variance and z^*_t are sampled with replacement from the empirical distribution of the standardised zero centred residuals $\hat{F}_z$. Analogously, same steps can be taken to construct bootstrap replicated time series $\hat{\boldsymbol{r}}^*_{1-p,t} = \{\hat{r}^*_{1-p,1}, \hat{r}^*_{1-p,2}, \ldots, \hat{r}^*_{1-p,T}\}$ for $\hat{\boldsymbol{r}}_{1-p,t}$.

Given the new time series, for both quantiles the AR(1)-GARCH(1,1) is re-fitted and corresponding parameter sets re-estimated such that we have $\hat{\boldsymbol{\theta}}^*_p = \{\hat{\varphi}^*_{0,p}, \hat{\varphi}^*_{1,p}, \hat{\omega}^*_p, \hat{\alpha}^*_p, \hat{\beta}^*_p\}$ and $\hat{\boldsymbol{\theta}}^*_{1-p} = \{\hat{\varphi}^*_{0,1-p}, \hat{\varphi}^*_{1,1-p}, \hat{\omega}^*_{1-p}, \hat{\alpha}^*_{1-p}, \hat{\beta}^*_{1-p}\}$ which are then used to forecast next period realised volatility as in equation (5.10) with $\hat{r}_{p,T+1}$, $\hat{r}_{1-p,T+1}$ and $\widehat{C}_p$ replaced by $\hat{r}^*_{p,T+1}$, $\hat{r}^*_{1-p,T+1}$ and $\widehat{C}^*_p$ respectively.

Repeating this process a large number of times B, where $b = 1, 2, \ldots, B$, will result in a distribution of next period realised volatility $\{\widehat{RV}^{(b)}_{T+1}\}^B_{b=1}$. Finally, the end quantiles of $100(1-\gamma)\%$ prediction interval for $\widehat{RV}_{T+1}$ are calculated as

$$\left[F^{-1}_{\widehat{RV}_{T+1}}\left(\frac{\gamma}{2}\right), F^{-1}_{\widehat{RV}_{T+1}}\left(\frac{1-\gamma}{2}\right)\right], \tag{5.12}$$

where $F^{-1}_{\widehat{RV}_{T+1}}$ is the empirical quantile function of $\widehat{RV}_{T+1}$.

5.6.3 HAR Model

One of the most simple, easy to implement and quite accurate realised volatility forecasting model is the Heterogenous AutoRegressive (HAR) model proposed by Corsi (2009). In this model the author makes use of additive processes with heterogeneous components in order to reproduce the stylised facts observed in time series of realised volatility, the most significant of which is long memory as depicted

by hyperbolically decaying autocorrelation function. Building on the Heterogenous Market Hypothesis of Muller et al. (1993) and the heterogeneous autoregressive conditional heteroskedasticity (HARCH) model proposed by Muller et al. (1997), Corsi introduced a HAR model that makes use of heterogenous realised volatility components in the form of daily, weekly and monthly realised volatilities in order to model and forecast realised volatility. The three component HAR model can be conveniently defined as

$$RV_{t+1} = \alpha_0 + \alpha_d RV_t + \alpha_w RV_{t-5:t} + \alpha_m RV_{t-22:t} + \epsilon_{t+1} \tag{5.13}$$

with $RV_{t+1-h:t} = \frac{1}{h}\sum_{k=0}^{h-1} RV_{t-k}$ and where ϵ_{t+1} is the Gaussian noise. For $h = 5$ and $h = 22$, $\frac{1}{h}\sum_{k=0}^{h-1} RV_{t-k}$ corresponds to average weekly and monthly realised volatilities respectively. Since the daily, weekly and monthly components comprising the HAR model can be calculated from historic data the parameters of the model can be readily estimated using the simple ordinary least square technique.

5.6.4 HAR-GARCH Model

After having observed the autoregressive conditional heteroskedastic nature of residuals obtained from the HAR model, Corsi et al. (2008) introduce HAR-GARCH model in which the residuals are linked with a conditional variance GARCH specification. Thus HAR-GARCH(1,1) model takes the form

$$\begin{aligned} RV_t = \alpha_0 + \alpha_d RV_{t-1} + \alpha_w RV_{t-5:t-1} + \alpha_m RV_{t-22:t-1} + \sigma_t z_t \\ \sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \\ \epsilon_t = \sigma_t z_t, \qquad z_t|\mathcal{F}_{t-1} \sim f(0,1) \end{aligned} \tag{5.14}$$

where $\mathcal{F}_{t-1}$ denotes the information set available up until time $t-1$.

In the empirical study to follow we apply the HAR-GARCH(1,1) characterised

by student-t error distribution with ν degrees of freedom, t_ν. In this model, realised volatility is modelled directly and one period ahead $100(1-\gamma)\%$ intervals are computed as

$$\left[\hat{RV}_{T+1} + \hat{\sigma}_{T+1} t_\nu^{-1}\left(\frac{\gamma}{2}\right), \hat{RV}_{T+1} + \hat{\sigma}_{T+1} t_\nu^{-1}\left(\frac{1-\gamma}{2}\right)\right], \tag{5.15}$$

where $t_\nu^{-1}(.)$ is the quantile function of the standardised t distribution with ν degrees of freedom and $\hat{RV}_{T+1}$ is the one-step ahead forecast of RV from the HAR-GARCH(1,1) model.

5.6.5 Quantile-HAR-GARCH Model

The construction of the HAR-GARCH model is build from the observation of very strong persistence of realised volatility lasting for a long period of time together with volatility clustering of the residuals of the original HAR model, (5.13). Based on this finding we propose an adapted extension of this model from a quantile perspective with the goal of creating prediction intervals for RV. In contrast to the HAR-GARCH that directly models RV, our proposed modification of the model henceforth termed Q_{RV}HAR-GARCH, employs in-sample daily quantiles of realised volatilities in order to project the end points of a prediction interval. Specifically, using the distribution of the difference of two quantiles (5.7) for each trading day we calculate the in-sample lower and upper bounds of a $100(1-\gamma)\%$ confidence intervals for $\{\widehat{RV}\}_{t=1}^{T}$ as $\{\widehat{RV}_{\gamma/2,t} = \mathcal{L}_t, \widehat{RV}_{1-\gamma/2,t} = \mathcal{U}_t\}_{t=1}^{T}$. Thereafter, these in-sample series are independently modelled using a HAR-GARCH(1,1) model as

$$\begin{aligned}
\mathcal{L}_t &= \alpha_{0,\mathcal{L}} + \alpha_{d,\mathcal{L}}\mathcal{L}_{t-1} + \alpha_{w,\mathcal{L}}\mathcal{L}_{t-5:t-1} + \alpha_{m,\mathcal{L}}\mathcal{L}_{t-22:t-1} + \sigma_{\mathcal{L},t} z_t \\
\sigma^2_{\mathcal{L},t} &= \omega_{\mathcal{L}} + \alpha_{\mathcal{L}}\epsilon^2_{\mathcal{L},t-1} + \beta_{\mathcal{L}}\sigma^2_{\mathcal{L},t-1} \qquad \epsilon_{\mathcal{L},t} = \sigma_{\mathcal{L},t} z_t
\end{aligned} \tag{5.16}$$

and

$$\begin{aligned} \mathcal{U}_t &= \alpha_{0,\mathcal{U}} + \alpha_{d,\mathcal{U}}\mathcal{U}_{t-1} + \alpha_{w,\mathcal{U}}\mathcal{U}_{t-5:t-1} + \alpha_{m,\mathcal{U}}\mathcal{U}_{t-22:t-1} + \sigma_{\mathcal{U},t}z_t \\ \sigma^2_{\mathcal{U},t} &= \omega_{\mathcal{U}} + \alpha_{\mathcal{U}}\epsilon^2_{\mathcal{U},t-1} + \beta_{\mathcal{U}}\sigma^2_{\mathcal{U},t-1} \qquad \epsilon_{\mathcal{U},t} = \sigma_{\mathcal{U},t}z_t. \end{aligned} \tag{5.17}$$

Finally, $100(1-\gamma)\%$ prediction intervals for RV

$$\Big[\mathcal{L}_{T+1}, \mathcal{U}_{T+1}\Big], \tag{5.18}$$

where $\mathcal{L}_{T+1}$ and $\mathcal{U}_{T+1}$ are forecasted quantiles from (5.16) and (5.17) respectively.

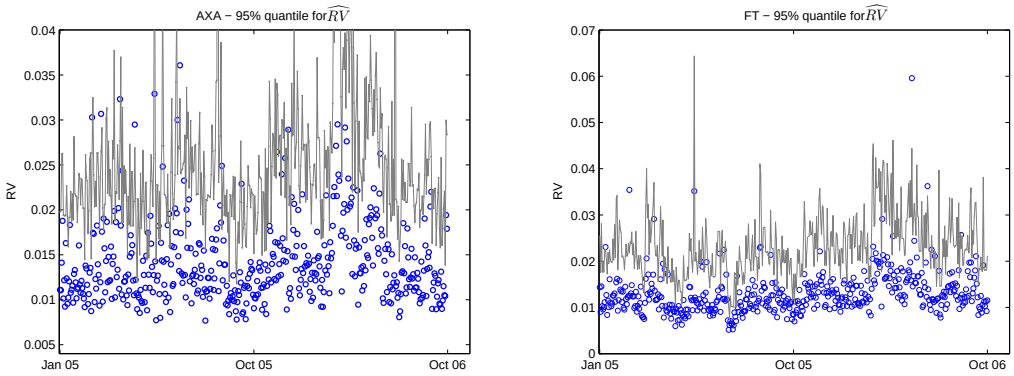

Figure 5.3: 95% quantiles of $\widehat{RV}$ calculated from the distribution of the difference of two quantiles, (5.7) (grey) and RV from equation (5.5) in (blue dots) for the first 473 days.

Figure 5.3 displays 95% quantile for $\widehat{RV}$ over a period of 473 days for both AXA and France Telecom, where for the former the four cases for which RV is greater than 0.04 are removed from the figure. Using RV as benchmark, in this period the estimated quantile are conservative with coverage levels 0.9771 and 0.9897 for AXA and FT respectively. Furthermore, the figure (especially AXA) exemplify the clustering effect exhibited in quantiles of $\widehat{RV}$ where periods of small(large) changes persist over time. Similarly, by way of illustration Figure 5.4 depicts sample autocorrelation functions (ACF) of realised volatilities for both AXA and France Telecom. On the same figure we also plot the ACFs of 5% and 95% quantiles of

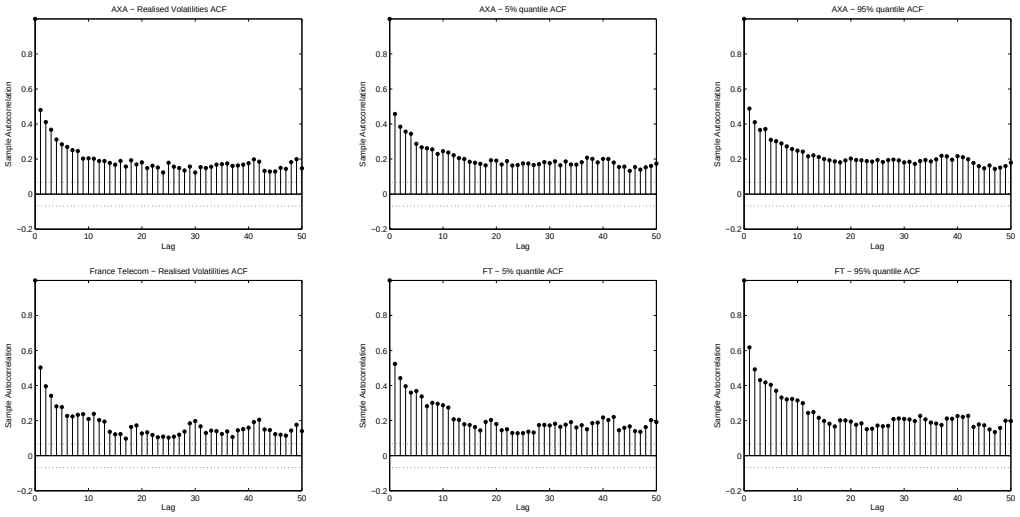

Figure 5.4: Sample autocorrelation functions (ACF) for realised volatilities together with ACFs for 5% and 95% quantiles for $\widehat{RV}$ calculated from the distribution of the difference of two quantiles, (5.7), for both AXA (top row) and France Telecom (bottom row).

a quantile-based realised volatility measure (5.5). It can be observed that, even at lag 50, the hyperbolic decay evident in the raw RV is also captured in the quantiles of both equities and thus supports the adaptation of the HAR-GARCH to model quantiles.

5.6.6 Quantile-HAR Model

Availability of covariates in the form of daily, weekly and monthly components calculated from the historical data permits the forecasting of next period realised volatility from the HAR model where the parameters are estimated by an easily implemented OLS procedure. By construction, the HAR model cannot be used to build prediction intervals since the next period estimates of standard deviation is not available. However, as a natural extension from the ordinary least square setting, it is possible to create prediction intervals of next period RV using quantile regression.

Although not GARCH-based, we propose a nonparametric modified extension of the original HAR model which we call Quantile-HAR (Q-HAR). Similar to the Q_{RV}HAR-GARCH, the proposed adaptation create $100(1-\gamma)\%$ prediction intervals

through independently forecasting the lower and upper bounds forming the band. Specifically, implementation of the Q-HAR utilises the quantile regression technique discussed in Section 4.1 of Chapter 4 in order to estimate the parameters of $100\gamma/2\%$ and $100(1-\gamma/2)\%$ regression quantiles. The fitted regression quantiles corresponding to the lower and upper bounds are,

$$\begin{aligned} RV_{t+1,\gamma/2} &= \alpha_{0,\gamma/2} + \alpha_{d,\gamma/2} RV_t + \alpha_{w,\gamma/2} RV_{t-5:t} + \alpha_{m,\gamma/2} RV_{t-22:t} \\ RV_{t+1,1-\gamma/2} &= \alpha_{0,1-\gamma/2} + \alpha_{d,1-\gamma/2} RV_t + \alpha_{w,1-\gamma/2} RV_{t-5:t} + \alpha_{m,1-\gamma/2} RV_{t-22:t}, \end{aligned} \tag{5.19}$$

with a $100(1-\gamma)\%$ prediction interval formed as

$$[\hat{RV}_{T+1,\gamma/2}, \hat{RV}_{T+1,1-\gamma/2}], \tag{5.20}$$

where $\hat{RV}_{T+1,\gamma/2}$ is the next prediction for the $100\gamma/2\%$ quantile of RV.

We note that Taylor and Bunn (1999) has employed quantile regression technique to build prediction intervals for exponential smoothing methods through regressing empirical fit errors in order to forecast error quantiles. A key distinct difference of this method with the QHAR is on the fact that for the latter a vector of explanatory variables are readily available and prediction intervals are built directly rather than fitting regression quantiles to a series of past forecast errors.

5.7 Evaluation of Interval Forecasts

In this section we present a number of tests that will be employed in evaluating the coverage accuracy of the proposed models.

One of the most widely used method for testing interval forecasts is the hit test introduced by Christoffersen (1998). Let $\{f_{t+1|t}\}_{t=1}^{\mathcal{T}}$ denote a sequence of next period forecasts of some variable of interest predicted using available information at

time t. Then, for lower and upper endpoints of an interval with $\theta \in [0,1]$ probability, $\{L_{t+1|t}\}_{t=1}^{\mathcal{T}}$ and $\{U_{t+1|t}\}_{t=1}^{\mathcal{T}}$, the test is build by defining a sequence of the so-called hit variable $\{Hit_{t+1}\}_{t=1}^{\mathcal{T}}$

$$Hit_{t+1} = I(L_{t+1|t} \leq f_{t+1|t} \leq U_{t+1|t}).$$

To evaluate the accuracy of models using the nominal coverage criterion Christoffersen (1998) demonstrate that this is equivalent to testing whether $\{Hit_{t+1}\}_{t=1}^{\mathcal{T}}$ is independent and identically Bernoulli distributed with parameter θ. That is,

$$H_0 : E(Hit_{t+1}) = \theta \qquad \text{against} \qquad H_1 : E(Hit_{t+1}) \neq \theta$$

where $E(Hit_{t+1}) = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} Hit_{t+1}$. The likelihood ratio used to conduct this test is given by

$$LR_{uc} = -2 \log \left(\frac{\ell(\theta)}{\ell(\hat{\theta})} \right) \tag{5.21}$$

where $H = \sum_{t=1}^{\mathcal{T}} Hit_{t+1}$, $\ell(\theta) = (1-\theta)^{n_0} \theta^{n_1}$, $\ell(\hat{\theta}) = (1-\hat{\theta})^{n_0} \hat{\theta}^{n_1}$ and $\hat{\theta} = n_1/(n_0+n_1)$ with $n_1 = H$ and $n_0 = \mathcal{T} - H$. Under the null hypothesis $H_0 : E(Hit_{t+1}) = \theta$, the likelihood ratio test is asymptotically chi-squared distributed with one degree of freedom, $\chi^2_{(1)}$.

While this simple unconditional test is useful, it does not differentiate between dependency of violations of forecasts from different models. To circumvent this and to devise a test that can detect dependency in the forecasts, Christoffersen (1998) further extends the unconditional likelihood ratio test (LR_{uc}) and introduce a likelihood ratio independence test (LR_{idn}) with a test statistic

$$LR_{ind} = -2 \log \left(\frac{\ell(\hat{\Lambda}_2)}{\ell(\hat{\Lambda}_1)} \right) \tag{5.22}$$

where

$$\ell(\hat{\Lambda}_1) = (1 - \theta_{01})^{n_{00}} (\theta_{01})^{n_{01}} (1 - \theta_{11})^{n_{10}} (\theta_{11})^{n_{11}}$$

and

$$\ell(\hat{\Lambda}_2) = (1-\theta_2)^{(n_{00}+n_{10})}(\theta_2)^{(n_{01}+n_{11})}.$$

The asymptotic distribution of this test is also $\chi^2_{(1)}$ where the quantity n_{ij} denotes the number of observations with value i followed by j with $\hat{\theta}_{ij} = n_{ij}/(n_{i0}+n_{i1})$ and θ_2 is estimated by $\hat{\theta}_2 = (n_{01}+n_{11})/\mathcal{T}$.

Finally, Christoffersen (1998) combined the unconditional and independence tests to construct a more powerful conditional coverage test with a test statistic

$$LR_{cc} = -2\log\left(\frac{\ell(\theta)}{\ell(\hat{\Lambda}_1)}\right) \tag{5.23}$$

that is under the null hypothesis asymptotically distributed as $\chi^2_{(2)}$.

For the empirical study we shall employ the unconditional and conditional coverage tests to assess the forecasting accuracy of the proposed models in constructing intervals for realised volatility by replacing $f_{t+1|t}$ with RV_{t+1}, out-of-sample RV.

5.8 Empirical Study

In this section we apply the methods described in the previous section to construct next day prediction intervals for realised volatility using the high frequency data outlined in section 5.2. By way of illustration, the dynamics of realised volatilities together with 5% and 95% quantiles of intraday returns for the two equities are depicted in Figure 5.5. Similarly, in Table 5.2 we present estimated parameters from fitting the fitted AR(1)-GARCH(1,1) model on 5% and 95% intraday quantile time series for both AXA and France Telecom.

Figure 5.5 clearly illustrate the relationship between extreme intraday quantiles and realised volatility. From the figure, it can be observed that periods of high volatilities are reflected in large extreme quantiles. Additionally, Figure 5.5 suggest a degree of presence of heteroscedastic effect in the intraday return quantiles. However,

such an effect may not be present for the whole spectrum of intraday time series quantiles, for instance the median time series (or those very close to it) are almost zeros and thus may not possess such characteristic. In this respect the usage of GARCH conditional variance specification provide flexibility in accounting for this important feature. Furthermore, this observation is illustrated through a careful investigation of Table 5.2 which reports a snapshot of parameters estimated using the first 473 days for both AXA and FT with a student-t innovation with ν degrees of freedom. From the figure it can be observed that there are difference in the estimated parameters of the 5% and 95% quantiles.

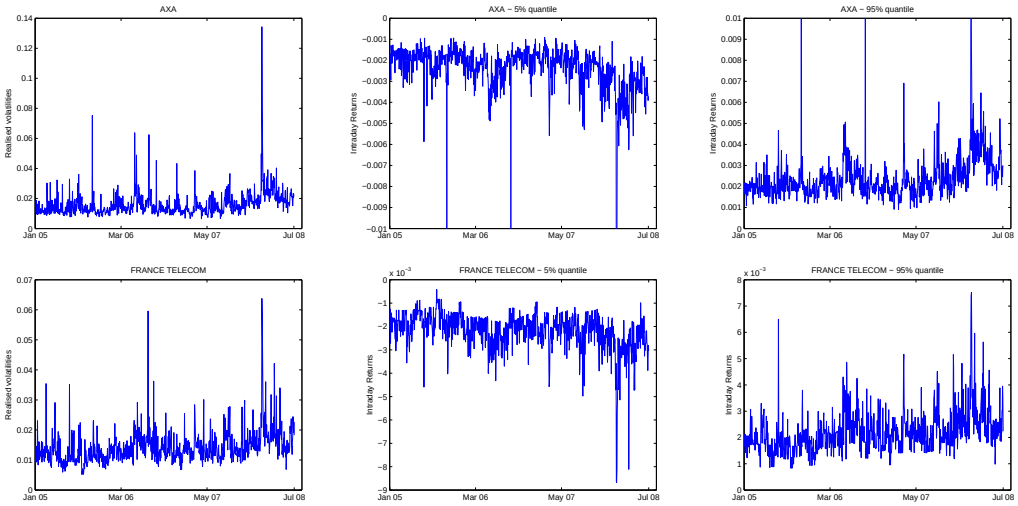

Figure 5.5: Realised volatilities together with intraday 5% and 95% quantiles calculated from high frequency data for AXA and France Telecom .

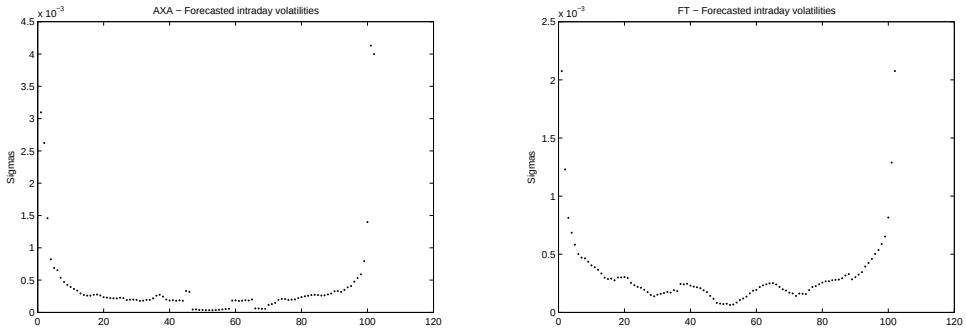

Figure 5.6: Intraday volatility forecasts for day 474 using the AR(1)-GARCH(1,1) model for AXA and France Telecom.

Table 5.2: Parameter estimates for the in-sample period January 2005 to September 2006 (473 observations). [Standard errors for parameter estimates are in braces].

	$\hat{\varphi}_0$	$\hat{\varphi}_1$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\nu}$
AXA						
$\hat{\boldsymbol{r}}_{5\%,t}$	-0.0012	0.3900	2.5588e-7	0.2464	0.3502	2.8756
	[8.6153e-5]	[0.0409]	[9.2543e-8]	[0.1437]	[0.1669]	[0.3009]
$\hat{\boldsymbol{r}}_{95\%,t}$	0.0013	0.3422	1.7098e-7	0.3344	0.2824	3.4818
	[9.8609e-5]	[0.0473]	[4.6333e-8]	[0.1336]	[0.1387]	[0.3584]
FT						
$\hat{\boldsymbol{r}}_{5\%,t}$	-0.0012	0.3975	3.0941e-8	0.0595	0.8404	7.0180
	[8.5470e-5]	[0.0412]	[8.0050e-9]	[0.0220]	[0.0069]	[1.0529e-4]
$\hat{\boldsymbol{r}}_{95\%,t}$	0.0011	0.4133	3.7270e-8	0.1125	0.7864	4.8737
	[8.0841e-5]	[0.0398]	[2.2008e-8]	[0.0491]	[0.0934]	[4.0726e-4]

Figure 5.6 displays intraday volatility forecasts made from fitting the AR(1)-GARCH(1,1) model to time series of intraday quantiles using the first 473 days. The figure gives a static presentation of almost constant intraday instantaneous volatilities, a result of which was used by Gonçalves and Meddahi (2009) to build confidence intervals for integrated volatility at a given time point. To re-emphasise, here we also call on this assumption by treating the forecasted intraday returns as identically distributed in order to construct prediction intervals for (5.5).

For the purpose of forecasting intraday quantiles the parameters of all models described in Section 5.6 are sequentially re-estimated daily on a moving window of 473 most recent observations and then we perform out-of-sample interval forecasts covering 400 days. It should be noted that, for the AR(1)-GARCH(1,1) outlined in subsection 5.6.1 this involves estimating the parameters M times daily, once for every quantile. Similarly, for the AR(1)-GARCH(1,1) bootstrap procedure the parameters are estimated $B+1$ times for 5% and 95% quantile time series, where the number of bootstraps B is set at 999.

Our main empirical findings are presented in Table 5.3 which displays the unconditional coverage $C = \mathcal{T}^{-1}\sum_{t=1}^{\mathcal{T}} Hit_t$ together with the test statistics for the LR_{uc} and LR_{cc} tests at both 90% and 95% significance levels. For 90% (95%)

significance level the critical values of the LR_{uc} and LR_{cc} tests are 2.7055 (3.8415) and 4.6052 (5.9915) respectively.

Table 5.3: 90% and 95% coverages together with LR_{uc} and LR_{cc} test statistics evaluated on the out-of-sample period, $\mathcal{T} = 400$ days.

	90%			95%		
	C	LR_{uc}	LR_{cc}	C	LR_{uc}	LR_{cc}
AXA						
AR(1)-GARCH(1,1)	0.8750	2.5947	2.8770	0.9300	3.0121	3.1581
Bootstrap	0.9250	3.0143	3.2055	0.9600	0.9014	1.1711
HAR-GARCH	0.0875	-	-	0.0925	-	-
Q_{RV}HAR-GARCH	0.8325	17.1966	30.4260	0.8600	46.8410	52.7918
Q-HAR	0.8975	0.0276	0.2577	0.9475	0.0518	2.4944
France Telecom						
AR(1)-GARCH(1,1)	0.9125	0.7219	10.8629	0.9325	5.2396	5.3979
Bootstrap	0.8625	5.6638	10.5301	0.9075	12.2959	12.2959
HAR-GARCH	0.0450	-	-	0.0650	-	-
Q_{RV}HAR-GARCH	0.8275	20.7452	31.6795	0.8750	35.9124	45.9526
Q-HAR	0.9075	0.2558	6.2250	0.9275	3.7655	9.9147

The main results for proposed procedures in light of their accuracy in covering next period realised volatility can be summarized as follows.

- While the HAR-GARCH has been shown by Corsi et al. (2008) to provide quite accurate forecasts of next period realised volatility this observation does not extend to prediction intervals. From table 5.3 we observe noticeable very poor nominal coverage levels for HAR-GARCH at both 90% and 95% which is reflected in the failure of LR_{uc} and LR_{cc} tests, with the hyphen sign indicating the tests statistics is infinite. Similarly, while the nominal coverages are much improved, though under covered, the Q_{RV}HAR-GARCH is also rejected by the LR_{uc} and LR_{cc} tests at both 90% and 95% significance levels.

- The computationally demanding bootstrap version of the AR(1)-GARCH(1,1) gives erratic results, in the sense that for the AXA dataset this approach produces slightly conservative intervals and is not rejected by both the LR_{uc} and LR_{cc}

tests at 90% and 95% significance levels. On the contrary, for France Telecom there is an under coverage at both significance levels as well as rejections from LR_{uc} and LR_{cc} tests.

- The two competing methods in terms of the three criteria are based on modelling intraday quantiles using AR(1)-GARCH(1,1) and the adapted HAR, Q-HAR. The latter method produce very accurate nominal coverage probabilities at both 90% and 95% levels, with an exception of an under coverage at 95% level for France Telecom. Furthermore, this model fails only two of the eight likelihood ratio tests, corresponding to 90% and 95% LR_{cc} test for France Telecom. Similarly, the AR(1)-GARCH(1,1) approach fails only two of the likelihood ratio tests corresponding to 90% LR_{cc} and 95% LR_{uc} for France Telecom. Additionally, the average nominal coverages of the AR(1)-GARCH(1,1) procedure at 90% and 95% levels are 0.8875 and 0.9313.

5.9 Chapter Summary

The main outcomes of the chapter can be summarised as:

- Using the result of Pearson and Tukey (1965) we demonstrate a link between the extreme quantile measures of volatility and realised volatility calculated from intraday returns. We then tie this link with the theory of order statistics to obtain a quantile-based estimator of realised volatility.

- An AR(1)-GARCH(1,1) model is used to model intraday quantiles calculated from high frequency data with the objective of projecting the distribution of intraday returns. Equipped with the density forecast we employ the distribution of the difference of two quantiles to calculate prediction intervals for RV.

- Two adapted extensions of the HAR model are presented which extends the functionality of the original model to building prediction intervals for realised volatility.

- The proposed procedures are applied to real data and we demonstrate quite accurate coverage prediction intervals for RV using the intraday returns quantile forecasting (AR(1)-GARCH(1,1)) method and Q-HAR model.

Chapter 6

Concluding Remarks

The thesis studies some theoretical and applied quantile-based methods relating to prediction, inference and risk measurement. Here, the main contributions of the thesis are summarised and a brief outline of possible future research topics are outlined.

6.1 Main Contributions

In chapter 2 two simple and easy-to-implement distribution transformation methods for constructing prediction intervals of future random variable are proposed. One of these methods, based on the normal transformation, is theoretically shown to provide exact prediction intervals while the other, utilising the exponential distribution is demonstrated to provide zero coverage error prediction intervals. Furthermore, the exponential distribution method is illustrated to admit any general distribution function and can therefore be used in a nonparametric framework.

Through exploiting the integral form representation of expected shortfall in chapter 3 a number of nonparametric kernel-based expected shortfall estimators are proposed and are numerically studied via Monte Carlo simulations. The notion of bias reduction is considered and under some theoretical conditions the bias reduced estimators are shown to outperform non-bias reduced ones, a fact confirmed by

numerical experiments.

In chapter 4 the problem of constructing simultaneous confidence bands for quantile regression functions when the covariates are restricted in region is tackled. Asymmetric Laplace distribution together with simulation algorithms form the building blocks for building the bands which are shown to attain good coverage probabilities. The procedure is shown to have an easy adaptation to form bands for interquantile regression functions. Furthermore, the simulation idea is utilised in creating quantile regression confidence bands for a classical ordinary least square model when the residuals are normally distributed and when this assumption is not tenable. The latter case is addressed using the well known Box-Cox transformation of the response variable.

In chapter 5 a practical approach of creating prediction intervals for realised volatility is considered. Exploiting a quantile-based presentation of volatility in conjunction with AR-GARCH we utilise rich information obtained from high frequency data by modelling and forecasting intraday quantiles of returns from which the predicted intervals of realised volatility are calculated using the theory of order statistics. Furthermore, we presented adaptations of the HAR model of Corsi (2009) and extend their functionality to building intervals forecasts for realised volatility. Using out-of-sample realised volatility as a benchmark we apply the proposed procedures to real data and demonstrate quite accurate coverage probabilities for intervals build from forecasting intradaily returns and the Q-HAR model.

6.2 Recommendations for Future Research

- The exponential transformation method proposed in chapter 2 is applicable for independent and identically distributed random variable. However, through filtering it may be possible to adapt this method to create a semi-parametric approach of constructing prediction intervals for financial time series such as asset prices.

- The accuracy of bias reduced estimators in chapter 3 maybe be thoroughly examined and enhanced when studied through future projection of expected shortfall. This can be tackled from, say, future scenarios generation.

- A natural progression of chapter 4 is the study of simultaneous tolerance bands for quantile regression. While this topic has already been explored in the classical linear regression model (see Limam and Thomas (1988) and references therein) it has yet to be extended to quantile regression. Similarly, the creation of simultaneous bands can also be studied in other quantile regression models, such as quantile regression model for survival data analysis.

- The work presented in chapter 5 raise a number of interesting questions for future consideration. Three of these are:

    1. Investigate the direct modelling of quantiles of integrated volatility as discussed by (Barndorff-Nielsen and Shephard, 2002; Gonçalves and Meddahi, 2008, 2009) with the aim of creating prediction intervals. On a related note, one can also investigate conditions under which the predictions of independently forecasted end points of prediction intervals will produce accurate coverage.

    2. Since there is a close relationship between volatility and value-at-risk forecasts, as an extension one can investigate the validity of the models in building intervals for value-at-risk.

    3. In keeping with the GARCH theme, there is scope to introduce a quantile-based GARCH model similar to the Conditional AutoRegressive Range (CARR) of Chou (2005).

Bibliography

Acerbi, C., C. Nordio, and C. Sirtori (2001). Expected shortfall as a tool for financial risk management. Quantitative finance papers, arXiv.org.

Acerbi, C. and D. Tasche (2001). Expected shortfall: a natural coherent alternative to value at risk. *Economic Notes 31*, 379–388.

Acerbi, C. and D. Tasche (2002). On the coherence of expected shortfall. *In: Szeg, G. (Ed.), Beyond VaR (Special Issue). Journal of Banking & Finance 26*, 1487–1503.

Al-Saidy, O. M., W. W. Piegorsch, R. W. West, and D. K. Nitcheva (2003). Confidence bands for low-dose risk estimation with quantal response data. *Biometrics 59*(4), pp. 1056–1062.

Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review 39*(4), 885–905.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2000). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96*, 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modelling and forecasting realised volatility. *Econometrica 71*, 579–625.

Artzner, P. (1999). Application of coherent risk measures to capital requirements in insurance. *North American Actuarial Journal 3*, 11–25.

Artzner, P., F. Delbaen, J. M. Eber, and D. Heath (1997). Thinking coherently. *Risk 10*, 68–71.

Artzner, P., F. Delbaen, J. M. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance 3*, 203–228.

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika 68*(1), 326–328.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of Royal Statistical Society 64*.

Basle Committee Banking Supervision (1996). *Amendment to the Capital Accord to incorporate Market Risks*.

Basle Committee on Banking Supervision (2003). *The New Basel Capital Accord*.

Blair, B. J., S.-H. Poon, and S. J. Taylor (2001). Forecasting s&p 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics 105*(1), 5–26.

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

Bowley, A. L. (1920). *Elements of Statistics*. New York: Charles Scribners Sons.

Bowman, A., P. Hall, and T. Prvan (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika 85*, 799–808.

Box, G. E. P. and D. R. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological) 26*(2), 211–252.

Buchinsky, M. (1995). Quantile regression, box-cox transformation model, and the u.s. wage structure, 1963-1987. *Journal of Econometrics 65*(1), 109–154.

Buchinsky, M. (1998). Recent advances in quantile regression models: A practical guideline for empirical research. *The Journal of Human Resources 33*(1), 88–126.

Canay, I. A. (2010). El inference for partially identified models: Large deviations optimality and bootstrap validity. *Journal of Econometrics 156*(2), 408 – 425.

Chamberlain, G. (1994). Quantile regression, censoring, and the structure of wages. In *Advancesin Econometrics: Sixth World Congress (C. Sims, ed.)*. Cambridge: Cambridge University Press.

Chatfield, C. (1993). Prediction intervals. *Journal of Business and Economic Statistics 11*, 121–135.

Chatfield, C. E. J. S. A. (2001). *Prediction Intervals, in Principles of ForecastingA Handbook for Researchers and Practitioners,*. Kluwer Academic. Boston.

Chen, S. X. (2008). Nonparametric Estimation of Expected Shortfall. *Journal of financial econometrics 6*(1), 87–107.

Chen, S. X. and C. Y. Tang (2005). Nonparametric Inference of Value-at-Risk for Dependent Financial Returns. *Journal of Financial Econometrics 3*(2), 227–255.

Cheng, M.-Y. and S. Sun (2006). Bandwidth selection for kernel quantile estimation. *Journal of the Chinese Statistical Association 44*(3), 271–295.

Chernozhukov, V., C. Hansen, and M. Jansson (2009). Finite sample inference for quantile regression models. *Journal of Econometrics 152*(2), 93–103.

Chou, R. (2005). Forecasting financial volatilities with extreme values: The conditional auto regressive range (carr) model. *Journal of Money Credit and Banking 37*, 561–582.

Christoffersen, P., B. Feunou, K. Jacobs, and N. Meddahi (2010). The Economic Value of Realized Volatility. *SSRN eLibrary*.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review 39*(4), 841–62.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Corsi, F., U. Kretschmer, S. Mittnik, and C. Pigorsch (2008). The volatility of realized volatility. *Econometric Reviews 27*(1), 46–78.

Crow, E. L. and M. M. Siddiqui (1967). Robust estimation of location. *Journal of the American Statistical Association 62*(318), 353–389.

Danelsson, J., G. Samorodnitsky, M. Sarma, B. N. Jorgensen, and C. G. D. Vries (2005). Subadditivity re-examined: the case for value-at-risk. Technical report, Preprint, London School of Economics. i=1.

David, H. A. (1970). *Order Statistics* (1st ed.). Wiley. New York.

David, H. A. (1981). *Order Statistics* (2nd ed.). Wiley. New York.

Delbaen, F., E. O. T. Hochschule, and Z. Urich (2000). Coherent risk measures on general probability spaces.

Dowd, K. (2002). *Measuring Market Risk*. John Wiley & Sons.

Drost, F. C. and T. E. Nijman (1993). Temporal aggregation of garch processes. *Econometrica 61*(4), 909–27.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*(4), 987–1007.

Engle, R. F. and S. Manganelli (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics 22*, 367–381.

Fermanian, J.-D. and O. Scaillet (2005). Sensitivity analysis of var and expected shortfall for portfolios under netting agreements. *Journal of Banking and Finance 29*(4), 927–958.

Galbraith, J., S. Zernov, and V. Zinde-Walsh (2001, November). Conditional quantiles of volatility in equity index and foreign exchange data. CIRANO Working Papers 2001s-61, CIRANO.

Galbraith, J. W. and V. Zinde-Walsh (1994). A simple noniterative estimator for moving average models. *Biometrika 81*(1), 143–155.

Galbraith, J. W. and V. Zinde-Walsh (1997). On some simple, autoregression-based estimation and identification techniques for arma models. *Biometrika 81*, 685–696.

Galbraith, J. W. and V. Zinde-Walsh (2000). Properties of estimates of daily garch parameters based on intra-day observations. Econometric Society World Congress 2000 Contributed Papers 1800, Econometric Society.

Galton, F. (1889). *Natural Inheritance*. Macmillian: London.

Geisser, S. (1993). *Predictive Inference: An Introduction*. London: Chapman and Hall.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics 131*(1-2), 59 – 95.

Giacomini, R., A. Gottschling, C. Haefke, and H. White (2008). Mixtures of t-distributions for finance and forecasting. *Journal of Econometrics 144*(1), 175 – 192.

Gilchrist, W. G. (2000). *Statistical Modelling with Quantile Functions*. London: Chapman & Hall.

Giot, P. and S. Laurent (2004). Modeling daily value-at-risk using realized volatility and arch type models. *Journal of Empirical Finance 11*, 379–398.

Gonçalves, S. and N. Meddahi (2008). Edgeworth correction for realised volatility. *Econometric Reviews 27*.

Gonçalves, S. and N. Meddahi (2009). Bootstrapping realized volatility. *Econometrica 77*(1), 283–306.

Hald, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. Wiley: New York.

Hall, P., L. Peng, and N. Tajvidi (1999). On prediction intervals based on predictive likelihood or bootstrap methods. *Biometrika 86*(4), 871–880.

Hall, P. and A. Rieck (2001). Improving coverage accuracy of nonparametric prediction intervals. *Journal Of The Royal Statistical Society Series B 63*(4), 717–725.

Hamada, M., V. Johnson, L. M. Moore, and J. Wendelberger (2004). Bayesian prediction intervals and their relationship to tolerance intervals. *Technometrics 46*(4), 452–459.

Hansen, P. R., Z. Huang, and H. Shek (2010). Realized GARCH: A Joint Model of Returns and Realized Measures of Volatility. *SSRN eLibrary*.

Hardle, W. K., Y. Ritov, and S. Song (2010). Partial linear quantile regression and bootstrap confidence bands. SFB 649 Discussion Papers SFB649DP2010-002, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.

Hardle, W. K. and S. Song (2010). Confidence bands in quantile regression. *Econometric Theory 26*(04), 1180–1200.

Hazen, A. (1914). Storage to be provided in impounding reservoirs for municipal water supply. *Transactions of the American Society of Civil Enginee 77*, 1529–1669.

Hinkley, D. V. (1975). On power transformations to symmetry. *Biometrika 62*(1), 101–111.

Hogg, R. V., A. Craig, and J. W. Mckean (2005). *Introduction to Mathematical Statistics* (6th ed.). Prentice Hall. New Jersey.

Holton, A. G. (2004). Defining risk. *Financial Analyst Journal. 60*, 19–25.

Hosking, J. R. M. and J. F. Wallis (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics 29*(3), 339–349.

Hothorn, T., F. Bretz, and P. Westfall (2008). Finite simultaneous inference in general parametric models. *Biometrical Journal 50*(3), 346–363.

Ingber, L. (1993). Simulated annealing: Practice versus theory. *Mathl. Comput. Modelling 18*, 29–57.

Isaacs, D., D. G. Altman, C. E. Tidmarsh, H. B. Valman, and A. D. Webster (1983). Serum immunoglobulin concentrations in preschool children measured by laser nephelometry: reference ranges for IgG, IgA, IgM. *Journal of Clinical Pathology 36*, 1193–1196.

Jondeau, E., S.-H. Poon, and M. Rockinger (2006). *Financial Modeling Under Non-Gaussian Distributions (Springer Finance)* (1 ed.). Springer.

Jones, M. C., J. S. Marron, and S. J. Sheather (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association 91*, 401–407.

Jones, M. C. and D. F. Signorini (1997). A comparison of higher-order bias kernel density estimators. *J. American Statistical Association 92*.

Jorion, P. (2001). *Value at Risk* (2nd Edition ed.). McGraw-Hill, New York.

Kanofsky, P. (1968). erivation of Simultaneous Confidence Intervals for Parametric Functions From a Parametric Confidence Region. *Sankhya 30*, 379–386.

Kirkpatrick, S., C. D. Gelatt, Jr., and M. P. Vecchi (1983). Optimization by simulated annealing. *Science 220*, 671–680.

Koenker, R. (2005). *Quantile Regression*. Cambridge Books. Cambridge University Press.

Koenker, R. and J. Bassett, Gilbert (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

Koenker, R. and J. A. F. Machado (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association 94*(448), pp. 1296–1310.

Koenker, R. and Q. Zhao (1996). Conditional quantile estimation and inference for arch models. *Econometric Theory 12*(05), 793–813.

Komunjer, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics 128*(1), 137–164.

Lawless, J. F. and M. Fredette (2005). Frequentist prediction intervals and predictive distributions. *Biometrika 92*(3), 529–542.

Lee, D. J. (2009). Testing parameter stability in quantile models: An application to the u.s. inflation process. Working papers 2009-26, University of Connecticut, Department of Economics.

Limam, M. M. T. and D. R. Thomas (1988). Simultaneous tolerance intervals for the linear regression model. *Journal of the American Statistical Association 83*(403), pp. 801–804.

Liu, W. (2010). *Simultaneous Inference in Regression*. Chapman and Hall.

Liu, W., M. Jamshidian, and Y. Zhang (2004). Multiple comparison of several linear regression models. *Journal of the American Statistical Association 99*(466), pp. 395–403.

Liu, W., M. Jamshidian, Y. Zhang, and J. Donnelly (2005). Simulation-based simultaneous confidence bands in multiple linear regression with predictor variables constrained in intervals. *Journal of Computational and Graphical Statistics 14*(2), 459–484.

Lunde, A. and P. R. Hansen (2005). A forecast comparison of volatility models: does anything beat a garch(1,1)? *Journal of Applied Econometrics 20*(7), 873–889.

Machado, J. A. F. and J. Mata (2000). Box-cox quantile regression and the distribution of firm sizes. *Journal of Applied Econometrics 15*(3), 253–274.

Manganelli, S. and R. F. Engle (2001). Value at risk models in finance. Working Paper Series 075, European Central Bank.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance 7*(1), 77–91.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. John Wiley and Sons, New York.

McAleer, M. and M. C. Medeiros (2008). Realized volatility: A review. *Econometric Reviews 27*(1-3), 10–45.

McNeil, A. J. (1999). Extreme value theory for risk managers. *Internal Modelling and CAD II published by RISK Books*, 93–113.

McNeil, A. J. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance 7*(3-4), 271–300.

Merton, R. C. (1980). On estimating the expected return on the market : An exploratory investigation. *Journal of Financial Economics 8*(4), 323–361.

Muller, U. A., M. M. Dacorogna, R. D. Dav, R. B. Olsen, O. V. Pictet, and J. E. von Weizscker (1997). Volatilities of different time resolutions – analyzing the dynamics of market components. *Journal of Empirical Finance 4*(2-3), 213 – 239. High Frequency Data in Finance, Part 1.

Muller, U. A., M. M. Dacorogna, R. D. Dave, O. V. Pictet, R. B. Olsen, and J. Ward (1993). Fractals and intrinsic time - a challenge to econometricians. Working Papers 1993-08-16, Olsen and Associates.

Nadaraya, E. A. (1964). Some new estimates for distribution functions. *Theory of Probability and its Applications 15*, 497–500.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business 53*(1), 61–65.

Parrish, R. S. (1990). Comparison of quantile estimators in normal sampling. *Biometrics 46*(1), 247–257.

Parzen, E. (1979). Nonparametric statistical data modeling. *J. American Statistical Association 92*.

Parzen, E. (1993). Change pp plot and continuous sample quantile function. *Communications in Statistics 22*, 3287–3304.

Parzen, E. (2004). Quantile probability and statistical data modeling. *Statistical Science*, 652–662.

Pascual, L., J. Romo, and E. Ruiz (2006). Bootstrap prediction for returns and volatilities in garch models. *Computational Statistics & Data Analysis 50*(9), 2293–2312.

Pearson, E. S. and J. W. Tukey (1965). A reliable data-based bandwidth selection method for kernel density estimation. *Biometrica 52*(3-4), 533–546.

Pearson, K. and A. Lee (1896). Mathematical contributions to the theory of evolution. *Proceedings of the Royal Society of London 60*, 273–283.

Pflug, G. C. and W. Römisch (2007). *Modeling, Measuring and Managing Risk*. World Scientific, Singapore.

Poon, S.-H. and C. W. J. Granger (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature 41*(2), 478–539.

Powell, J. (1999). Estimation of monotonic regression models under quantile restrictions. In *Nonparametric and Semiparametric Methods in Econometrics*, pp. 357–384. New York: Cambridge University Press.

R Documentation (2010). Pearson dataset. `http://wiener.math.csi.cuny.edu/UsingR/Data/father.son.html`.

Rao, R. C. (1973). *Linear Statistical Inference and Its Applications*. John Wiley & Sons Inc.

Reiss, R. D. (1989). *Approximate Distributions of Order Statistics, with Applications to Nonparametric Statistics*. New York: Springer.

Ruppert, D. (2006). *Statistics and finance, an introduction*. Springer.

Sakia, R. M. (1992). The box-cox transformation technique: A review. *Journal of the Royal Statistical Society. Series D (The Statistician) 41*(2), 169–178.

Sarhan, A. E. and B. G. Greenberg (1962). *Contributions to Order Statistics*. Wiley: New York.

Scaillet, O. (2004). Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance 14*(1), 115–129.

Shao, X.-D., Y.-J. Lian, and L.-Q. Yin (2009). Forecasting value-at-risk using high frequency data: The realized range model. *Global Finance Journal 20*, 128 – 136.

Sheather, S. J. and M. C. Jones (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological) 53*(3), 683–690.

Sheather, S. J. and J. S. Marron (1990). Kernel quantile estimators. *J. American Statistical Association 85*, 410–416.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. York: Chapman & Hall.

Steinhorst, R. K. and D. C. Bowden (1971). Discrimination and confidence bands on percentiles. *Journal of the American Statistical Association 66*(336), 851–854.

Sun, J., J. Raz, and J. Faraway J (2000). Simultaneous confidence bands for growth and response curves. *Statistica Sinica 9*(3), pp. 679–698.

Sweeting, T. J. (1999). Coverage probability bias, objective bayes and the likelihood principle. *Biometrika 88*, 657–675.

Tay, A. S., K. F. Wallis, and C. C. Al (2000). Density forecasting: A survey. *Journal of Forecasting 19*, 235–254.

Taylor, J. W. (2005). Generating Volatility Forecasts from Value at Risk Estimates. *Management Science 51*(5), 712–725.

Taylor, J. W. and D. W. Bunn (1999). A Quantile Regression Approach to Generating Prediction Intervals. *Management Science 45*(2), 225–237.

Thomas, D. L. and D. R. Thomas (1986). Confidence bands for percentiles in the linear regression model. *Journal of the American Statistical Association 81*(395), 705–708.

Tippett, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population. *Biometrika 17*(3-4), 364–387.

Tukey, J. W. (1965). Which part of the sample contains the information? *Proceedings of the National Academy of Sciences of the United States of America 53*(1), 127–134.

Turner, D. L. and D. C. Bowden (1977). Simultaneous confidence bands for percentile lines in the general linear model. *Journal of the American Statistical Association 72*, 886–889.

Turner, D. L. and D. C. Bowden (1979). Sharp confidence bands for percentile lines and tolerance bands for the simple linear model. *Journal of the American Statistical Association 74*(368), 885–888.

Černý, V. (1985). Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications 45*(1), 41–51.

Visser, M. P. (2008). Garch parameter estimation using high-frequency data. MPRA Paper 9076, University Library of Munich, Germany.

Wand, M. and M. Jones (1995). *Kernel Smoothing*. Chapman and Hall, London.

Weiss, A. A. (1986). Asymptotic theory for arch models: Estimation and testing. *Econometric Theory 2*(01), 107–131.

Working, H. and H. Hotelling (1929). Applications of the theory of error to the interpretation of trends. *Journal of the American Statistical Association 24*(165), 73–85.

Xiao, Z. and R. Koenker (2009). Conditional quantile estimation for garch models. Boston College Working Papers in Economics 725, Boston College Department of Economics.

Yu, K. (1997). *Smooth Regression Quantile Estimation.* Ph. D. thesis, unpublished, The Open University.

Yu, K., Z. Lu, and J. Stander (2003). Quantile regression: Applications and current research areas. *The Statistician 52*, 331–350.

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

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

Zhang, M.-H. and Q.-S. Cheng (2005). An approach to var for capital markets with gaussian mixture. *Applied Mathematics and Computation 168*(2), 1079 – 1085.