QUANTILE REGRESSION AND EXTREMES

By

Raka Mandal

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Statistics – Doctor of Philosophy

2020

ABSTRACT

QUANTILE REGRESSION AND EXTREMES

By

Raka Mandal

The estimation of conditional quantiles of a response variable is of prime interest in many statistical applications. Quantile regression (QR) is a powerful and popular method used extensively to estimate conditional quantiles of a response Y in the presence of a covariate X. Moreover, QR can quantify the effect of covariates at different quantile levels. While modeling a rare event, quantiles at high or low tails are of particular interest. In such cases, QR has inevitable shortfalls. Since we have fewer data at tails, QR estimates suffer from high variability. The performance deteriorates further when the underlying distribution is heavy-tailed. Estimation of extreme quantiles is therefore challenging, especially when the data comes from a heavy-tailed distribution.

Extreme value theory (EVT) provides mathematical tools to analyze rare events. In practice, EVT can be used to assess the probability of more extreme events than any previously observed occurrences. The scope of application of EVT includes financial risk assessment, extreme climate modeling, network anomaly detection, etc. Statistical modeling using EVT can be carried out in two approaches: Block maxima and Peak over Threshold (POT). In the block maxima approach, Generalized extreme value (GEV) distribution is fitted to a series of maximums derived from the observations. In the POT approach, one fits Generalized Pareto (GP) distribution to observations exceeding a certain threshold.

An important problem while using the POT approach is the choice of the threshold. Models based on EVT use asymptotic arguments to approximate the tail behavior. Hence, the choice of the threshold is crucial in order to fit the GP distribution to the data. POT approach is a popular tool for the estimation of extreme quantiles in heavy-tailed data wherein the excesses over a threshold are modeled as a function of the covariate X. However,

the efficiency of POT is severely compromised when the threshold itself varies as a function of the covariate.

This dissertation proposes an integrated approach for estimating extreme conditional quantiles from a heavy-tailed distribution. We begin with the case where the threshold does not vary as a function of the covariate X. Using the POT approach to model the scaled conditional excess, we propose an estimator for high conditional quantile. We establish large sample properties of our estimator in the context of the GP distribution. Through numerical investigations, we demonstrate the superiority of our method over QR at higher quantiles. However, this method is not adequate when the threshold itself varies as a function of the covariate X. In order to circumvent this issue, we propose an extension of our approach, which uses standard QR to extract information on the threshold and then model the residuals as a GP distribution with covariate dependent parameters. We establish the asymptotic properties of our method in the context of the GP distribution. Extending further, we thoroughly study the asymptotic performance of this approach for a wide class of heavy-tailed distributions using numerical simulations. Using simulation studies, we compare our approach with existing methods used in estimation of high conditional quantiles for heavy tailed distributions. As an application, we have implemented our method for the task of precipitation downscaling with data obtained from the Vancouver International Airport weather station. We have also demonstrated how our method of covariate adaptive threshold selection can be implemented in practice.

In memory of my grandmother Lila Ghosh.

ACKNOWLEDGEMENTS

This Ph.D. journey has been challenging, both in my professional and personal life, particularly for the last two years. I want to take this opportunity to thank everyone who has given me immense support through this journey.

First of all, I would like to thank my doctoral advisors, Dr. Tapabrata Maiti and Dr. Shrijita Bhattacharya, who has introduced me to the world of research. I am thankful to Dr. Maiti for believing in my potential and giving me several opportunities during my time at Michigan State University. His constant guidance has helped me to grow as a researcher and become successful in my career. It would not have been possible for me to write this dissertation to I am today without his counsel and support. I am thankful to Dr. Bhattacharya for introducing me to the magnificent domain of Extreme Value Theory. She has been an excellent mentor, and working with her has been rewarding. Her timely advice on my career plans, constant encouragement during stressful times, and constructive criticisms have aided me in improving and prepare myself for the next phase of my life. Not to mention, she has provided me with vital financial support for this work.

I would like to thank Prof. Andrew Finley and Prof. Hyokyoung (Grace) Hong for being part of my dissertation committee. Their helpful suggestions and recommendations have been crucial for the completion of this dissertation.

I would also like to thank the Department of Statistics and Probability at Michigan State University (MSU) to provide me with this incredible opportunity to pursue my academic endeavor. East Lansing has been my home away from home for the past seven years. I consider myself fortunate to stay here and enjoy the picturesque campus of MSU. Be it Spring, Summer, Fall, or Winter, my every venture into the nooks and corners of this campus has been full of exciting discoveries. During my stressful days, I found solace by the river banks of the Red Cedar River just outside my department. Thank you, WKAR-FM, for fueling my mornings with classical melodies. Writing this thesis would have been mundane

without you. Thank you, Dr. Tawa Sina, for inspiring me and making me believe in myself every time I had given up. I consider you my unofficial advisor and a guarding angel on this mission. I would also like to thank my friends and fellow Ph.D. students in the department Rejaul Karim, Abdhi Sarkar, Metin Eroglu, Yi-Chen Zhang, Siddhartha Nandy, Atreyee Majumder, and Asish Banik. They have helped me big time during my tenure at MSU.

I have been blessed with a wonderful circle of friends who have always been caring and supportive through this challenging phase of life. This acknowledgment would be incomplete without mentioning a few of them. Thank you, Suman Chakraborty, for your constant guidance like the north star to a sailor. Thank you, Sohini Raha, for being my lifeline during the darkest hours of my life. Thank you, Sujatro Chakladar, for your immense patience and endless encouragement. Thank you, Rejaul Karim, for being a friend everyone shall have in their life. Thank you, Shreya Saha and Ankita Bhattacharya, for opening the new avenue of spirituality in my life. Thank you, Sejuti Dasgupta, for giving me the comfort of a home far away from my family. Thank you, Moumanti Podder, for your continual supply of courage and confidence. Last but not least, thank you, Sebanti Sengupta, Snigdha Panigrahi, Arpita Mukherjee, Saswati Saha, Shrijita Bhattacharya, Saptarshi Mukherjee, Portia Bannerjee, Shruti Saripalli, Rohosen Bandopadhyay, for making this journey easier. I apologize if I have missed anyone.

I want to end this acknowledgment with thanking my family members, though a mere thank you is not enough for every little thing they have done for me. I would like to thank my parents, Tripti Mandal and Prasanta Mandal, for their unconditional love, endless support, and countless sacrifices. They have devoted their life to my well being. Thank you, Maa, for nurturing my scientific curiosity since childhood. You have introduced me to the world of mathematics, and that is why I came this far in my academic venture. Thank you, Baba, for always believing in my potential and giving me the liberty to pursue my dreams, be it my career choice or my hobbies. You are an impeccable human being, and I always look up to you. I must thank my little sister, Trisha Mandal, for being always available for me, in

spite of staying on a different continent. Thank you, Chottu, for countless times you have woken me up with your call so that I could reach to my classes in time. I cannot express how hard it has been to stay away from all of you. Last but not least, thank you, Didu, for blessing me with elated childhood. I know you would have been proud and happy if you were still with us.

.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In many practical applications, we are interested in events that are rare but have very significant consequences. For example, a sudden heavy rainfall can abruptly increase the volume of water inside a dam. The water then released can cause flooding in the lowland areas. Moreover, depending on the architecture and other attributes, the dam can break if the water level exceeds a certain threshold. Hence it is vital to take into account the extreme climate pattern while constructing the dam. Suppose that rainfall data from the past 50 years from the location is available. The project aims to build a dam that can sustain for the next 100 years. How can we predict an event that might happen once in 100 years with observations from the past 50 years? Extreme Value Theory provides the theoretical justification of extrapolation beyond a range of observations. One can derive natural estimators of relevant quantities like extreme quantiles, return levels, etc. We can expect these estimators to have desirable analytical properties. In the example given above, we might be interested in predicting 0.99 quantile of the daily rainfall at the construction location.

The scope of application of Extreme Value Theory (EVT) is vast. For example, Tawn (1992) used EVT to estimate probabilities of annual maximum hourly sea-levels. This information is necessary to accurately predict the height of sea defenses in the coastal areas. Jansen & Vries (1991) have successfully applied EVT to analyze financial data and concluded that the market crash of 1987 was not an outlier. Gilli & këllezi (2006) demonstrated the application of extreme value distributions for risk assessment with major stock market indices. Scarf & Laycock (1996) studied the analysis of corrosion extremes with EVT. The main difference between traditional statistical modeling and EVT is that we generally model the mean in the former while EVT models the extreme quantile, maxima, or some higher order statistic. In this article, we are going to focus on methods for extreme tail quantiles.

Several approaches have been proposed to address extreme tail quantile estimation based on EVT. In the nonregression setup, the estimation of tail quantiles using EVT is based on the assumption that observations are independently and identically distributed. For example, Weissman (1978) approached the high quantile estimation problem where only the k largest observations from a sample of size n are available, $k < n$. The choice of the optimal sample fraction k is vital to minimize the quantile estimate's asymptotic mean square error. As pointed out by Li et al. (2010), one can focus on the bias reduction of the parameter estimates that produce a reduced bias quantile estimator in turn. The first class of bias reduction methods uses optimal k at the same rate as the theoretical optimal rate of the tail index estimator (e.g., Beirlant et al. (2002), Gomesa & Martins (2002)) while the second class of methods uses a larger order of k than the theoretical optimal rate (e.g., Feuerverger et al. (1999), Peng & Qi (2004), Caeiro & Gomes (2009)). Although these methods are developed for the nonregression setup, they are a good starting point based on which many extensions for the regression framework can be adapted.

In a regression framework, we want to incorporate information on a covariate X into our analysis. The foundation of regression quantiles has been laid by Koenker & Bassett (1978), which uses liner quantile functions to explore the relationship between the response Y and a covariate or a set of covariates X. The scope of regression quantiles proposed by Koenker & Bassett (1978) is not just limited to tail quantiles. With quantile regression, we can study the impact of X on different quantiles of the response distribution. There are many advantages to choosing quantile regression over mean regression. Firstly, quantile regression gives us a complete picture of the conditional response distribution. Secondly, it is robust to outliers. Thirdly, it is a distribution free approach so we can apply it to non-Gaussian models. Abrevaya (2001) used regression quantiles to study the effects of several factors linked to newborns' lower birthweights. Umantsev & Chernozhukov (2001) used regression quantiles to model Value-at-Risk, a measure of market risk widely used in finance. Buchinsky (1994) examined the effects of education and experience at different quantiles of the wage

distribution using quantile regression.

Although quantile regression serves as a basis, the large sample theory of quantile regression does not apply far enough into tails. For details, see Chernozhukov (2005). Moreover, data sparsity at tails further amplifies the variability of the estimate. Naturally, behavior at tails is contingent on the nature of the underlying distribution. Therefore, one has to build an alternative model for tails. Many authors have proposed quantile regression or a variation of it to predict extreme quantiles. For example, Friederichs & Hense (2007b) proposed censored quantile regression to predict high precipitation. Taylor (2008) used exponentially weighted quantile regression to analyze stock market data. Quantile regression methods tailored for higher quantiles have been proposed for modeling extreme weather phenomenon like tropical cyclones (Elsner & Jagger (2008), Jagger & Elsner (2008)), severe snowstorms (Velthoen et al. (2019a)) and heatwaves (Kyselý (2002)). Chernozhukov (2005) used the extreme value theory framework by Feigin & Resnick (1994) to develop a theoretical understanding of the linear quantile regression models at the tail. Chernozhukov (2005) has shown that the asymptotic distribution of extreme quantiles depends on the *tail index*[1] and the regression design. This *tail index* is a crucial indicator of the *domain of attraction* [2] of the underlying process. Not all quantile estimation methods can capture the true quantile function if the distribution belongs to the *heavy-tailed*[3] domain of attraction.

To the best of our knowledge, there are very few methods that have been tailored for estimating extreme quantiles in heavy tailed models. However, as pointed out by Wang & Li (2013), the existing techniques can be broadly grouped into four classes. The first class of methods uses fully parametric models developed using EVT. Generally, these models fit the Generalized Extreme Value (GEV) or the Generalized Pareto (GP) distribution to the data. Here, the information on the covariate X can be incorporated via a combination of the location, scale, or the shape parameters of the GEV or GP distributions. These methods

[1]See definition 2.3.1
[2]See definition 2.3.1
[3]See definition 2.3.1

usually require a choice of covariate dependent tuning parameter, e.g., threshold value. More details can be found in Davison & Smith (1990), Davison & Ramesh (2000), Hall & Tajvidi (2000), Wang & Tsai (2009), and Chavez-Demoulin & Davison (2005). The second class of models uses local estimation to extend the EVT methods in the regression framework. Here, the high conditional quantiles of Y given $X = x$ is estimated using the observations in a small neighborhood of x. For details, see Gardes & Girard (2010), Daouia et al. (2011), Gardes & Girard (2010), and Velthoen et al. (2019b). The efficiency of this class of method depends on the richness of the data in the neighborhood of x. The third class of models uses quantile regression directly to estimate extreme quantiles. More details can be found in Bremnes (2004), Friederichs & Hense (2007a) and Jagger & Elsner (2008). The fourth and final class of methods uses a combination of quantile regression and EVT while imposing some suitable assumptions of the tail behavior of the underlying distribution. Chernozhukov & Du (2006) and Wang et al. (2012) proposed methods in which the intermediate conditional quantiles are estimated using QR and used to extrapolate to the extreme quantiles. These methods are developed assuming that the target quantile functions are linear in the covariates and the distribution is in the maximum domain of attraction of a heavy tailed distribution.

In this article, we propose another approach for estimating high conditional quantiles using EVT and quantile regression, where we use the GP distribution as the basis of our model. Pickands (1975) first suggested using the GP distribution to model the upper tail of a distribution and Davison & Smith (1990) extended this method for the regression setup. There are several advantages of using GP distribution to model tails. Firstly, by Pickands–Balkema–de Haan theorem, the exceedance over a high threshold of an arbitrary heavy tail distribution can be well approximated by GP. Hence, regardless of the original distribution, GP can approximate the tail behavior if it is heavy tailed. Secondly, the analytic behavior of the parameter estimates of GP is well studied and they have desirable asymptotic properties. This is useful for studying the behavior of different statistics based on the parameter estimates. Nonetheless, approximation by GP is contingent upon the selection of an appropriate

threshold. In a non-regression settings, the threshold is usually determined via model diagnostics or some nonparametric tests which assess the goodness of the fit (Langousis et al. (2016)). These methods can be adapted in a regression framework when the threshold is not dependent on the covariate X. However, the problem becomes challenging when the threshold changes with X. As far as we know, the problem of covariate adaptive threshold selection has not been addressed. In this dissertation we propose a solution to this problem which in turn facilitates the estimation of extreme conditional quantiles.

Our contribution in this article can be summarised as following. Our approach provides consistent estimation of quantiles for the GP family of distributions featuring a non negative tail index. Through numerical investigations, we show that our method can overcome the shortcomings of the standard Quantile Regression (Koenker & Bassett (1978)) far in the tail when the distribution is heavy tailed in nature. Moreover, we propose a generalization of our method which can accommodate a large class of distributions in the heavy tail domain of attraction. Selection of a covariate adaptive threshold is a nontrivial problem in this case. Through several examples, we demonstrate the choice of an optimal threshold for implementing our method. Our numerical investigations also show that our method provides a robust estimation of extreme quantiles for heavy tailed distributions. Wang et al. (2012) has proposed a semi-parametric approach for estimating higher conditional quantiles in the context of heavy tailed distributions. We demonstrate the superiority of our approach over Wang et al. (2012) in the context of a Pareto or a Generalized Pareto distribution. When the underlying distribution is in the heavy-tailed domain of attraction, we show that with appropriate choice of the threshold, our approach is as efficient as Wang et al. (2012). Our approach is computationally efficient than Wang et al. (2012) which requires estimation of a number of intermediate regression quantiles. Moreover,it is also easy to implement in practice and we demonstrate it though an application in the real data. The contents of the rest of the thesis is organized as below.

Chapter 2 proposes a method for estimating conditional quantiles from a regression model

where the scale is modeled as a function of the covariate X while the location is not dependent X. Since the threshold is independent of the covariate, a direct adaptation of the POT approach in Davison & Smith (1990) is shown to outperform the quantile regression at high conditional quantiles. We establish the asymptotic properties of the proposed estimator when the true model is GP. Through numerical investigations, we establish the efficiency of our method over standard Quantile Regression for the case when the observations are distributed according to a GP distribution.

Estimating conditional quantiles with GP when the threshold itself varies as a function of the covariate is challenging. Chapter 3 proposes an extension of our method in Chapter 2, which can be used for a covariate adaptive threshold selection. We show that our proposed extension offers a significant improvement over our proposal in chapter 2 and standard Quantile regression through numerical investigations. Also, we establish the asymptotic properties of our estimator in the context of the GP distribution. The selection of a proper threshold is vital for applying our method in practice where we do not know the exact underline distribution. Here, we give a guideline to choose an appropriate threshold when the underlying distribution is a heavy tailed distribution from the Fréchet domain of attraction. We also show that our proposed method can significantly reduce computational complexity, which is usually associated with the optimization of the quantile loss function. Extending further, we thoroughly study the asymptotic performance of our approach for a wide class of heavy tailed distributions via numerical simulations. We compare the performance of our approach with Wang et al. (2012) in estimating conditional quantiles for a variety of distributions in the heavy tailed domain of attraction. Finally, we implement our method to predict extreme precipitation with the data obtained from the Vancouver International Airport weather station. We end this dissertation with some discussions regarding our proposed method in chapter 4. Here we also propose some future directions that can be pursued.

CHAPTER 2

QUANTILE REGRESSION AND ESTIMATION OF CONDITIONAL QUANTILES

2.1 Introduction

The estimation of high conditional quantiles of a response variable is of prime interest in many statistical applications. Quantile regression is a powerful and popular method for this task. One can estimate the effects of the covariates at different conditional quantiles of the response using quantile regression. For example, Friederichs & Hense (2007b) used meteorological information obtained from global climate model projections to predict localized high precipitation. Lower infant birth weight can lead to severe health conditions. Abrevaya (2001) used quantile regression to quantify the effects of maternal behavior, demographics, etc. on different quantiles of the birth weight distribution. Umantsev & Chernozhukov (2001) analyzed factors associated with high risk in finance. Although the scope of quantile regression applications is vast, Chernozhukov (2005) pointed out that the large sample theory of quantile regression is not applicable sufficiently far in the tails. Hence, we should proceed with caution while using quantile regression at tails.

The literature of modern extreme value theory has provided us with elegant tools to quantify rare events. In the nonregression framework, the problem is familiar and well explored. For details, see Weissman (1978), Li et al. (2010), and Embrechts et al. (1997). In the regression framework, one can model the extreme quantiles by fitting a fully parametric model using the generalized extreme value or the generalized Pareto distribution. See Coles et al. (2001) for details.

In this chapter, we explore different approaches for estimating conditional quantiles. We start with quantile regression method proposed by Koenker & Bassett (1978) in section 2.2. In section 2.3, we introduce fundamental results from the extreme value literature. In section

2.4 we introduce the statistical modeling framework using GP that serves as an basis for our work. In section 2.5, we propose our method for estimating conditional quantile using POT approach. We provide large sample properties of the estimator in the context of the GP distribution. In section 2.6 we demonstrate the efficiency of our method over standard quantile regression. From our numerical investigation, we can see that quantile regression estimates are not adequate for estimating higher conditional quantiles. Our proposed method can be adapted for estimating high conditional quantiles, given the data follows GP distribution.

2.2 Quantile Regression

Consider a random variable Y with distribution function F and let $0 < \tau < 1$. Let $Q_Y(\tau) = \inf\{y : F(y) \geq \tau\}$ be the τ quantile of Y and $E(Y)$ is the expected value of Y. Then

$$E(Y) = \underset{c}{\operatorname{argmin}}\, E(Y - c)^2$$

Or we can say that the mean $E(Y)$ minimizes the square error loss. Consider the function $\rho_\tau : \mathbb{R} \to [0, \infty)$ such that

$$\rho_\tau(y) = y\big(\tau - \mathbb{I}_{[y<0]}\big) = \begin{cases} y\tau & \text{if } y \geq 0 \\ (\tau - 1)y & \text{if } y < 0 \end{cases}$$

Where $\tau \in (0, 1)$ is a fixed number and $\mathbb{I}_{[y<0]}$ is the indicator function. The function ρ_τ in figure 2.1 is often referred to as quantile loss function or tilted absolute loss function. This is a convex function differentiable everywhere except at 0. Like mean, quantile of a random variable can also be derived as a solution to a minimization problem. Following Koenker (2005) one can easily verify that

$$Q_Y(\tau) = \underset{c}{\operatorname{argmin}}\, E\big(\rho_\tau(Y - c)\big) \tag{2.1}$$

$Q_Y(\tau)$ is the minimizer of the quantile error loss ρ_τ. When $\tau = 0.5$, minimizing equation 2.1 is equivalent to minimizing $E|Y - c|$ which is the absolute error loss. In this case, the solution is the median of Y. In the regression paradigm, one explores the relationship

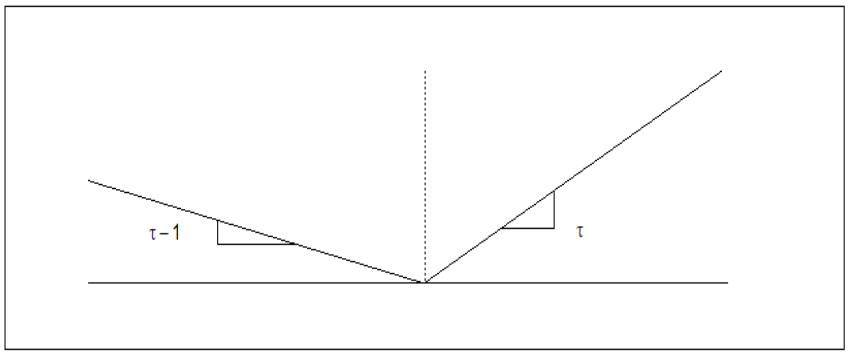

Figure 2.1: Quantile Loss Function

between the response Y and the covariate X. In order to do so, the covariate information must be incorporated into the response in some form. In classical regression, one models the the conditional mean of the response. When the conditional mean is a parametric function of the covariates, i.e, $E(Y|X) = f(X, \beta)$ one can estimate the conditional mean function $E(Y|X)$ by minimizing the square error loss function $E(Y - f(X, \beta))^2$. Suppose we have sample observations (y_i, x_i), $i = 1, \ldots, n$. The estimation of the mean function can be achieved by the method of least squares. One can minimize the sample analog of the squared error loss

$$\sum_{i=1}^{n} \big(y_i - f(x_i, \beta)\big)^2$$

The method of least squares enjoys computational feasibility since the loss function is differentiable. However, it does not give more insight into the conditional distribution of the response. Also, it lacks robustness as it is sensitive to outliers. Koenker & Bassett (1978) introduced quantile regression as a robust alternative to mean regression. Quantile regression can be viewed as an extension of classical mean model regression where the conditional quantile is modeled as a function of the covariates. When the conditional quantile is modeled as a parametric function of the covariates, i.e., $Q_{Y|X}(\tau) = g(X, \beta)$ one can estimate it

by minimizing $E\rho_\tau(Y - g(X, \beta))$. With sample observations (y_i, x_i), $i = 1, \ldots, n$ one can minimize the sample quantile error loss at τ quantile

$$\sum_{i=1}^{n} \rho_\tau\big(y_i - g(x_i, \beta)\big) \tag{2.2}$$

By choosing different values of τ, one can obtain more insight into the conditional distribution beyond the mean. Quantile regression methods are robust and do not rely on the Gaussian error assumption. Since the quantile loss function is not differentiable, optimization of the objective function in 2.2 is not straightforward. Koenker & Bassett (1978) has shown that for linear quantiles, i.e., when

$$Q_{Y|X}(\tau) = g(X, \beta) = \alpha(\tau) + X^T\beta(\tau)$$

one can estimate the τ quantile at $X = x$ as

$$\begin{aligned} \hat{Q}_{Y|x}(\tau) &= \hat{\alpha}(\tau) + x^T\hat{\beta}(\tau) \\ \hat{\alpha}(\tau), \hat{\beta}(\tau) &= \underset{\alpha,\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(y_i - \alpha - x_i^T\beta) \end{aligned} \tag{2.3}$$

The minimization of equation 2.3 can be reformulated as a linear programming problem and solved accordingly. Moreover, the optimization has to be done separately for every choice of τ. Nonetheless, this classical quantile regression model can be adapted for a broad class of regression problems. In Chapter 1, we have discussed how the presence of extreme observations in the data can increase the challenge of consistent estimation of higher quantiles by several manifolds. Extreme Value Theory (EVT) provides us some powerful tools to deal with the analysis of rare events. To understand the challenges, we start with a few basics of EVT.

2.3 Extreme Value Theory: Definitions and Results

Extreme value theory deals with rare events. Rare events are events that happen with low probability near the upper or lower endpoints of the distribution function F. Intuitively, the asymptotic behavior of the maximum (or minimum) can give us more insight into the

upper (or lower) tail of the distribution. Let $Y_1, Y_2, \ldots$ be a sequence of independent random variables with common distribution function F and let $M_n = \max(Y_1, \ldots, Y_n)$. Let $y^+ = \sup\{y : F(y) < 1\}$, the upper endpoint of the distribution function F. Then $P(M_n < y) = \{F(y)\}^n$ for any y in the domain of F and M_n converges to y^+ in probability as $n \to \infty$. Now,

$$\lim_{n\to\infty} \{F(y)\}^n = \begin{cases} 0 & \text{if} \quad y < y^+ \\ 1 & \text{if} \quad y \geq y^+ \end{cases}$$

Hence, in order to obtain a nondegenerate limit distribution, proper centering and scaling of M_n are necessary. Suppose that there exists sequence of constants $a_n > 0$ and b_n such that

$$\lim_{n\to\infty} P((M_n - b_n)/a_n < y) = \lim_{n\to\infty} F^n(a_n y + b_n) = G(y)$$

for some distribution function G and for all y in the domain of F. The natural question that follows is what the possibilities of the limit are? We know from Central limit theorems that with proper centering and scaling, the sum $\sum_{i=1}^n Y_i$ of iid random variables converge to normal distribution. The answer is given by the following theorem, which is also a significant result in the extreme value literature.

Theorem 2.3.1. *(Fisher-Tippet-Gnedenko) Let $Y_1, Y_2, \ldots$ be independent random variables with common distribution function F and let $M_n = max(Y_1, \ldots, Y_n)$. If there exists a sequence of constants $a_n > 0$, b_n such that $(M_n - b_n)/a_n$ converges to a non degenerate distribution G, i.e,*

$$\lim_{n\to\infty} G_n(y) = \lim_{n\to\infty} P((M_n - b_n)/a_n < y) = \lim_{n\to\infty} F^n(a_n y + b_n) = G(y)$$

then, the only three possible forms of the limit are:

$$\begin{aligned} Gumble : G_0(y) &= exp(-e^{-y}), \quad y \in \mathbb{R} \\ Frechet : G_{1,\alpha}(y) &= exp(-y^{-\alpha}), \quad y \geq 0, \alpha > 0 \\ Weibull : G_{2,\alpha}(y) &= exp(-(-y)^{-\alpha}), \quad y \leq 0, \alpha < 0 \end{aligned} \tag{2.4}$$

Theorem 2.3.1 says that if the limiting distribution of $F^n(a_n y + b_n)$ exists, then there are only three possible forms of the limit as given above. The limits of 2.4 can be collectively thought as a member of a single family of distributions known as Generalized Extreme Value (GEV) distributions. We say that Y follows standard GEV distribution with parameter ξ, $Y \sim \text{GEV}(\xi)$ if

$$P(Y < y) = G_\xi(y) = \exp[-(1+\xi y)^{-1/\xi}], \quad 1 + \xi y > 0, \xi \neq 0 \tag{2.5}$$

Note that the support of Y is dependent on ξ. We have $-1/\xi < y < \infty$ for $\xi > 0$ while $-\infty < y < -1/\xi$ for $\xi < 0$. This naturally leads to two very different sub-classes of distributions within the GEV family. The parameter ξ is a crucial indicator of the sub-class and it is known as the extreme value index or tail index. For $\xi = 0$, the cdf in 2.5 is not defined. Rather, we use the limit $\xi \to 0$ which leads us to the Gumble distribution as mentioned in equation 2.4.

$$\lim_{\xi \to 0} G_\xi(y) = G_0(y) = \exp(-e^{-y}), \quad y \in \mathbb{R}$$

Again, we can see that the support has changed to entire $\mathbb{R}$ from the $\xi \neq 0$ cases. This concludes the third sub-class in GEV family: Gumble. In fact, the other two limits: Fréchet and Weibull belong to the two different sub-classes mentioned above. With a little reparametrization $\alpha = 1/\xi$ and $Y^* = 1 + \xi Y$ where $Y \sim \text{GEV}(\xi)$ we can see that $Y^* \sim G_{1,\alpha}$ for $\alpha > 0$ and $Y^* \sim G_{2,\alpha}$ for $\alpha < 0$.

We have seen that the behaviour of the GEV distribution is governed by the value of the tail index ξ. Any change of location and scale does not affect ξ. Hence, the entire GEV family can be summarised by introducing two additional parameters: $\tilde{\mu}$ as location and $\tilde{\sigma} > 0$ as scale. Let $\tilde{Y} = (Y - \tilde{\mu})/\tilde{\sigma}$. Then $\tilde{Y} \sim \text{GEV}(\tilde{\mu}, \tilde{\sigma}, \xi)$ with

$$P(\tilde{Y} < y) = \begin{cases} \exp\Big[-\Big(1 + \xi(\frac{y-\tilde{\mu}}{\tilde{\sigma}})\Big)^{-1/\xi} \Big] & \text{if } \xi \neq 0 \\ \exp\Big[-\exp\Big(\frac{y-\tilde{\mu}}{\tilde{\sigma}}\Big) \Big] & \text{if } \xi = 0 \end{cases} \tag{2.6}$$

Where $1 + \xi(\frac{y-\tilde{\mu}}{\tilde{\sigma}}) > 0$. When Y is standard GEV with $\mu = 0$ and $\sigma = 1$, we simply say $Y \sim \text{GEV}(\xi)$. The following definitions are useful to keep in mind for future use.

Definition 2.3.1. *(Domain of attraction) Suppose that* $Y_1, Y_2, \ldots \overset{iid}{\sim} F$ *and* $M_n = max(Y_1, \ldots, Y_n)$. *If*

$$\lim_{n\to\infty} P((M_n - b_n)/a_n) < y) = \lim_{n\to\infty} F^n(a_n y + b_n) = G(y)$$

for some constants $a_n > 0$ *and* b_n *and* G *is a non-degenerate distribution, we say* F *belongs to the domain of attraction of* G.

From theorem 2.3.1, we know that there are three possible domains of attraction of F, namely Fréchet ($\xi > 0$), Weibull ($\xi < 0$) or Gumble $\xi = 0$. Moreover, F can belong to only one of the three sub-classes. In practice, we can use the value of the tail index ξ to identify the domain of attraction. Hence we adopt the notation $F \in \mathcal{D}(G_\xi)$ to indicate that F belongs to the domain of attraction of G with tail index ξ.

Definition 2.3.2. *(Heavy-tailed distribution)* F *is a heavy-tailed distribution if* $F \in \mathcal{D}(G_\xi)$ *for some tail-index* $\xi > 0$.

When F is known, one can easily find the quantile by inverting the cdf. Here we are interested in a higher quantile τ close to 1 when F is unknown. We will now give an outline of approximating a higher quantile of Y using EVT. Suppose $u \in \mathbb{R}$. One can interpret $Y > u$ as an extreme event for high values of u. Consider the following quantity:

$$1 - F_u(y) = P(Y > y + u | Y > u) = \frac{P(Y > y + u)}{P(Y > u)} = \frac{1 - F(y + u)}{1 - F(u)}$$

The above quantity can be interpreted as the conditional exceedance probability over a value u. We say $F_u(y) = P(Y \leq y + u | Y > u)$ is the *excess* cdf of Y over a threshold u. If $Y_u = Y | Y > u$, then $P(Y_u < y + u) = F_u(y)$. The following theorem gives us the limiting distribution of Y_u as $u \to \infty$.

Theorem 2.3.2. *(Pickands–Balkema–de Haan) Let* $Y_1, Y_2, \ldots$ *be independent random variables with common distribution function* F. *Define the conditional distribution of excess as* $Y - u | Y > u \sim F_u$.

$$F_u(y) = P(Y - u \leq y | Y > u) = \frac{F(u + y) - F(u)}{1 - F(u)}$$

If $F \in \mathcal{D}(G_\xi)$ for some ξ satisfying theorem 2.3.1 then for large enough u, F_u can be approximated as

$$H(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-\frac{y}{\sigma}} & \text{if } \xi = 0 \end{cases} \tag{2.7}$$

for some $\sigma > 0$.

Theorem 2.3.2 is also known as the second theorem in extreme value literature, while theorem 2.3.1 by Fisher is known as the first theorem. The distribution H in theorem 2.3.2 is known as generalized Pareto (GP) distribution with scale σ and tail index ξ. The standard GP distribution has scale $\sigma = 1$. We say $Y \sim \text{GP}(\xi)$ if

$$P(Y > y) = \begin{cases} (1 + \xi y)^{-1/\xi} & \text{if } \xi \neq 0 \\ e^{-y} & \text{if } \xi = 0 \end{cases} \tag{2.8}$$

where $y \geq 0$ if $\xi > 0$ and $0 \leq y \leq -1/\xi$ if $\xi < 0$. Just like in GEV, the range of Y depends on ξ and leads to three different sub-classes within the GP family. Consider the following reparametrization: $Y^* = 1 + \xi Y$ and $\alpha = 1/\xi$. Then, for $\alpha > 0$, $P(Y^* > y) = y^{-\alpha}, y \geq 1$ which is the standard Pareto distribution. When $\xi < 0$, we have a bounded distribution on $0 \leq y \leq -1/\xi$ with $P(Y^* > y) = y^{-1/\xi}$. Finally, for $\xi = 0$, $P(Y > y) = e^{-y}$ is standard exponential distribution with rate 1. The entire GP family can be represented by introducing two additional parameters: σ for scale and μ for location. Let $\tilde{Y} = (Y - \mu)/\sigma$ then we say $\tilde{Y} \sim GP(\mu, \sigma, \xi)$ and

$$P(\tilde{Y} > y) = \begin{cases} \left[1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-1/\xi} & \text{if } \xi \neq 0 \\ \exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right] & \text{if } \xi = 0 \end{cases}$$

Suppose $F \in \mathcal{D}(G_\xi)$ satisfies theorem 2.3.1 with $\lim_{n\to\infty} F^n(a_n y + b_n)$ following $\text{GEV}(\tilde{\mu}, \tilde{\sigma}, \xi)$. According to theorem 2.3.2, for large values of u the distribution of the excess Y_u is well approximated by $\text{GP}(\mu = 0, \sigma, \xi)$ with $\sigma = \tilde{\sigma} + \xi(u - \tilde{\mu})$. The important thing to note here is that the shape parameter ξ in GP is in fact the tail index in GEV. The scale parameter σ is a function of GEV parameters and u.

2.4 Statistical Modeling of Extreme Values with GP

Let Y denotes the univariate random response variable and $X = (X_1, \ldots, X_p)$ denotes the p dimensional random vector of covariates. $(y_i, x_i); i = 1, \ldots, n$ are random samples from the joint distribution of Y and X. We also assume that $Y \sim F_{Y|x}$. In the previous section we have seen that if $F \in \mathcal{D}(G_\xi)$ then for a large enough value u, the distribution of the *excess* $e_u = Y - u|Y > u$ can be approximated as

$$\bar{F}_u(y) = P(Y - u \leq y|Y > u) \approx \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}, \quad y \geq 0, (1 + \xi y/\sigma) > 0$$

for some $\sigma > 0$.In a parametric regression setup, one has to incorporate the covariates' information in the parameters of the conditional distribution $F_{Y|x}$. This can be done in several ways. Let us begin with the following model:

$$Y = (\theta + X^T\gamma)\epsilon \tag{2.9}$$

where $\theta \in \mathbb{R}$ and $\gamma = (\gamma_1, \ldots, \gamma_p) \in \mathbb{R}^p$ are unknown parameters. $X^T\gamma = \gamma_1 X_1 + \cdots + \gamma_p X_p$ with $\theta + X^T\gamma > 0$. We assume that $\epsilon \sim \text{GP}(\xi)$ with some $\xi > 0$. Clearly, $Y|x$ follows GP distribution with $\mu = 0$, tail index ξ and scale $\sigma = \sigma(x) = \theta + x^T\gamma$.

Our goal is to estimate quantiles from model 2.9. Let $0 < \tau < 1$ and $Q_{Y|x}(\tau) = F^{-1}_{Y|x}(\tau)$. Then from 2.8, we have

$$Q_{Y|x}(\tau) = (\theta + x^T\gamma)Q_\epsilon(\tau) = \frac{(\theta + x^T\gamma)}{\xi}\Big[(1 - \tau)^{-\xi} - 1\Big] \tag{2.10}$$

Here $Q_\epsilon(\tau) = \frac{1}{\xi}[(1 - \tau)^{-\xi} - 1]$ is the τ quantile of standard GP distribution. We need to estimate θ, γ and ξ from the data. We can obtain these parameters from maximizing the log

likelihood

$$
\begin{aligned}
l(\theta,\gamma,\xi) &= -\sum_{i=1}^{n} dF(y_i|x_i) \\
&= -\sum_{i=1}^{n} \log \frac{1}{\sigma(x_i)}\Big(1+\frac{\xi y_i}{\sigma(x_i)}\Big)^{-\frac{1}{\xi}-1} \\
&= \sum_{i=1}^{n} \log \sigma(x_i) + (1+\frac{1}{\xi})\sum_{i=1}^{n} \log\Big(1+\frac{\xi y_i}{\sigma(x_i)}\Big) \\
&= \sum_{i=1}^{n} \log(\theta+x_i^T\gamma) + (1+\frac{1}{\xi})\sum_{i=1}^{n} \log\Big(1+\frac{\xi y_i}{\theta+x_i^T\gamma}\Big)
\end{aligned}
\tag{2.11}
$$

Note that 2.11 is valid if $\sigma(x) = \theta + x_i^T\gamma > 0$. Also, there is no analytical solution for 2.11. Since the range of GP depends on its parameters, numerical optimization of 2.11 needs separate treatment for $\xi < 0$. From now onward, we will only focus on the methods of parameter estimation for $\xi \geq 0$. Given $\xi \geq 0$, the support of GP is unbounded above. We numerically optimize 2.11 with the constraints $\xi > 0$ and $\theta + \gamma x_i > 0$ for $i = 1, \ldots, n$. Let

$$
\hat{\theta},\hat{\gamma},\hat{\xi} = \underset{\theta,\gamma,\xi}{\operatorname{argmin}}\, l(\theta,\gamma,\xi) = \underset{\theta,\gamma,\xi}{\operatorname{argmin}} \Big\{ \sum_{i=1}^{n} \log(\theta+x_i^T\gamma) + (1+\frac{1}{\xi})\sum_{i=1}^{n} \log\Big(1+\frac{\xi y_i}{(\theta+x_i^T\gamma)}\Big)\Big\}
$$

We propose the following estimator for model 2.9 by plugging in the estimators in 2.10 we propose to estimate the conditional quantile as

$$
\hat{Q}_{Y|x}(\tau) = \frac{(\hat{\theta}+x^T\hat{\gamma})}{\hat{\xi}}\Big[(1-\tau)^{-\hat{\xi}}-1\Big] \tag{2.12}
$$

Note that by theorem 2.3.2, we can reasonably approximate quantiles from model 2.9 by 2.12 as long as $Y|x \sim F \in \mathcal{D}(G_\xi)$ for some $\xi > 0$. What happens when we extend model 2.9 as

$$
Y = \phi + (\theta+\gamma X)\epsilon \tag{2.13}
$$

where $\phi \in \mathbb{R}$ and $\theta + \gamma X > 0$ and $\epsilon \sim \mathrm{GP}(\xi)$ with $Y|x \sim \mathrm{GP}(\phi, \theta + x^T\gamma, \xi)$. Now we have to estimate an additional parameter ϕ. We begin with our proposal in the following section.

2.5 Proposed Method: GP for Scale Models

Let us assume the following extension of the regression model:

$$Y = \phi + (\theta + X^T\gamma)\epsilon \tag{2.14}$$

Where $\phi \in \mathbb{R}$ and $\gamma = (\gamma_1, \ldots, \gamma_p) \in \mathbb{R}^p$ are unknown parameters. $X^T\gamma = \gamma_1 X_1 + \cdots + \gamma_p X_p$. We assume that $\epsilon \sim \text{GP}(\xi)$ for some $\xi > 0$. Clearly, here $Y|x$ follows GP distribution with tail index ξ, scale $\sigma(x) = \theta + x^T\gamma$ and location $\mu = \phi$. Hence

$$P(Y > y|x) = \left[1 + \frac{\xi(y-\phi)}{\sigma(x)}\right]^{-1/\xi}$$

Note that in model 2.14, the true τ quantile is given by

$$Q_{Y|x}(\tau) = \phi + (\theta + x^T\gamma)Q_\epsilon(\tau) = \phi + \frac{(\theta + x^T\gamma)}{\xi}\Big[(1-\tau)^{-\xi} - 1\Big] \tag{2.15}$$

We need to estimate the parameters ϕ, γ and ξ. If the true distribution of ϵ is GP, we can numerically optimize 2.16 under the constraints $\xi > 0$, $\theta + x_i^T\gamma > 0$ and $y_i - \phi > 0$ for all $i = 1, \ldots, n$.

$$\begin{aligned}\hat{\phi}, \hat{\theta}, \hat{\gamma}, \hat{\xi} &= \underset{\phi,\theta,\gamma,\xi}{\text{argmin}}\, l(\phi, \theta, \gamma, \xi) \\ &= \underset{\phi,\theta,\gamma,\xi}{\text{argmin}} \left[n \log(\theta + x_i^T\gamma) + (1 + \frac{1}{\xi}) \sum_{i=1}^{n} \log\left(1 + \frac{\xi(y_i - \phi)}{\theta + x_i^T\gamma}\right)\right]\end{aligned} \tag{2.16}$$

Then we can plug in estimates from 2.16 in 2.15 to obtain $\hat{Q}_{Y|x}(\tau)$. Note that the last condition $y_i - \phi > 0$ adds n many more constraints in 2.16. In model 2.9 we had only $n + 1$ constraints. Moreover, in many practical applications, the true distribution of ϵ might not be GP. It can be some distribution in $\mathcal{D}(G_\xi)$ with $\xi > 0$. Our aim is to find a reasonable estimator for quantiles in regression models in a more general setup. We propose our method in the following section.

2.5.1 Method of Estimation

Suppose that we have a random sample $(y_i, x_i), i = 1, \ldots, n$ from the distribution of (Y, X) where $y_i \in \mathbb{R}$ and $x_i = (x_{i1}, \ldots, x_{ip}) \in \mathbb{R}^p$. $Y|x \sim F$ where $F \sim \text{GP}(\xi)$ for some $\xi > 0$. Let

τ_c be a fixed value such that $0 < \tau_c < 1$. Let $y_{(1)} = min(y_1, \ldots, y_n)$. We fit a GP model to $\tilde{y_i} = y_i - y_{(1)}; i = 1, \ldots, n$. Clearly, $\tilde{y_i} \geq 0$. We can estimate θ, γ and ξ from the restricted likelihood as:

$$\hat{\theta}, \hat{\gamma}, \hat{\xi} = \underset{\theta,\gamma,\xi}{\operatorname{argmin}}\, l(\theta, \gamma, \xi) = \underset{\theta,\gamma,\xi}{\operatorname{argmin}} \left[\sum_{i:\tilde{y_i}>0} \log(\theta + x_i^T\gamma) + (1 + \frac{1}{\xi}) \sum_{i:\tilde{y_i}>0} \log\left(1 + \frac{\xi\tilde{y_i}}{\theta + x_i^T\gamma}\right)\right] \tag{2.17}$$

with only $n + 1$ constraints: $\theta + x_i^T\gamma > 0$ and $\xi > 0$. Here we do not estimate ϕ directly form the GP likelihood. Our advantage is a huge gain in efficiency as we avoid n additional constraints as described in 2.16. Finally, we propose the τ Quantile estimate for 2.15 as:

$$\begin{aligned} \hat{Q}_{Y|x}(\tau) &= y_{(1)} + \frac{(\hat{\theta} + x^T\hat{\gamma})}{\hat{\xi}}\left[(1-\tau)^{-\hat{\xi}} - 1\right] \\ &= y_{(1)} + \frac{(\hat{\theta} + x^T\hat{\gamma})}{\hat{\xi}}\left[(1-\tau)^{-\hat{\xi}} - 1\right] \end{aligned} \tag{2.18}$$

Algorithm 1 can be used to find quantile estimates with our proposed method.

Algorithm 1: GP for scale family

1. Get $\tilde{y_i} = y_i - min(y_1, \ldots, y_n)$
2. Get $\hat{\theta}, \hat{\gamma}, \hat{\xi}$ from 2.17
3. Given $X = x$, estimate the τ quantile as

$$\hat{Q}_{Y|x}(\tau) = y_{(1)} + \frac{(\hat{\theta} + x^T\hat{\gamma})}{\hat{\xi}}\left[(1-\tau)^{-\hat{\xi}} - 1\right]$$

In the following proposition, we establish the consistency of algorithm 1 for GP distribution under the following compactness assumption on the covariate x.

Assumption (A1): $|x_i| \leq M$ for some $M > 0$.

Proposition 2.5.1. *Let*

$$y_i = \phi + (\theta + x_i^\top\gamma)\epsilon_i, \quad \epsilon_i \overset{iid}{\sim} GP(\xi).$$

Let $\hat{\phi} = Y_{(1)}$. *With* $\widetilde{x}_i = [1, x_i^\top]$ *and* $\widetilde{\gamma} = [\theta, \gamma]$, *let*

$$\hat{\gamma}, \hat{\xi} = \underset{c,d}{\operatorname{argmin}} \left[\sum_{i=1}^{n} \log(\widetilde{x}_i^\top c) + \left(1 + \frac{1}{d}\right) \sum_{i=1}^{n} \log\left(1 + \frac{d(y_i - \hat{\phi})}{\widetilde{x}_i^\top c}\right) \right] \tag{2.19}$$

Estimate the τ *quantile of* Y *as*

$$\hat{Q}_{Y|x}(\tau) = \hat{\phi} + \frac{\widetilde{x}^\top \hat{\gamma}}{\hat{\xi}} \left((1-\tau)^{-\hat{\xi}} - 1 \right)$$

Then,

1. $$\sqrt{n}((\hat{\gamma}, \hat{\xi}) - (\widetilde{\gamma}, \xi)) \implies N(0, \Sigma)$$

2. $$\sqrt{n}(\hat{Q}_{Y|x}(\tau) - Q_{Y|x}(\tau)) \implies N(0, \sigma^2)$$

Proof. We first show that $Y_{(1)}$ is a consistent estimate of ϕ.

$$\begin{aligned} P(\sqrt{n}|Y_{(1)} - \phi| > \delta) &= P(Y_{(1)} > \phi + \delta/\sqrt{n}) = \prod_{i=1}^{n} P((\widetilde{x}_i^\top \widetilde{\gamma})\epsilon_i > \delta/\sqrt{n}) \\ &= \prod_{i=1}^{n} \left(1 + \frac{\xi\delta}{n(\widetilde{x}_i^\top \widetilde{\gamma})}\right)^{-1/\xi} \leq \prod_{i=1}^{n} \left(1 + \frac{\xi\delta}{\sqrt{n} M_{\widetilde{\gamma}}}\right)^{-1/\xi} \\ &\approx e^{-\sqrt{n}\delta/M_{\widetilde{\gamma}}} \to 0, n \to \infty \end{aligned}$$

where $M_{\widetilde{\gamma}} = M \max|\widetilde{\gamma}_j|$ by assumption (A1). Thus, $\hat{\phi} - \phi = o_P(1/\sqrt{n})$.

First, suppose ϕ is fixed, then

$$\hat{\gamma}_0, \hat{\xi}_0 = \underset{\widetilde{\gamma},\xi}{\operatorname{argmin}} \left[\sum_{i=1}^{n} \log \widetilde{x}_i^T \widetilde{\gamma} + (1 + \frac{1}{\xi}) \sum_{i=1}^{n} \log\left(1 + \frac{\xi(y_i - \phi)}{\theta + \widetilde{x}_i^T \widetilde{\gamma}}\right) \right] \tag{2.20}$$

Then, the MLE is consistent and satisfies

$$\sqrt{n}((\hat{\gamma}_0, \hat{\xi}_0) - (\widetilde{\gamma}, \xi)) \implies N(0, I(\gamma, \xi)^{-1})$$

where

$$I(\gamma, \xi) = \begin{bmatrix} \frac{1}{\xi} \sum_{i=1}^{n} \frac{\widetilde{x}_i \widetilde{x}_i^\top}{(\widetilde{x}_i^\top \gamma)^2} & \frac{-1}{(1-\xi)(1-2\xi)} \sum_{i=1}^{n} \frac{\widetilde{x}_i}{\widetilde{x}_i^\top \gamma} \\ \frac{-1}{(1-\xi)(1-2\xi)} \sum_{i=1}^{n} \frac{\widetilde{x}_i^\top}{\widetilde{x}_i^\top \gamma} & \frac{2}{(1-\xi)(1-2\xi)} \end{bmatrix} \tag{2.21}$$

Thus, for ϕ fixed $(\hat{\gamma}_0, \hat{\xi}_0)$ is the solution of the equation

$$\nabla_{\widetilde{\gamma},\xi} \sum_{i=1}^{n} \log \widetilde{x}_i^T \gamma + (1 + \frac{1}{\xi}) \log \left(1 + \frac{\xi(y_i - \phi)}{\theta + \widetilde{x}_i^T \gamma} \right) = 0$$

which implies

$$(\hat{\gamma}_0, \hat{\xi}_0) = (h_1(\phi), h_2(\phi))$$

Now,

$$(h_1(\hat{\phi}), h_2(\hat{\phi})) \approx (h_1(\phi) + (\hat{\phi} - \phi)h_1'(\phi), h_2(\phi) + (\hat{\phi} - \phi)h_2'(\phi))$$

Therefore,

$$\sqrt{n}((h_1(\hat{\phi}), h_2(\hat{\phi})) - (\widetilde{\gamma}, \xi)) \approx \sqrt{n}((h_1(\hat{\phi}), h_2(\hat{\phi})) - (\widetilde{\gamma}, \xi)) + \sqrt{n}(\hat{\phi} - \phi)(h_1'(\phi), h_2'(\phi))$$

As shown previously, $\sqrt{n}(\hat{\phi} - \phi) = o_P(1)$, thus

$$\sqrt{n}((\hat{\gamma}, \hat{\xi}) - (\gamma, \xi)) \implies N(0, \Sigma)$$

where $\Sigma = (h_1'(\phi), h_2'(\phi)) I(\gamma, \xi)^{-1} (h_1'(\phi), h_2'(\phi))^\top$ with $I(\gamma, \xi)$ same as in (2.21).

For part (2), note that

$$\begin{aligned}
&\sqrt{n}(\hat{Q}_{Y|x}(\tau) - Q_{Y|x}(\tau)) \\
&= \sqrt{n}(\hat{\phi} - \phi) + \sqrt{n} \left(\frac{\widetilde{x}^\top \hat{\gamma}}{\hat{\xi}} \left((1-\tau)^{-\hat{\xi}} - 1 \right) - \frac{\widetilde{x}^\top \widetilde{\gamma}}{\hat{\xi}} \left((1-\tau)^{-\hat{\xi}} - 1 \right) \right) \\
&= \sqrt{n}(\hat{\phi} - \phi) + \sqrt{n}(g(\hat{\gamma}, \hat{\xi}) - g(\widetilde{\gamma}, \xi)) = o_P(1) + \sqrt{n}(g(\hat{\gamma}, \hat{\xi}) - g(\widetilde{\gamma}, \xi))
\end{aligned}$$

Now,

$$\sqrt{n}(g(\hat{\gamma}, \hat{\xi}) - g(\widetilde{\gamma}, \xi)) \approx \sqrt{n}((\hat{\gamma}, \hat{\xi}) - (\widetilde{\gamma}, \xi))^\top \nabla_{\widetilde{\gamma},\xi} g(\widetilde{\gamma}, \xi)$$

Since $\sqrt{n}((\hat{\gamma}, \hat{\xi}) - (\widetilde{\gamma}, \xi)) \implies N(0, \Sigma)$, thus

$$\sqrt{n}(g(\hat{\gamma}, \hat{\xi}) - g(\widetilde{\gamma}, \xi)) \implies N(0, \sigma^2)$$

where $\sigma^2 = \nabla_{\widetilde{\gamma},\xi} g(\widetilde{\gamma}, \xi)^\top \Sigma \nabla_{\widetilde{\gamma},\xi} g(\widetilde{\gamma}, \xi)$. This completes the proof. □

2.6 Simulation Study

We want to assess our proposed GP scale model's performance in estimating quantiles from regression models of the form $Y = \phi + (\theta + X^T\gamma)\epsilon$. In this case, the quantile is linear in covariates. Throughout this section and later, we will concentrate on estimating quantile functions of the form:

$$Q_{Y|x}(\tau) = \alpha(\tau) + x^T\beta(\tau)$$

where $\alpha(\tau)$ is the slope and $\beta(\tau) = \{\beta_1(\tau), \ldots, \beta_p(\tau)\}$ is the intercept. For our regression model

$$\begin{aligned} Y &= \phi + (\theta + X^T\gamma)\epsilon, \quad \epsilon \sim F \\ Q_{Y|x}(\tau) &= \phi + (\theta + x^T\gamma)Q_\epsilon(\tau) \\ \alpha(\tau) &= \phi + Q_\epsilon(\tau), \quad \beta(\tau) = \gamma Q_\epsilon(\tau) \end{aligned}$$

Since the quantiles are linear, quantile regression can be used without violating any model assumptions. Note that both the slope and the intercept of the target quantile are affected by the change in the quantile level τ. This effect is very significant at higher values of τ, particularly if $\epsilon \in \mathcal{D}(G_\xi)$ for $\xi > 0$. Hence we want to see how different values of τ and ξ affect the quantile estimates. We start with the simple case: $\phi = 0$ (case-I). In this case, we can directly fit GP distribution to the data with location $\mu = 0$, scale $\sigma(x) = \theta + x^T\gamma > 0$ and shape $\xi > 0$. Since we assume that $\phi = 0$, we do not estimate it from the data. Both the slope and the intercept of the target quantile can be estimated from a GP fit to the data. When $\phi \neq 0$ (case-II), we use the scaled GP method as proposed in section 2.5, where we estimate ϕ as the minimum value in the data. We compare our method (GP.scale) with quantile regression (QR) for both case-I and II. We generate our data as follows:

$$y_i = \phi + (1 + 0.9x_i)e_i, i = 1, \ldots, n$$

Here $x_i \sim U(-1, 1)$ and e_i are independent random variables, $\theta = 1$ and $\gamma = 0.9$. Clearly, $1 + 0.9x_i > 0$ for all $i = 1, \ldots, n$. We choose two different values $\phi = 0$ and 2. For both

$\phi = 0$ and $\phi = 2$, we simulate e_i from two distributions: exponential with mean $\lambda = 1$ (exp(1)) and standard GP with $\xi = 0.5$ (GP(0.5)). Standard exponential is the limit of the GP distribution in 2.3.2 for $\xi = 0$. We use two sample sizes $n = 500$ and 1000. The Monte Carlo sample size for all of the cases is $K = 500$.

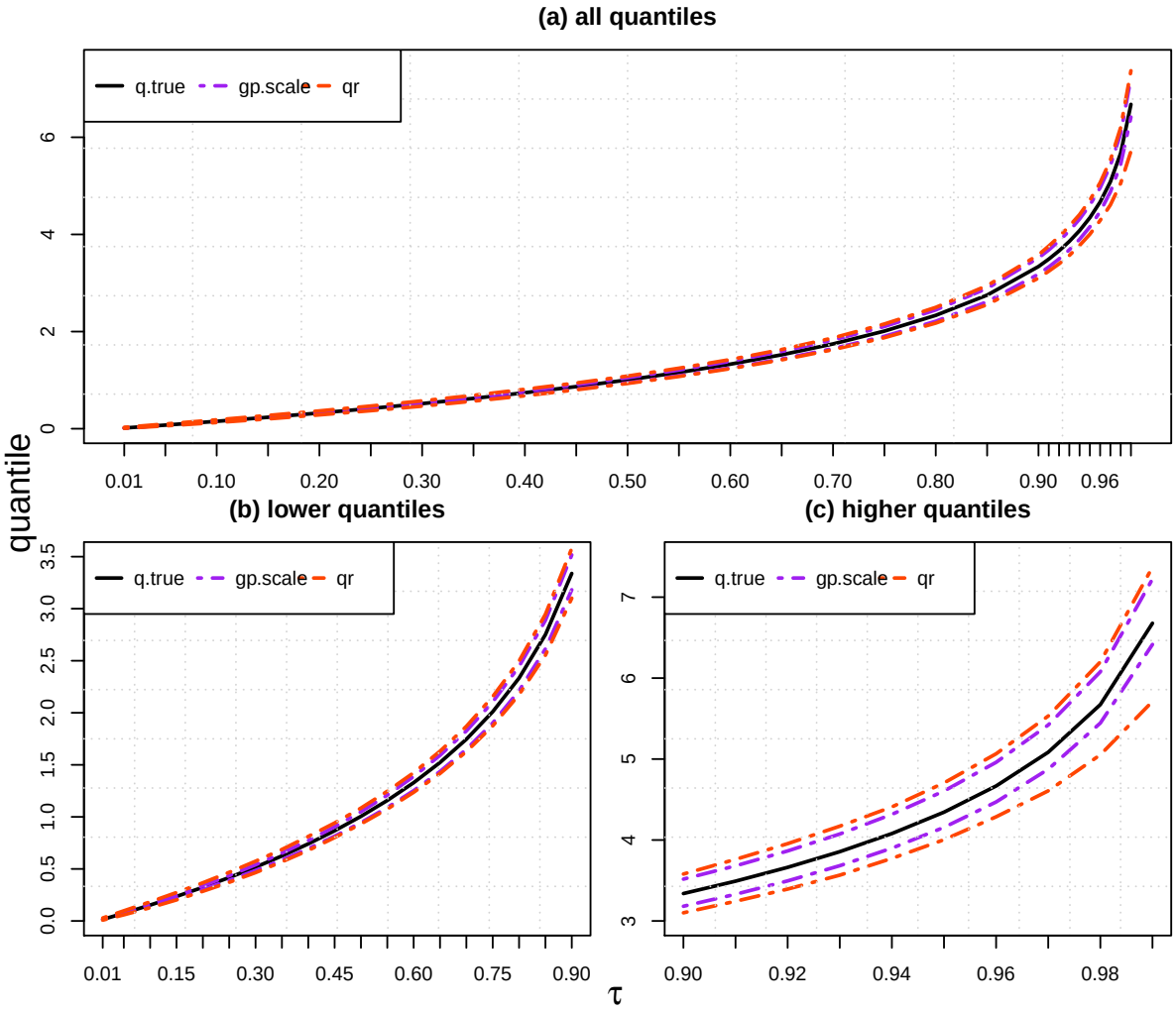

Figure 2.2: Quantile estimates for $\phi = 0$, exp(1) at $x = 0.5$ and quantile levels τ. The black line is the true quantile function. Average $\pm$ standard error curves of the estimates are plotted for QR (orange) and GP.scale (violet). The sample size is $n = 500$.

First we look at the τ quantile estimates for case-I with $e_i \sim \exp(1)$. We use the Monte

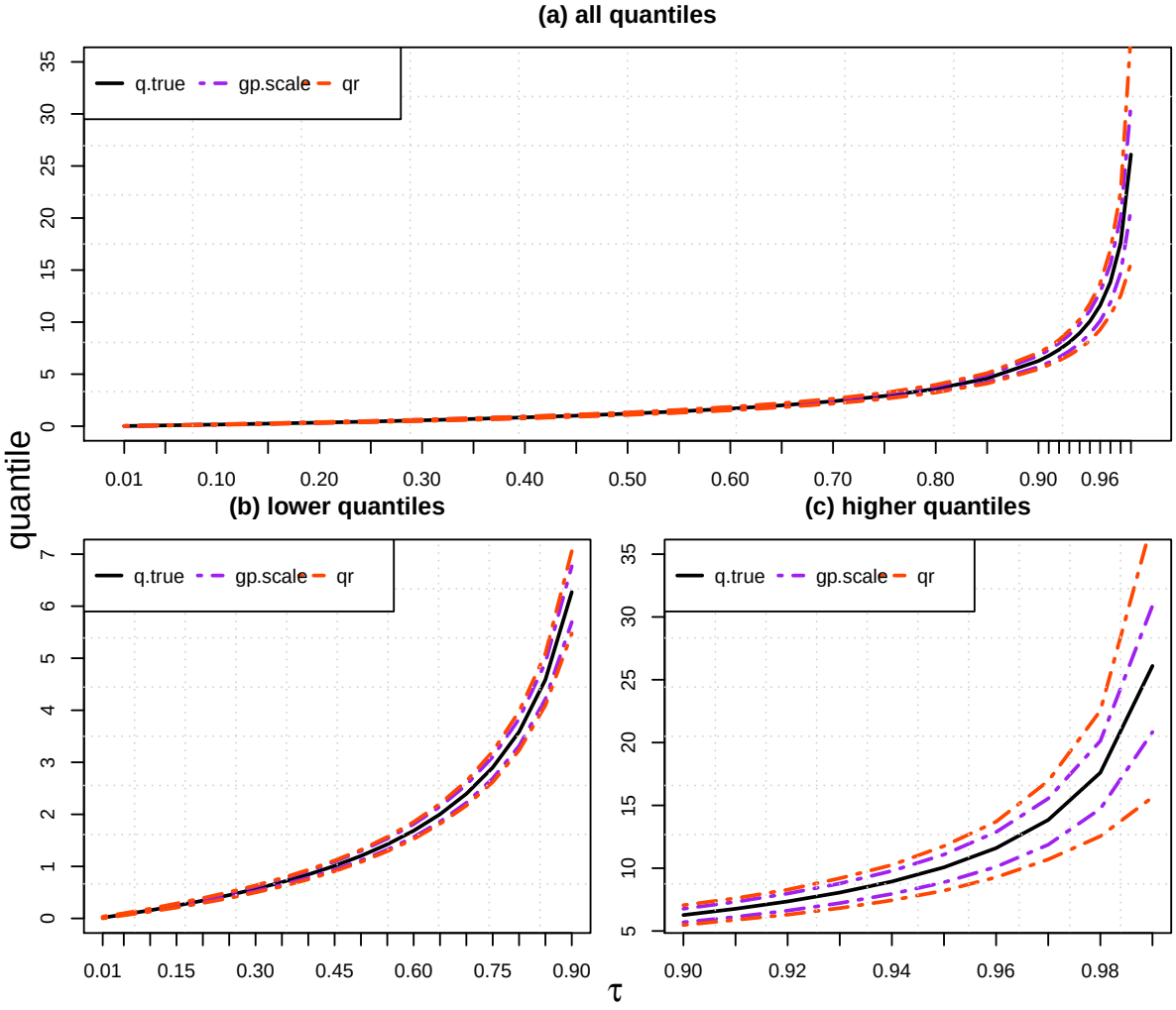

Figure 2.3: Quantile estimates for $\phi = 0$, GP(0.5) at $x = 0.5$ and quantile levels τ. The black line is the true quantile function. Average $\pm$ standard error curves of the estimates are plotted for QR (orange) and GP.scale (violet). The sample size is $n = 500$.

Carlo average estimate for the τ quantile

$$\bar{Q}(\tau) = \frac{1}{K} \sum_{i=1}^{K} \hat{Q}_{Y|x}(\tau)$$

The black line in figure 2.2 is the true quantile (q.true) function. Average estimates from both of the methods are very close to q.true. For better visualization, we have included only average $\pm$ standard error (se) curves for both QR (orange) and GP.scale (violet). We plot for all quantiles $0 < \tau < 1$ in figure 2.2 (a). Both GP.scale and QR estimates have less

Table 2.1: $\phi = 0$, GP(0.5): rel.bias(se) of quantile estimates for $\phi = 0$ at $x = 0.5$ and quantile levels τ.

τ	$n = 500$		$n = 1000$	
	GP.scale	QR	GP.scale	QR
0.1	0.07(0.06)	0.14(0.11)	0.05(0.03)	0.11(0.08)
0.3	0.06(0.05)	0.09(0.08)	0.05(0.03)	0.07(0.05)
0.5	0.06(0.05)	0.08(0.06)	0.04(0.03)	0.06(0.04)
0.7	0.06(0.04)	0.08(0.06)	0.04(0.03)	0.06(0.04)
0.9	0.07(0.05)	0.1(0.08)	0.05(0.04)	0.07(0.05)
0.93	0.08(0.06)	0.12(0.09)	0.06(0.04)	0.08(0.06)
0.95	0.09(0.06)	0.14(0.11)	0.06(0.05)	0.09(0.07)
0.97	0.11(0.08)	0.18(0.14)	0.07(0.06)	0.11(0.09)
0.99	0.15(0.12)	0.29(0.31)	0.1(0.08)	0.17(0.15)

variability in lower quantiles. For better visibility, we separate our quantiles into two regions: $\tau \in (0.9, 1)$ (higher quantiles) and $\tau \in (0, 0.9)$ (lower quantiles). In figure 2.2 (b) we can see that both QR and GP.scale approximate lower quantiles pretty well. At higher quantiles, QR has more variability than GP.scale. To see if the behavior changes with heavier than exponential tails, we proceed to case-I where $e_i \sim$ GP(0.5). In figure 2.3 (a) we can see that GP(0.5) quantiles have a much higher range than exp(1). We do not have any issues with QR at lower quantiles. In figure 2.3 (c) we can see that QR estimates have very high standard error, particularly for $\tau > 0.95$. Note that from figures 2.2 and 2.3 we do not get any idea about the bias of the estimators. So we look at the relative bias from K iterations.

$$rel.bias = \frac{1}{K}\sum_{i=1}^{n}\Big|1 - \frac{\hat{Q}_{Y|x,k}(\tau)}{Q_{Y|x}(\tau)}\Big|$$

We have enumerated the relative bias for case-I, GP(0.5) in table 2.1. Our true quantile values are very close to 0 for $\tau \approx 0$. That leads to a higher relative bias at lower quantiles. At higher quantiles, we expect the rel.bias to be large. The relative bias of case-I, exp(1) is given in table A.1 in the appendix. As we can see, looking at the relative bias is not enough. Hence, we use mean square error (mse), which can account for both the bias and the variance.

$$mse = \frac{1}{K}\sum_{i=1}^{n}\Big(Q_{Y|x}(\tau) - \hat{Q}_{Y|x,k}(\tau)\Big)^2 \tag{2.22}$$

Figure 2.4 (a) gives us mse of quantile estimates for QR and GP.scale for all quantiles.

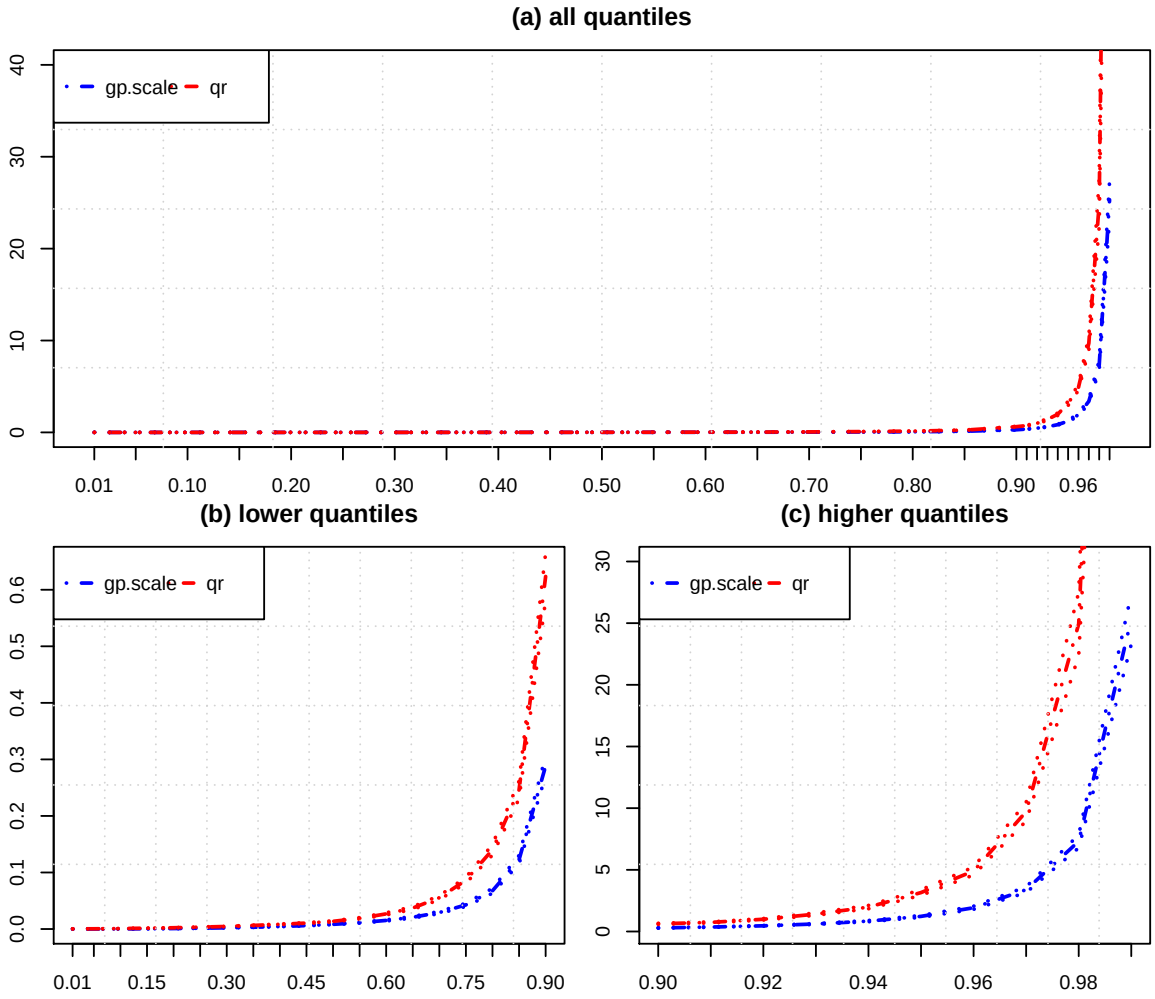

Figure 2.4: mse(se) for $\phi = 0$, GP(0.5) at $x = 0.5$ and quantile levels τ. The sample size is $n = 500$.

We can see that the mse is less than 0.6 at lower quantiles for both of the methods. After 0.95 quantile, mse for both QR and GP.scale increases at a very high rate.But relative to QR, GP.scale is more stable, particularly at higher quantiles. Note that we have very few observations after 0.9 quantile. With a sample size of $n = 500$ we do not expect to do any better. We have summarised these results in table 2.2 for some quantile levels. Clearly for $n = 500$, mse of QR starts to explode after $\tau = 0.95$. When we increase our sample size to 1000, the performance of QR improves, but it does not surpass GP.scale. Note that

Table 2.2: $\phi = 0$, GP(0.5): mse(se) of quantile estimates for $\phi = 0$ at $x = 0.5$ and quantile levels τ.

τ	$n = 500$		$n = 1000$	
	GP.scale	QR	GP.scale	QR
0.1	0(0)	0(0)	0(0)	0(0)
0.3	0(0)	0(0)	0(0)	0(0)
0.5	0.01(0)	0.01(0)	0(0)	0.01(0)
0.7	0.03(0)	0.06(0)	0.02(0)	0.03(0)
0.9	0.29(0.02)	0.62(0.04)	0.15(0.01)	0.33(0.02)
0.93	0.61(0.04)	1.41(0.1)	0.31(0.02)	0.67(0.04)
0.95	1.23(0.08)	3.16(0.23)	0.61(0.04)	1.35(0.09)
0.97	3.39(0.23)	9.72(0.76)	1.61(0.12)	3.77(0.26)
0.99	25.32(2.04)	120.95(19.78)	11.29(0.91)	33.62(3.03)

Table 2.3: $\phi = 2$, GP(0.5): rel.bias(se) of quantile estimates for $\phi = 2$ at $x = 0.5$ and quantile levels τ.

τ	$n = 500$		$n = 1000$	
	GP.scale	QR	GP.scale	QR
0.1	0.01(0)	0.01(0.01)	0(0)	0.01(0.01)
0.3	0.01(0.01)	0.02(0.02)	0.01(0.01)	0.01(0.01)
0.5	0.02(0.02)	0.03(0.02)	0.02(0.01)	0.02(0.02)
0.7	0.03(0.02)	0.04(0.03)	0.02(0.02)	0.03(0.02)
0.9	0.05(0.04)	0.08(0.06)	0.04(0.03)	0.06(0.04)
0.93	0.06(0.05)	0.09(0.07)	0.05(0.03)	0.07(0.05)
0.95	0.07(0.06)	0.11(0.08)	0.05(0.04)	0.08(0.06)
0.97	0.09(0.07)	0.14(0.1)	0.06(0.05)	0.1(0.08)
0.99	0.14(0.11)	0.26(0.25)	0.1(0.07)	0.17(0.14)

the rel.bias in table 2.1 could not reflect so many details as the magnitude of the bias is proportional to the value of the true quantile. We do not have this issues when we use mse. For case-I, exp(1), the magnitude of the bias is much lower than GP(0.5) at all quantiles. These results are given in the appendix figure A.1 and table A.2.

Now we move to case-II where $\phi = 2$. First we simulate $e_i \sim$ GP(0.5). We want to see how our estimates fluctuate at higher quantiles ($\tau > 0.9$). We use two sample sizes $n = 500$ and 1000. Figure 2.5 gives us the quantile function and the estimates. Like in case-I, both QR and GP.scale estimates improves with increasing sample size. In table 2.3, we can see that the relative bias is an increasing function of quantile level τ for both of the methods. Now

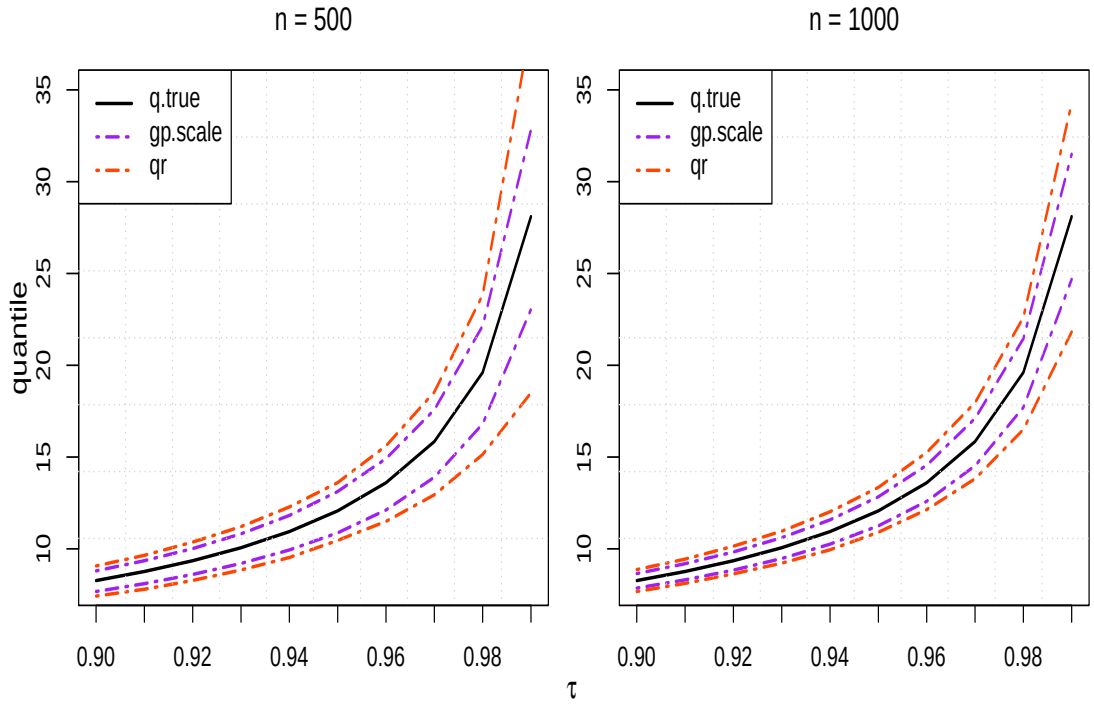

Figure 2.5: Quantile estimates for $\phi = 2$ at $x = 0.5$ and quantile levels τ. The black line is the true quantile function. Average$\pm$ standard error curves of the estimates are plotted for QR (orange) and GP.scale (violet).

the true quantile value is always more than $\phi = 2$ since $Q_{Y|x}(\tau) = 2 + (1 + 0.9x)Q\epsilon(\tau) \geq 2$ for all $\tau \in [0, 1]$. We can see that for all quantile values, GP.scale has lower rel.bias compared to QR, even when we increase our sample size to 1000. Also, QR has very high relative bias at $\tau = 0.97$ and 0.99. The mse in figure 2.6 conveys the same picture as relative bias and it is easier to interpret figure 2.6 than table 2.3. Moreover, we can see that mse for both case-I and case-II reflects the magnitude of the bias. Hence from now onward, we will only use mse as a measure of comparing different estimates. We have summarised mse values at some quantile levels for case-II, GP(0.5) in table 2.4. It is evident from here that using GP.scale is preferred over QR at higher quantiles. From our simulation studies, we see that the tails are approximated well with GP.scale when observations are form a scale model 2.9 and the underlying distribution is GP or exponential. Motivated by this we extend our method for the location-scale family of regression models in the next chapter.

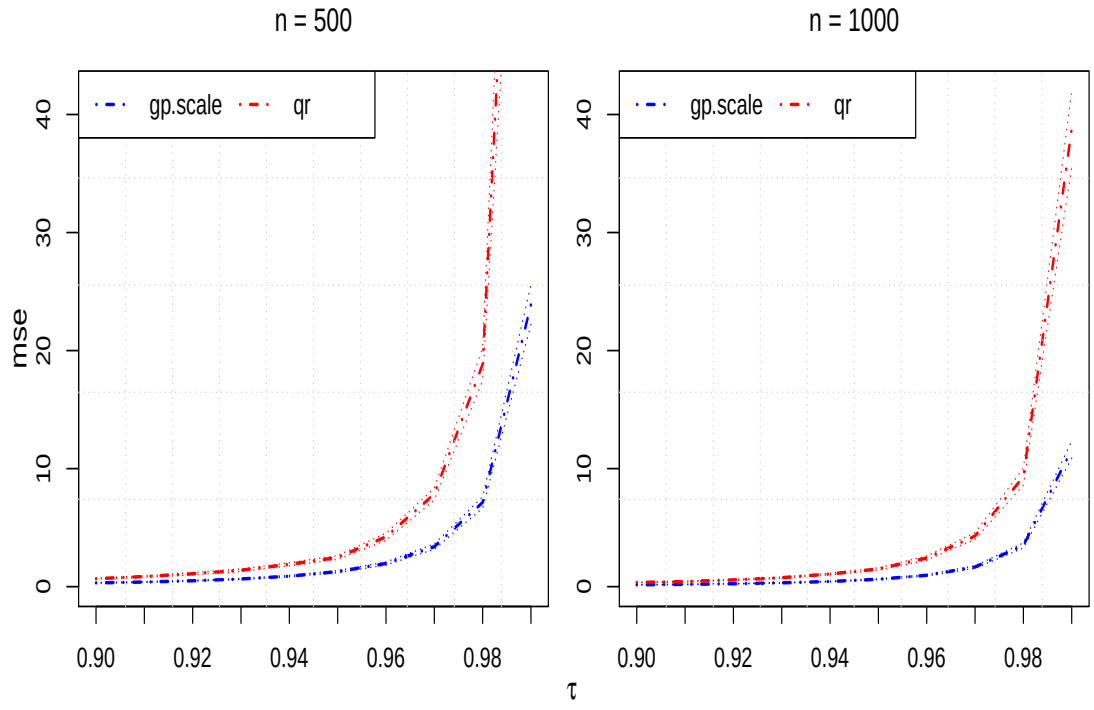

Figure 2.6: mse±se for $\phi = 2$, GP(0.5) and quantile levels τ.

Table 2.4: $\phi = 2$, GP(0.5): mse(se) of quantile estimates for $\phi = 2$ at $x = 0.5$ and quantile levels τ.

τ	$n = 500$		$n = 1000$	
	GP.scale	QR	GP.scale	QR
0.5	0.01(0)	0.01(0)	0(0)	0.01(0)
0.75	0.05(0)	0.09(0.01)	0.02(0)	0.04(0)
0.9	0.32(0.02)	0.67(0.05)	0.16(0.01)	0.36(0.02)
0.93	0.65(0.04)	1.39(0.1)	0.32(0.02)	0.76(0.05)
0.95	1.27(0.08)	2.47(0.16)	0.63(0.04)	1.49(0.1)
0.97	3.39(0.22)	7.84(0.51)	1.66(0.1)	4.29(0.29)
0.99	23.9(1.68)	100.11(14.26)	11.6(0.73)	38.78(3.24)

CHAPTER 3

ESTIMATION OF EXTREME CONDITIONAL QUANTILES WITH COVARIATE ADAPTIVE THRESHOLD SELECTION

3.1 Introduction

The study of the tail behavior of a distribution is helpful for the analysis of rare events. Often, Generalized Pareto (GP) distribution is used for modeling the tail of a distribution. This approach is popularly known as the Peak Over Threshold (POT) approach. In the POT approach, a GP distribution is commonly fitted to the tail region of the data using only the observations exceeding a certain threshold. A fundamental result in the extreme value literature by Pickands (1975) pointed out that GP distribution can be used to model the tail, as long as the underlying distribution satisfy certain regularity conditions. Let Z be a random variable with distribution function F. Pickands (1975) showed that the distribution of the conditional excess, $Z - u|Z > u$ can be approximated by GP distribution as long as F belongs to the maximum domain of attraction of an extreme value distribution. It is hard to find distributions which do not belong to the maximum domain of attraction. Hence, the POT approach can be used even when F is unknown. It is important to keep in mind that the GP approximation is valid when u is large. As pointed out by Song & Song (2012a), selecting a low threshold value will lead to poor approximation of the excess by GP and hence the resulting estimates will have high bias. On the other hand, when the threshold is too high, the effective sample size used for estimating the parameters is small which leads to high variance. Thus, choosing an optimal threshold is necessary for the bias-variance tradeoff. Once the threshold is determined, the rest of the parameters: scale and shape are estimated from the GP fit to the scaled excess data. Langousis et al. (2016) have nicely summarised the available methods for threshold selection. Davison & Smith (1990) introduced statistical modeling of peaks over threshold with GP distribution where

the parameters of the GP distibution are allowed to vary as a function of covariates. Some of the more recent works which use covariate-based GP distribution for modeling peaks over threshold include Park & Kim (2016), Song & Song (2012b), del Castillo & Serra (2015). A comprehensive review of GP modeling where the parameters of the GP vary as a function of the covariates has been presented in Coles et al. (2001). In all these works, the threshold of which GP is fit is chosen among the set of empirical quantiles of the response variable. To the best of our knowledge, the problem of choosing the threshold as a function of the covariate has not been addressed in the literature. The main focus of this chapter is to develop a method which can choose a covariate dependent threshold while still suitably estimating the covariate dependent parameters of the GP distribution.

In this chapter, we approach the conditional quantile estimation problem where the threshold can vary across different covariates. Integrating QR and EVT, we propose a consistent estimation of the quantile function. This method is quite flexible in the sense that it can be easily adapted for distributions in the heavy tail domain of attraction. Although there are relatively few works which address the issue of estimation of high conditional quantiles for heavy tailed distributions, the one by Wang et al. (2012) presents a very promising method for this purpose. Their method uses standard QR to estimate intermediate quantiles and then employs EVT to extrapolate this information to estimate extreme quantiles. Their approach is semi-parametric and uses the Hill estimator (see Hill (1975)) for estimating the tail behavior. We compare our parametric approach based on the GP distribution to that of Wang et al. (2012) in terms of accuracy of estimation for both quantiles and the tail index. The parametric based approach is the best method when the true observations are indeed from a Pareto or Generalized Pareto distribution. For other distributions in the Fréchet domain of attraction, the method is fairly competitive and comes with several advantages in terms of practical implementation.

The rest of this chapter is organized as follows. In section 3.2 we propose our method (QR adjusted GP) for location-scale family of regression models. In 3.2.1, we establish asymptotic

properties of QR adjusted GP in the context of the GP distribution. We also demonstrate efficiency of our approach over standard QR and our previous method (GP.scale, section 2.5) through numerical investigations in 3.2.2. In section 3.3 we propose a generalization of our method for heavy tailed distributions. We give an outline of the analytical properties of this extension in 3.3.1. In section 3.4, we explore conditional quantile estimation in the heavy tail domain of attraction through simulation study. In 3.4.1 we review high conditional quantile estimation method of Wang et al. (2012) where the slopes of the quantile function also changes with quantile levels. We compare our method with Wang et al. (2012) for different distributions in the heavy-tail domain of attraction in 3.4.2. From our findings, we conclude that our proposed method better or as good as Wang et al. (2012). Finally, in section 3.5 we apply our method for the precipitation downscaling task with rainfall data.

3.2 Proposed Method: QR Adjusted GP for Location-Scale Models

In chapter 2, we have seen that our proposed method is suitable for the following family of regression models:

$$Y = \phi + (\theta + X^T\gamma)\epsilon \tag{3.1}$$

where $\phi \in \mathbb{R}$ and $\theta + X^T\gamma > 0$, and $\epsilon \sim \text{GP}(\mu = 0, \sigma = 1, \xi)$. In this case, $Y|X = x \sim \text{GP}(\mu = \phi, \sigma(x) = \theta + x^T\gamma, \xi)$. Hence, the conditional distribution of Y depends on x only through the scale parameter $\sigma(x) = \theta + x^T\gamma$. Now we concentrate on the general case:

$$Y = \phi + X^T\rho + (\theta + X^T\gamma)\epsilon \tag{3.2}$$

Here the conditional distribution of Y given $X = x$ depends on x through both the scale $\sigma(x) = \theta + x^T\gamma$ and the location $\mu(x) = \phi + x^T\rho$. Clearly, the model 3.1 is a special case of 3.2 for $\rho = 0$. Let $\tilde{y}_i = (y_i - \mu(x_i))/\sigma(x_i)$. If we try to fit the GP distribution to the

observations directly, the parameters in 3.4 can be estimated as

$$\begin{aligned}\hat{\phi},\hat{\rho},\hat{\theta},\hat{\gamma},\hat{\xi} &= \underset{\phi,\rho,\theta,\gamma,\xi}{\operatorname{argmin}}\, l(\phi,\rho,\theta,\gamma,\xi) \\ &= \underset{\phi,\rho,\theta,\gamma,\xi}{\operatorname{argmin}} \left[\sum_{i:\tilde{y}_i>0} \log(\sigma(x_i)) + (1+\frac{1}{\xi}) \sum_{i:\tilde{y}_i>0} \log\Big(1+\xi\frac{(y_i-\mu(x_i))}{\sigma(x_i)}\Big)\right] \\ &= \underset{\phi,\rho,\theta,\gamma,\xi}{\operatorname{argmin}} \left[\sum_{i:\tilde{y}_i>0} \log(\theta+x_i^T\gamma) + (1+\frac{1}{\xi}) \sum_{i:\tilde{y}_i>0} \log\Big(1+\xi\frac{(y_i-\phi-x_i^T\rho)}{\theta+x_i^T\gamma}\Big)\right]\end{aligned} \tag{3.3}$$

We have to minimize the loss function in 3.3 such that $\xi > 0$, $\sigma(x_i) > 0$ and $y_i - \mu(x_i) > 0$ for all $i = 1, \ldots, n$. Like in the optimization problem of 2.11 in the previous chapter, there is no analytical solution to this. We need to solve this with $2n+1$ many constraints numerically. Note that n many constraints come from the condition $y_i - \mu(x_i) > 0$. Our proposed method can reduce the computational challenge by removing these constrains. We discuss the details in the next section.

3.2.1 Method of Estimation

We assume the following location-scale regression model

$$Y = \phi + X^T\rho + (\theta + X^T\gamma)\epsilon, \tag{3.4}$$

where $\phi, \theta \in \mathbb{R}$, $\rho = (\rho_1, \ldots, \rho_p)$ and $\gamma = (\gamma_1, \ldots, \gamma_p)$ in $\mathbb{R}^p$ are unknown parameters. $\theta + X^T\gamma > 0$ for all X in its domain. $X^T\gamma = \gamma_1 X_1 + \cdots + \gamma_p X_p$ and $X^T\rho = \rho_1 X_1 + \cdots + \rho_p X_p$ are the linear combinations of covariates in equation 3.4. Given $X = x$ and a quantile level $0 < \tau < 1$, the conditional quantile is given by

$$Q_{Y|x}(\tau) = \phi + x^T\rho + (\theta + x^T\gamma)Q_\epsilon(\tau)$$

Suppose ϵ follows standard GP distribution with tail index $\xi > 0$, i.e., $\epsilon \sim \text{GP}(\xi)$. Then Y given $X = x$ follows GP distribution with location $\mu(x) = \phi + x^T\rho$, scale $\sigma(x) = \theta + x^T\gamma > 0$, and tail index $\xi > 0$. The conditional quantile is

$$Q_{Y|x}(\tau) = \phi + x^T\rho + \frac{(\theta + x^T\gamma)}{\xi}\Big[(1-\tau)^{-\xi} - 1\Big] \tag{3.5}$$

Through our numerical investigation in section 2.6, we have seen that for the scale family of models $Y = \phi + (\theta + X^T\gamma)\epsilon$, $\epsilon \sim \text{GP}(\xi)$, quantile regression can reasonably approximate any lower quantile $\tau \in (0, 0.9)$. But quantile regression suffers from very high variability, particularly when $\tau > 0.95$. Although we need an alternative of quantile regression when τ is close to 1, we can use the information from quantile regression estimates a lower quantiles. For covariate free threshold $\mu = \phi$, we can use the first order statistic $Y_{(1)}$ to estimate μ. Motivated by this, we propose using the quantile regression estimate at quantile level $\tau_c = \frac{1}{n}$ as our threshold.

$$\begin{aligned} \hat{\mu}(x) &= \widehat{\phi + x^T}\rho = \hat{\text{QR}}_{Y|x}\Big(\frac{1}{n}\Big) \\ \hat{\text{QR}}_{Y|x}\Big(\frac{1}{n}\Big) &= \hat{\alpha}\Big(\frac{1}{n}\Big) + x^T\hat{\beta}\Big(\frac{1}{n}\Big) \end{aligned} \tag{3.6}$$

We obtain $\hat{\alpha}(\frac{1}{n})$ and $\hat{\beta}(\frac{1}{n})$ from equation 2.3. Suppose

$$e\Big(\frac{1}{n}, x\Big) = Y - \hat{\text{QR}}_{Y|x}\Big(\frac{1}{n}\Big) = Y - \hat{\mu}(x)$$

Then mimicking the procedure in 2.17, we estimate the rest of the parameters θ, γ, and ξ as

$$\begin{aligned} \hat{\theta}, \hat{\gamma}, \hat{\xi} &= \underset{\theta,\gamma,\xi}{\text{argmin}}\, l(\theta, \gamma, \xi) \\ &= \underset{\theta,\gamma,\xi}{\text{argmin}} \left[\sum_{i:e(\frac{1}{n},x_i)>0} \log(\theta + x_i^T\gamma) + (1 + \frac{1}{\xi}) \sum_{i:e(\frac{1}{n},x_i)>0} \log\Big(1 + \xi\frac{(y_i - \hat{\mu}(x_i))}{\theta + x_i^T\gamma}\Big) \right] \\ &= \underset{\theta,\gamma,\xi}{\text{argmin}} \left[\sum_{i:e(\frac{1}{n},x_i)>0} \log(\theta + x_i^T\gamma) + (1 + \frac{1}{\xi}) \sum_{i:e(\frac{1}{n},x_i)>0} \log\Big(1 + \xi\frac{e(\frac{1}{n}, x_i)}{\theta + x_i^T\gamma}\Big) \right] \end{aligned} \tag{3.7}$$

Finally, using 3.7 and 3.6, we can estimate the quantile function in 3.5 as

$$\hat{Q}_{Y|x}(\tau) = \hat{\text{QR}}_{Y|x}\Big(\frac{1}{n}\Big) + \frac{\hat{\theta} + x^T\hat{\gamma}}{\hat{\xi}}\Big[(1 - \tau)^{-\hat{\xi}} - 1\Big]$$

The algorithm of the estimation procedure is outlined below.

In the following proposition, we establish the consistency of algorithm 2 for GP distribution under the compactness assumption (A1) on the covariate x.

Algorithm 2: QR adjusted GP

1. Get $\hat{\mathrm{QR}}_{Y|x_i}\left(\frac{1}{n}\right) = \hat{\alpha}\left(\frac{1}{n}\right) + x_i^T\hat{\beta}\left(\frac{1}{n}\right)$ from 3.6.
2. Get $e(\frac{1}{n}, x_i) = y_i - \hat{\mathrm{QR}}_{y_i|x_i}\left(\frac{1}{n}\right)$
3. Get $\hat{\xi}, \hat{\theta}$ and $\hat{\gamma}$ from 3.7
4. Given $X = x$ and $0 < \tau < 1$, estimate the τ quantile as

$$\hat{Q}_{Y|x}(\tau) = \hat{\mathrm{QR}}_{Y|x}\left(\frac{1}{n}\right) + \frac{\hat{\theta} + x^T\hat{\gamma}}{\hat{\xi}}\left[(1-\tau)^{-\hat{\xi}} - 1\right]$$

Proposition 3.2.1. *Let* $y_i = \phi + x_i^\top\rho + (\theta + x_i^\top\gamma)\epsilon_i$, $\epsilon_i \overset{iid}{\sim} GP(\xi)$.
With $\widetilde{x}_i = [1, x_i^\top]$, $\widetilde{\gamma} = [\theta, \gamma]$ *and* $\widetilde{\rho} = [\phi, \rho]$, *let*

$$\widehat{\widetilde{x}_i^\top\widetilde{\rho}} = \hat{QR}_{Y|x_i}\left(\frac{1}{n}\right)$$

and

$$\hat{\gamma}, \hat{\xi} = \underset{c,d}{\operatorname{argmin}}\left[\sum_{i=1}^{n}\log(\widetilde{x}_i^\top c) + \left(1 + \frac{1}{d}\right)\sum_{i=1}^{n}\log\left(1 + \frac{d(y_i - \widehat{\widetilde{x}_i^\top\widetilde{\rho}})}{\widetilde{x}_i^\top c}\right)\right] \tag{3.8}$$

and estimate the τ *quantile of* Y *as*

$$\hat{Q}_{Y|x}(\tau) = \widehat{\widetilde{x}_i^\top\widetilde{\rho}} + \frac{\widetilde{x}^\top\hat{\gamma}}{\hat{\xi}}\left((1-\tau)^{-\hat{\xi}} - 1\right)$$

Then,

1. $$\sqrt{n}((\hat{\gamma}, \hat{\xi}) - (\widetilde{\gamma}, \xi)) \implies N(0, \Sigma)$$

2. $$\sqrt{n}(\hat{Q}_{Y|x}(\tau) - Q_{Y|x}(\tau)) \implies N(0, \sigma^2)$$

Proof. We first show that $\hat{\mathrm{QR}}_{Y|x_i}(1/n)$ is a consistent estimate of $\widetilde{x}_i^\top\tilde{\rho}$.

In this direction, note that by relation (7.1) in Chernozhukov (2005), we have

$$\sqrt{n}(\hat{\mathrm{QR}}_{Y|x_i}(1/n) - Q_{Y|x_i}(1/n)) \sim N\left(0, \frac{1/n(1-/n)}{f^2(F^{-1}(1/n))}(E(XX'))^{-1}\right)$$

where f and F denote the density and cumulative distribution function of $\mathrm{GP}(\xi)$ respectively. Since $F^{-1}(1/n) = (1/\xi)((1-1/n)^{-1/\xi} - 1) \to 0$, $f^2((1/\xi)((1-1/n)^{-1/\xi} - 1)) \to 1$. Hence,

$$\sqrt{n}(\hat{\mathrm{QR}}_{Y|x_i}(1/n) - Q_{Y|x_i}(1/n)) = o_P(1)$$

Also,

$$\sqrt{n}(Q_{Y|x_i}(1/n) - \widetilde{x}_i^\top \widetilde{\rho}) = \frac{\sqrt{n}\widetilde{x}_i^\top \widetilde{\gamma}}{\xi}\left(\left(1 - \frac{1}{n}\right)^{-1/\xi} - 1\right) \lesssim \frac{\sqrt{n}M_{\widetilde{\gamma}}}{n} = o_P(1)$$

where $M_{\widetilde{\gamma}} = M \max_j |\gamma_j|$ by assumption (A1). Hence,

$$\sqrt{n}(\hat{\mathrm{QR}}_{Y|x_i}(1/n) - \widetilde{x}_i^\top \widetilde{\rho}) = o_P(1)$$

First, suppose $\widetilde{x}_i^\top \widetilde{\rho}$ is known, then

$$\hat{\gamma}_0, \hat{\xi}_0 = \underset{\widetilde{\gamma},\xi}{\mathrm{argmin}}\left[\sum_{i=1}^n \log \widetilde{x}_i^T\widetilde{\gamma} + (1+\frac{1}{\xi})\sum_{i=1}^n \log\left(1 + \frac{\xi(y_i - \widetilde{x}_i^\top\widetilde{\rho})}{\theta + \widetilde{x}_i^T\widetilde{\gamma}}\right)\right] \tag{3.9}$$

Then,

$$\sqrt{n}((\hat{\gamma}_0, \hat{\xi}_0) - (\widetilde{\gamma}, \xi)) \implies N(0, I(\widetilde{\gamma}, \xi)^{-1})$$

where $I(\widetilde{\gamma}, \xi)$ is same as defined in (2.21). Thus, for $\widetilde{x}_i^\top \widetilde{\rho}$ known $(\hat{\gamma}_0, \hat{\xi}_0)$ is the solution of the equation

$$\nabla_{\widetilde{\gamma},\xi}\sum_{i=1}^n \log \widetilde{x}_i^T\gamma + (1+\frac{1}{\xi})\log\left(1 + \frac{\xi(y_i - \widetilde{x}_i^\top\widetilde{\rho})}{\theta + \widetilde{x}_i^T\gamma}\right) = 0$$

which implies

$$(\hat{\gamma}_0, \hat{\xi}_0) = (h_1(\widetilde{x}_i^\top\widetilde{\rho}), h_2(\widetilde{x}_i^\top\widetilde{\rho}))$$

Now, by Taylor expansion, we have

$$\begin{aligned}
&(h_1(\widehat{\widetilde{x}_i^\top\widetilde{\rho}}), h_2(\widehat{\widetilde{x}_i^\top\widetilde{\rho}})) \\
&\approx (h_1(\widetilde{x}_i^\top\widetilde{\rho}), h_2(\widetilde{x}_i^\top\widetilde{\rho})) + ((\widehat{\widetilde{x}_i^\top\widetilde{\rho}} - \widetilde{x}_i^\top\widetilde{\rho})h_1'(\widetilde{x}_i^\top\widetilde{\rho}), h_2(\widetilde{x}_i^\top\widetilde{\rho}) + (\widehat{\widetilde{x}_i^\top\widetilde{\rho}} - \widetilde{x}_i^\top\widetilde{\rho})h_2'(\widetilde{x}_i^\top\widetilde{\rho}))
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sqrt{n}((h_1(\widehat{\widetilde{x}_i^\top\widetilde{\rho}}), h_2(\widehat{\widetilde{x}_i^\top\widetilde{\rho}})) - (\widetilde{\gamma}, \xi)) \\
&\approx \sqrt{n}((h_1(\widetilde{x}_i^\top\widetilde{\rho}), h_2(\widetilde{x}_i^\top\widetilde{\rho})) - (\widetilde{\gamma}, \xi)) + \sqrt{n}(\widehat{\widetilde{x}_i^\top\widetilde{\rho}} - \widetilde{x}_i^\top\widetilde{\rho})(h_1'(\widetilde{x}_i^\top\widetilde{\rho}), h_2'(\widetilde{x}_i^\top\widetilde{\rho}))
\end{aligned}$$

As shown previously, $\sqrt{n}(\widehat{\widetilde{x}_i^\top\widetilde{\rho}} - \widetilde{x}_i^\top\widetilde{\rho}) = o_P(1)$, thus

$$\sqrt{n}((\hat{\gamma}, \hat{\xi}) - (\gamma, \xi)) \implies N(0, \Sigma)$$

where $\Sigma = (h_1'(\phi), h_2'(\phi))I(\gamma, \xi)^{-1}(h_1'(\phi), h_2'(\phi))^\top$ with $I(\gamma, \xi)$ same as in (2.21). For part (2), note that

$$\begin{aligned}
&\sqrt{n}(\hat{Q}_{Y|x}(\tau) - Q_{Y|x}(\tau)) \\
&= \sqrt{n}(\widehat{\widetilde{x}_i^\top\widetilde{\rho}} - \widetilde{x}_i^\top\widetilde{\rho}) + \sqrt{n}\left(\frac{\widetilde{x}^\top\hat{\gamma}}{\hat{\xi}}\left((1-\tau)^{-\hat{\xi}} - 1\right) - \frac{\widetilde{x}^\top\widetilde{\gamma}}{\xi}\left((1-\tau)^{-\xi} - 1\right)\right) \\
&= \sqrt{n}(\widehat{\widetilde{x}_i^\top\widetilde{\rho}} - \widetilde{x}_i^\top\widetilde{\rho}) + \sqrt{n}(g(\hat{\gamma}, \hat{\xi}) - g(\widetilde{\gamma}, \xi)) = o_P(1) + \sqrt{n}(g(\hat{\gamma}, \hat{\xi}) - g(\widetilde{\gamma}, \xi))
\end{aligned}$$

Now,

$$\sqrt{n}(g(\hat{\gamma}, \hat{\xi}) - g(\widetilde{\gamma}, \xi)) \approx \sqrt{n}((\hat{\gamma}, \hat{\xi}) - (\widetilde{\gamma}, \xi))^\top \nabla_{\widetilde{\gamma}, \xi} g(\widetilde{\gamma}, \xi)$$

Since $\sqrt{n}((\hat{\gamma}, \hat{\xi}) - (\widetilde{\gamma}, \xi)) \implies N(0, \Sigma)$, thus

$$\sqrt{n}(g(\hat{\gamma}, \hat{\xi}) - g(\widetilde{\gamma}, \xi)) \implies N(0, \sigma^2)$$

where $\sigma^2 = \nabla_{\widetilde{\gamma}, \xi} g(\widetilde{\gamma}, \xi)^\top \Sigma \nabla_{\widetilde{\gamma}, \xi} g(\widetilde{\gamma}, \xi)$. This completes the proof. □

3.2.2 Simulation Study: GP Distribution

We want to assess performance of our method when the data is simulated from the following regression model

$$\begin{aligned}
&Y = \phi + X^T\rho + (\theta + X^T\gamma)\epsilon, \quad \epsilon \sim \text{GP}(\xi) \\
&Q_{Y|x}(\tau) = \phi + x^T\rho + (\theta + x^T\gamma)Q_\epsilon(\tau) \\
&\alpha(\tau) = \phi + \theta Q_\epsilon(\tau), \quad \beta(\tau) = \rho + \gamma Q_\epsilon(\tau)
\end{aligned} \tag{3.10}$$

Again, we see that quanitles are linear in x. So we can use quantile regression in this case without violating any model assumption. We simulate $(x_i, y_i), i = 1, \ldots, n$ iid observations from 3.10. We use $\phi = \theta = 1$, $\rho = (1, 1)$ and $\gamma = (0.9, 0)$ so that

$$y_i = 1 + x_{1i} + x_{2i} + (1 + 0.9x_{1i})e_i$$

We simulate $x_{ij} \sim U(-1, 1)$ for $j = 1, 2$ so that $1 + 0.9x_{1i} > 0$ for all $i = 1, \ldots, n$. We want to see how QR and GP.scale perform compared to our proposed method QR adjusted GP (QR.GP) for estimating higher quantiles above 0.9. We simulate e_i from two distributions: exponential with mean $\lambda = 1$ (exp(1)) and standard GP with $\xi = 0.5$ and 0.8 (GP(ξ)). We use two sample sizes $n = 500$ and 1000. From proposition 3.2.1, we know that $\tau_c = \frac{1}{n}$ leads to consistent estimation of higher quantiles with QR.GP when the true distribution is GP. Here we want to validate our proposition through simulations. We also investigate the performance of QR.GP and QR.Scale with varying threshold. For that purpose, we choose τ_c from a grid of values in (0,1). When the data is generated from a GP or exponential distribution, values of τ_c close to 0 would be the best choice. We start with the case $e_i \sim$ GP(0.5). We use mse as our measure of assessment and our monte carlo sample size is $K = 500$.

The results for sample size $n = 500$ is summarised in figure 3.1. We can see in 3.1 (a) that GP.scale is not appropriate in this case. It fails to capture the effect of the covariate x at at any given higher quantile $0.9 \leq \tau \leq 0.99$. QR.GP performs better than QR for threshold $\tau_c = 0.01$ at any quantile level $0.9 \leq \tau \leq 0.99$. In 3.1 (b), we can see how mse of QR.GP changes for different values of τ_c when the quantile level is $\tau = 0.93$. QR.GP has lower mse than QR for all τ_c except at $\tau_c = 0.95$. When we increase the quantile level to $\tau = 0.97$, QR.GP has lower mse than QR for all values of τ_c (figure 3.1 (c)).

Now we increase our sample size to $n = 1000$. In figure 3.2 (a) we can see that even with an increase in the sample size, QR.GP outperforms QR at $\tau_c = 0.01$. From figure 3.2 (b) and (c), we get an understanding of the effect of τ_c on the estimated quantile by QR.GP. The mse is an increasing function of τ_c in this case. This is bound to happen since the effective sample size used for estimation decreases when we increase τ_c. Note that according

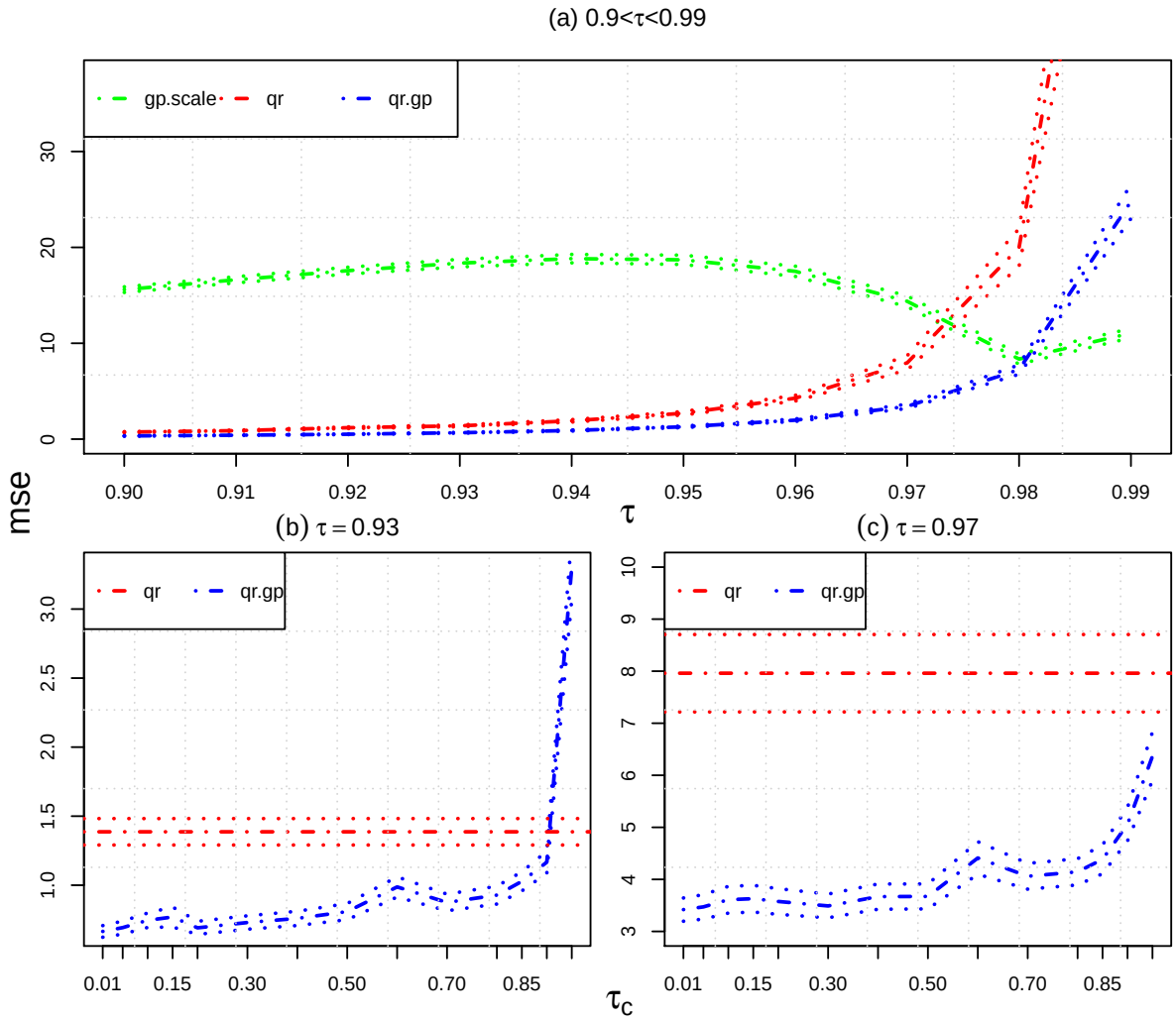

Figure 3.1: mse±se for, GP(0.5) at $x = (0.5, 0.5)$ and quantile levels τ. The sample size is $n = 500$. (a): $\tau_c = 0.01$ for QR.GP (b) and (c): $0 < \tau_c < 1$

to algorithm 2, given τ_c, we only use $\{y_i : y_i - x_i^T \hat{\text{QR}}(\tau_c) > 0\}$ to estimate the scale and the shape parameters. We get the best performance from QR.GP when $\tau_c = 0.01$. We can comfortably choose any τ_c as long as $\tau_c \leq 0.85$. Any choice of τ_c higher than 0.85 leads to inconsistent estimation of GP parameters as we have very few observations above the threshold.

Sparsity of the observations at tail of a distribution is associated with higher values of the tail index ξ. Next we simulate samples of size $n = 1000$ from GP with tail index $\xi = 0.8$. In figure 3.3 (a) we can see that the magnitude of the mse is higher than the previous case

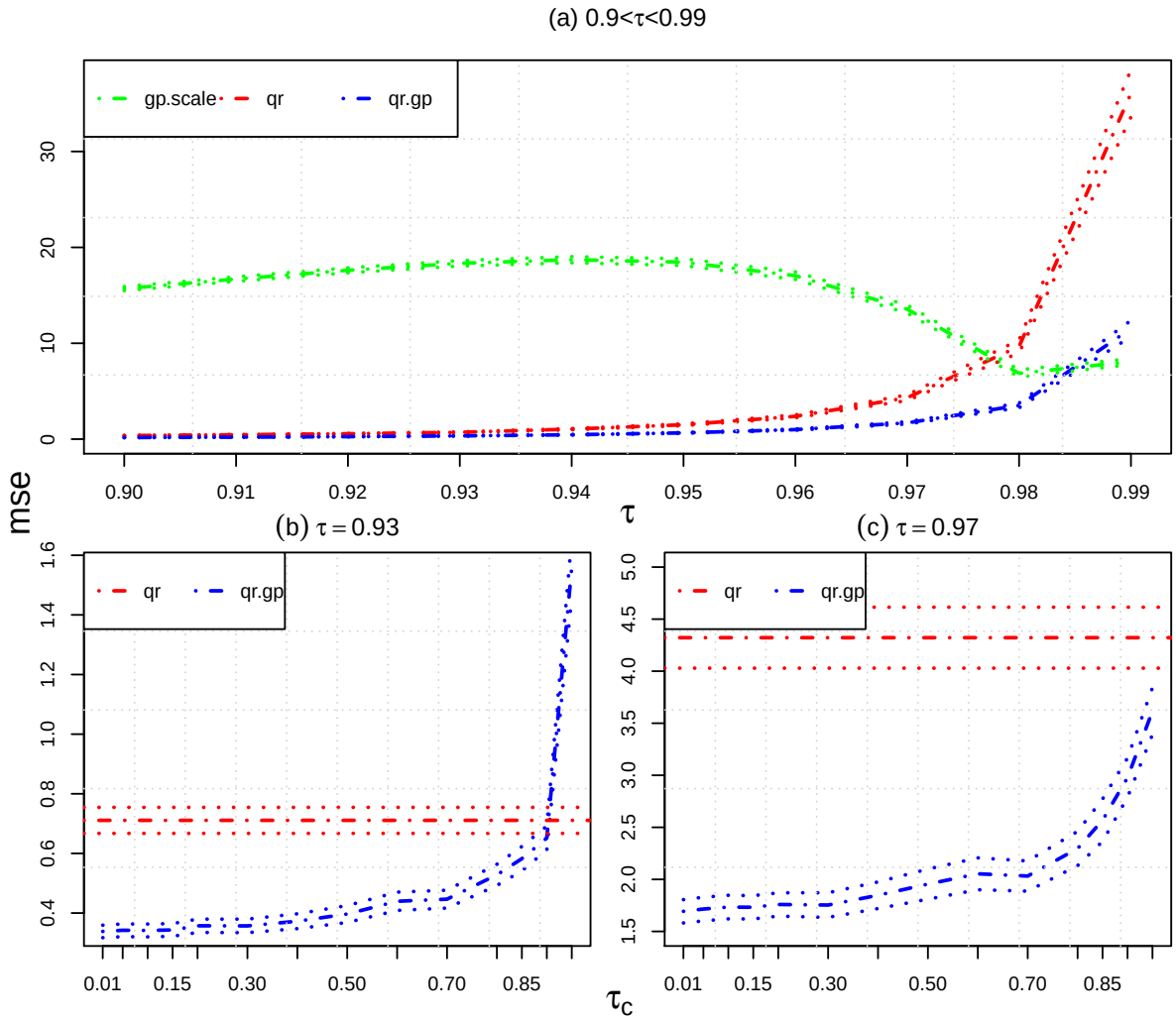

Figure 3.2: mse±se for, GP(0.5) at $x = (0.5, 0.5)$ and quantile levels τ.The sample size is $n = 1000$. (a): $\tau_c = 0.01$ for QR.GP. (b) and (c): $0 < \tau_c < 1$

$\xi = 0.5$ for all the methods. Nonetheless, we still outperform QR. From figure 3.3 (b) and (c), we can see that lower τ_c produces lower mse. Note that, our choice of τ_c is still 0.01 even when we increase ξ from 0.5 to 0.8. We end this section with our final simulation study for the exponential case, i.e., $e_i \sim \exp(1)$. This is an asymptotic case of GP(ξ)) when $\xi \to 0$. Hence, we start with larger sample size $n = 1000$. After experimenting with different τ_c, we have seen that $\tau_c \approx 0$ is optimal for QR.GP in this case. We have also seen that for $\tau_c \leq 0.15$ we do reasonably well and our performance deteriorates when we increase τ_c. Hence in figure 3.4 (a), we have plotted mse of QR.GP at $\tau_c = 0.01$. Although QR has

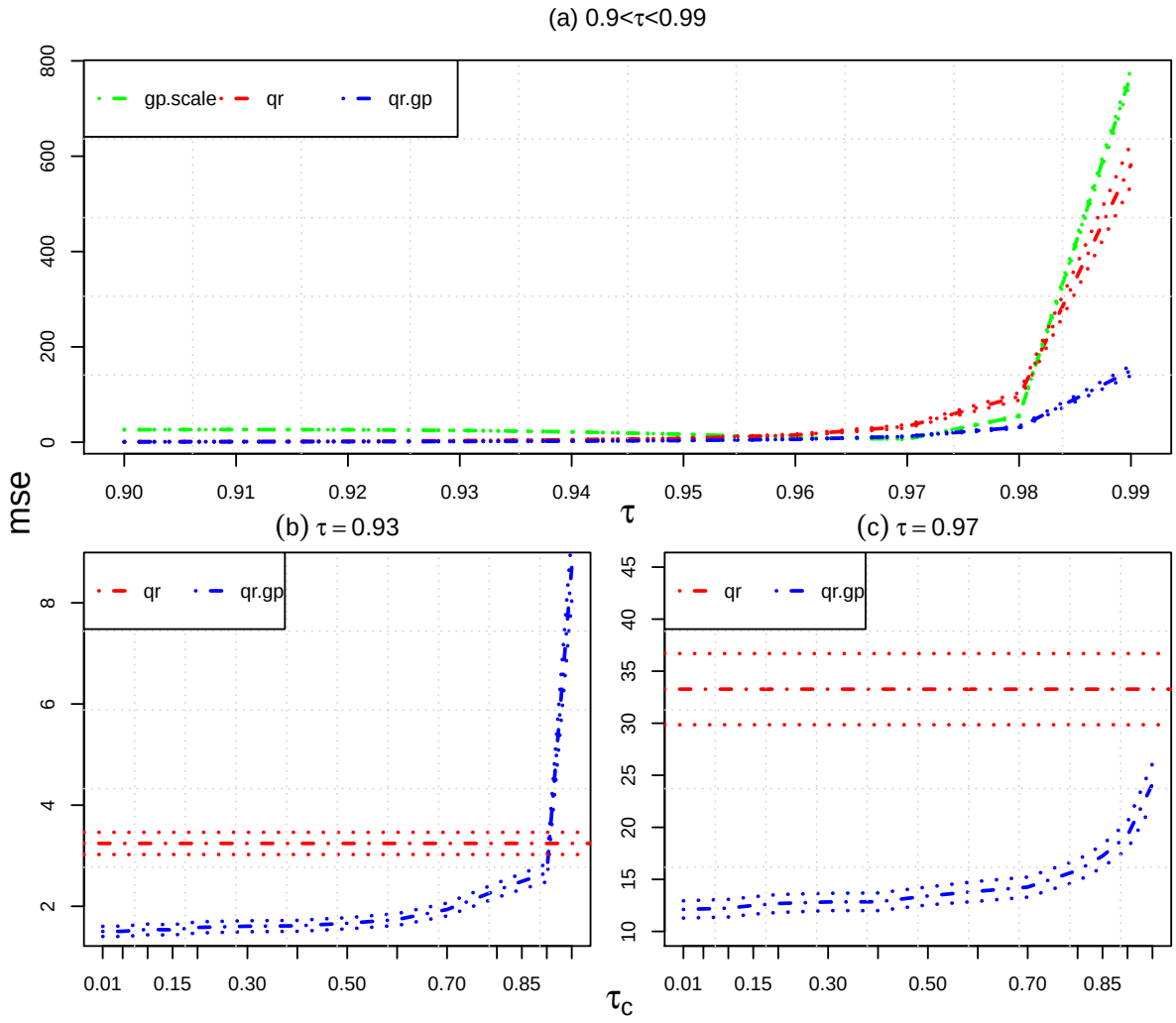

Figure 3.3: mse±se for, GP(0.8) at $x = (0.5, 0.5)$ and quantile levels τ. The sample size is $n = 1000$. (a): $\tau_c = 0.01$ for QR.GP. (b) and (c): $0 < \tau_c < 1$

lower mse than QR.GP, the magnitude of the difference is not much (less than 1.5 at higher quantiles). When we increase our sample size to $n = 2000$, we outperform QR at all the higher quantiles.

From our simulation studies, we can conclude that when the data is generated from GP or exponential distribution, optimal threshold for the QR.GP is close to 0. Performance of the QR deteriorates with increasing quantile level τ and higher values of the tail index ξ. Now we want to investigate how QR.GP behaves when the underlying distribution is a heavy tailed distribution other than GP.

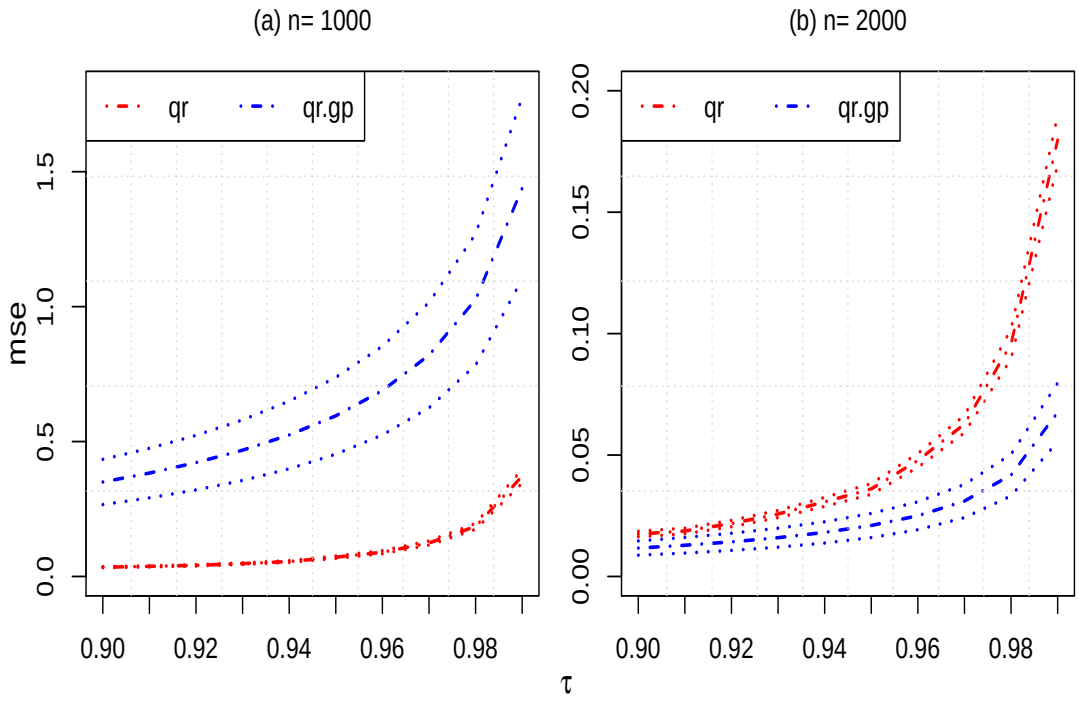

Figure 3.4: mse±se for, exp(1) at $x = (0.5, 0.5)$ and quantile levels τ. $\tau_c = 0.01$ for QR.GP.

3.3 Extension to the Nonparametric Problems and Threshold Selection

So far, we have only focused on the GP distribution, i.e., $\epsilon \sim \text{GP}(\xi)$, $\xi \geq 0$. Now we want to generalize for the case when ϵ is any heavy tailed distribution, i.e., $\epsilon \sim F$ where $F \in \mathcal{D}(G_\xi)$, $\xi > 0$. When we do not know F, we can use GP approximation to compute a higher quantile. First we focus on the nonregression case. Let $Y \sim F$ where $F \in \mathcal{D}(G_\xi)$ for some $\xi > 0$. Then, by Theorem 2.3.2, for a large enough value u, $Y|Y > u$ is approximately distributed according to a GP distribution. Hence, for any $y > u$,

$$P(Y > y|Y > u) = \left(1 + \frac{\xi(y-u)}{\sigma}\right)^{-1/\xi}$$

Let y_τ be the τ quantile of Y such that $y_\tau \geq u$. Then

$$\begin{aligned} P(Y > y_\tau) &= P(Y > u)P(Y > y_\tau | Y > u) \\ &= \bar{F}(u)\Big(1 + \frac{\xi(y_\tau - u)}{\sigma}\Big)^{-\frac{1}{\xi}} \\ &= 1 - \tau \\ \text{therefore,} \quad y_\tau = Q_Y(\tau) &= u + \frac{\sigma}{\xi}\Big[\Big(\frac{1-\tau}{\bar{F}(u)}\Big)^{-\xi} - 1\Big] \end{aligned} \tag{3.11}$$

Clearly, if we know F, we can substitute the exact expression for $\bar{F}(u)$ in 3.11. In practice we do not know F so we estimate $\bar{F}(u) = E(I(Y > u))$ with $\hat{p}_u = \sum_{i=1}^{n} I(y_i > u)/n$. Hence we can estimate the quantile in 3.11 as

$$\hat{Q}_Y(\tau) = u + \frac{\hat{\sigma}}{\hat{\xi}}\Big[\Big(\frac{1-\tau}{\hat{p}_u}\Big)^{-\hat{\xi}} - 1\Big] \tag{3.12}$$

where $\hat{\sigma}$ and $\hat{\xi}$ are estimated as

$$\hat{\sigma}, \hat{\xi} = \underset{\sigma,\xi}{\operatorname{argmin}} \sum_{i:y_i>u} \log(\sigma) + (1 + \frac{1}{\xi}) \sum_{i:y_i>u} \log\Big(1 + \xi\frac{(y_i - u)}{\sigma}\Big)$$

An important thing to note here is that we do not estimate u from the data. Moreover, 3.11 is an approximation of the true quantile for large u.

3.3.1 Method of Estimation

We assume model 3.4 with $\epsilon \sim F$, $F \in \mathcal{D}(G_\xi)$ for some $\xi > 0$. Let $0 < \tau_c < 1$ be an intermediate quantile level such that τ_c is not too close to 1. Here we use $\hat{\mu}(x)$ as our threshold.

$$\begin{aligned} \hat{\mu}(x) &= \hat{\mathrm{QR}}_{Y|x}(\tau_c) \\ \hat{\mathrm{QR}}_{Y|x}(\tau_c) &= \hat{\alpha}(\tau_c) + x^T\hat{\beta}(\tau_c) \end{aligned} \tag{3.13}$$

We obtain $\hat{\alpha}(\tau_c)$ and $\hat{\beta}(\tau_c)$ from equation 2.3. In 3.2.1 we used $\tau_c = \frac{1}{n}$. Here we want to choose τ_c appropriately so that $Y - \hat{\mu}(x)$ approximately follows GP distribution. Let

$e(\tau_c, x) = Y - \hat{\text{QR}}_{Y|x}(\tau_c)$. Then

$$\begin{aligned} e(\tau_c, x) &= \phi + x^T\rho + (\theta + x^T\gamma)\epsilon - \hat{\text{QR}}_{Y|x}(\tau_c) \\ &= \phi + x^T\rho + (\theta + x^T\gamma)\epsilon - \hat{\alpha}(\tau_c) - x^T\hat{\beta}(\tau_c) \\ &= \phi - \hat{\alpha}(\tau_c) + x^T(\rho - \hat{\beta}(\tau_c)) + (\theta + x^T\gamma)\epsilon \end{aligned}$$

We assume that given τ_c, $e(\tau_c, x)|e(\tau_c, x) > 0 \approx \text{GP}(\mu = 0, \sigma(x) = \theta + x^T\gamma, \xi)$. Hence, we can estimate the rest of the unknown parameters θ, γ and ξ by fitting a scaled GP model to the conditional residuals $e(\tau_c, x_i)|e(\tau_c, x_i) > 0$ for $i = 1, \ldots, n$.

$$\begin{aligned} \hat{\theta}, \hat{\gamma}, \hat{\xi} &= \underset{\theta,\gamma,\xi}{\text{argmin}}\, l(\theta, \gamma, \xi) \\ &= \underset{\theta,\gamma,\xi}{\text{argmin}} \sum_{i:e(\tau_c,x_i)>0} \log(\theta + x_i^T\gamma) + (1 + \frac{1}{\xi}) \sum_{i:e(\tau_c,x_i)>0} \log\left[1 + \xi\frac{e(\tau_c, x_i)}{\theta + x_i^T\gamma}\right] \end{aligned} \tag{3.14}$$

We need to minimize the loss function in 3.14 such that $\xi > 0$ and $\theta + x_i^T\gamma > 0$ for all $i = 1, \ldots, n$. Using $\hat{\text{QR}}_{Y|x}(\tau_c)$ to estimate $\mu(x) = \phi + x^T\rho$, we have reduced n constraints in our optimization problem. Let τ be a high quantile level such that $Q_{e(\tau_c,x)}(\tau) > 0$. Then from 3.11 we have

$$Q_{e(\tau_c,x)}(\tau) \approx \frac{\sigma(x)}{\xi}\Big[\Big(\frac{1-\tau}{P(e(\tau_c, x) > 0)}\Big)^{-\xi} - 1\Big]$$

Let $\hat{p}_0(\tau_c, n) = \frac{1}{n}\sum_{i=1}^n I[e(\tau_c, x_i) > 0]$. Following 3.12 we can estimate the τ quantile of the residual $e(\tau_c, x)$

$$\hat{Q}_{e(\tau_c,x)}(\tau) = \frac{\hat{\theta} + x^T\hat{\gamma}}{\hat{\xi}}\Big[\Big(\frac{1-\tau}{\hat{p}_0(\tau_c, n)}\Big)^{-\hat{\xi}} - 1\Big] \tag{3.15}$$

Finally, combining 3.15 and 3.13, we can estimate the conditional quantile of Y given x at a quantile level τ as

$$\begin{aligned} \hat{Q}_{Y|x}(\tau) &= \hat{\text{QR}}_{Y|x}(\tau_c) + \hat{Q}_{e(\tau_c,x)}(\tau) \\ &= \hat{\alpha}(\tau_c) + x^T\hat{\beta}(\tau_c) + \frac{\hat{\theta} + x^T\hat{\gamma}}{\hat{\xi}}\Big[\Big(\frac{1-\tau}{\hat{p}_0(\tau_c, n)}\Big)^{-\hat{\xi}} - 1\Big] \end{aligned}$$

The algorithm for estimation is outlined below.

Algorithm 3: QR adjusted GP for location-scale family

start: Set $0 < \tau_c \approx 0$; fix x, $\delta > 0$;

I. Obtain $\hat{\mathrm{QR}}_{Y|x_i}(\tau_c) = \hat{\alpha}(\tau_c) + x_i^T \hat{\beta}(\tau_c)$ from 3.13.

II. Get $e(\tau_c, x_i) = y_i - \hat{\mathrm{QR}}_{y_i|x_i}(\tau_c)$

III. Get $\hat{\xi}, \hat{\theta}$ and $\hat{\gamma}$ from optimizing 3.14 and $\hat{p}_0(\tau_c, n) = \frac{1}{n}\sum_{i=1}^{n} I[e(\tau_c, x_i) > 0]$

IV. For $0 < \tau < 1$ estimate τ quantile as

$$\hat{Q}_{Y|x}(\tau) = \hat{\mathrm{QR}}_{Y|x}(\tau_c) + \frac{\hat{\theta} + x^T\hat{\gamma}}{\hat{\xi}}\Big[\Big(\frac{1-\tau}{\hat{p}_0(\tau_c, n)}\Big)^{-\hat{\xi}} - 1\Big]$$

V. Set $\tau_c = \tau_c + \delta$ such that $0 < \tau_c < 1$ and repeat steps I-IV

end: Stop when $\tau_c \approx 0.95$

3.3.2 Discussion: Threshold Selection for Heavy Tailed Distributions

According to proposition 3.2.1, if the underlying distribution is GP, QR adjusted GP with $\tau_c = \frac{1}{n}$ leads to a consistent quantile estimator for a quantile level τ close to 1. Here we propose an outline to generalize our method when the underline distribution is in the heavy-tail domain of attraction. Let $Z_1, \ldots, Z_n$ be an independent and identically distributed sample with common distribution function $F \in \mathcal{D}(G_\xi)$. Then, given $Z_{(n-k,n)}$, the order statistics differences

$$(\tilde{Z}_1, \ldots, \tilde{Z}_k) = (Z_{(n-k+1)} - Z_{(n-k,n)}, \ldots, Z_{(n,n)} - Z_{(n-k,n)}) \tag{3.16}$$

follow approximately GP distribution with tail index ξ. Using theorem 2.3.2, we can find the parameters of GP by maximizing the log likelihood

$$\hat{\sigma}, \hat{\xi} = \underset{\gamma,\xi}{\operatorname{argmin}} \left[\sum_{i:\tilde{z}_i>0} \log(\sigma) + \left(1 + \frac{1}{\xi}\right) \sum_{i:\tilde{z}_i>0} \log\left(1 + \frac{\xi\tilde{z}_i}{\sigma}\right)\right]$$

In Theorem 3.4.2 of de Haan & Ferreira (2010), it has been shown $z_i \overset{d}{=} \frac{c_\xi}{(k/n)^\xi} + \frac{w_i}{\sqrt{k}(k/n)^\xi}$ where $W = (w_1, \ldots, w_k)$ follows multivariate normal distribution. This allowed them to

prove

$$\sqrt{k}((\hat{\xi}, \hat{\sigma}/(k/n)^{-\xi}) - (\xi, 1)) \implies N(\mu, \Sigma)$$

Under the same set of assumptions, in Theorem 3.1 of Diebolt et al. (2007), it has been shown that for

$$\hat{Q}_Z(\tau) = z_{(n-k,n)} + \frac{\hat{\sigma}}{\hat{\xi}} \left(\left(\frac{1-\tau}{k/n} \right)^{-\hat{\xi}} - 1 \right)$$

where

$$Q_Z(\tau) = F_Z^{-1}(\tau)$$

then

$$\sqrt{k}(k/n)^{\xi}(\hat{Q}_Z(\tau) - Q_Z(\tau)) \implies N(m, s^2).$$

For a regression model of the form,

$$Y_i = \underbrace{\phi + x_i^\top \rho}_{\widetilde{x}_i^\top \widetilde{\rho}} + \underbrace{(\theta + x_i^\top \gamma)}_{\widetilde{x}_i^\top \widetilde{\gamma}} \epsilon_i, \quad \epsilon_i \in \mathcal{D}(G_\xi), \quad \epsilon_i \ iid$$

We consider

$$E_i = Y_i - \hat{Q}_{Y|x_i} \left(1 - \frac{k}{n} \right)$$

where k is the same which ensures that given $E_{(n-k,n)}$

$$(E_{(n-k+1)} - E_{(n-k,n)}, \ldots, E_{(n,n)} - E_{(n-k,n)})$$

are approximately independent and identically distributed as GP with tail index ξ.

By Theorem 5.1 in Chernozhukov (2005),

$$\sqrt{k}(k/n)^{\xi} \left(\hat{Q}_{Y|x_i} \left(1 - \frac{k}{n} \right) - \widetilde{x}_i^\top \widetilde{\rho} - \widetilde{x}_i^\top \widetilde{\gamma} F_\epsilon^{-1}(1 - k/n) \right) \implies N(0, \sigma_X^2)$$

Therefore,

$$E_i = \widetilde{x}_i^\top \widetilde{\gamma}(\epsilon_i - F_\epsilon^{-1}(1 - k/n)) + O_P \left(\frac{1}{\sqrt{k}} \left(\frac{n}{k} \right)^{\xi} \right)$$

Applying Theorem 5.1 in Chernozhukov (2005) with covariate $x = 1$ (i.e. only constant term), we get

$$\sqrt{k}(k/n)^{\xi} \left(\epsilon_{(n-k,n)} - F_\epsilon^{-1}(1 - k/n) \right) \implies N(0, \sigma_\epsilon^2)$$

By assumption (A1), $\widetilde{x}_i^\top \widetilde{\gamma} \leq M \max_j |\gamma_j| = M_{\widetilde{\gamma}}$, therefore

$$E_i = \widetilde{x}_i^\top \widetilde{\gamma}(\epsilon_i - \epsilon_{(n-k,n)}) + O_P\left(\frac{1}{\sqrt{k}}\left(\frac{n}{k}\right)^{\xi}\right)$$

Let $i_0, i_1, i_2, \ldots, i_k$ be the indices corresponding to $\epsilon_{(n-k+1,n)}, \ldots, \epsilon_{(n,n)}$, then

$$S = \{i : E_i > 0\} \approx \{i_1, \ldots, i_k\}$$

Since, $\epsilon_{(n-k+1,n)} - \epsilon_{(n-k,n)}, \ldots, \epsilon_{(n,n)} - \epsilon_{(n-k,n)}$ are iid observations from GP distribution. This implies $E_{i_1}, \cdots, E_{i_k}$ are iid observations from GP distribution with scale $\widetilde{x_i}^\top \widetilde{\gamma}$ and tail index ξ.

Therefore, we can fit GP distribution on the residuals and estimate the parameters as

$$\hat{\gamma}, \hat{\xi} = \underset{\gamma,\xi}{\operatorname{argmin}} \left[\sum_{i:e_i>0} \log(\widetilde{x}_i^\top \gamma) + \left(1 + \frac{1}{\xi}\right) \sum_{i:e_i>0} \log\left(1 + \frac{\xi e_i}{\widetilde{x}_i^\top \gamma}\right)\right]$$

Note that e_i closely share the asymptotic behavior of $\tilde{z}_i$'s as defined in (3.16), i.e. $e_i = O_P(1/\sqrt{k}(k/n)^{\xi})$. However, the proof of asymptotic normality of the e_i's is too involved and is thereby omitted in this dissertation.

Mimicking the steps of Theorem 3.4.2 in de Haan & Ferreira (2010), the rate of convergence of the tail index $\hat{\gamma}$ and the scale coefficient $\hat{\gamma}$ can be determined.

Similary, mimicking the steps of Theorem 3.1 in Diebolt et al. (2007), one can determine the convergence rate of $\hat{Q}_{Y|X=x}(\tau)$. For the non-parametric regime, we explore the performance of our method from a numerical standpoint. Intricate details on the exact convergence rates of the tail index and the quantiles from a theoretical standpoint have been avoided.

However, the above formulation does give us an insight into the choice of τ_c. Note that we vary $\tau_c = 1 - \frac{k}{n}, k = 1, \ldots, n$ and choose k such that estimation of ξ stabilizes. Figure 3.6 displays the tail index estimate for t-distributed errors with by applying the GP-fit to errors and the proposed QR-GP method to the regression output y's. Note that the plot stabilizes at $\tau_c \approx 0$, which means that $k = 0$. Indeed the tail index estimation based on our approach for regression model closely agrees with the estimates if GP had been directly fit to the errors. Also, the estimate of the tail index agree with the true value.

Figure 3.6 displays the tail index estimate for t-distributed errors and burr-distributed errors by applying the GP-fit to errors and the proposed QR-GP method to the regression output y's. For t-distributed errors, for τ_c lying between 0.75-0.9, the plot stabilizes, thus one can choose $1 - k/n \approx 0.75$. For Burr-distributed errors, for τ_c lying between 0.4-0.6, the plot stabilizes, thus one can choose $1 - k/n \approx 0.4$. Note that the region of stabilization completely agree to the case if a GP model had been fit to the original errors. Therefore, this approach gives an alternative to GP modeling when both location and scale are functions of the covariate.

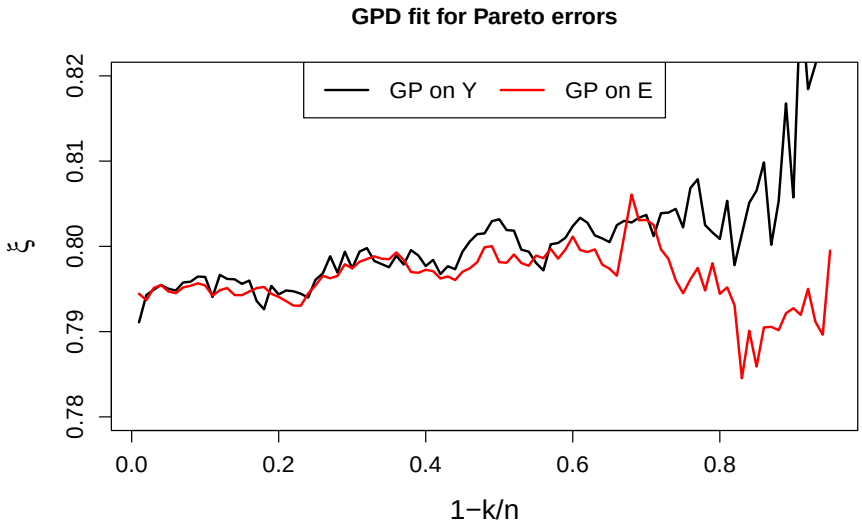

Figure 3.5: Tail index estimates with changing values of τ_c. Red line corresponds to the estimate of ξ when GP is fit to the error ϵ_i. Black line corresponds to the estimate when our proposed method is fit to y_i. True tail index is 0.8.

3.4 Simulation Study: Nonparametric Problem

In section 3.2.2, our focus has been on the cases $\epsilon \sim$GP(ξ) for $\xi \geq 0$. We have assessed the efficacy of our method for higher quantile levels $0.9 \leq \tau \leq 0.99$. In section 3.3, we have proposed a guideline to extend our method for distributions in the heavy tail domain of attraction. Also, we have seen that selecting thresholds of the form $\tau_c = 1 - \frac{k}{n}$ would lead to consistent estimation of the tail index ξ. The value of k is determined by the nature of the

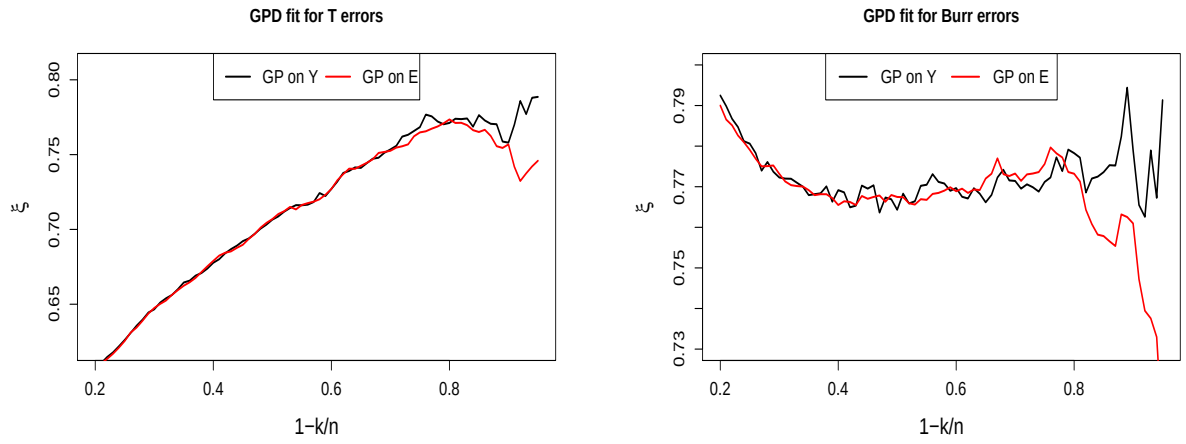

Figure 3.6: Tail index estimates with changing values of τ_c. Red line corresponds to the estimate of ξ when GP is fit to the error ϵ_i. Black line corresponds to the estimate when our proposed method is fit to y_i. True tail index is 0.8. *Left*: T-distributed errors. *Right:* Burr-distributed errors

underlying distribution. Let $\hat{\xi} = \hat{\xi}(k)$ be the estimated tail index for threshold $\tau_c = 1 - \frac{k}{n}$. We have demonstrated through numerical simulations that when we select k such that $\hat{\xi}(k)$ is stable, we predict ξ accurately. The estimation of the tail index is an important problem in the extreme value literature. There have been several alternative approaches to estimate the tail index, namely: method of moments, Hill estimator, Pickands estimator, etc.(Haan & Ferreira (2006)). Hill (1975) has proposed a nonparametric estimator for $\xi > 0$ using upper ordered statistics. Suppose that $Z_1, \ldots, Z_n$ are n iid random variables with distribution F, which is unknown. An equivalent condition for theorem 2.3.1 says that if $F \in \mathcal{D}(G_\xi)$ for some $\xi > 0$ if and only if

$$\lim_{u \to \infty} \frac{\bar{F}(uz)}{\bar{F}(u)} = z^{-\frac{1}{\xi}} \quad \xi > 0, z > 1 \tag{3.17}$$

i.e., as $u \to \infty$, the conditional *excess* $P(Z > uz | Z > u)$ behaves like Pareto. Meaning that for large u, the scaled *excess* Z/u conditioned on $Z > u$ is approximately Pareto with tail index ξ. Theorem 1.2.2 in Haan & Ferreira (2006) gives an equivalent form of of the

condition in 3.17

$$\lim_{u\to\infty} \frac{\int_u^\infty (1-F(z))\frac{dz}{z}}{1-F(u)} = \xi$$

Further simplification leads to

$$\lim_{u\to\infty} \frac{\int_u^\infty (\log z - \log u) dF(z)}{1-F(u)} = \xi$$

Let $Z_{(i)}$ be the i^{th} ordered observation with $Z_{(n)} = max(Z_1, \ldots, Z_n)$. Replacing u by $Z_{(n-k)}$ and F by the empirical distribution function, the Hill estimator for ξ based on k upper order statistics $Z_{(n-k+1)}, \ldots Z_{(n)}$ is

$$\hat{\xi}_H(k) = \frac{1}{k}\sum_{i=1}^{k} \log\Big[\frac{Z_{(n-i+1)}}{Z_{(n-k)}}\Big] \tag{3.18}$$

Let $k = k(n)$ such that $k(n) \to \infty, k(n)/n \to 0$, and $k(n+1)/k(n) \to 1$ as $n \to \infty$. Then, by the theorem 3.2.4 in Haan & Ferreira (2006), $\hat{\xi}_H(k) \to \xi$ in probability. More details about the derivation of the Hill estimator and its asymptotic properties can be found in Haan & Ferreira (2006), Hill (1975). Also, due to the nonparametric nature of the estimator, this method can be easily adapted as long as the underlying distribution is heavy tailed. But different choices of k leads to different values of $\hat{\xi}_H(k)$; hence it is crucial to choose k appropriately for consistent estimation of ξ. Wang et al. (2012) combined quantile regression and the Hill estimator to estimate conditional quantiles from a heavy tailed distribution. We outline the method of estimating higher quantiles by Wang et al. (2012) in the following section.

3.4.1 Review of an Existing Method for Extreme Quantile Estimation

In Wang et al. (2012) the authors have proposed methods for regression models with quantile functions $Q_{Y|x}(\tau) = \alpha(\tau) + x^T\beta(\tau)$ where $Y|x \sim \mathcal{D}(G_\xi)$ for some $\xi > 0$. Let $0 < \tau_c < 1$ be a fixed constant, close to 1. The authors have assumed the linear quantile function only for higher quantiles $\tau \in [\tau_c, 1]$.

$$Q_{Y|x}(\tau) = \alpha(\tau) + x\beta(\tau) \text{ for } \tau \in [\tau_c, 1]$$

$Q_{Y|x}(\tau)$ has no specific form for $\tau \in (0, \tau_c)$. Let $\tau_c < \tau_j = j/(n+1)$ be a sequence of quantiles where $j = n-k, \ldots, m$ and $m < n$ is an integer close to n. Suppose $[a]$ denotes the integer part of a. The authors assume $m = n - [n^\eta]$ with $\eta > 0$ being a small constant such that $n^\eta < k$. They also assume that as $n \to \infty$, $k = k(n) \to \infty$, and $k(n)/n \to 0$. Consider the sequence of quantile levels $\tau_c < \tau_{n-k} <, \cdots < \tau_m < 1$. The authors use quantile regression to obtain a reasonable quantile estimate at the intermediate quantile level τ_j for $j = n-k, \ldots, m$. Let

$$\begin{aligned} q_j =& \hat{\alpha}(\tau_j) + x^T \hat{\beta}(\tau_j) \\ \hat{\alpha}(\tau_j), \hat{\beta(\tau_j)} =& \underset{\alpha,\beta}{\operatorname{argmin}} \rho_{\tau_j} \sum_{i=1}^{n} (y_i - \alpha - x_i^T \beta) \end{aligned}$$

be the QR estimate at τ_j. Then one might consider q_j as an upper ordered statistic of a sample from $F_{Y|x}$. The authors have used q_j as a replacement for $Y_{(j)}$ in equation 3.18. Using $k + (n-m)$ many upper quantiles, they have proposed to estimate the tail index ξ as

$$\hat{\xi}_H(k,n) = \hat{\xi} = \frac{1}{k - [n^\eta]} \sum_{j=[n^\eta]}^{k} \log \frac{q_{n-j}}{q_{n-k}} \tag{3.19}$$

Let U be the tail quantile function, i.e.,

$$U(t) = \inf\{z : F(z) \geq 1 - 1/t\}, \quad t \in [1, \infty)$$

$U(t)$ is the $1 - 1/t$ quantile. When $F \in \mathcal{D}(G_\xi)$ and $\xi > 0$, we have for $z > 0$

$$\lim_{t \to \infty} \frac{U(tz)}{U(t)} = z^\xi$$

let $1 - 1/t = \tau_{n-k} = 1 - (k+1)/(n+1)$ and $1 - 1/tz = \tau_n$ so that as $n \to \infty$, $\tau_n \to 1$. Then for large n,

$$\frac{U(tz)}{U(t)} = \frac{Q(\tau_n)}{Q(\tau_{n-k})} \approx \Big(\frac{1 - \tau_{n-k}}{1 - \tau_n}\Big)^\xi \tag{3.20}$$

Combining 3.19 and 3.20, the conditional quantile is estimated as

$$\hat{Q}_{Y|x}(\tau_n) = \Big(\frac{1 - \tau_{n-k}}{1 - \tau_n}\Big)^{\hat{\xi}} q_{n-k} \tag{3.21}$$

Before we get into our simulation model, we would like to briefly discuss the optimal choice of tail sample fraction determined by $k = k(n)$. As noted by Wang et al. (2012), the choice of k is a very important problem in statistical applications of extreme value theory. By optimally choosing k we identify the tail sample fraction that may have the most relevant information on the tail behavior. Garrido & Lezaud (2013) explained the influence of k on $\hat{\xi}_H(k)$. For small values of k, the variance of the estimate is large, while increasing k leads to higher bias. The trade off between the bias and the variance comes at some intermediate value of k. We would like to see how the method proposed by Wang et al. (2012) (QR.Hill) behaves for different values of k. We will discuss more on the choices for k and η during our numerical investigation. Wang et al. (2012) has also proposed a variation of their method for $\beta(\tau) = \beta$, i.e., when the slopes are constant at higher quantiles. That is a special case of our simulation model in 3.10 with $\gamma = 0$. We restrict ourselves to the general case, i.e., $\gamma \neq 0$. In section 3.4.2, we compare our proposed QR.GP method with QR.Hill for heavy tailed data at higher quantiles. We have outlined the algorithm for QR.Hill in algorithm 4. We use algorithm 3 to implement QR.GP.

3.4.2 Simulation Study

We simulate from the following model

$$\begin{aligned} Y &= \phi + X^T\rho + (\theta + X^T\gamma)\epsilon \\ \epsilon &\sim F \in \mathcal{D}(G_\xi), \quad \xi > 0 \end{aligned} \tag{3.22}$$

We have the same model as in 3.10. We now explore different distributions in the domain of attraction of the Fréchet distribution. The conditional τ quantile of 3.22 is given by

$$Q_{Y|x}(\tau) = \phi + x^T\rho + (\theta + x^T\gamma)Q_\epsilon(\tau) \tag{3.23}$$

For all $0 < \tau < 1$. Hence the true quantile function 3.23 is linear in x. Moreover, both the slope $\alpha(\tau) = \phi + \theta Q_\epsilon(\tau)$ and the intercept $\beta(\tau) = \rho + \gamma Q_\epsilon(\tau)$ vary with τ. Thus, we satisfy the model assumption 2.10 in Wang et al. (2012) with any $\tau \in (0, 1)$. We simulate

Algorithm 4: QR.Hill

start: Set $0 < \tau_c \approx 1$; fix x, $\delta > 0$;

I. Set $k = n - [n\tau_c]$ and $m = n - [n^\eta]$. Select k upper quantiles $\tau_j = \frac{j}{n+1}$ for $j = n-k, \ldots, m$. Get

$$q_j = \hat{\mathrm{QR}}_{Y|x}(\tau_j) = \hat{\alpha}(\tau_j) + x^T \hat{\beta}(\tau_j)$$

II. Estimate tail-index ξ as

$$\hat{\xi} = \frac{1}{k - [n^\eta]} \sum_{j=[n^\eta]}^{k} \log \frac{q_{n-j}}{q_{n-k}}$$

III. Estimate τ quantile as

$$\hat{Q}_{Y|x}(\tau) = q_{n-k}\Big(\frac{1-\tau_{n-k}}{1-\tau}\Big)^{\hat{\xi}}$$

IV. Set $\tau_c = \tau_c - \delta$ such that $0 < \tau_c < 1$ and repeat steps I-IV

end: Stop when $\tau_c \approx 0.01$

$(x_i, y_i), i = 1, \ldots, n$ iid observations from 3.22

$$y_{ij} = 1 + x_{1i} + x_{2i} + (1 + 0.9x_{1i})e_i$$

Here $\phi = \theta = 1, \rho = (1, 1)$ and $\gamma = (0.9, 0)$. We simulate $x_{ij} \sim U(-1, 1)$ for $j = 1, 2$ and $e_i \sim F$. We choose three example distributions F in the Fréchet domain: Pareto, absolute t and Burr. See Table B.1 in the appendix for a list of well known distributions in the heavy-tail domain. The true τ quantile at x is

$$Q_{Y|x}(\tau) = 1 + x_1 + x_2 + (1 + 0.9x_1)(Q_\epsilon(\tau)) \tag{3.24}$$

Pareto: We start our analysis with standard Pareto distribution Pa(α). Note that Pareto distribution is a special case of standard GP distribution. If $Y \sim \mathrm{Pa}(\alpha)$ and $Y^* = \alpha(Y - 1)$ then $Y^* \sim \mathrm{GP}(\mu = 0, \sigma = 1, \xi = \frac{1}{\alpha})$. Pa($\alpha$) has tail index $\xi = \frac{1}{\alpha}$. Tails of GP and Pareto behave very similarly. In section 3.2.2, we have looked into quantiles up to 0.99 from GP with $\xi = 0.5$ and 0.8.

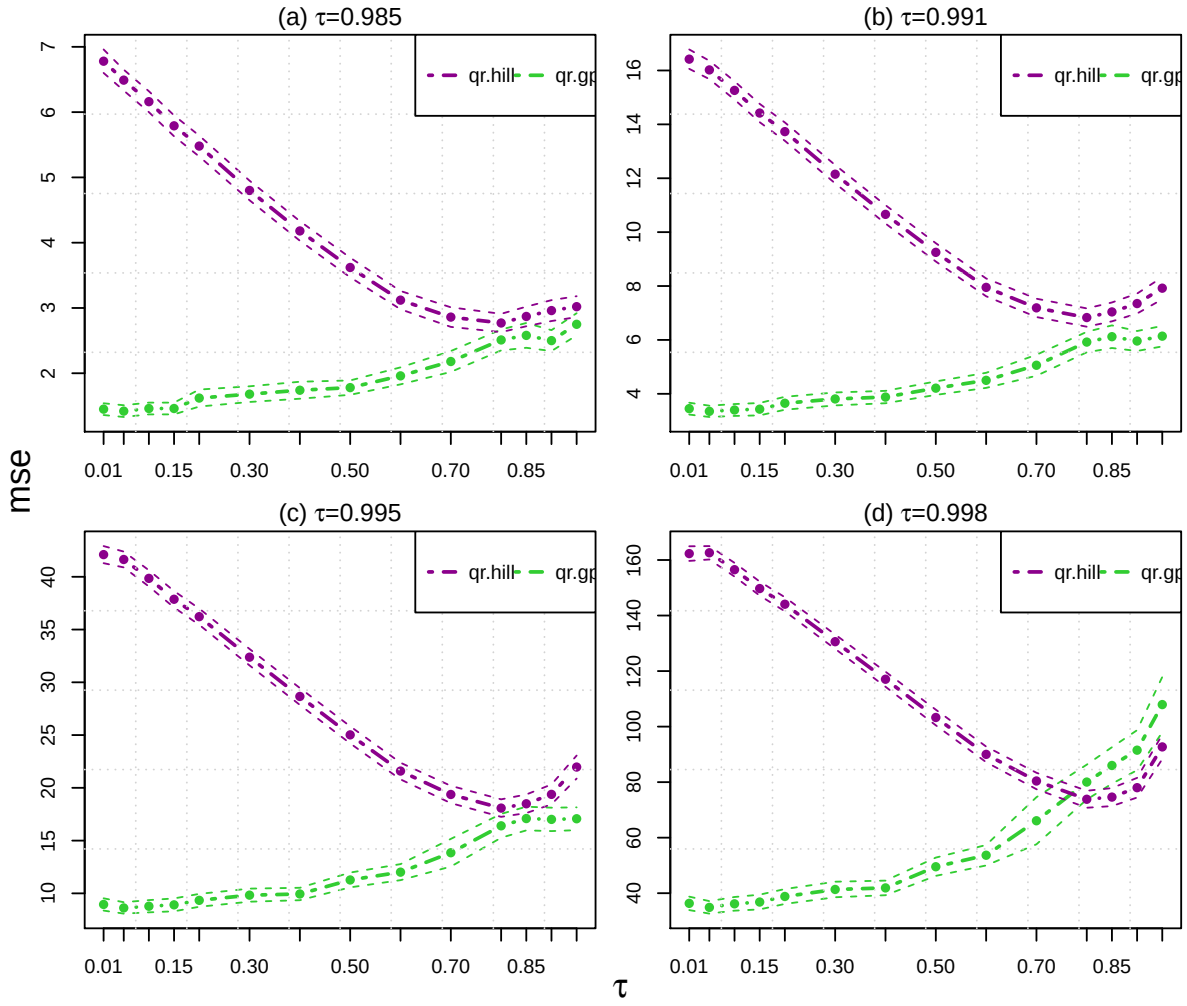

Figure 3.7: mse±se for Pareto(0.5) at $x = (0.5, 0.5)$ and quantile levels τ. The sample size is $n = 1000$ and $0 < \tau_c < 1$

Now we will focus on the four extreme quantile levels $\tau \in (0.985, 0.991, 0.995, 0.998)$. We compare our method QR.GP with QR.Hill for different values of tuning parameter $\tau_c = 1 - \frac{k}{n}$ for $k = 1, \ldots, n$. We choose from a grid of values $\tau_c \in (0.01, 0.05, \ldots, 0.95)$. When we look into extreme quantiles, the data is even more sparse. So we choose sample sizes $n \geq 1000$ in our simulations. Our monte carlo sample size is $K = 500$ for all of the cases. we are going to compare QR.GP (algorithm 3) with QR.Hill (algorithm 4) for different choices of τ_c. We choose $m = n - [n^{0.1}]$ for QR.Hill as per the recommendation by Wang et al. (2012).

Figure 3.7 summarises mse of the quantile estimate at $x = (0.5, 0.5)$ for sample size

$n = 1000$ and tail index $\xi = 0.5$. Clearly, the optimal threshold for QR.GP is $\tau_c = 0.01$. Our result is consistent with our findings in section 3.2.2. We can see that we outperform QR.Hill regardless of the choice of threshold at $\tau = 0.985, 0.991$ and 0.995 quantiles. For QR.Hill, optimal τ_c is between 0.8 and 0.9. Now we increase our sample size to 2000 and compare QR.Hill with QR.GP at $\tau = 0.998$. In figure 3.8 (b) we can see that QR.GP has lower mse than QR.Hill at all threshold values. With increasing sample size, we can see that QR.GP can outperform QR.Hill at quantile level $\tau = 0.998$.

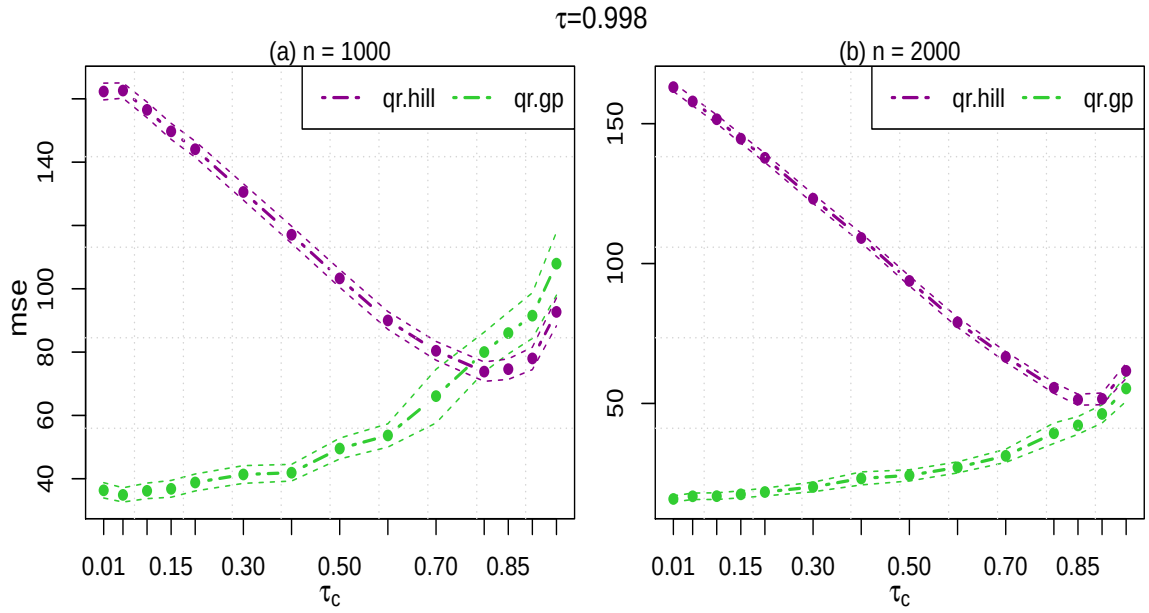

Figure 3.8: mse±se for Pareto(0.5) at $x = (0.5, 0.5)$ at quantile $\tau = 0.998$ and sample sizes (a) $n = 1000$, (b) $n = 2000$

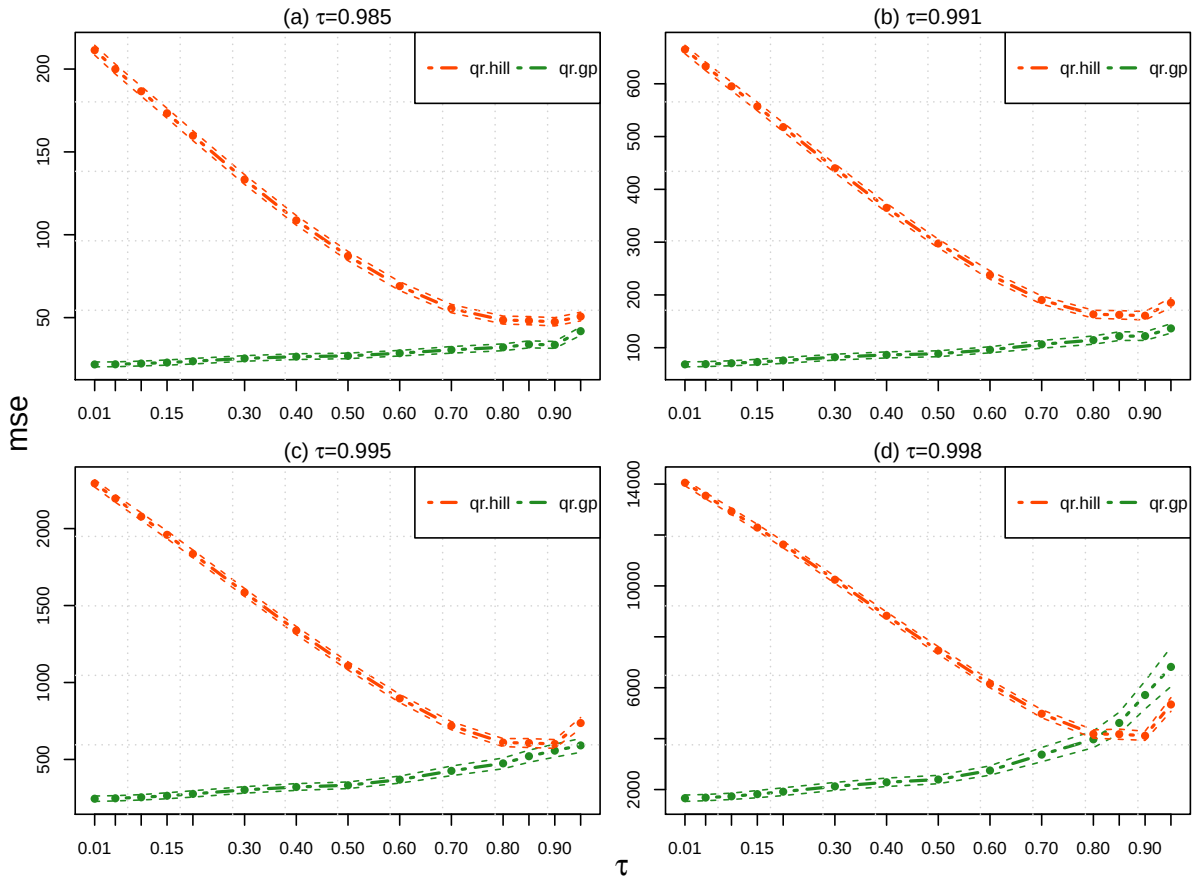

Figure 3.9: mse±se for Pareto(0.8) at $x = (0.5, 0.5)$ and quantile levels τ. The sample size is $n = 2000$ and $0 < \tau_c < 1$

We now look into the case when $\xi = 0.8$. We choose the sample size 2000 as higher values of ξ is associated with even more data sparsity. From figure 3.9 (a), (b) and (c) we can clearly see that QR.GP is uniformly better that QR.Hill at all but quantile $\tau = 0.998$.In fact, we can see in figure 3.11 that we achieve uniform dominance at quantile$\tau = 0.998$ for $n = 5000$. For $\tau = 0.985$, QR.GP is still uniformly better than QR.Hill even when we increase our sample size form 2000 to 5000. We can see this in figure 3.10.

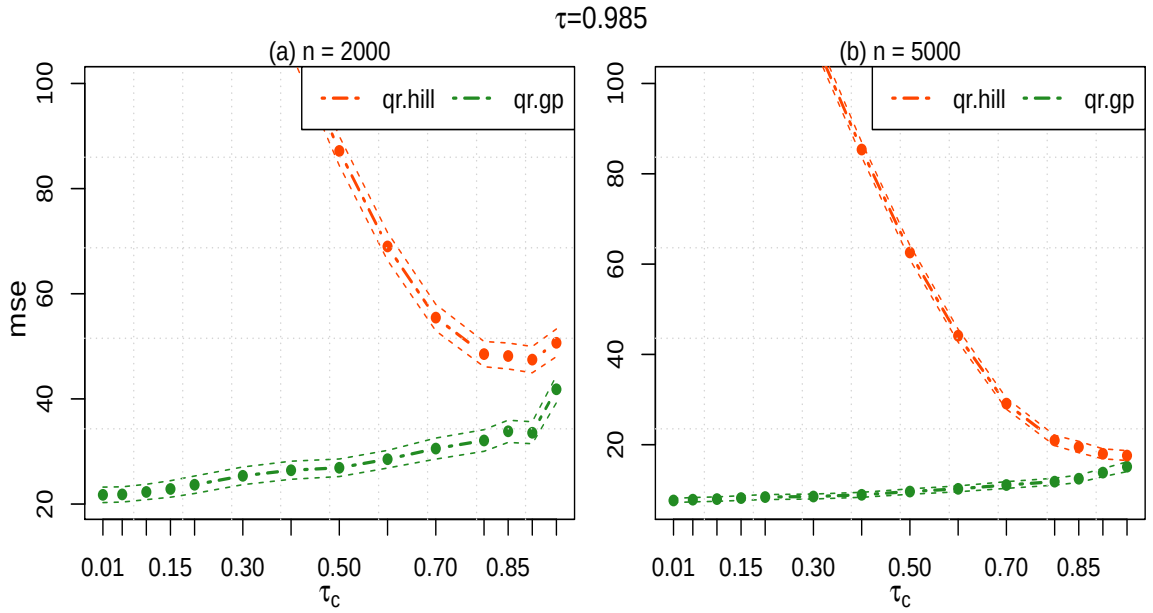

Figure 3.10: mse±se for Pareto(0.8) at $x = (0.5, 0.5)$ at quantile $\tau = 0.985$ and sample sizes (a) $n = 2000$, (b) $n = 5000$

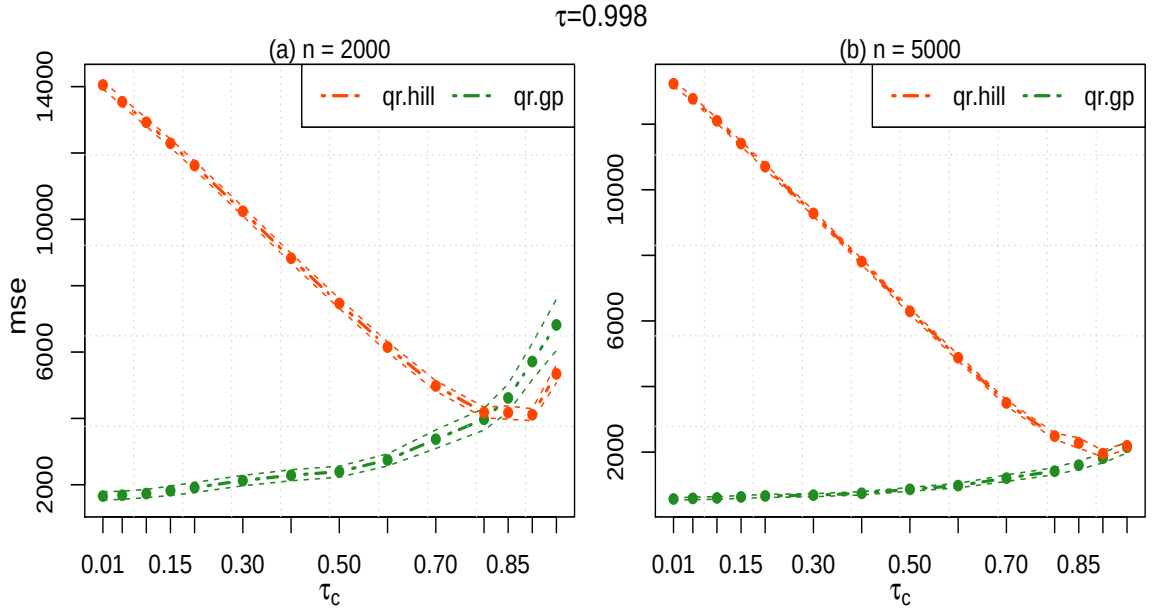

Figure 3.11: mse±se for Pareto(0.8) at $x = (0.5, 0.5)$ at quantile $\tau = 0.998$ and sample sizes (a) $n = 2000$, (b) $n = 5000$

Absolute t: Next we simulate ϵ from the absolute t distribution. The tail index for t distribution is $\xi = \frac{1}{\nu}$ where ν is the degrees of freedom. We choose $\xi = 0.5$ and 0.8. Note that now the underlying distribution is asymptotically GP with tail index ξ. Here we explore the optimal choice for τ_c in the case of absolute t distribution. From our numerical investigations, we have seen that τ_c around 0.75 works best with QR.GP. For QR.Hill, the optimal is 0.7. In figure 3.12 we demonstrate their performance at their corresponding optimal across lower quantiles $0.9 \leq \tau \leq 0.99$ for sample sizes $n = 1000$ and 2000.

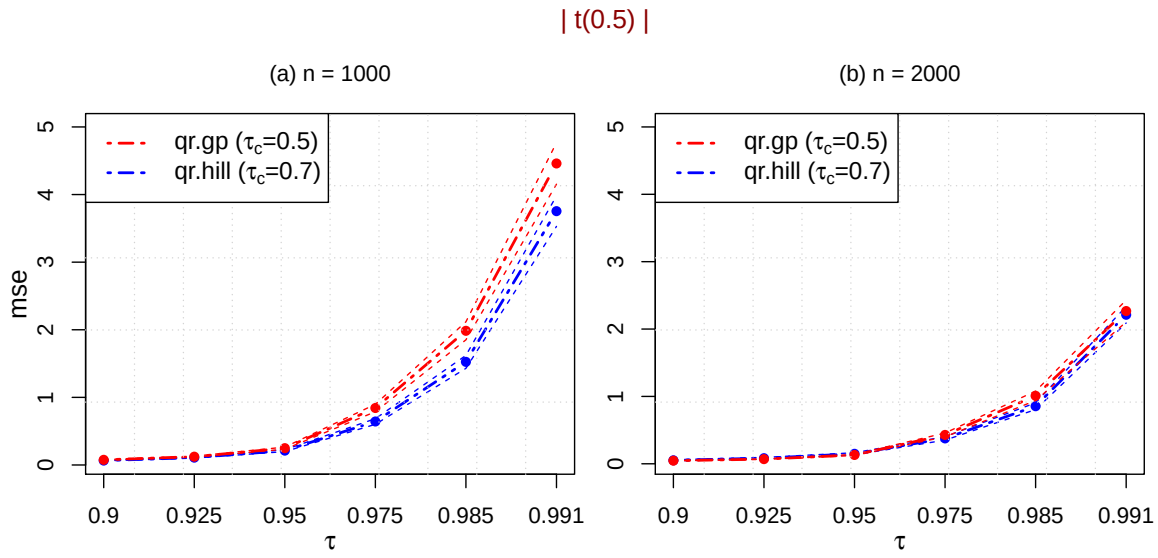

Figure 3.12: mse±se for absolute t with $\xi = 0.5$ at $x = (0.5, 0.5)$ and quantiles $0.9 \leq \tau \leq 0.991$. Sample sizes (a) $n = 1000$, (b) $n = 2000$

We can see in figure 3.12 (a) that both QR.GP an QR.Hill are consistent in the sense that they have low mse when $\tau \leq 0.95$. For $n = 1000$, QR.Hill outperforms QR.GP at $\tau > 0.95$. When we increase our sample size to $n = 2000$, QR.Hill cannot outperform QR.GP at any quantile level between 0.9 and 0.991. Next we move into higher quantiles. Figure 3.13 illustrates QR.GP and QR.Hill across different choices of τ_c for $\tau = 0.985, 0.991$.

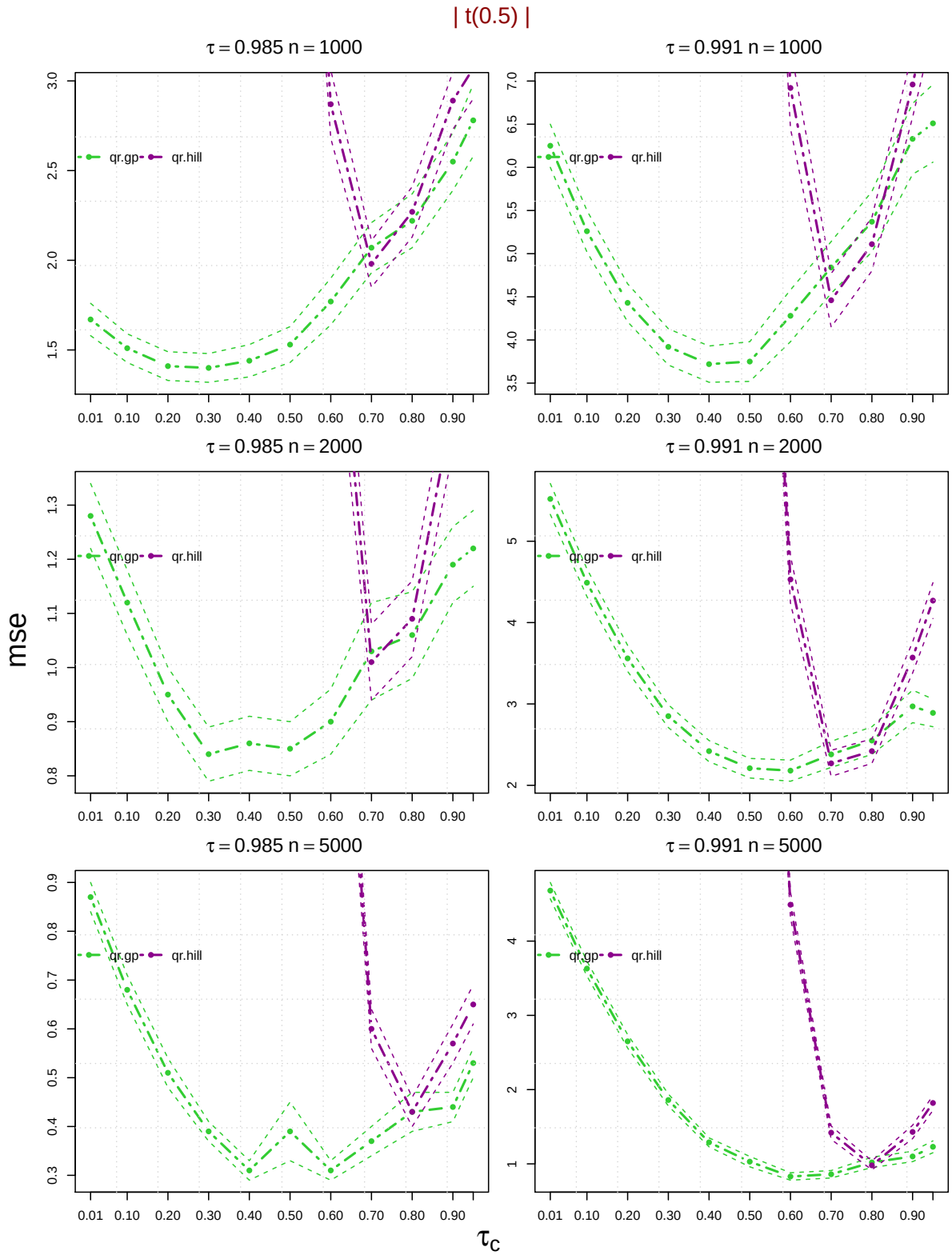

Figure 3.13: mse±se for absolute t with $\xi = 0.5$ at $x = (0.5, 0.5)$. Quantile values $\tau = 0.985, 0.991$ and sample sizes $n = 1000, 2000$ and 5000

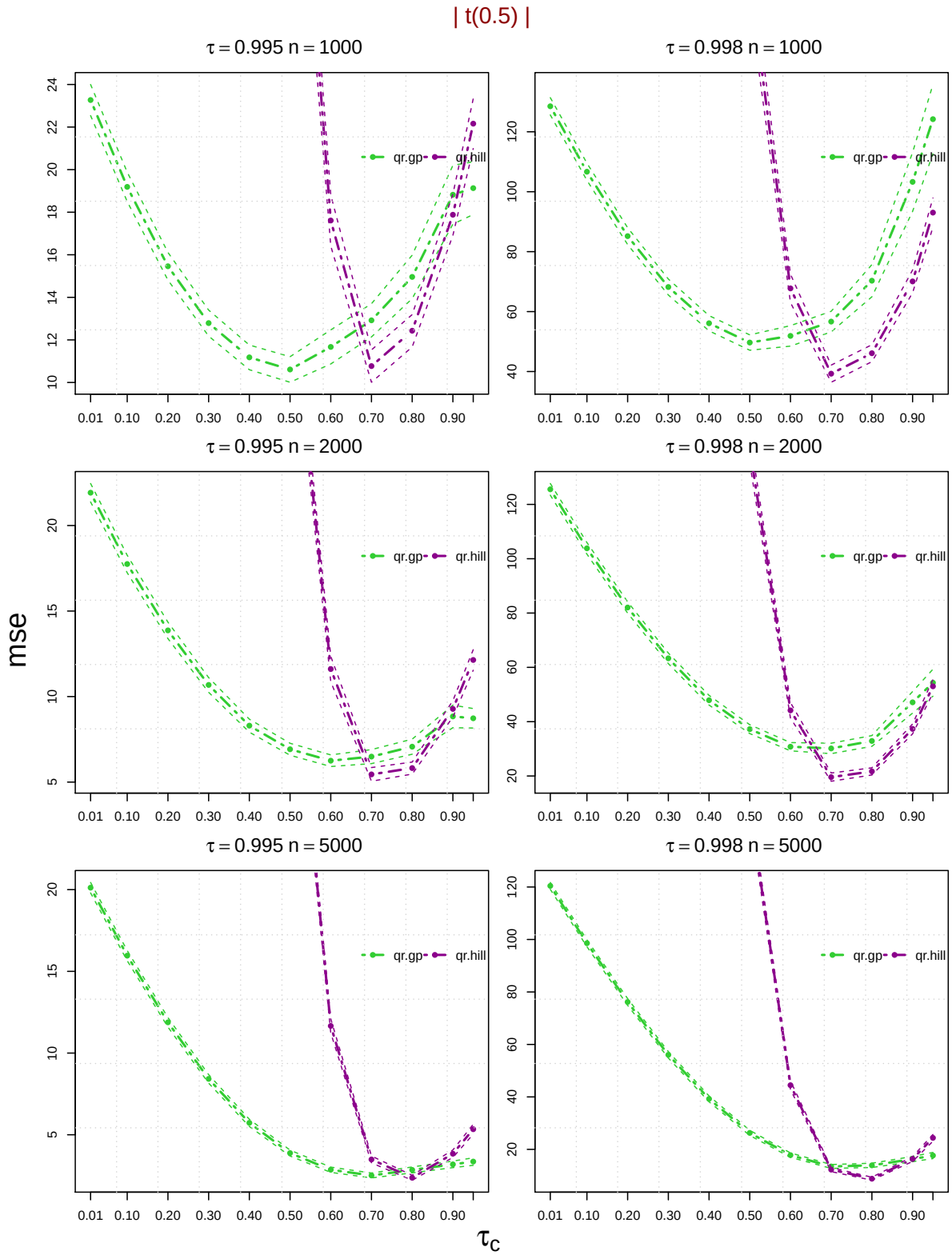

Figure 3.14: mse±se for absolute t with $\xi = 0.5$ at $x = (0.5, 0.5)$. Quantile values $\tau = 0.995, 0.998$ and sample sizes $n = 1000, 2000$ and 5000

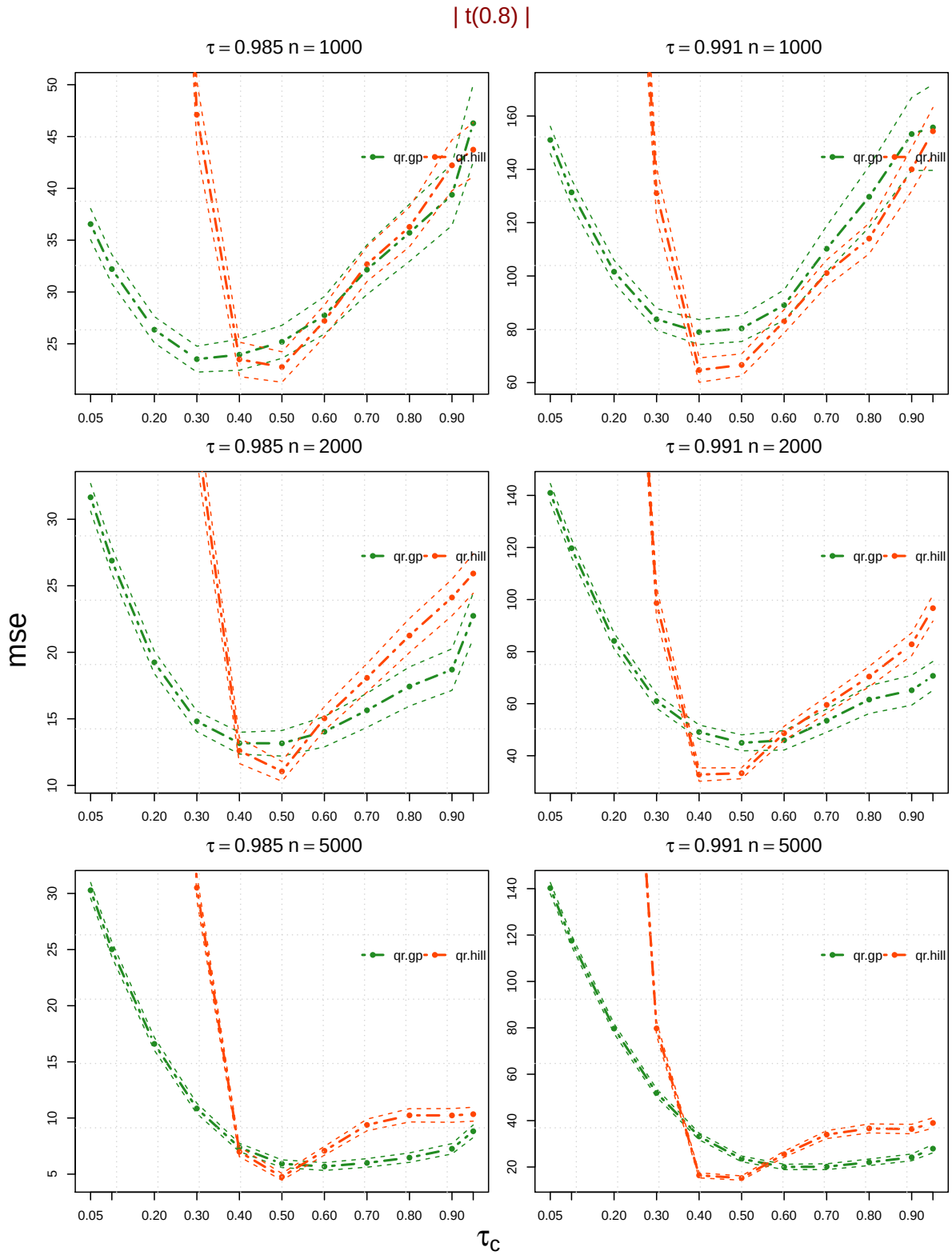

Figure 3.15: mse±se for absolute t with $\xi = 0.8$ at $x = (0.5, 0.5)$. Quantile values $\tau = 0.985, 0.991$ and sample sizes $n = 1000, 2000$ and 5000

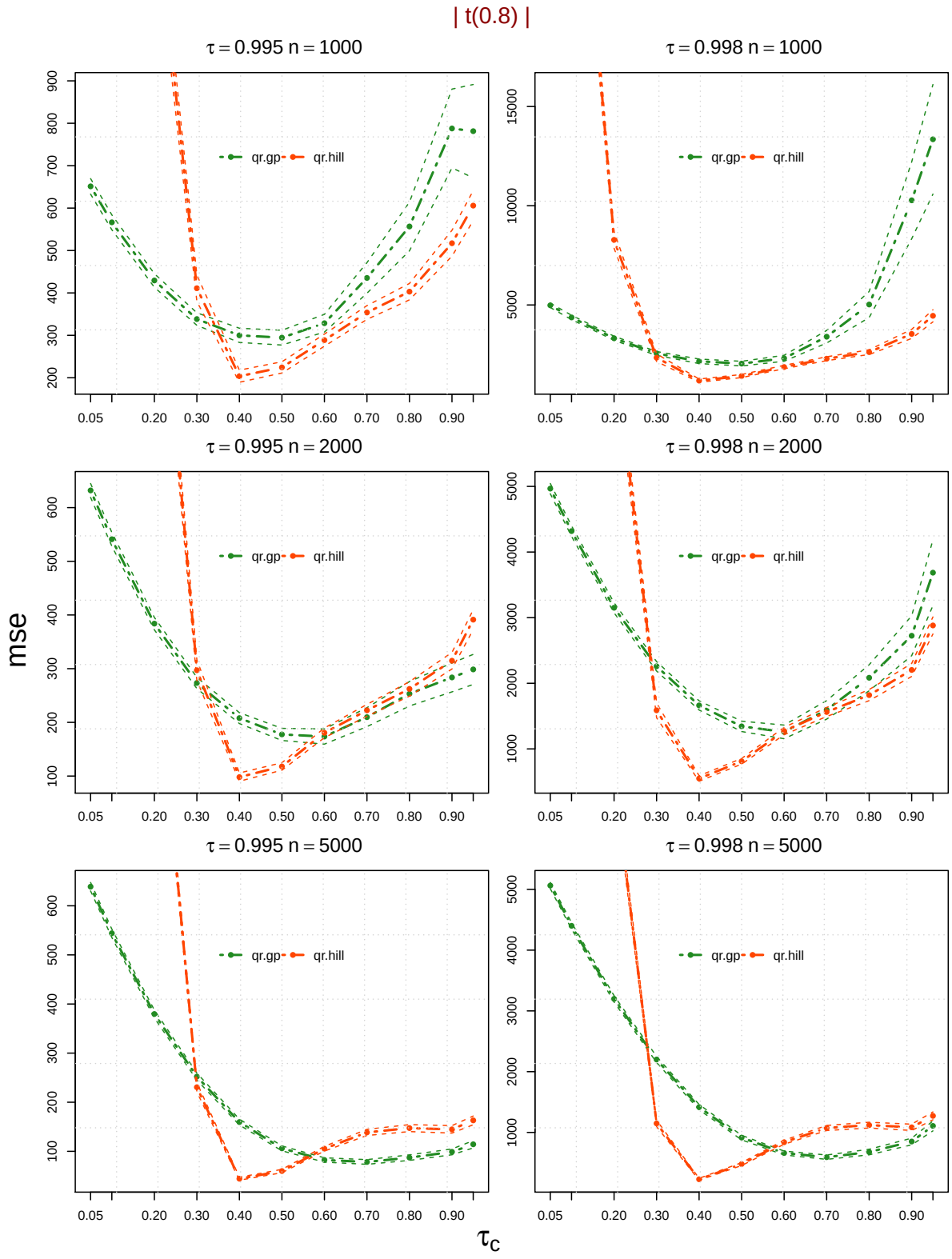

Figure 3.16: mse±se for absolute t with $\xi = 0.8$ at $x = (0.5, 0.5)$. Quantile values $\tau = 0.995, 0.998$ and sample sizes $n = 1000, 2000$ and 5000

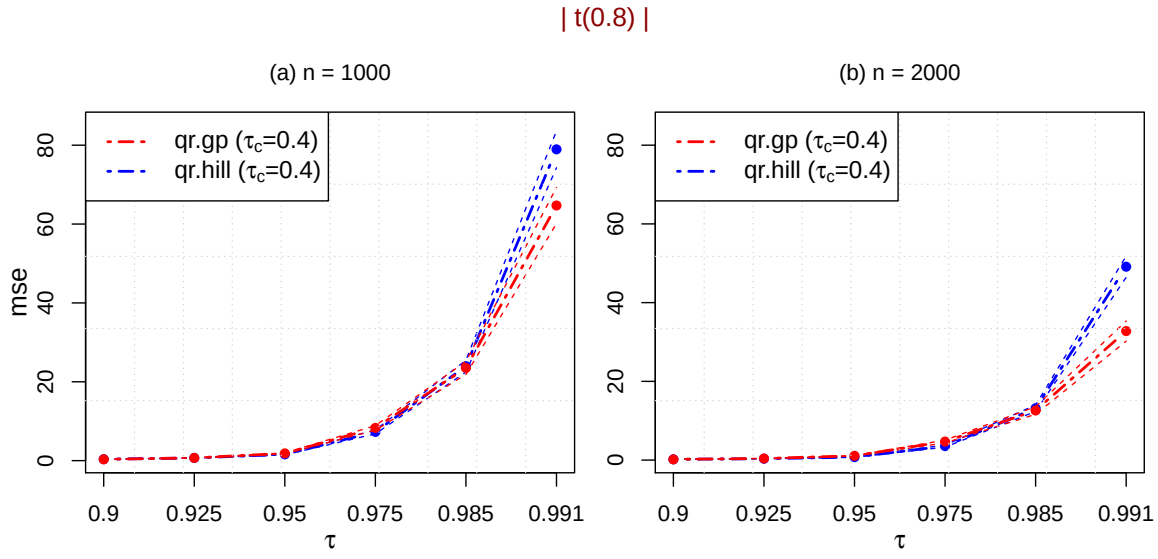

Figure 3.17: mse±se for absolute t with $\xi = 0.8$ at $x = (0.5, 0.5)$ and quantiles $0.9 \leq \tau \leq 0.991$. Sample sizes (a) $n = 2000$, (b) $n = 5000$

We achieve uniform dominance at quantile 0.985 with sample size $n = 2000$. We do so for quantile 0.991 with $n = 5000$. In figure 3.14 we can see that for $\tau = 0.995$ and 0.998, QR.GP is comparable to QR.Hill. An important observation we have is that QR.GP is more robust to the choice of τ_c than QR.Hill. Little deviation from optimal τ_c results to huge fluctuation in the mse of QR.Hill. In comparison, our mse is bounded and converges to 0 as we increase n for all $\tau_c \in (0, 1)$. Next we want to see how changing values of ξ can affect our predictor. We increase ξ to 0.8. From our numerical investigations, we saw that $\tau_c = 0.4$ is optimal for both QR.GP and QR.Hill. Figure 3.17 shows mse at optimal values of $\tau_c = 0.4$ at quantiles $0.9 < \tau < 0.99$. QR.GP does a better job at higher values of τ. Note that for QR.Hill, the optimal τ_c is 0.7 when $\xi = 0.5$ while for $\xi = 0.8$, optimal $\tau_c = 0.4$. The optimal threshold τ_c is between 0.4 and 0.5 for QR.GP. We plot mse for different choices of τ_c and quantiles $\tau = 0.985$ and 0.991 in figure 3.15. We can see that any τ_c between 0.4 and 0.6 works well for QR.GP. Figure 3.16 shows large sample behavior of the quantile estimates at quantiles $\tau = 0.995$ and 0.998. With increasing sample size, QR.GP can approximate the tail well for thresholds $0.4 \leq \tau_c < 1$.

Burr: We simulate from the Burr distribution with tail index 0.5. The results for quantiles $\tau = 0.985$ and 0.991 are summarised in figure 3.18. Optimal τ_c for QR.GP is close to 0.5. Any $\tau_c > 0.5$ works well when the sample size is 5000. For QR.Hill, the optimal τ_c is around 0.6. In case of $|t(0.5)|$ we have seen in figure 3.13 that with increasing sample size, QR.GP can approximate the 0.985 quantile for all $\tau_c \in (0, 1)$. We notice the same characteristic here. Moreover, with increasing sample size QR.GP and QR.Hill both are comparable at their optimal. For quantiles 0.991 and 0.995, we can see from figure 3.19 that QR.GP is comparable with QR.Hill at their optimal. Now we simulate from Burr with tail index 0.8. From figure 3.20 we can see that the optimal τ_c is 0.5 for QR.GP whereas optimal τ_c for QR.Hill is 0.7. For higher quantiles $\tau = 0.995, 0.998$, we can see in figure 3.21 that both QR.GP and QR.Hill are comparable at their corresponding optimals.

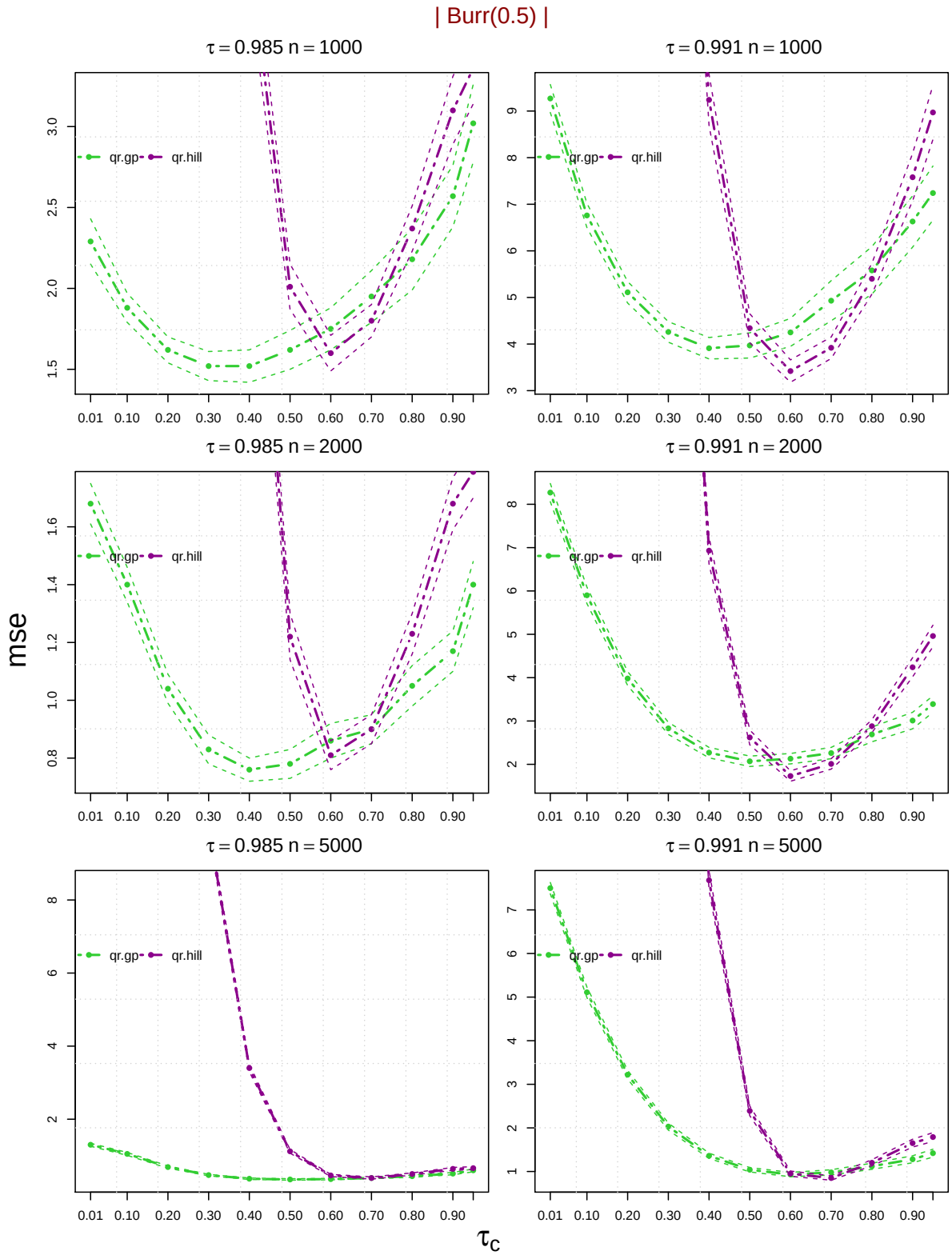

Figure 3.18: mse±se for Burr with $\xi = 0.5$ at $x = (0.5, 0.5)$. Quantile values $\tau = 0.985, 0.991$ and sample sizes $n = 1000, 2000$ and 5000

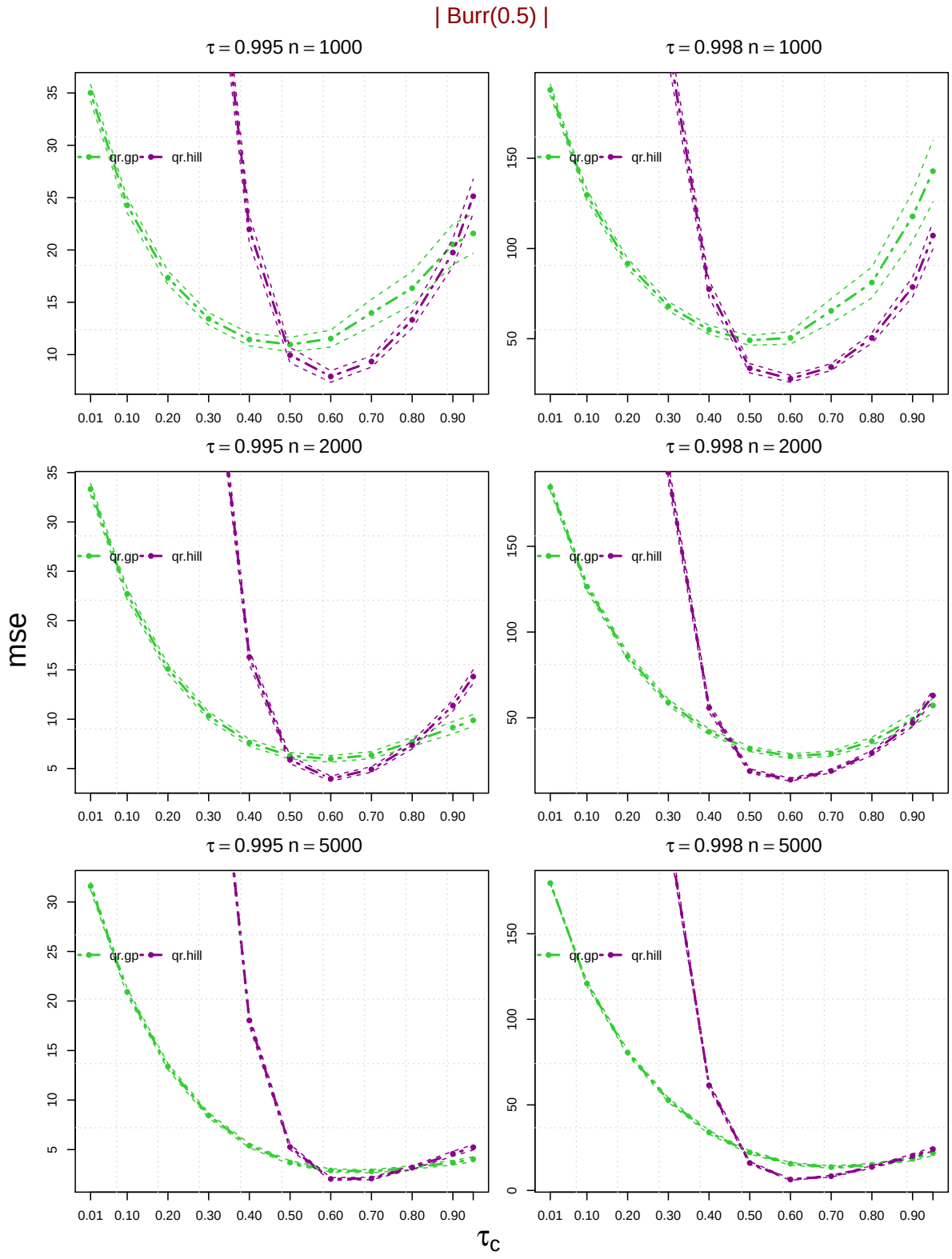

Figure 3.19: mse±se for Burr with $\xi = 0.5$ at $x = (0.5, 0.5)$. Quantile values $\tau = 0.995, 0.998$ and sample sizes $n = 1000, 2000$ and 5000

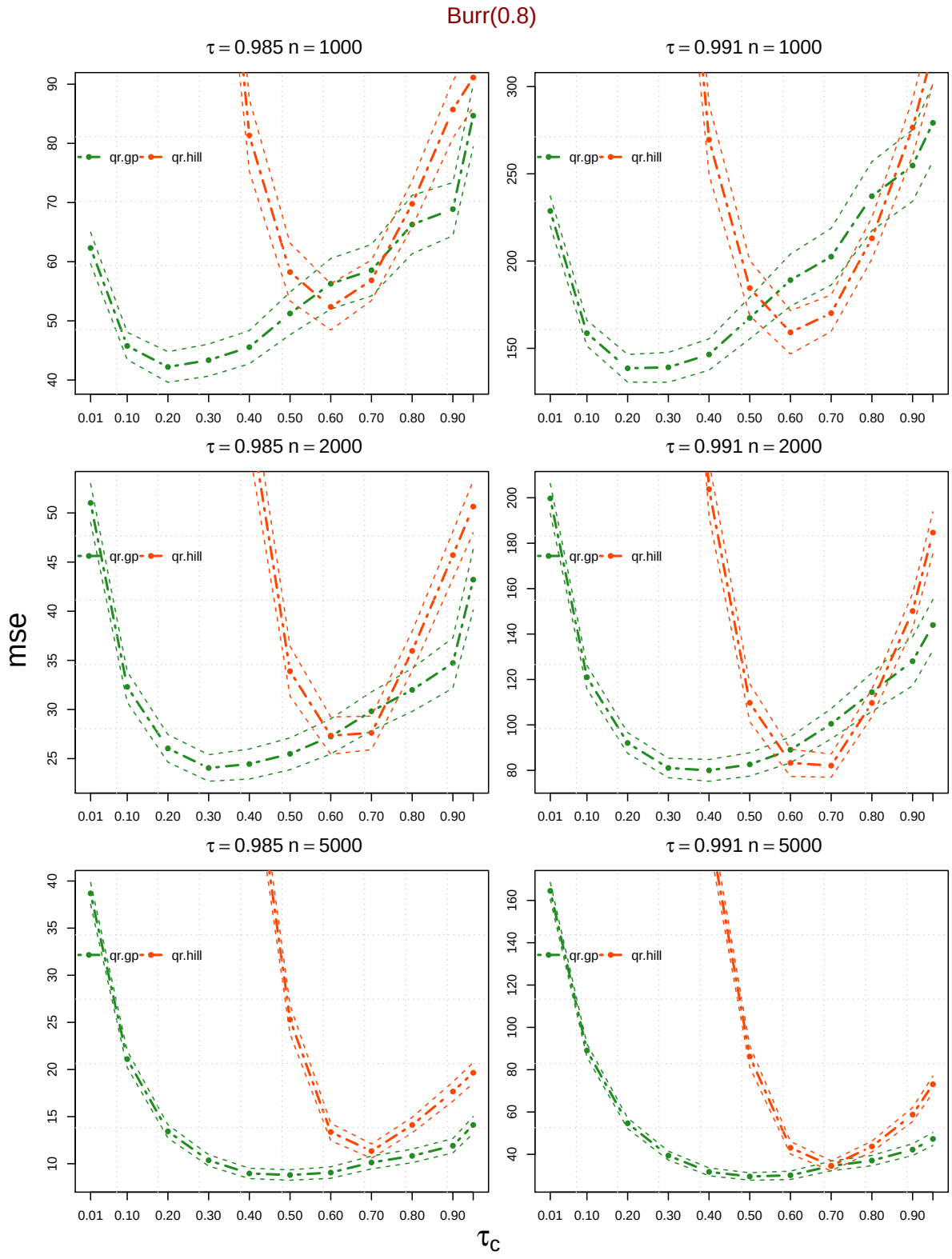

Figure 3.20: mse±se for Burr with $\xi = 0.8$ at $x = (0.5, 0.5)$. Quantile values $\tau = 0.985, 0.991$ and sample sizes $n = 1000, 2000$ and 5000

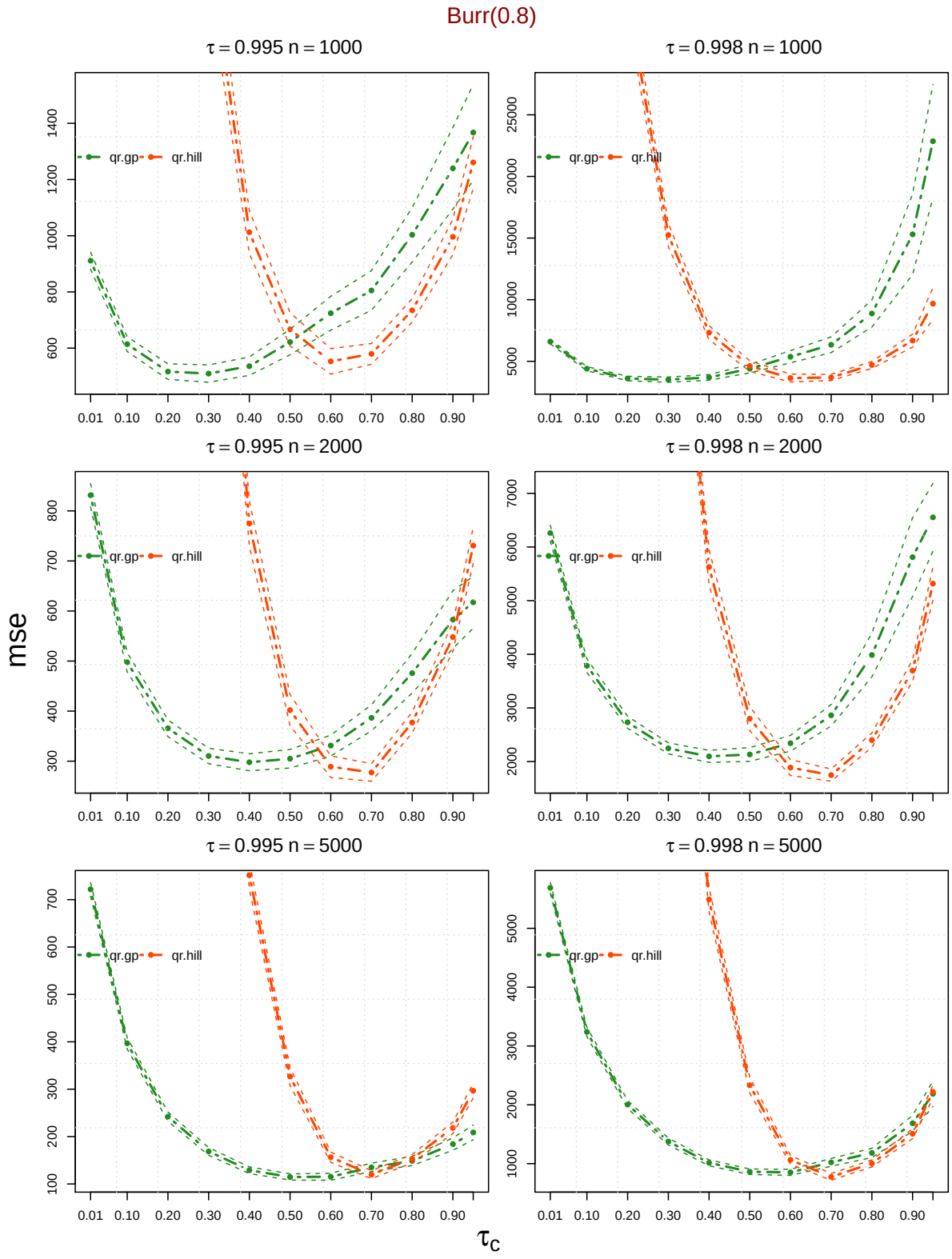

Figure 3.21: mse±se for Burr with $\xi = 0.8$ at $x = (0.5, 0.5)$. Quantile values $\tau = 0.995, 0.998$ and sample sizes $n = 1000, 2000$ and 5000

3.5 Application to precipitation downscaling

According to the National Oceanic and Atmospheric Administration (NOAA), "Statistical downscaling encompasses the use of various statistics-based techniques to determine relationships between large-scale climate patterns resolved by global climate models and observed local climate responses." Statistical downscaling techniques aim to quantify the relationship between global scale climate model outputs and local scale climate variables. Statistical downscaling methods are useful for making predictions about local climate variables like precipitation, temperature, wind speed and direction, air quality, etc.. In this section, we want to apply our proposed method to predict extreme quantiles of precipitation. Here we are not looking for the best set of predictors form the global scale model outputs that explains regional precipitation. Instead, we use relevant variables like sea level pressure, specific humidity, etc. which contain necessary information and use them as our predictors. Using a pre-determined set of covariates, we assess the predictive performance of three different methods: QR, QR.GP, and QR.Hill in the context of the estimation of high conditional quantiles.

Cannon (2011) have used historic precipitation data obtained from the Vancouver International Airport (WMO station 71892, 49°12′N, 123°10.8′W) weather station to demonstrate an application of their proposed method: quantile regression with neural network. This dataset is publicly available with the R package *'qrnn'*. The dataset consists of daily precipitation records of 30 years, from 1971-2000. From figure 3.22, we can see that the distribution of the precipitation is unimodal and skewed. Moreover, out of the total 10958 days, there are 5978 dry days. Precipitation for these days is recorded as 0mm. We did not encounter any missing values in this dataset. The maximum precipitation is 89.4mm while less than 1% of the observations are above 31mm.

We use three global climate model outputs: daily sea-level pressure, 500-hPa geopotential height, and 700-hPa specific humidity as covariates. The predictor variables are obtained from the NCEP/NCAR reanalysis project by Kalnay et al. (1996) from the nearest grid

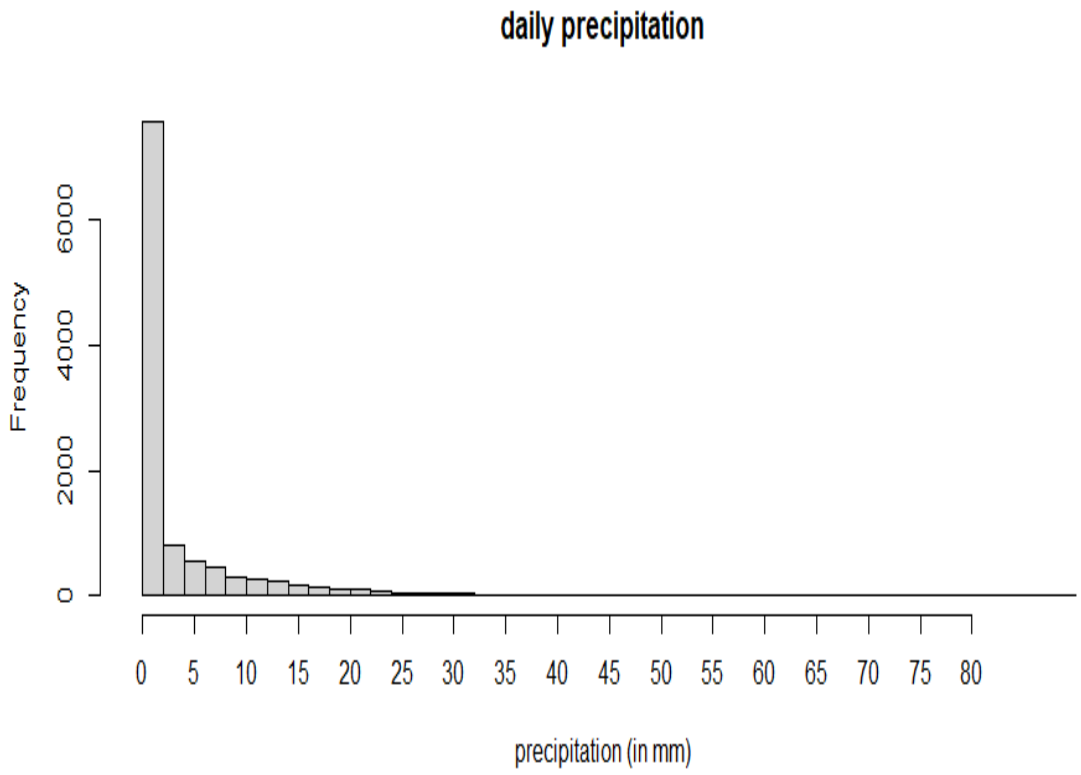

Figure 3.22: Daily precipitation at the Vancouver International Airport, 1971-2000

point: 50°N, 122.5°W. We also include the sine and cosine of the day of the year as our predictors. We standardize each of these variables before the analysis.

In real data, we do not know the true quantiles. In order to compare the predictive performance of a quantile estimation method, we need a reference quantile estimate. Following the suggestion by Cannon (2011), we use monthly sample quantiles as our reference quantile. Given a quantile level τ, let $Q_{Y(m,i)}(\tau)$ be the monthly sample quantile for the day i of the month m. Then $Q_{Y(m,i)}(\tau)$ is the sample τ quantile based on the observations from month m, and we assume the same value for all the days of month m.

Here our aim is to compare the performance of three different quantile estimators: QR, QR.GP, and QR.Hill. For all three methods, we use the same set of covariates as described earlier in this section. We can directly fit QR as there is no threshold selection involved. For

both QR.Hill and QR.GP, we need to choose a threshold.

Using a grid of threshold values τ_c in $(0.01, \ldots, 0.8)$, we fit QR.GP to the data following algorithm 3. In figure 3.23, we have plotted different values of the estimated tail index against the thresholds. The tail index estimate stabilizes around $\tau_c \approx 0.7$. We choose $\tau_c = 0.68$ (red dotted line in figure 3.23) for which $\hat{\xi} = 0.18$. Next, we fit QR.Hill to the data using four different threshold values $\tau_c = 0.8, 0.85, 0.9$, and 0.95. To assess the goodness of fit, we look into mean square prediction error at τ quantile

$$MSPE(\tau) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Q}_{Y|x_i}(\tau) - Q_{Y(m,i)}(\tau)\right)^2$$

where $\hat{Q}_{Y|x_i}(\tau)$ is the predicted conditional τ quantile given x_i and $Q_{Y(m,i)}(\tau)$ is the monthly sample quantile for day i. Note that for both QR.GP and QR.Hill, we can only predict

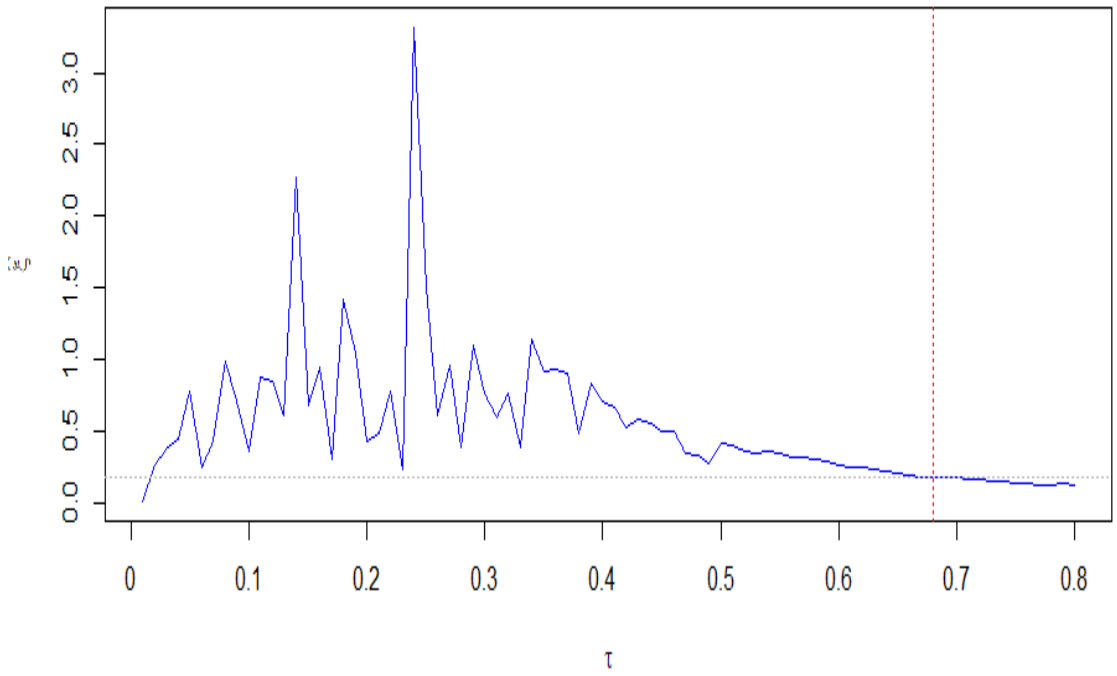

Figure 3.23: Estimated tail index ($\hat{\xi}$) at different choices of $\tau_c \in (0, 1)$. Method of estimation is QR.GP

quantiles τ above the threshold τ_c. We choose $\tau \in \{0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.985, 0.99, 0.995\}$. We use only quantile levels above 0.95 for the case $\tau_c = 0.95$ (QR.Hill). Figure 3.24 shows the mean square prediction error for different quantile estimates. We can see that the mspe for QR.GP is lower than QR at all quantile levels $\tau > 0.9$. Although the performance of QR and QR.GP are not very different. Note that the tail index estimated by QR.GP is $\hat{\xi} = 0.18$ which indicates that the data is not very heavy tailed. We have seen from our numerical investigations that QR performs poorly when ξ is 0.5 or above. But for $\xi = 0$ (exponential), QR is comparable with QR.GP. Moreover, we have a decent sample size $n = 10958$ which gives QR some benefit. Hence we are not very surprised that the performance of QR closely follows our method. On the other hand, predictive performance of QR.Hill is very sensitive to the choice of τ_c. Higher thresholds tend to work better at higher quantile levels. For QR.Hill, we cannot agree upon a universal threshold value that gives us the lowest mspe at all quantile levels above 0.9.

Implementation of the QR.Hill method in practice is challenging. Firstly, For a given value of the predictor x, Wang et al. (2012) proposed to choose τ_c such that $\hat{\xi}|x$ is stable. This makes the choice of tail index dependent on the covariate. Thus, we cannot choose a optimal τ_c for all x. Meanwhile, with QR.GP we can still select a covariate adaptive threshold for a fixed τ_c as we use $\text{QR}(\tau_c|x)$ to estimate the threshold. Secondly, due to the nature of QR.Hill, one has to compute QR estimates for $n(1 - \tau_c)$ many quantiles. This is computationally challenging, particularly when τ_c is low or the sample size n is high. For this data, with the sample size of $n = 10958$ implementing QR.Hill is very time consuming. Compared to that we compute QR only once for each choice of τ_c.

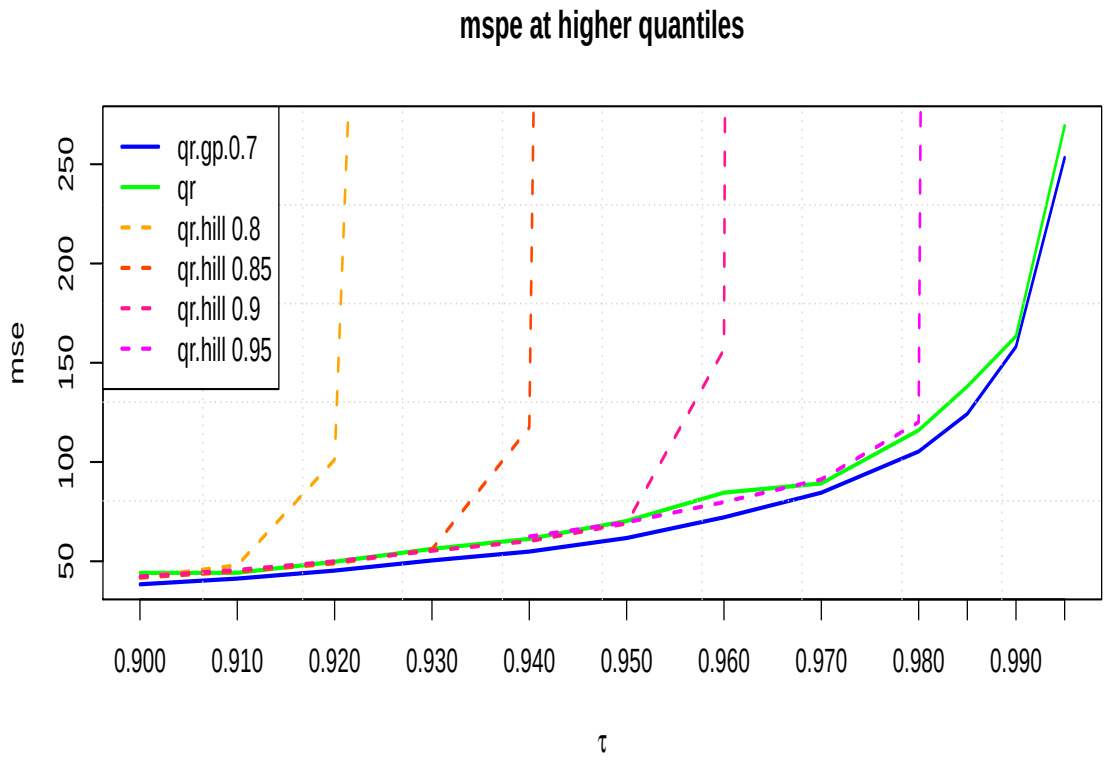

Figure 3.24: Mean square prediction error (MSPE) of different quantile estimates at quantile levels $0.9 < \tau < 0.995$. Thresholds: 0.7 for QR.GP(blue); 0.8, 0.85, 0.9 and 0.95 for QR.Hill (dotted lines)

CHAPTER 4

DISCUSSIONS AND FUTURE DIRECTIONS

Estimation of extreme conditional quantile is of prime interest in many fields of science. Quantile Regression is useful to quantify effects of the covariates at different quantile levels. When the underlying distribution is heavy tailed in nature, Quantile Regression estimates suffer form high variability particularly at higher quantiles.

In this dissertation, we have developed methods tailored for estimating high conditional quantiles for heavy tailed distributions. We have used the GP distribution to model the tail behavior. Selection of an appropriate threshold is necessary for fitting a GP model. The threshold selection problem itself is challenging: a higher threshold leads to high variance while a lower threshold is associated with higher bias. The efficiency of the GP model is severely compromised when the threshold itself varies as a function of the covariate. The common practice is to use an empirical quantile of the response variable as a threshold which is not covariate adaptive. Combining Quantile Regression with Extreme Value theory, we have proposed a novel method (QR.GP) of selecting a covariate adaptive threshold which in turn produce a consistent estimator for higher quantiles. We establish desirable asymptotic properties: consistency and asymptotic normality of QR.GP when the distribution is belongs to the GP family. We have shown that with proper thresholding, QR.GP can be adapted for a wide class of distribution in the heavy tailed domain of attraction. Although there are relatively few works which address the issue of estimation of high conditional quantiles for heavy tailed distributions, the one by Wang et al. (2012) presents a very promising approach. We have compared QR.GP approach which is fully parametric with that of Wang et al. (2012). Our findings show that when the underlying distribution is Pareto or Generalized Pareto, QR.GP provides consistent estimation of higher quantiles and has lower mean square error that Wang et al. (2012). From our numerical investigations, we show that the value of the threshold in QR.GP is different for different distributions in the heavy tailed

domain. Nonetheless, we have proposed a guideline for choosing the threshold when we implement QR.GP in practice. Compared to Wang et al. (2012) our method is more robust, computationally much faster and easy to interpret and apply in real data.

In our work, we have focused on the linear quantile functions. As pointed out by Wang & Li (2013), the linearity assumption is restrictive. Wang & Li (2013) have used Box-Cox power transformation in which the quantiles of the transformed response variable is linear in the covariates. An extension of work could be modeling some power transformation of the response variable with GP distribution. Our method features a constant tail index. In future, we can look at a possible extension where the tail index is also a function of the covariate. Establishing asymptotic properties of our method is challenging when the underlying distribution is a heavy tailed distribution other than Pareto or GP. This is an open ended problem and one can explore further in this direction.

APPENDICES

APPENDIX A

APPENDIX FOR CHAPTER 2

Supplemental figures and tables:

Table A.1: $\phi = 0$, exp(1): rel.bias(se) of quantile estimates for $\phi = 0$ at $x = 0.5$ and quantile levels τ.

τ	$n = 500$		$n = 1000$	
	GP.scale	QR	GP.scale	QR
0.1	0.04(0.03)	0.14(0.11)	0.03(0.03)	0.11(0.08)
0.3	0.04(0.03)	0.08(0.06)	0.03(0.03)	0.06(0.05)
0.5	0.04(0.03)	0.06(0.05)	0.03(0.02)	0.05(0.04)
0.7	0.04(0.03)	0.06(0.04)	0.03(0.02)	0.04(0.03)
0.9	0.04(0.03)	0.06(0.04)	0.03(0.02)	0.04(0.03)
0.93	0.04(0.03)	0.06(0.05)	0.03(0.02)	0.05(0.04)
0.95	0.04(0.03)	0.07(0.05)	0.03(0.02)	0.05(0.04)
0.97	0.04(0.03)	0.07(0.06)	0.03(0.03)	0.05(0.04)
0.99	0.05(0.04)	0.1(0.08)	0.04(0.03)	0.07(0.05)

Table A.2: $\phi = 0$, exp(1): mse(se) of quantile estimates for $\phi = 0$ at $x = 0.5$ and quantile levels τ.

τ	$n = 500$		$n = 1000$	
	GP.scale	QR	GP.scale	QR
0.1	0(0)	0(0)	0(0)	0(0)
0.3	0(0)	0(0)	0(0)	0(0)
0.5	0(0)	0.01(0)	0(0)	0(0)
0.7	0.01(0)	0.01(0)	0(0)	0.01(0)
0.9	0.03(0)	0.06(0)	0.02(0)	0.03(0)
0.93	0.04(0)	0.09(0.01)	0.02(0)	0.05(0)
0.95	0.05(0)	0.12(0.01)	0.03(0)	0.07(0)
0.97	0.08(0.01)	0.21(0.01)	0.05(0)	0.12(0.01)
0.99	0.18(0.01)	0.71(0.05)	0.1(0.01)	0.36(0.02)

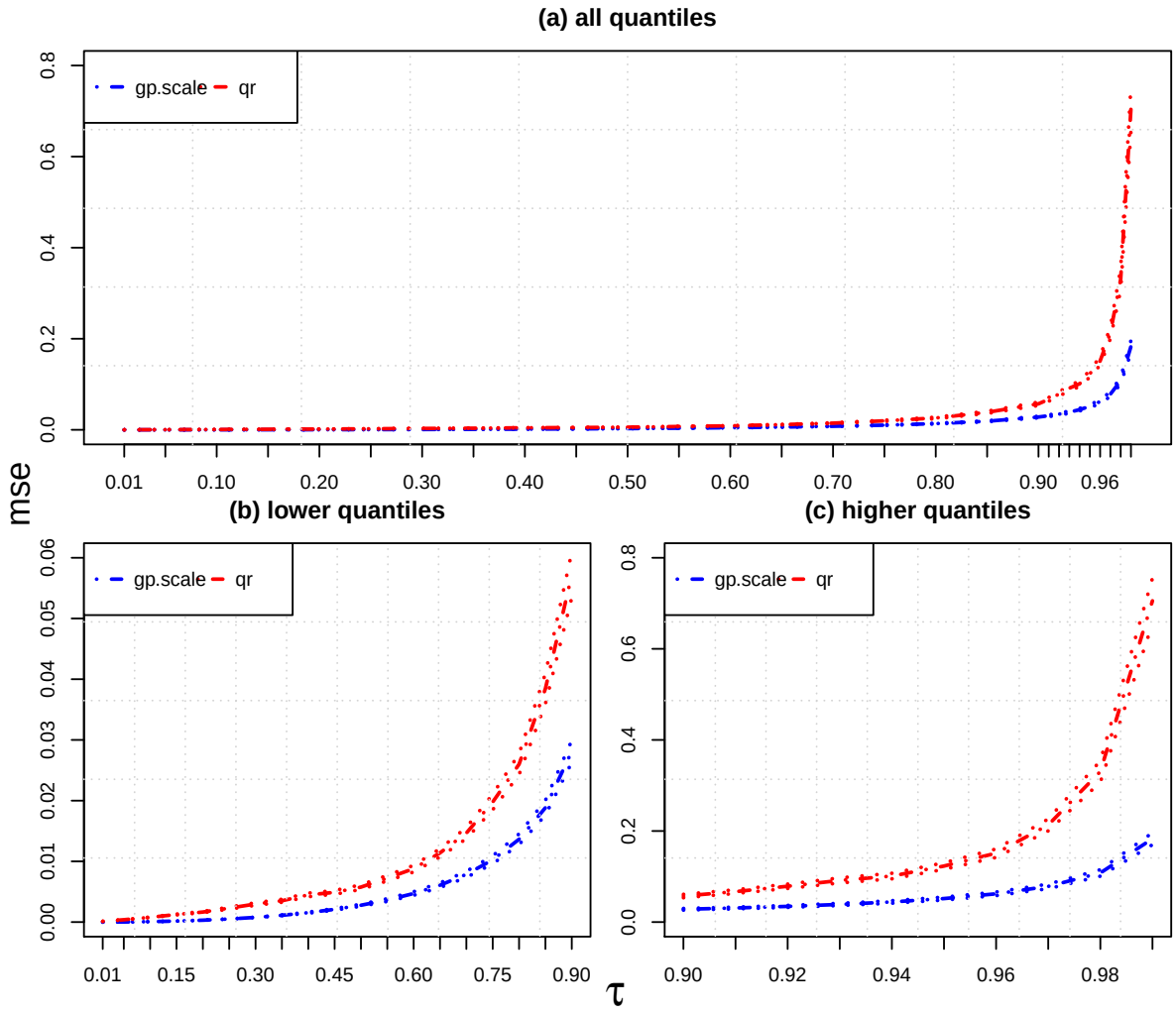

Figure A.1: mse(se) for $\phi = 0$, exp(1) at $x = 0.5$ and quantile levels τ. The sample size is $n = 500$.

APPENDIX B

APPENDIX FOR CHAPTER 3

Table B.1: Probability distributions in heavy-tail domain

Distribution	$1 - F(y)$	ξ	Parameters	$F^{-1}(\tau)$
$\mathrm{GP}(\mu, \sigma, \xi)$	$\left(1 + \frac{\xi(y-\mu)}{\sigma}\right)^{-\frac{1}{\xi}}$	ξ	$\mu < y, \sigma, \xi > 0$	$\mu + \frac{\sigma}{\xi}\left[(1-\tau)^{-\xi} - 1\right]$
$\mathrm{GP}(0, 1, \xi)$	$(1 + \xi y)^{-\frac{1}{\xi}}$	ξ	$0 < y, \xi > 0$	$\frac{1}{\xi}\left[(1-\tau)^{-\xi} - 1\right]$
$\exp(\lambda)$	$\exp(-\lambda y)$	0	$y > 0, \lambda > 0$	$-\frac{1}{\lambda}\log(1-\tau)$
$\mathrm{Pa}(\alpha)$	$y^{-\alpha}$	$\frac{1}{\alpha}$	$y > 1, \alpha > 0$	$(1-\tau)^{-\frac{1}{\alpha}}$
$\mathrm{Frechet}(\alpha)$	$1 - \exp(-y)^{-\alpha}$	$\frac{1}{\alpha}$	$y > 0, \alpha > 0$	$(-\log(\tau))^{\frac{1}{\alpha}}$
$\mathrm{Burr}(\eta, \lambda, \nu)$	$1 - \left(\frac{\eta}{\eta + y^{-\nu}}\right)^{\lambda}$	$\frac{1}{\nu}$	$y > 0, \eta, \lambda, \nu > 0$	$\left[\eta\left(\tau^{-\frac{1}{\lambda}} - 1\right)\right]^{-\frac{1}{\nu}}$
$\lvert\mathrm{T}(n)\rvert$	$2\left(1 - F_{t(n)}(y)\right)$	$\frac{1}{n}$	$y > 0, n > 0$	$F^{-1}_{t(n)}\left(\frac{1+\tau}{2}\right)$

BIBLIOGRAPHY

BIBLIOGRAPHY

Abrevaya, J. 2001. "the effect of demographics and maternal behavior on the distribution of birth outcomes,". *Empirical Economics* 26. 247.

Beirlant, Jan, Goedele Dierckx, A Guillou & C Staăricaă. 2002. On exponential representations of log-spacings of extreme order statistics. *Extremes* 5(2). 157–180.

Bremnes, John Bjørnar. 2004. Probabilistic wind power forecasts using local quantile regression. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology* 7(1). 47–54.

Buchinsky, Moshe. 1994. Changes in the u.s. wage structure 1963-1987: Application of quantile regression. *Econometrica* 62(2). 405–458. http://www.jstor.org/stable/2951618.

Caeiro, Frederico & M Ivette Gomes. 2009. Semi-parametric second-order reduced-bias high quantile estimation. *Test* 18(2). 392–413.

Cannon, Alex J. 2011. Quantile regression neural networks: Implementation in r and application to precipitation downscaling. *Computers and Geosciences* 37(9). 1277 – 1284. doi:https://doi.org/10.1016/j.cageo.2010.07.005.

del Castillo, Joan & Isabel Serra. 2015. Likelihood inference for generalized pareto distribution. *Computational Statistics Data Analysis* 83. 116 – 128. doi:https://doi.org/10.1016/j.csda.2014.10.014. http://www.sciencedirect.com/science/article/pii/S0167947314003065.

Chavez-Demoulin, V. & A. Davison. 2005. Generalized additive modeling of sample extremes. *Journal of the Royal Statistical Society Series C* 54. 207–222. doi:10.1111/j.1467-9876.2005.00479.x.

Chernozhukov, Victor. 2005. Extremal quantile regression. *Ann. Statist.* 33(2). 806–839. doi:10.1214/009053604000001165. https://doi.org/10.1214/009053604000001165.

Chernozhukov, Victor & Songzi Du. 2006. Extremal quantiles and value-at-risk .

Coles, Stuart, Joanna Bawa, Lesley Trenner & Pat Dorazio. 2001. *An introduction to statistical modeling of extreme values*, vol. 208. Springer.

Daouia, Abdelaati, Laurent Gardes, Stéphane Girard & Alexandre Lekina. 2011. Kernel estimators of extreme level curves. *Test* 20(2). 311–333.

Davison, A. C. & R. L. Smith. 1990. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)* 52(3). 393–425. doi:10.1111/j.2517-6161.1990.tb01796.x. https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1990.tb01796.x.

Davison, AC & NI Ramesh. 2000. Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62(1). 191–208.

Diebolt, Jean, Armelle Guillou & Pierre Ribereau. 2007. Asymptotic normality of extreme quantile estimators based on the peaks-over-threshold approach. *Communications in Statistics - Theory and Methods* 36(5). 869–886. doi:10.1080/03610920601036317. `https://doi.org/10.1080/03610920601036317`.

Elsner, Kossin J. P, J. B & T. H Jagger. 2008. "the increasing intensity of the strongest tropical cyclones,". *Nature* 455. 92. doi:https://doi.org/10.1038/nature07234.

Embrechts, Paul, Claudia Klüppelberg & Thomas Mikosch. 1997. Modelling extremal events, volume 33 of applications of mathematics.

Feigin, Paul D. & Sidney I. Resnick. 1994. Limit distributions for linear programming time series estimators. *Stochastic Processes and their Applications* 51(1). 135 – 165. doi:https://doi.org/10.1016/0304-4149(94)90022-1. `http://www.sciencedirect.com/science/article/pii/0304414994900221`.

Feuerverger, Andrey, Peter Hall et al. 1999. Estimating a tail exponent by modelling departure from a pareto distribution. *The Annals of Statistics* 27(2). 760–781.

Friederichs, P & A Hense. 2007a. Statistical downscaling of extreme precipitation events using censored quantile regression. *Monthly weather review* 135(6). 2365–2378.

Friederichs, P & A. Hense. 2007b. "statistical downscaling of extreme precipitation events using censored quantile regression,". *Monthly Weather Review* 135. 2365.

Gardes, Laurent & Stephane Girard. 2010. Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels. *Extremes* 13(2). 177–204.

Garrido, Myriam & Pascal Lezaud. 2013. Extreme value analysis : an introduction .

Gilli, Manfred & Evis këllezi. 2006. An application of extreme value theory for measuring financial risk. *Computational Economics* 27. 207–228. doi:10.1007/s10614-006-9025-7.

Gomesa, M Ivette & M João Martins. 2002. "asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes* 5(1). 5–31.

Haan, Laurens & Ana Ferreira. 2006. *Extreme value theory: An introduction*. doi:10.1007/0-387-34471-3.

de Haan, Laurens & Ana Ferreira. 2010. *Extreme value theory: An introduction (springer series in operations research and financial engineering)*. Springer 1st edn.

Hall, Peter & Nader Tajvidi. 2000. Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science* 153–167.

Hill, Bruce M. 1975. A simple general approach to inference about the tail of a distribution. *Ann. Statist.* 3(5). 1163–1174. doi:10.1214/aos/1176343247. https://doi.org/10.1214/aos/1176343247.

Jagger, T. H & J. B Elsner. 2008. "modeling tropical cyclone intensity with quantile regression,". *International Journal of Climatology* 29. 1351. doi:https://doi.org/10.1002/joc.1804.

Jansen, Dennis & Casper Vries. 1991. On the frequency of large stock returns: Putting booms and busts into perspective. *The Review of Economics and Statistics* 73. 18–24. doi:10.2307/2109682.

Kalnay, Eugenia, M. KANAMITSU, R. KISTLER, William Collins, D. DEAVEN, Gandin LS, M. IREDELL, Satyajit Saha, G. White, J. WOOLLEN, Yuejian Zhu, Muthuvel Chelliah, W. EBISUZAKI, Wayne Higgins, John Janowiak, K. C, C. ROPELEWSKI, Julian Wang & A. LEETMAA. 1996. The nmc/ncar 40-year reanalysis project. bull am meteorol soc. *Bulletin of the American Meteorological Society.* 77. doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Koenker, Roger. 2005. *Quantile regression* Econometric Society Monographs. Cambridge University Press. doi:10.1017/CBO9780511754098.

Koenker, Roger & Gilbert Bassett. 1978. Regression quantiles. *Econometrica* 46(1). 33–50.

Kyselý, Jan. 2002. Probability estimates of extreme temperature events: Stochastic modelling approach vs. extreme value distributions. *Studia Geophysica et Geodaetica* 46. 93–112. doi:10.1023/A:1019841717369.

Langousis, Andreas, Antonios Mamalakis, Michelangelo Puliga & Roberto Deidda. 2016. Threshold detection for the generalized pareto distribution: Review of representative methods and application to the noaa ncdc daily rainfall database. *Water Resources Research* 52(4). 2659–2681.

Li, D, L Peng & J Yang. 2010. "bias reduction for high quantiles,". *Journal of Statistical Planning and Inference* 140. 2433.

Park, Myung Hyun & Joseph H.T. Kim. 2016. Estimating extreme tail risk measures with generalized pareto distribution. *Computational Statistics Data Analysis* 98. 91 – 104. doi:https://doi.org/10.1016/j.csda.2015.12.008. http://www.sciencedirect.com/science/article/pii/S0167947315003163.

Peng, Liang & Yongcheng Qi. 2004. Estimating the first-and second-order parameters of a heavy-tailed distribution. *Australian & New Zealand Journal of Statistics* 46(2). 305–312.

Pickands, James, III. 1975. Statistical inference using extreme order statistics. *Ann. Statist.* 3(1). 119–131. doi:10.1214/aos/1176343003. https://doi.org/10.1214/aos/1176343003.

Scarf, Philip A. & Patrick J. Laycock. 1996. Estimation of extremes in corrosion engineering. *Journal of Applied Statistics* 23(6). 621–644. doi:10.1080/02664769623982. https://doi.org/10.1080/02664769623982.

Song, Jongwoo & Seongjoo Song. 2012a. A quantile estimation for massive data with generalized pareto distribution. *Computational statistics and data analysis* 56(1). 143–150.

Song, Jongwoo & Seongjoo Song. 2012b. A quantile estimation for massive data with generalized pareto distribution. *Computational Statistics Data Analysis* 56(1). 143 – 150. doi:https://doi.org/10.1016/j.csda.2011.06.030. http://www.sciencedirect.com/science/article/pii/S0167947311002477.

Tawn, Jonathan A. 1992. Estimating probabilities of extreme sea-levels. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 41(1). 77–93. http://www.jstor.org/stable/2347619.

Taylor, James W. 2008. Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Econometrics* 6(3). 382–406. https://EconPapers.repec.org/RePEc:oup:jfinec:v:6:y:2008:i:3:p:382-406.

Umantsev, Len & Victor Chernozhukov. 2001. Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics* 26(1). 271–292. https://EconPapers.repec.org/RePEc:spr:empeco:v:26:y:2001:i:1:p:271-292.

Velthoen, Jasper, Juan-Juan Cai, Geurt Jongbloed & Maurice Schmeits. 2019a. Improving precipitation forecasts using extreme quantile regression. *Extremes* 22(4). 599–622. doi:10.1007/s10687-019-00355-1.

Velthoen, Jasper, Juan-Juan Cai, Geurt Jongbloed & Maurice Schmeits. 2019b. Improving precipitation forecasts using extreme quantile regression. *Extremes* doi:10.1007/s10687-019-00355-1.

Wang, Hansheng & Chih-Ling Tsai. 2009. Tail index regression. *Journal of the American Statistical Association* 104(487). 1233–1240.

Wang, Huixia Judy & Deyuan Li. 2013. Estimation of extreme conditional quantiles through power transformation. *Journal of the American Statistical Association* 108(503). 1062–1074.

Wang, Huixia Judy, Deyuan Li & Xuming He. 2012. Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association* 107(500). 1453–1464. doi:10.1080/01621459.2012.716382.

Weissman, Ishay. 1978. "estimation of parameters and large quantiles based on the k largest observations,". *Journal of the American Statistical Association* 73. 812.