Parametric quantile regression based on the generalised gamma distribution

Thesis

Parametric Quantile Regression Based on the Generalised Gamma Distribution

by

Angela Noufaily
Maitrise in Statistics
MSc in Pure Mathematics

A thesis
submitted to The Open University
in fulfilment of the requirements for the degree of
Doctor of Philosophy

Department of Mathematics and Statistics
Faculty of Mathematics, Computing & Technology
The Open University
Walton Hall, Milton Keynes, MK7 6AA
United Kingdom

March 2011

Date of Submission: 3 December 2010
Date of Award: 9 March 2011

Abstract

Quantile regression offers an extension to regression analysis where a modified version of the least squares method allows the fitting of quantiles at every percentile of the data rather than the mean only. Using the well-known three-parameter generalised gamma distribution to model variation in data, we present a parametric quantile regression study for positive univariate reference charts. The study constitutes an overall package that includes all different stages of parametric modeling starting from model identification to parameter estimation, model selection and finally model checking.

We improve on earlier work by being the first to formulate the iterative approach to solution of the likelihood score equations of the generalised gamma distribution in such a way that the individual equations involved are uniquely solvable and far from being problematic as a number of authors have suggested. We conduct likelihood ratio tests to choose the best model within the three-, four-, five- and six-parameter generalised gamma family obtained by making the parameters linearly (or loglinearly) dependent on a univariate covariate. Quantiles are plotted accordingly and asymptotic theory for obtaining the expressions for confidence bands around them is given. Based on the chi-square goodness-of-fit test, we suggest a test statistic that checks the goodness of the generalised gamma model for given data. We validate the whole theoretical process computationally via simulations. Lastly, we demonstrate the different steps of the proposed modeling procedure through two main applications; one is environment-related and the other health-related.

In a parallel fashion, inspired by the generalised gamma distribution, we introduce an alternative three-parameter distribution with useful statistical properties. We explore briefly maximum likelihood estimation and asymptotic theory of the alternative distribution and we compare it computationally to the generalised gamma.

Dedicated to my parents,

my sisters and my brother.

Acknowledgements

I would have never pursued my PhD at the Open University if the hand of God did not bring me there. Thank God for this thesis and for putting the right people on my way to accomplish this work. I take this opportunity to thank all those people who contributed and helped me along the way.

I express my deepest gratitude to my supervisor Prof. Chris Jones whose guidance and insight were invaluable all through my PhD. I could have never asked for better advice, more professionalism and greater encouragement. In addition to the weekly meetings, he was always very welcoming any other time I knocked on his office, with his usual smile and readiness to answer all my questions. He is an excellent supervisor that I will always look up to all through my research life.

To the Open University, I am very grateful for the funding. I offer many thanks to the members of the Statistics Department for providing a great working environment. I also take the opportunity to thank Prof. Paddy Farrington for being the initial contact who encouraged me to study at the Open University and Prof. Paul Garthwaite who provided me with constructive ideas that I included in my thesis. I extend my gratitude to my second supervisor Dr. Heather Whitaker and to Dr. Karim Anaya-Izquierdo who offered advice with my **R** codes.

Further acknowledgements go to my colleagues Swarup De, Fadlalla Elfadaly, Doyo Gragn Enki and Steffen Unkel for being a great help in the process.

I am grateful to Dr. Rissa De la Paz and Dr. Yoseph Araya for being great friends along the way. I also thank Yoseph for supplying me with the water table depth data set. Furthermore, I acknowledge all the help offered from my friends and relatives who encouraged me throughout my PhD.

I would like to thank Brendan McGovern whose support kept me going through the hardest times.

To my parents, Isabelle and Pierre Noufaily, thank you for all the encouragement and care you offered. Christine, Karine, Berna and Joe, each one of you contributed in your own way to my progress. I dedicate this thesis to my family. Your presence helped me make the transition from where I came to where I am. I acknowledge your support and prayers and I love you.

Contents

Chapter 1

Introduction

As the title of the thesis indicates, two main statistical elements are involved in this study: parametric quantile regression (PQR) and the generalised gamma (GG) distribution. PQR represents the approach followed in this study and the GG is the model used in this framework. As for the target, it is principally to construct reference charts for positive response data conditional on a univariate covariate.

There is a need in various fields such as pharmacology, health, econometrics and many others to obtain reference charts that can specify the "common" range in data. Given positive data points from a response distribution conditional on a univariate covariate, the following questions arise: What best shape or model can be attributed to the data, and what limiting curves define its percentiles?

In a regression equation, a dependent variable Y is modeled as a function of an independent variable X, corresponding parameters, and a random variable error term which represents unexplained variation in the dependent variable. Most commonly this error term is considered as normally distributed. However, what if the unexplained variation in the dependent variable is not symmetric? From here comes the search for distributions with parameters that control skewness. The three-parameter GG is our proposed distribution, applicable to response data that take positive values only.

In regression analysis, one curve (the mean curve) is fitted to the whole data. To allow the fitting of quantiles for reference charts, we extend our study to quantile regression (QR), while using the GG distribution to model variation in data. As Koenker (2005) mentions:

> Quantile regression is intended to offer a comprehensive strategy for completing the regression picture.

By complementing the exclusive focus of classical least squares regression on the conditional mean, it offers a strategy for examining how covariates influence the location, scale and shape of the entire response distribution. The literature shows that very little work has been done on PQR especially the part that involves using nonsymmetric distributions for model-fitting. On that, we quote Gilchrist (2008) saying:

> In contrast with the wealth of theory associated with the classical Normal-based regression, this paper has given little theory. In fact, there is little theory available [on parametric quantile regression]. (...) There are a number of theoretical and practical issues to be addressed in this topic. (...) It is recommended that regression be revisited from these perspectives.

Driven by the aforementioned motives and the will to explore and improve on a field that has been a subject of interest to researchers but yet needs revisiting, we perform a rounded study on PQR using the GG distribution while passing through all different stages of parametric modeling. Of course, we perform model identification, parameter estimation and model selection and we suggest a goodness-of-fit test, all of which we combine in one modeling package. Using the concept of QR, this thesis provides an overall modeling package that can be applied to positive univariate data with a single covariate.

For given positive univariate regression data, our overall modeling package involves:

1. Specifying the GG PQR model by linking its parameters to the covariate.
2. Estimating the three-, four-, five-, and six-parameter GG using maximum likelihood (ML).
3. Selecting the best subset from within the GG family using the likelihood ratio test (LRT).
4. Providing expressions for the corresponding quantiles and plotting them.
5. Providing expressions for the confidence intervals (CIs) around the quantiles and plotting them.
6. Suggesting a goodness-of-fit test for our GG model using the concept of quantiles.

The overall study was validated using a grand simulation study involving various data sets from different cases of the GG distribution. It is also illustrated through applications of reference charts to two data sets. The first data set involves an environmental issue that studies the effect of flux on water table depth, while the second one addresses a health issue where weights of individuals are analysed for given heights. All these ideas are put together in eight chapters starting with the current introduction. Chapter 2 offers some basic definitions and properties of the three-parameter GG distribution along with maximum likelihood estimation (MLE) of the parameters. The GG distribution is put in a QR context in Chapters 3 and 4. Chapter 3 offers general definitions and a literature review related to QR. GG quantile shapes are also discussed thoroughly in the presence of a covariate. In Chapter 4, we present MLE of the four-, five- and six-parameter GG along with the asymptotics. We explain how we use LRTs to choose the best GG model, we give expressions for the CIs around the estimated quantile and we suggest a goodness-of-fit test. Chapters 5 and 6 present respectively a simulation study that validates some of the theory in the thesis and applications to real data sets. An alternative generalised gamma distribution is introduced in Chapter 7. Chapter 8 presents a final set of conclusions.

Chapter 2

The Univariate Generalised Gamma Distribution

The GG distribution is a key ingredient in this thesis from which the whole story begins. We consider the univariate continuous three-parameter GG distribution applicable to data that take positive values only. Being a distribution that encompasses many of the life distributions as special cases, the GG has the additional property that allows it to target especially skewness in data. It has two shape parameters that control both tails of the distribution. These parameters allow the density function to have various interesting shapes and hence to model different aspects in data.

In this chapter, we present the basic definitions and properties related to this useful distribution such as the probability density function (pdf), the cumulative distribution function and the moments which we cover in Section 2.1. In Section 2.2, we list some of its most popular special cases and we discuss random variate generation from the GG. Section 2.3 offers a literature review on previous researchers' attempts at estimating the GG parameters followed by our contribution to the MLE of the GG in Section 2.4. This is validated via simulations in Section 2.5. Finally, asymptotics of the ML estimators are given in Section 2.6 and asymptotic correlations between parameters are discussed.

2.1 The Generalised Gamma Density Function, Cumulative Distribution Function and Moments

Let us start by defining the gamma function at a point z by

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt.$$

We also define the first derivative of the logarithm of the gamma function at a point z to be

$$\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)} = \frac{\int_0^\infty \log t e^{-t} t^{z-1} dt}{\int_0^\infty e^{-t} t^{z-1} dt}.$$

A selection of four of the most significant versions of the GG pdfs introduced by researchers is summarised below. These pdfs will be referred to later in the thesis according to their corresponding version.

- Version 1, $GG(\theta, \alpha, \beta)$, by Stacy (1962):

  $$f(t) = \frac{\beta}{\theta^\alpha \Gamma\left(\frac{\alpha}{\beta}\right)} t^{\alpha-1} e^{-\left(\frac{t}{\theta}\right)^\beta}, \qquad t > 0$$

  where θ is a scale parameter and α and β are shape parameters. β, α, and t are all positive.

  Setting $k = \alpha/\beta$, we get:

- Version 2, $GG(\theta, k, \beta)$, by Stacy and Mirham (1965):

  $$f(t) = \frac{\beta}{\theta^{k\beta} \Gamma(k)} t^{k\beta-1} e^{-\left(\frac{t}{\theta}\right)^\beta}, \qquad t > 0$$

  where θ is a scale parameter and k and β are shape parameters. k, β, θ, and t are all positive.

- Version 3, $GG(\gamma, \sigma, q)$, by Prentice (1974):

  $$f(y) = \begin{cases} \frac{|q|}{\sigma\Gamma(q^{-2})} \exp\left\{w_1 q^{-2} - e^{w_1}\right\} & \text{if } q \neq 0 \\ \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y-\gamma}{\sigma}\right)^2\right\} & \text{if } q = 0, \end{cases}$$

where $w_1 = (y-\gamma)\sigma^{-1}q + \mu^*$, $y = \log t \in \mathbb{R}$, $\mu^* = \psi(q^{-2})$ and $q = k^{-1/2}$. γ is a location parameter related to the parameters in Version 2 by $\gamma = \log\theta + \mu^*/\beta$, σ is a scale parameter $\sigma = q/\beta$, and q is a shape parameter. w_1, γ and $q \in \mathbb{R}$ while σ is positive.

- Version 4, $GG(\mu, \sigma, k)$, by Lawless (1980):

$$f(y) = \frac{k^{k-\frac{1}{2}}}{\sigma\Gamma(k)} \exp\left\{\sqrt{k}w - ke^{\frac{w}{\sqrt{k}}}\right\},$$

where $w = (y-\mu)/\sigma$ and $y = \log t \in \mathbb{R}$. μ is a location parameter related to the parameters in Version 2 by $\mu = \log\theta + \log k/\beta$, σ is a scale parameter $\sigma = 1/(\beta\sqrt{k})$, and k is a shape parameter. w and $\mu \in \mathbb{R}$ while σ and k are positive.

Note that Version 2 is a reparametrisation of Version 1, and Versions 3 and 4 are obtained by reparametrisations and transformations applied to Version 2; in particular, Versions 3 and 4 pertain to the equivalent log-GG distribution, the distribution of $Y = \log T$ where $T \sim GG$. What about the relationship between Versions 3 and 4?

Prentice's Version 3 (for $q > 0$) is equivalent to

$$\begin{aligned} f(y;\gamma,\sigma,q) &= \frac{q}{\sigma\Gamma(q^{-2})} \exp\left\{\frac{(y-\gamma)}{\sigma q} + \mu^* q^{-2} - \exp\left(\frac{q(y-\gamma)}{\sigma} + \mu^*\right)\right\} \\ &= \frac{q}{\sigma\Gamma(q^{-2})} \exp\left\{\frac{y}{\sigma q} - \frac{\gamma}{\sigma q} + \mu^* q^{-2} - \exp\left(\frac{yq}{\sigma} - \frac{\gamma q}{\sigma} + \mu^*\right)\right\}. \end{aligned}$$

If we let $k = q^{-2}$, Lawless's Version 4 is equivalent to

$$\begin{aligned} f(y;\mu,\sigma,q) &= \frac{1}{q^{2q^{-2}-1}\sigma\Gamma(q^{-2})} \exp\left\{\frac{(y-\mu)}{\sigma q} - \frac{1}{q^2}\exp\left(\frac{q(y-\mu)}{\sigma}\right)\right\} \\ &= \frac{q}{\sigma\Gamma(q^{-2})} \exp\left\{\frac{y}{\sigma q} - \frac{\mu}{\sigma q} - \log\left(q^{2q^{-2}}\right) - \exp\left(\frac{yq}{\sigma} - \frac{\mu q}{\sigma} + \log\left(q^{-2}\right)\right)\right\}. \end{aligned}$$

As a result, Versions 3 and 4 are equivalent iff the following equality holds

$$\log\left(q^{-2}\right) - \frac{\mu q}{\sigma} = \mu^* - \frac{\gamma q}{\sigma}.$$

We may conclude that Versions 3 and 4 are equivalent iff $k = q^{-2}$ and $\mu = \gamma + (\sigma/q)\left(\log q^{-2} - \mu*\right)$.

It is now time to illustrate how the GG pdf versions actually look and how they behave for different values of the parameters by presenting some pdf plots. Figures 2.1 and 2.2 present plots of the GG pdf Version 1. In the former we plot the pdfs for fixed values of θ and β and we vary α. We take the particular case of a $GG(\theta = 2, \alpha, \beta = 1.5)$ for $\alpha \in \{0.1, 0.25, 0.5, 1, 2, 5, 25\}$. We notice that for $\alpha < 1$, the distribution rises to infinity at zero. When $\alpha > 1$, the distribution is bell-shaped with mode at $\theta((\alpha - 1)/\beta)^{1/\beta}$, and at $\alpha = 1$ it decreases monotonically from $\beta/(\theta\Gamma(1/\beta))$ to zero (at infinity). Analogously, in the latter figure, we fix θ and α and we vary β. This time, we consider the particular $GG(\theta = 2, \alpha = 3, \beta)$ for $\beta \in \{0.75, 0.9, 1, 1.5, 2, 3, 4\}$. As $\alpha > 1$ (which is the more interesting two-tailed case), all pdfs are bell-shaped. The distribution becomes more and more flat as β increases.

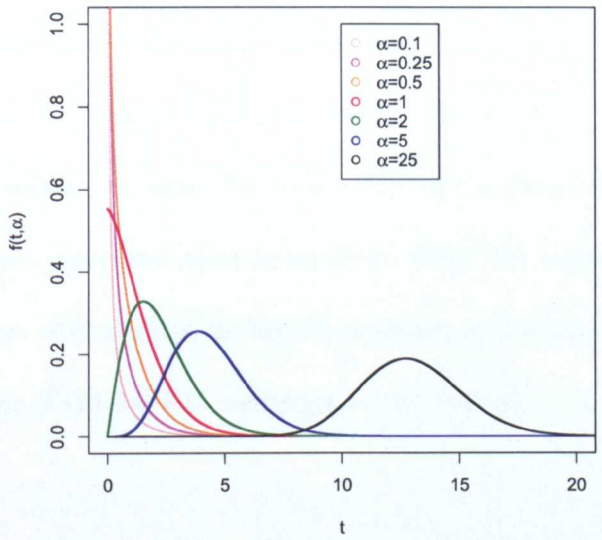

Figure 2.1: GG pdf Version 1 plotted at seven α values for fixed θ and β.

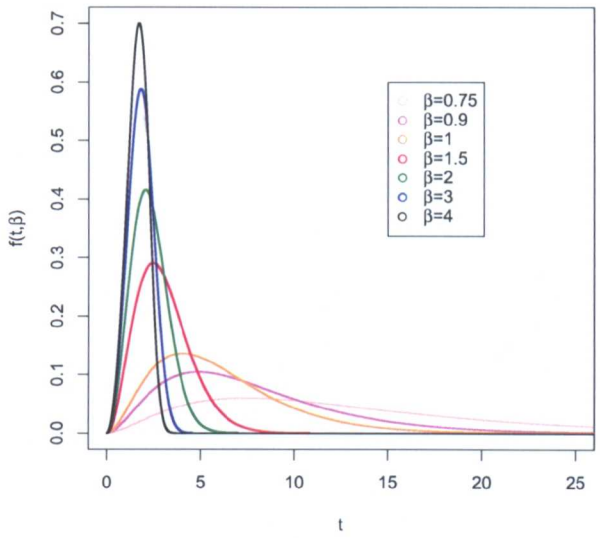

Figure 2.2: GG pdf Version 1 plotted at seven β values for fixed θ and α.

Let us now look at the GG pdf Version 2 and 4 in Figures 2.3 and 2.4, respectively. Figure 2.3 presents a plot of the $GG(\theta = 2, k, \beta = 3)$ for $k \in \{0.1, 0.25, 0.5, 1, 2, 5, 25\}$ while Figure 2.4 plots the $GG(\mu = 2, \sigma = 3, k)$ for $k \in \{0.1, 0.25, 0.5, 1, 2, 5, 25\}$. In Figure 2.3, as k increases, the distribution becomes more spiked and shifts more towards the right hand side of the positive real line. Of course, as $\beta = 3$, for $k = 0.1$ and $k = 0.25$ which correspond to $\alpha < 1$, the pdfs go to infinity at zero. We notice that in Figure 2.4, the pdf shape changes as k increases and becomes more and more symmetric. Also, for large values of k the shape of the distribution changes only slightly. In fact, it is shown in Lawless (1980) that as k tends to infinity the distribution of the log-GG converges to the normal.

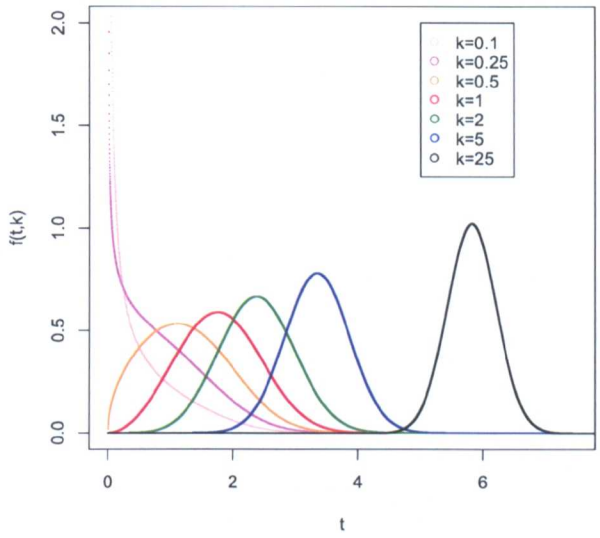

Figure 2.3: GG pdf Version 2 plotted at seven k values for fixed θ and β.

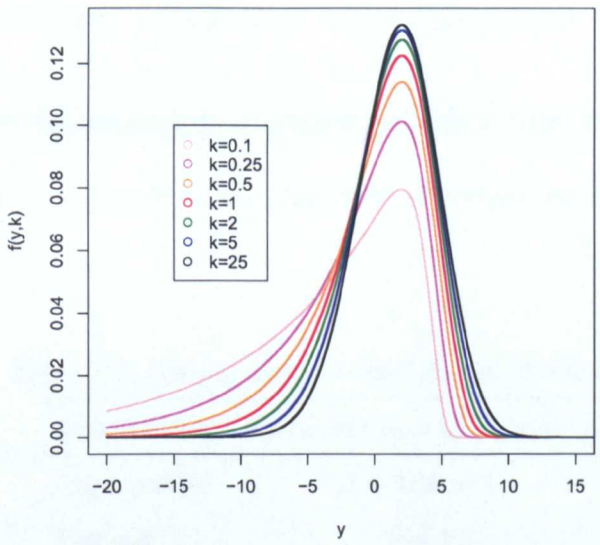

Figure 2.4: GG pdf Version 4 plotted at seven k values for fixed μ and σ.

The regularised gamma function, as defined by Abramowitz and Stegun (1965), is given by

$$P(k,r) = \frac{\Gamma_r(k)}{\Gamma(k)} = \frac{\int_0^r e^{-t}t^{k-1}dt}{\int_0^\infty e^{-t}t^{k-1}dt}, \tag{2.1}$$

where $\Gamma(k)$ is the well-known gamma function and $\Gamma_r(k)$ is the incomplete gamma function. As in Johnson *et al.* (1994), and referring to the pdf Version 2, the r^{th} moments and the cumulative distribution function of a GG distribution are respectively

$$E(t^r) = \theta^r \frac{\Gamma\left(\frac{r}{\beta}+k\right)}{\Gamma(k)}$$

and

$$F(t) = P\left(k, \left(\frac{t}{\theta}\right)^\beta\right) = \frac{\Gamma_{\left(\frac{t}{\theta}\right)^\beta}(k)}{\Gamma(k)}.$$

2.2 Special Cases and Random Variate Generation from a Generalised Gamma Distribution

For particular values of the parameters, a random variable T from a GG distribution with pdf as in Versions 2 and 4 becomes one of the known life distributions as shown in Tables 2.1 and 2.2 respectively.

Table 2.1: Special cases of the GG pdf Version 2.

Distribution	Particular case of Version 2
Exponential	$\beta = 1; k = 1$
Gamma	$\beta = 1$
Weibull	$k = 1$
Chi-square	$\theta = 2; \beta = 1; k = \nu/2$
Lognormal	$k \to \infty$

Table 2.2: Logged version of distributions in Table 2.1 — special cases of the GG pdf Version 4.

Distribution	Particular case of Version 4
Log-Gamma	$\sigma\sqrt{k} = 1$
Extreme value	$k = 1$
Normal	$k \to \infty$

A wide range of known life distributions are particular cases of the GG distribution, or simpler versions of it. From here comes its importance and richness. As a matter of fact, one could generate values of a GG distribution from values of some simpler life distribution (such as the gamma distribution) by simple variable transformations as will be illustrated in the next paragraph.

Denote by $G(\theta, k)$ the gamma distribution with scale parameter θ and shape parameter k. Given a random variable Z from a gamma distribution with scale parameter θ^β and shape parameter k, $Z \sim G(\theta^\beta, k)$, with pdf

$$f(z) = \frac{1}{(\theta^\beta)^k \Gamma(k)} z^{k-1} e^{-(\frac{z}{\theta^\beta})},$$

the transformation $T = Z^{\frac{1}{\beta}}$ allows obtaining a random variable T from a $GG(\theta, k, \beta)$ distribution with pdf

$$f(t) = \frac{\beta}{\theta^{k\beta}\Gamma(k)} t^{k\beta-1} e^{-(\frac{t}{\theta})^\beta}.$$

In short, if $Z \sim G(\theta^\beta, k)$, then $Z^{1/\beta} \sim GG(\theta, k, \beta)$. For example, to simulate 100 values from a GG distribution in Version 2, e.g. $GG(\theta = 0.5, k = 3, \beta = 2)$, one has to start with 100 values of a $G(\theta = 0.5^2, k = 3)$ random variables and take their square roots.

In fact, it is shown by Roberts (1971) that a necessary condition for a random variable $|T|^\beta$, $\beta \neq 0$, to be a $G(\theta, k)$, is that the density of T be $f(t) = h(t) \mid t \mid^{k\beta-1} \exp\{-(1/\theta)|t|^\beta\}$ where $h(t) + h(-t) = |\beta|(1/\theta)^k/\Gamma(k)$ is a constant for all $t \in \mathbb{R}$. Also, Johnson and Kotz (1972) studied power transformations of $G(\theta, k)$ that generate the Stacy (1962) GG distributions.

As we just explained, the GG is a three-parameter univariate unimodal continuous life distribution that takes as special cases some of the known life distributions such as the exponential, lognormal, Weibull, and gamma distributions. Being a rich family of distributions, it was a subject of interest to several researchers. We present now a literature review on what some researchers, through history, have done for the purpose of estimating its parameters.

2.3 The History of the Generalised Gamma Distribution

In this section, we go through the history of the GG distribution, in particular that concerning parameter estimation. We divide the history into three main subsections, the first displaying the birth of the GG distribution and the earliest contributions, the second highlighting the two most significant works of Prentice (1974) and Lawless (1980) that we have used and extended, and the third explaining the most recent contributions.

2.3.1 From Amoroso (1925) to Stacy (1973)

The basic story really begins with Stacy (1962) whose main concern was to introduce a generalisation of the gamma distribution. The specific form was suggested by Liouville's extension to Dirichlet's integral formula. In this form it may also be regarded as a special case of a function introduced by Amoroso (1925) and studied further by D'Addario (1932). The latter fitted such a distribution to a data set of income rates. Little interest was taken between 1925 and 1962 in this family of distributions until Stacy (1962) suggested the basic pdf form of what became known as the GG distribution. The pdf of the GG as first introduced by Stacy (1962) is Version 1 of Section 2.1. It is the density of the distribution of random variable T:

$$f(t) = \frac{\beta}{\theta^{\alpha}\Gamma\left(\frac{\alpha}{\beta}\right)} t^{\alpha-1} e^{-\left(\frac{t}{\theta}\right)^{\beta}}, \qquad t > 0, \tag{2.2}$$

where θ is a scale parameter and α and β are shape parameters. θ, β and α are all positive. The generalisation (2.2) is accomplished by supplying a positive parameter β as an exponent in the exponential factor of the gamma distribution which has density

$$h(t) = \frac{1}{\theta^{\alpha}\Gamma(\alpha)} t^{\alpha-1} e^{-\frac{t}{\theta}}, \qquad t > 0, \tag{2.3}$$

and corresponds to $\beta = 1$. The GG is also obtained by a generalisation to the Weibull distribution which has density

$$g(t) = \frac{\beta}{\theta^{\beta}} t^{\beta-1} e^{-(\frac{t}{\theta})^{\beta}}, \qquad t > 0,$$

and corresponds to the special case $\alpha = \beta$. So the GG is a clever concatenation and extension of two of the most popular life distributions. Distributions for some functions of independent GG random variables are given by Stacy such as the distribution of the sum of independent GG random variables.

Following the introduction of this promising distribution, several researchers were motivated to study it further especially regarding MLE of its parameters, a task that has been reported to be difficult and complicated. Parr and Webster (1965) considered MLE of the three-parameter GG distribution. This was done by the traditional way of differentiating the loglikelihood with respect to the three parameters, setting the resulting derivatives equal to zero, and solving for each parameter.

For a known location parameter, given $t_1, ..., t_n$ independent observations from a GG distribution with pdf as in (2.2), the loglikelihood function is

$$n \log \beta - n\alpha \log \theta - n \log \Gamma(\alpha/\beta) + (\alpha - 1) \sum_{i=1}^{n} \log t_i - \frac{1}{\theta^{\beta}} \sum_{i=1}^{n} t_i^{\beta}.$$

Maximising the loglikelihood with respect to α, θ and β yields

$$\theta = \left\{ \left(\frac{\beta}{n\alpha} \right) \sum_{i=1}^{n} t_i^{\beta} \right\}^{\frac{1}{\beta}}, \tag{2.4}$$

$$\log\beta - \log n - \log\alpha + \log\left(\sum_{i=1}^{n} t_i^{\beta}\right) + \psi\left(\frac{\alpha}{\beta}\right) - \left(\frac{\beta}{n}\right)\sum_{i=1}^{n}\log t_i = 0 \tag{2.5}$$

and

$$\log\beta - \log n - \log\alpha + \log\left(\sum_{i=1}^{n} t_i^{\beta}\right) + \psi\left(\frac{\alpha}{\beta}\right) + \frac{\beta}{\alpha} - \left(\frac{\beta}{\sum_{i=1}^{n} t_i^{\beta}}\right)\sum_{i=1}^{n} t_i^{\beta}\log t_i = 0, \tag{2.6}$$

where the ψ function is the digamma function defined in Abramowitz and Stegun (1965) as being the derivative of the logarithm of the well-known gamma function. Parr and Webster (1965) suggest solving equations (2.5) and (2.6) simultaneously for α and β , a task that is not always successful. They also, correctly, say that a closed form solution is not available.

Along the lines of Stacy (1962), Stacy and Mirham (1965) introduce an alternative version of the generalisation of the gamma distribution and discuss parameter estimation using a modified method of moments technique for the distribution of $\log T$. They consider a reparametrised form of the pdf (2.2) whereby they set $k = \alpha/\beta$ and replaced α by k. The pdf in this case is

$$f(t) = \frac{|\beta|}{\theta^{k\beta}\Gamma(k)} t^{k\beta-1} e^{-\left(\frac{t}{\theta}\right)^{\beta}}, \qquad t > 0, \tag{2.7}$$

where θ is a scale parameter and k and β are now the shape parameters. k and θ are positive. This family of distributions differs from Version 2 of Section 2.1 in that it allows negative values for the parameter β. Thus, it includes the distribution of T^{-1} whenever it includes the distribution of a random variable T. The suggested method of moments leads to simultaneous equations for which closed form solutions were not available. A graphical solution was proposed.

Harter (1967) suggested that the addition of a location parameter, d, to the pdf form (2.7) enhances its usefulness. An iterative procedure for MLE of the resulting four-parameter distribution is developed from complete and censored samples. Iterative procedures for solving the score equations were suggested, such as the rule of false position, the Newton Raphson algorithm and the gradient method. Numerical examples were given and the iterative procedure was applied to all four cases starting with the case where the four parameters are unknown, to

the cases where any three parameters, any two parameters, and any one parameter is unknown. We note that the addition of a location parameter in this case allows the shifting of the distribution to the negative real line which might not be very appropriate especially since the aim of such life distributions is to deal with data that take positive values only.

Also, referring to the pdf form (2.7) with $\beta > 0$, Hager and Bain (1970) suggested another approach for MLE of the parameters θ, β, and k. Given $t_1, ..., t_n$ independent observations from (2.7), the loglikelihood function is

$$n\log\beta - n\log\Gamma(k) - nk\beta\log\theta + (k\beta - 1)\sum_{i=1}^{n}\log t_i - \sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta}. \tag{2.8}$$

Differentiating (2.8) with respect to θ, β, and k, in turn, yields the score equations

$$\frac{\beta}{\theta}\left\{-nk + \sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta}\right\} = 0; \tag{2.9}$$

$$\frac{n}{\beta} + k\sum_{i=1}^{n}\log\left(\frac{t_i}{\theta}\right) - \sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta}\log\left(\frac{t_i}{\theta}\right) = 0; \tag{2.10}$$

$$-n\psi(k) + \beta\sum_{i=1}^{n}\log\left(\frac{t_i}{\theta}\right) = 0. \tag{2.11}$$

Equation (2.9) yields

$$\theta = \left(\frac{\sum_{i=1}^{n} t_i^{\beta}}{nk}\right)^{\frac{1}{\beta}}. \tag{2.12}$$

Substituting this expression for θ into (2.10) yields

$$k = \frac{-1}{\beta\left\{\frac{1}{n}\sum_{i=1}^{n}\log t_i - \frac{\sum_{i=1}^{n} t_i^{\beta}\log t_i}{\sum_{i=1}^{n} t_i^{\beta}}\right\}}. \tag{2.13}$$

Substituting into (2.11) yields an equation in β:

$$H(\beta) = -\psi(k) + \beta\frac{\sum_{i=1}^{n} t_i}{n} - \log\left(\sum_{i=1}^{n} t_i^{\beta}\right) + \log(nk) = 0. \tag{2.14}$$

Solving this equation, we obtain the estimates $\widehat{\theta}$, $\widehat{\beta}$, and $\widehat{k}$ of θ, β, and k by iteration. Extensive investigations of $H(\beta)$ conducted by Hager (1970) in his doctoral thesis indicated that it is

not always possible to determine if $H(\beta) = 0$ has a root. There are conflicting reports in the literature as to the possibility of solving $H(\beta) = 0$ and as to the well behaviour of the score equations. Hager and Bain (1970) mention that they haven't been able to find meaningful estimates in some cases. Nevertheless, they suggest using another technique that will be explained in what follows. For given values of k, they report that θ_k and β_k can be calculated without difficulty using the following reasoning:

- Fix $k = m$; the corresponding parameters are β_m and θ_m.
- Set $w = (t/\theta)^\beta$. By taking the variance of the log on both sides of this equation, and knowing that $\text{Var}(\log w) = \psi'(m)$, the following is deduced

$$\beta_m = \left(\frac{\psi'(m)}{Var(\log t)}\right)^{\frac{1}{2}},$$

  where $\text{Var}(\log t)$ would be replaced by the sample variance.
- From (2.11), obtain the expression of $\widehat{\theta}$ as a function of β:

$$\theta_m = e^{\left(\overline{\log t} - \frac{\psi(m)}{\beta_m}\right)},$$

  where $\overline{\log t} = \sum_{i=1}^n \log t_i/n$.

Note that the estimated parameters are based on the idea that $k = m$ is fixed. It is also indicated that the solution for θ_k and β_k is the same whether the parameter k is really equal to m or not or unknown, so the estimates can be used as starting values when the sampling is from a GG with all parameters unknown.

In a similar but extended approach, Stacy (1973) suggests that the score equations can be manipulated so as to obtain equations (2.12), (2.13) and

$$\psi(k) - \log k = n^{-1} \sum_{i=1}^{n} \log\left(\frac{t_i^\beta}{\sum_{i=1}^n t_i^\beta/n}\right).$$

In his paper, Stacy (1973) suggests fixing β in equation (2.13) to obtain an estimate for k as well as setting $z_i = t_i^\beta$ and $\lambda = \theta^\beta$. Consequently, the z_i values may be regarded as if obtained from a gamma distribution where t is replaced by z and β by λ. This leads to a case similar to that of the MLE of a two-parameter gamma distribution where k and λ are to be estimated instead of k and β. Parameter estimation depends on the fact that β is fixed in the first place.

2.3.2 The Prentice (1974) and Lawless (1980) Approaches

By extending and reparametrising the distribution of the logarithm of the GG random variable of Stacy (1962), Prentice (1974) showed that its underlying models are all embraced by a single parametric family. A regression generalisation is given. Prentice (1974) suggests that if survival time is given by T and $Y = \log T$, then the GG model of Stacy (1962) can be written in the linear form $y = \log\theta + \beta^{-1}v$ where the pdf of v is $\{1/\Gamma(k)\}\exp\{kv - e^v\}$ and the transformation $w_1 = k^{1/2}(v - \psi(k))$ leads to a standard normal for w_1 as $k \to \infty$. Manipulation of the model was completed in general by setting $q = k^{-c}$ (c being some positive constant). The resulting density function for $q = k^{-1/2}$ is Version 3 of Section 2.1,

$$f(y) = \begin{cases} \frac{|q|}{\sigma\Gamma(q^{-2})}\exp\{w_1 q^{-2} - e^{w_1}\} & \text{if } q \neq 0 \\ \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{y-\gamma}{\sigma}\right)^2\right\} & \text{if } q = 0, \end{cases} \tag{2.15}$$

where $w_1 = (y - \gamma)\sigma^{-1}q + \mu^*$, $y = \log t \in \mathbb{R}$, $\mu^* = \psi(q^{-2})$. γ is a location parameter $\gamma = \log\theta + \mu^*/\beta$, σ is a scale parameter $\sigma = q/\beta$, and q is a shape parameter. w_1, γ and $q \in \mathbb{R}$ while σ is positive.

MLE was studied via simulations in some special cases. Given $Y_1, ..., Y_n$ random variables from (2.15), the loglikelihood in the case $q \neq 0$ is

$$n\log|q| - n\log\sigma - n\log\left(\Gamma\left(q^{-2}\right)\right) + \sum_{i=1}^{n}\left(\frac{y_i - \gamma}{q\sigma}\right) + nq^{-2}\mu^* - \sum_{i=1}^{n}\exp\left\{\frac{q(y_i - \gamma)}{\sigma} + \mu^*\right\}.$$

It is indicated that direct manipulation of the likelihood equations ($\hat{q} \neq 0$) gives $\hat{\gamma} = \bar{y}$. The loglikelihood equations for σ and q, $q \neq 0$, are respectively

$$\sum_{i=1}^{n} \{(e^{v_i} - q^{-2})u_i\} = n \tag{2.16}$$

and

$$\sum_{i=1}^{n} \{(e^{v_i} - q^{-2})u_i\} - 2\sigma^{*2}q^{-2} \sum_{i=1}^{n} (e^{v_i} - q^{-2}) = n, \tag{2.17}$$

where $u_i = (y_i - \bar{y})q/\sigma$, $v_i = u_i + \mu^*$ and $\sigma^* = \psi'(q^{-2})$, ψ' being the derivative of the digamma function known as the trigamma function. Substituting (2.16) into (2.17), we deduce that $\sum_{i=1}^{n} (e^{v_i} - q^{-2}) = 0$. According to Prentice (1974), a simple Newton Raphson search for $\hat{\sigma}$ at fixed q values along with a tabulation of the maximised loglikelihood at these values gives an adequate determination of the loglikelihood estimates.

The final model (2.15) can be written as $y = \gamma + \sigma w_1$ where the error pdf has the form

$$f(w_1) = \begin{cases} \frac{|q|\left(q^{-2q^{-2}}\right)}{\Gamma(q^{-2})} \exp\{w_1 q^{-1} - e^{w_1 q}\} & \text{if } q \neq 0 \\ \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}w_1^2\right\} & \text{if } q = 0, \end{cases} \tag{2.18}$$

The effect of regression variables $x = (x_1, ..., x_r)$ in loglinear form gives $y = a + xb + \sigma w_1$ with w_1 following (2.18) and $b' = (b_1, ..., b_r)$ the regression coefficients. This model was referred to as the log-gamma regression model for y by Farewell and Prentice (1977) who study applications of the suggested model.

Lawless (1980) presented exact inference procedures for obtaining CIs or tests of significance for the parameters, quantiles and the reliability function of the logarithm of the GG distribution from uncensored samples when k is assumed known. The model was reparametrised in a way similar to the Prentice (1974) one. The proposed pdf has the form

$$f(y) = \frac{k^{k-\frac{1}{2}}}{\sigma\Gamma(k)} \exp\left\{\sqrt{k}w - ke^{\frac{w}{\sqrt{k}}}\right\}, \qquad y \in \mathbb{R}, \tag{2.19}$$

where $w = (y - \mu)/\sigma$ and $y = \log t \in \mathbb{R}$. μ is a location parameter $\mu = \log\theta + \log k/\beta$, σ is a scale parameter $\sigma = 1/(\beta\sqrt{k})$, and k is a shape parameter. w and $\mu \in \mathbb{R}$ while σ and k are positive. This is Version 4 of Section 2.1. Given $Y_1, ..., Y_n$ random variables from (2.19), the loglikelihood function is

$$n\left\{-\log(\sigma) + \left(k - \frac{1}{2}\right)\log(k) - \log\Gamma(k) + \sqrt{k}\frac{\overline{Y} - \mu}{\sigma} - \frac{k}{n}\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)\sum_{i=1}^{n}\exp\left(\frac{Y_i}{\sigma\sqrt{k}}\right)\right\}. \tag{2.20}$$

Differentiating with respect to μ and σ and setting the derivatives equal to zero resulted in the score equations from which expressions for μ and σ as a function of k were deduced:

$$e^{\mu} = \left\{\frac{1}{n}\sum_{i=1}^{n} e^{\frac{y_i}{\sigma\sqrt{k}}}\right\}^{\sigma\sqrt{k}} \tag{2.21}$$

and

$$\frac{\sum_{i=1}^{n} y_i e^{\left(\frac{y_i}{\sigma\sqrt{k}}\right)}}{\sum_{i=1}^{n} e^{\left(\frac{y_i}{\sigma\sqrt{k}}\right)}} - \overline{y} - \frac{\sigma}{\sqrt{k}} = 0. \tag{2.22}$$

The estimates $(\widehat{\mu}, \widehat{\sigma}, \widehat{k})$ of (μ, σ, k) were then obtained by taking fixed k values, then solving equation (2.22) iteratively for σ, and finally obtaining μ from equation (2.21). As the paper indicates, such results are important for two main reasons. The first is that good inference procedures are difficult to obtain with k assumed unknown. Making inference conditional on k, but doing this for a range of plausible k values, gives an informative picture. The second reason is that often a model with a particular value of k is actually analysed. Accuracy of some approximations from standard large sample ML methods were discussed. The method was programmed in FORTRAN by Hogg *et al.* (1982) and an application to real data was presented. Summary and further discussions were given in Lawless (1982) and later in the 2^{nd} edition in Lawless (2003). It is concluded that asymptotic normal approximations to the distributions of the ML estimates although somehow more convenient, produce less accurate

results than the likelihood ratio methods which produce somehow accurate results for moderate samples, but unacceptable results for small sample sizes.

2.3.3 Most Recent Contributions and Numerical Methods

Lefante and Turner (1985) derive the average likelihood (defined as the integral of the likelihood function over the parameter space) using a uniform prior when sampling from one-parameter members of a GG distribution.

Taking a loglinear model with one covariate and a GG model for the error and a noninformative prior based on the Jeffreys rule considering uncensored data, Achkar and Bolfarine (1986) found the posterior densities for the parameters of interest. They suggest using Lawless's (1982) approach for MLE. They claim that since many standard survival distributions are particular cases of the GG model, their proposed Bayesian method is very useful for discriminating between possible models used in data analysis.

Considering the difficulties in MLE of the GG distribution, Wingo (1987) proposed finding roots for (2.14) using the Root Isolation Method. The Root Isolation Method is a process of finding real intervals for the real roots of a function such that each interval contains exactly one real root and every real root is contained in some interval. Using the self-contained Fortran subroutine, ROOT, was suggested for this purpose. This subroutine returns the root intervals each of which is of width ϵ. Dividing the sum of the left and the right interval endpoints by two, yields the roots of $H(\beta)$ which is defined in equation (2.14). For each one of those roots, $k(\beta)$ and $\theta(\beta)$ are easily calculated from equations (2.13) and (2.12) respectively to obtain triplets $(\theta(\hat{\beta}), \hat{\beta}, k(\hat{\beta}))$. Each of those triplets is replaced in the loglikelihood and the one that renders the highest likelihood is considered as the ML parameter estimate. The case where $H(\beta)$ does not have root intervals is also taken into consideration.

DiCiccio (1987), following Lawless (1980), considers for the logarithm of the GG model adjustments to the usual likelihood ratio methods designed to improve their accuracy, particularly in small samples. Comparison between exact and approximate results is given.

Cohen and Whitten (1988) considered the loglikelihood function (2.8), while replacing the scale parameter θ by an alternative $\delta = \theta^{\beta}$. With this change, the loglikelihood function is

$$n\log\beta - n\log\Gamma(k) - nk\log\delta - \frac{1}{\delta}\sum_{i=1}^{n} t_i^{\beta} + (k\beta - 1)\sum_{i=1}^{n}\log t_i. \tag{2.23}$$

Differentiating (2.23) with respect to δ, k, and β in turn, yields the score equations

$$0 = -\frac{nk}{\delta} + \frac{1}{\delta^2}\sum_{i=1}^{n} t_i^{\beta}; \tag{2.24}$$

$$0 = -n\log\delta - n\psi(k) + \beta\sum_{i=1}^{n}\log t_i; \tag{2.25}$$

$$0 = \frac{n}{\beta} - \frac{1}{\delta}\sum_{i=1}^{n} t_i^{\beta}\log t_i + k\sum_{i=1}^{n}\log t_i. \tag{2.26}$$

Equation (2.24) yields

$$\delta = \frac{\sum_{i=1}^{n} t_i^{\beta}}{nk}. \tag{2.27}$$

Substituting this expression for δ into (2.26), yields

$$\sum_{i=1}^{n} t_i^{\beta} - \frac{nk\sum_{i=1}^{n} t_i^{\beta}\log t_i}{\frac{n}{\beta} + k\sum_{i=1}^{n}\log t_i} = 0. \tag{2.28}$$

Equation (2.25) yields

$$\beta = \frac{n\log\delta + n\psi(k)}{\sum_{i=1}^{n}\log t_i}. \tag{2.29}$$

Cohen and Whitten (1988) indicated that a solution does not exist in closed form and suggested the following:

1. Set $D(\delta, k, \beta) = -n\log\delta - n\psi(k) + \beta\sum_{i=1}^{n}\log t_i$.

2. Start with a first approximation β_1 for β.

3. For given β_1, compute an approximation k_1 for k from equation (2.28).

4. Substitute k_1 and β_1 into equation (2.27) to obtain the corresponding approximation δ_1 for δ.

5. Substitute the three approximations β_1, k_1, and δ_1 into equation (2.29). If the equation is satisfied, then $\widehat{\delta} = \delta_1$, $\widehat{k} = k_1$, $\widehat{\beta} = \beta_1$ and the calculations are complete. Otherwise, select a second approximation β_2, and repeat steps (2), (3) and (4) until a pair of approximations β_i and β_j are found in a sufficiently narrow interval such that

$$D(\widehat{\delta}_i, \widehat{k}_i, \widehat{\beta}_i) > 0 > D(\widehat{\delta}_j, \widehat{k}_j, \widehat{\beta}_j)$$

or

$$D(\widehat{\delta}_i, \widehat{k}_i, \widehat{\beta}_i) < 0 < D(\widehat{\delta}_j, \widehat{k}_j, \widehat{\beta}_j).$$

6. Finally, use the obtained estimates $\widehat{\beta}$ and $\widehat{\delta}$ to calculate $\widehat{\theta} = \widehat{\delta}^{\frac{1}{\beta}}$.

Later contributions include the work by Rao *et al.* (1991) who designed expressions for moments of order statistics from the GG distribution. More recent research was made in this regard such as the contribution by Balakrishnan (1995) using order statistics as a basis of estimation.

Wong (1993) presented two computational approaches for the MLE method suggested by Stacy and Mirham (1965) and developed computer programs for the computational procedures. In the first approach, two of the score equations are used for parameter estimation, whereas in the second one, an attempt to find all solutions to the system of three equations is presented.

Taking into account the computational difficulties in estimating a GG model, Hirose (2000) mentions that often enlarging the parameter space makes the numerical estimation more stable. A reparametrisation of the four-parameter Harter (1967) model is given. The reparametrisation

involves a new set of parameters (o, p, q, s) such that $k = q^{-2}$, $\beta = 1/s$, $d = o - p/|qs|$ and $\theta = p/|qs|$. The new model was called the extended four-parameter GG distribution. The continuation method in Allgower and Georg (1990) was suggested for solving the likelihood equations.

Tsionas (2001) considers a Bayesian analysis of the generalised four-parameter gamma distribution. Posterior inference from the pdf given in Stacy and Mirham (1965) was performed using numerical methods organised around Gibbs sampling.

Similarly to the work of Farewell and Prentice (1977), Ghilgaber (2005) analyses survival data from a GG perspective. The paper has a number of purposes. It describes how a range of parametric models such as the exponential, Weibull, and lognormal may be embedded in a single parametric framework, and how each competing model may be assessed relative to a more comprehensive one. Cox's proportional hazard estimation was also described. The final form of y in Farewell and Prentice (1977) was referred to as the extended generalised gamma EGG. Five models for t were included as special cases of the EGG model. Likelihood ratio statistics corresponding to various tests for special cases of the EGG model were presented. A natural question arises as to which procedure to use when one is confronted with a specific data analysis problem.

Balakrishnan and Peng (2006) presented a procedure to obtain ML estimates of the known parameters in the GG frailty model. The form of the pdf used is similar to the one suggested by Prentice (1974) and an approximate likelihood function is given.

Huang and Hwang (2006) claim that although the Prentice (1974) procedure is efficient, it is still quite complicated. They propose a simpler method using the GG characterisation and the moment estimation approach. They claim that their approach is efficient for small samples.

Kokkinakis and Nandi (2007) propose the GG distribution as a model basis for a family of flexible score functions for blind source separation, a promising application of Independent Component Analysis (ICA). Instead of solving the score equations to obtain the ML estimates, they maximise the likelihood using the Nelder-Mead method, a general-purpose optimisation procedure.

Also recently, Cox *et al.* (2007) presented a taxonomy of the hazard functions of the GG family. Using the Prentice (1974) parametrisation, they applied the proposed taxonomy to study the survival after a diagnosis of clinical AIDS during different eras of HIV therapy. For such computations, algorithms are now available in standard statistical packages such as **R** (2009) (Development Core Team), Stata and SAS. The aim was to consider regression models involving all three parameters and to compare them with all of the two-parameter subfamilies of the GG distribution (i.e. Weibull, lognormal, and gamma distributions) as well as with the semi-parametric proportional hazard models.

According to Gomes *et al.* (2008), estimation of the parameters of the GG is still an open topic; thus they propose a new heuristic approach to parameter estimation of the GG distribution using an iterative method. This routine was implemented in the SPLUS software. Because of the difficulties in applying the moments and MLE methods, they propose a new extension of these methods that uses goodness-of-fit tests to measure the degree of agreement between the distribution of a data sample and the theoretical distribution. The whole novelty of their procedure is based on the idea of transforming a gamma distribution to a GG distribution.

In their paper, Cooray and Ananda (2008) mention the following: "There is difficulty in developing inference procedures with the GG distribution, especially the ML parameter estimation in which the iteration method such as Newton-Raphson did not work." Alternatively, they derive the two-parameter generalised half-normal distribution (GHN), a special case of the

GG pdf form (2.7) (by taking $k = 1/2$ in the pdf Version 2 and setting $\eta = \theta 2^{1/\beta}$ and $\rho = \beta/2$) that inherits some of its significant properties. They argue that the computational difficulties faced with the GG do not affect this GHN distribution.

Song (2008) presents a fast and globally convergent algorithm for estimating the three-parameter GG distribution. Previous approaches for estimation are mentioned and their disadvantages and difficulties are highlighted. In sharp contrast to the previous methods, his method is constructed by raising the random variable to certain specially chosen powers and by expressing the shape parameters as a scale-free function of the shape parameter through appropriate expectation and derivative operations so that the resulting sample scale independent shape estimation SISE equations are completely independent of the gamma and polygamma functions which made computations easier according to the paper. We note that this paper does not use ML for estimation. It uses rather what is interpreted as "adaptive fractional moment methods or more generally as methods of nonlinear estimating equations".

A good reference for a summary of some of the previously mentioned approaches and some more studies on the estimation of a GG is the book by Johnson *et al.* (1994). Also, for more references and approaches involving the estimation of a GG model, we could mention Hager *et al.* (1971), Ortega *et al.* (2003), and Ortega *et al.* (2009).

2.4 Maximum Likelihood Estimation for the Distribution of the Log of a GG Random Variable

All attempts mentioned in the literature review to estimate the GG distribution showed computational difficulties and complications. Procedures based on MLE often assumed that the shape parameter (or one of the parameters) is considered known. Other procedures such as the method of moments have well-known disadvantages. With the rise of computers, recent investi-

gations most commonly developed algorithms based on MLE procedures proposed by previous researchers along with computational enhancements as well as applications and comparisons with simpler models. In this section, we present our contribution to MLE of a GG distribution overcoming the difficulties that came up in the literature. An iterative method for MLE of its parameters following the Lawless (1980) approach is developed considering all parameters unknown and none of them fixed. Sections 2.4 and 2.5 present a grander version of the work done on the three-parameter GG in Noufaily and Jones (2010) which rehabilitates MLE for the GG as being a perfectly reasonable and efficient approach.

2.4.1 Parameter Estimation Following Lawless (1980)

Having mentioned some of the basic properties of the GG distribution, and having presented previous researchers' attempts at estimating its parameters, we present in this section the approach we have followed for estimating the parameters of the GG distribution using pdf Version 4. We extended the work done by Lawless (1980) to solve the score equations in a simple and yet efficient way taking all parameters to be unknown and none of them fixed. For that, we look at the ML parameter estimation problem approached through the distribution of $Y_i = \log T_i$ parametrised as in Version 4

$$f(y) = \frac{k^{k-\frac{1}{2}}}{\sigma\Gamma(k)} \exp\left\{\sqrt{k}w - ke^{w/\sqrt{k}}\right\}, \qquad y \in \mathbb{R}$$

where $w = (y-\mu)/\sigma$. Note that $-\infty < \mu < \infty$ and $\sigma, k > 0$.

As in equation 2.20, the loglikelihood is

$$l = n\left\{-\log(\sigma) + \left(k-\frac{1}{2}\right)\log(k) - \log\Gamma(k) + \sqrt{k}\frac{\overline{Y}-\mu}{\sigma} - \frac{k}{n}\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)\sum_{i=1}^{n}\exp\left(\frac{Y_i}{\sigma\sqrt{k}}\right)\right\}. \tag{2.30}$$

Let

$$S_j \equiv \frac{1}{n}\sum_{i=1}^{n} Y_i^j \exp\left(\frac{Y_i}{\sqrt{k}\sigma}\right); \quad j = 0, 1, 2.$$

Differentiating the loglikelihood with respect to μ, σ and k in turn yields the score equations

$$0 = \frac{n\sqrt{k}}{\sigma}\left\{\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right) S_0 - 1\right\}, \tag{2.31}$$

$$0 = \frac{n\sqrt{k}}{\sigma^2}\left\{\exp\left(\frac{-\mu}{\sqrt{k}\sigma}\right)(S_1 - \mu S_0) - \frac{\sigma}{\sqrt{k}} - \overline{Y} + \mu\right\} \tag{2.32}$$

and

$$0 = n\left\{\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)\left(\frac{S_1 - \mu S_0}{2\sigma\sqrt{k}} - S_0\right) + \log k + 1 - \frac{1}{2k} - \psi(k) + \frac{\overline{Y} - \mu}{2\sigma\sqrt{k}}\right\}. \tag{2.33}$$

As in Lawless (1980), (2.31) yields

$$\exp(\mu) = S_0^{\sigma\sqrt{k}}, \tag{2.34}$$

an expression for μ in terms of k and σ. Using this, we can reduce (2.32) to

$$R(\sigma) \equiv \frac{S_1}{S_0} - \overline{Y} - \frac{\sigma}{\sqrt{k}} = 0. \tag{2.35}$$

Again following Lawless, we think of this as an equation in σ (note that S_0 and S_1 also depend on σ) for any k and solve it numerically. This is easy because we can show that $R(\sigma)$ is monotone decreasing:

$$\frac{\partial R(\sigma)}{\partial \sigma} = \frac{1}{\sqrt{k}}\left\{\frac{S_1^2 - S_0 S_2}{\sigma^2 S_0^2} - 1\right\};$$

both terms inside the curly brackets are negative, the first one by the Cauchy Schwartz inequality. Also, $\lim_{\sigma\to 0} R(\sigma) = Y_{\max} - \overline{Y} > 0$ where $Y_{\max}$ is the maximum of Y_i, $i = 1, ..., n$, and $\lim_{\sigma\to\infty} R(\sigma) = -\infty$. So there must be precisely one value of $\sigma > 0$ (for any k) for which $R(\sigma) = 0$. In fact, we can speed up our search a little by using the fact that this root will lie

in a certain interval. This is because the interpretation of S_1/S_0 as a mean of Y values implies that $S_1/S_0 < Y_{\max}$ and hence that if σ_0 is the root of $R(\sigma) = 0$, then it satisfies

$$0 < \sigma_0 = \sqrt{k}\left(\frac{S_1}{S_0}\bigg|_{\sigma_0} - \overline{Y}\right) < \sqrt{k}\{Y_{\max} - \overline{Y}\}. \tag{2.36}$$

We now turn our attention to (2.33). Using (2.34) and (2.35) to remove the quantities involving $\exp(-\mu/\sigma\sqrt{k})$, S_0 and S_1, we find that (2.33) reduces to

$$T(k) \equiv \frac{\overline{Y} - \mu}{\sigma\sqrt{k}} + \log(k) - \psi(k) = 0.$$

How does this behave as a function of k for fixed μ and σ? For small k, $\psi(k) \sim -\frac{1}{k}$ which goes to (minus) infinity faster than either of the other two terms and hence $\lim_{k\to 0} T(k) = \infty$. According to Abramowitz and Stegun (1965), for large k,

$$\psi(k) \sim \log(k) - \frac{1}{2k}$$

and so $\lim_{k\to\infty} T(k) = 0$. This limit is reached from either the positive or negative side, depending on the sign of $\overline{Y} - \mu$ (since $(\overline{Y} - \mu)/\sigma\sqrt{k}$ is the dominant term).

Now, we can apply Jensen's inequality to (2.34) [to the power $1/(\sqrt{k}\sigma)$, pretending that Y is a random variable which selects one of $Y_1/(\sqrt{k}\sigma), ..., Y_n/(\sqrt{k}\sigma)$ with probability $\frac{1}{n}$] to obtain

$$\exp\left(\frac{\mu}{\sqrt{k}\sigma}\right) = S_0 = \text{average}(\exp(Y)) > \exp(\text{average}(Y)) = \exp\left(\frac{\overline{Y}}{\sqrt{k}\sigma}\right)$$

or

$$\mu > \overline{Y}. \tag{2.37}$$

It follows that the limit is reached from the negative side.

So, it comes down to the behaviour of

$$T_L(k) \equiv \log(k) - \psi(k) - \frac{L}{\sqrt{k}} \tag{2.38}$$

where $L = (\mu - \overline{Y})/\sigma > 0$ provided that $\mu > \overline{Y}$ or follows from (2.37). $T_L(k)$ proves to have a single root k_0 in the interval $(1/4L^2, 1/L^2)$ as we now explain.

According to Dang and Weerakkoda (2000), if k_0 is a solution to $F(k) = \log(k) - \psi(k) - g = 0$, where $g > 0$, then

$$\frac{1}{2k_0} < g < \frac{1}{k_0}.$$

By applying this to $T_L(k)$ at $k = k_0$, we obtain

$$\frac{1}{2k_0} < \frac{L}{\sqrt{k_0}} < \frac{1}{k_0}. \tag{2.39}$$

Finally, rearranging inequalities (2.39), we find that

$$\frac{1}{4L^2} < k_0 < \frac{1}{L^2}. \tag{2.40}$$

2.4.2 Iterative Algorithm

We have now proved that each of the equations (2.35) and (2.38) has a unique root, the latter provided that $L > 0$. We will solve (2.34), (2.35) and (2.38) simultaneously and iteratively to obtain ML estimates, $\hat{\mu}$, $\hat{\sigma}$ and $\hat{k}$, of μ, σ and k, respectively. Our suggested algorithm is the following:

1. Set the iteration number i to 0 and obtain an initial guess for $L = L_0 > 0$.

2. Set $i = i + 1$.

3. For given L_{i-1}, compute $\widehat{k}_i$ by solving $T_L(k) = 0$ where $T_L(k)$ is given by (2.38) using either the bisection method or the Newton Raphson algorithm (we have used the former).

4. Replace the obtained $\widehat{k}_i$ in S_0 and S_1 and later in $R(\sigma)$ to compute $\widehat{\sigma}_i$ by solving (2.35) using the bisection method or the Newton Raphson algorithm (we have used the latter).

5. Substitute $\widehat{k}_i$ and $\widehat{\sigma}_i$ into (2.34) to obtain the corresponding $\widehat{\mu}_i$.

6. Use these estimates to obtain L_i and to compute the value of the loglikelihood function.

7. Repeat steps 2, 3, 4, 5 and 6 until desired accuracy of the likelihood is achieved.

Note the position of step 5 to guarantee the positivity of L_1.

When implementing the above algorithm in **R**, we added some computational devices to avoid crashes caused by large parameter values. We now display some of the computational tricks we have introduced to the aforementioned iterative algorithm to help **R** computationally and to increase the probability of the algorithm's success significantly.

- For $k > 171$, **R** reports infinity as an answer for $\Gamma(k)$, a fact that would cause our algorithm to crash. To overcome this computational difficulty, we set up a condition whenever $k > 171$. We used Stirling's formula to approximate the gamma function for large k. In reference to Abramowitz and Stegun (1965), Stirling's formula states that $\Gamma(k)$ can be approximated by $e^{-k}k^{k-1/2}\sqrt{2\pi}$ plus an error term which we ignored. By replacing this approximation for $\Gamma(k)$ in the loglikelihood, we will have avoided encountering infinite responses for large k values. As a result, the loglikelihood takes the form

$$n\left\{-\log(\sigma)+k-\frac{1}{2}\log(2\pi)+\sqrt{k}\frac{\overline{Y}-\mu}{\sigma}-\frac{k}{n}\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)\sum_{i=1}^{n}\exp\left(\frac{Y_i}{\sigma\sqrt{k}}\right)\right\}.$$

- Another common computational difficulty is faced when an exponential term is so large that **R** reports infinity as an answer. We have encountered this problem while computing the values of S_i. For example, to avoid the "infinities" in $R(\sigma)$, we suggest multiplying the ratio S_1/S_0 by the term $\exp(-Y_{max})/\exp(-Y_{max})$ where Y_{max} is the maximal Y_i $\forall$ $i = 1, ...n$. As a result, the ratio S_1/S_0 takes the form

$$\frac{y_1\exp((Y_1-Y_{\max})/\sigma\sqrt{k})+Y_2\exp((Y_2-Y_{\max})/\sigma\sqrt{k})+\cdots...+Y_n\exp((Y_n-Y_{\max})/\sigma\sqrt{k})}{\exp((Y_1-Y_{\max})/\sigma\sqrt{k})+\exp((Y_2-Y_{\max})/\sigma\sqrt{k})+\cdots...+\exp((Y_n-Y_{\max})/\sigma\sqrt{k})},$$

where the terms inside the exponential are, now, either negative or zero, but never infinity. Similarly, to compute μ using equation (2.34), we multiplied S_0 by $\exp(-Y_{\max})/\exp(-Y_{\max})$ and considered a logarithm on both sides of the equation; hence, avoiding infinities.

We note that our theoretical results above are very much still partial in the sense that they do not guarantee convergence of our algorithm nor uniqueness of the ML estimates. These issues will be explored numerically in Section 2.5 where simulation studies show that, computationally, MLE of the three-parameter GG is generally straightforward.

2.4.3 Bounds on Maximum Likelihood Estimates

Let $\hat{\mu}$, $\hat{\sigma}$ and $\hat{k}$ denote ML estimates of μ, σ and k respectively. Then, the inequalities (2.36) and (2.37) obtained in Subsection 2.4.1 for use at intermediate stages of the ML algorithm also apply to the ML estimates themselves. We therefore have the reassurance that

$$\overline{Y} < \hat{\mu} < Y_{\max}. \tag{2.41}$$

From (2.40),

$$\frac{1}{4}\frac{\hat{\sigma}^2}{(\hat{\mu}-\overline{Y})^2} < \hat{k} < \frac{\hat{\sigma}^2}{(\hat{\mu}-\overline{Y})^2}$$

and from (2.36),

$$0 < \hat{\sigma}^2 < \hat{k}(Y_{\max}-\overline{Y})^2$$

which combine to give

$$\hat{k}(\hat{\mu}-\overline{Y})^2 < \hat{\sigma}^2 < \hat{k}\min\left[\{2(\hat{\mu}-\overline{Y})\}^2, (Y_{\max}-\overline{Y})^2\right]. \tag{2.42}$$

2.5 Simulation Study Using the Iterative Algorithm in Subsection 2.4.2

2.5.1 Validity of Iterative Algorithm and Comparison with BFGS and Nelder-Mead

Based on the iterative algorithm proposed in Subsection 2.4.2, we developed a program in **R** that computes the estimates of a three-parameter GG distribution. We present now a simulation study where the aim is to show the validity and reliability of our suggested algorithm for MLE of the three-parameter GG regarding its behavior for different initial values and relative to other optimisation procedures. For this purpose, we compare it with other general-purpose optimisation methods such as the Nelder-Mead method introduced by Nelder and Mead (1965) and the Broyden-Fletcher-Goldfarb-Shanno method (BFGS) explained in Nocedal and Wright (1999). This is done via simulations. The **R** function, `optim`, is a general-purpose optimiser based on the Nelder Mead, quasi-Newton and conjugate-gradient algorithms. We used it to maximise the likelihood of the GG distribution for the Nelder-Mead and the BFGS methods.

We simulated 100 data sets from a GG and ran our program, Nelder Mead (from `optim`), and BFGS (from `optim`) for each. Then, we compared the results. A detailed explanation of the comparison we have done is presented in the next paragraph.

To start with, we simulated a set of "real parameters" consisting of 100 triplets (k, σ, and μ) from each of which we simulated a GG data set. We also simulated another set of "initial parameters" consisting of 100 triplets (k_0, σ_0, and μ_0.) Note that the sets "real parameters" and "initial parameters" are available in Tables A.1 and A.2 respectively of the Appendix. For every data set, we ran the three mentioned methods (our program, Nelder-Mead, and BFGS) 100 times each time using one of the triplets from the set of "initial parameters". Thus, we ran every method 10000 times in total. k and σ are positive and a logical range for their values

would be around the interval (0,6); therefore, we used a $G(\theta = 1, k = 2)$ to simulate k and σ for both sets of "real parameters" and "initial parameters". μ can be negative, therefore, we used a standard normal distribution to simulate its values for both sets of "real parameters" and "initial parameters". Note that, for our program, k_0 is not needed if $L_0 > 0$; if, by the random mechanisms above, $L_0 < 0$, we set $k_0 = 1/L_0^2$ and proceed from Step 4. This exercise was repeated with the same set of real and initial parameters for $n = 200$ and $n = 500$.

We estimated the parameters for each of the hundred data sets using our program, BFGS and Nelder-Mead. For each, we specified two stopping criteria; the convergence criterion relative tolerance (reltol) and the maximum number of iterations (maxit). What is meant by 'iteration' in our program is the implementation of the algorithm for steps 1 (or 2) till 7, that lead to calculating the resulting likelihood. The term 'iteration' in **R** generally refers to implementations that lead to an evaluation of the likelihood. This definition also depends on the method used for optimisation. The reltol is defined in **R** as being a set value for which the algorithm stops if it is unable to increase the evaluated likelihood by a factor of reltol(|likelihood| + reltol) at a step. For our program, the algorithm stops if it is unable to increase the evaluated likelihood by a factor of reltol at an iteration. For our program, we set reltol to be 10^{-8} and maxit to be 2000, whereas for both BFGS and Nelder-Mead, we set reltol to be 10^{-16} and maxit to be 10^9. After experimenting for a while, we decided that 2000 iterations are most of the time (excluding a few cases) enough to make our program converge to the global maximum. Increasing the number of iterations is very reasonable when dealing with one data set; however, it could mean taking much more time when dealing with hundreds of trials as in our study. For our program only, in addition to the stopping criteria mentioned above, we set two other stopping criteria for the Newton-Raphson algorithm and the bisection method used to solve the likelihood equations.

We allowed the root-finding methods to stop when they reached a relative accuracy of 10^{-4} and to report an error when more than 1000 iterations were needed for convergence. We considered a program to have failed to reach the global maximum of the likelihood if one of the following occurred:

a) The program crashed and reported an error.

b) A "non-global maximum" is attained rather than the global one. In most cases, this only means that the program has been stopped by reaching maxit or reltol when an increase in the number of iterations allowed or a decrease in the value of reltol would solve the problem (at a cost, of course, in time taken). In a few cases, the non-global maximum appears to be a local maximum of the likelihood surface.

c) The rendered value (by **R**) of the ML is, essentially, equal to negative infinity (-Inf).

With regard to item (b), the global maximum is taken to be the maximal likelihood attained for a data set over the 300 times its likelihood was maximised (by each of 3 methods from 100 sets of initial parameter values); the global maximum was said to be achieved in another run if the ML value was within 0.01 of the overall maximum. Regarding item (c), which happens very rarely, it is a result of the computed likelihood being very small (as close as zero) so that the loglikelihood is negative infinity. It happens basically when the algorithm cannot handle the maximisation properly and so results in unreasonable values.

In short, for each of the 100 generated data sets we ran the three methods 100 times, each time using a triplet of the simulated "initial parameters" keeping record of the likelihood and the estimated parameter values at every time. Tables 2.3 and A.3 present the results for the $n = 200$ and $n = 500$ cases respectively. Each row represents information obtained after running the three methods 100 times (300 in total) for a data set. "Likemax" is the value of the maximal likelihood attained for a data set among the 300 times. $\hat{k}$, $\hat{\sigma}$, and $\hat{\mu}$ are values of the

parameter estimates leading to "Likemax". " Our Program", "BFGS", and "NM" represent the number of times (out of 100) each of our program, BFGS, and Nelder-Mead respectively fails to reach "Likemax" up to an error equal to 0.01. The values in Tables 2.3 and A.3 are rounded to three decimal digits. We observe closeness between the real parameters and their estimates. In very rare cases, as in GG56, we observe that k is highly over-estimated though the resulting model is very similar to the real one since the GG is very similar for a range of large k values.

As for the behaviour of the algorithm for different initial parameters, Table 2.4 presents a summary of the total number of failures of the programs together with a closeup on the number of times each of the mentioned reasons contributes to the failures. Since algorithms generally cope less well with larger data sets, it is gratifying to observe little difference in performance between the two sample sizes. Overall, our program and the Nelder-Mead algorithm attain the global maximum likelihood the highest proportion of times (98.7% and 98.9% when $n = 200$), a little ahead of the BFGS algorithm (90.9% when $n = 200$). When $n = 500$, our program never crashed nor went to negative infinity; Nelder-Mead displayed a different pattern of failures. Local maxima of the likelihood, on the rare occasions they were observed, are both far from the global maximum and have much smaller values of the likelihood. We suspect that the reasons behind the occurrence of case (b) are either because a local maximum is attained or because larger stopping criteria are needed to ensure that the programs have converged to the global maximum. For example, increasing the value of maxit in our program (i.e. maxit=5000) would decrease slightly the total number of failures, but also unfortunately takes more time.

Based on the given results, we deduce that MLE of the three-parameter GG is manageable. Running one of the algorithms several times for different initial values ensures that a reasonable GG model is obtained. This fact will be explored more in further simulation studies where the quality of estimation will be analysed for the GG conditional distribution as well.

Table 2.3: Number of times each method fails to reach the maximal likelihood for $n = 200$.

Data	Likemax	$\hat{\mu}$	$\hat{\sigma}$	$\hat{k}$	Our Progam	BFGS	NM
GG1	-603.409	-0.339	4.658	2.781	1	32	2
GG2	-474.756	0.506	1.579	0.239	1	7	1
GG3	-477.518	2.195	2.202	0.871	1	12	1
GG4	-688.949	0.735	5.760	0.531	2	18	3
GG5	-358.865	0.569	1.167	0.688	1	8	1
GG6	-183.735	-0.279	0.570	2.652	1	0	1
GG7	-323.014	0.331	1.043	1.030	1	3	1
GG8	-481.662	-0.286	1.745	0.293	2	4	1
GG9	-402.972	-0.323	1.685	2.220	1	9	1
GG10	-362.185	0.808	1.438	5.806	1	12	1
GG11	-447.563	0.193	2.180	4.207	1	13	1
GG12	-529.599	-1.484	3.322	5.861	1	15	1
GG13	-335.679	-0.607	1.147	1.323	2	4	1
GG14	-407.534	0.533	1.761	3.136	2	6	1
GG15	-427.762	-1.016	1.826	1.374	1	5	1
GG16	-470.745	0.091	2.293	1.550	2	11	1
GG17	-446.064	-0.361	1.824	0.729	1	1	1
GG18	-464.722	0.520	2.339	3.013	1	12	1
GG19	-496.295	0.046	2.284	0.635	1	8	1
GG20	-566.459	-2.445	2.237	0.174	3	0	3
GG21	-177.735	-0.083	0.565	4.059	2	6	1
GG22	-337.657	1.756	1.176	1.517	1	16	1
GG23	-366.301	0.741	1.356	1.507	1	4	1
GG24	-294.208	0.933	1.024	5.835	1	8	1
GG25	-608.285	0.076	4.741	2.491	1	28	2
GG26	-204.559	0.465	0.567	0.914	1	5	1

Table 2.3 Continued

Data	Likemax	$\hat{\mu}$	$\hat{\sigma}$	$\hat{k}$	Our Progam	BFGS	NM
GG27	-603.346	0.501	4.442	1.523	2	28	2
GG28	-439.581	0.628	1.118	0.148	1	1	1
GG29	-556.788	-0.572	3.426	1.200	1	16	1
GG30	-415.145	-1.582	1.615	0.883	1	1	1
GG31	-308.643	-1.574	0.978	1.091	1	1	1
GG32	-403.932	-0.307	1.281	0.387	1	1	1
GG33	-296.926	-2.948	0.951	1.395	1	1	2
GG34	-648.889	1.313	5.134	0.819	2	8	2
GG35	-331.291	-1.198	1.226	4.895	2	5	1
GG36	-530.215	0.459	3.139	1.857	0	16	1
GG37	-493.105	-0.076	2.665	2.482	1	4	1
GG38	-525.317	0.249	2.978	1.391	2	17	1
GG39	-222.575	0.725	0.537	0.447	1	4	1
GG40	-474.738	0.544	2.494	4.058	1	14	1
GG41	-211.773	-1.434	0.490	0.388	1	0	0
GG42	-562.656	-0.518	3.862	3.851	2	21	1
GG43	-377.637	-0.057	1.500	2.586	0	10	1
GG44	-536.139	-0.738	3.175	1.526	1	13	1
GG45	-571.445	1.205	3.820	1.666	1	26	2
GG46	-365.512	-0.327	1.475	8.207	0	9	1
GG47	-267.595	0.207	0.810	1.235	3	1	1
GG48	-426.992	-0.591	1.592	0.592	1	3	1
GG49	-464.409	0.346	2.333	2.967	1	12	1
GG50	-419.638	-0.448	1.571	0.664	1	3	1
GG51	-119.814	-1.544	0.407	2.057	1	0	1
GG52	-456.288	0.833	2.289	4.848	1	20	1

Table 2.3 Continued

Data	Likemax	$\hat{\mu}$	$\hat{\sigma}$	$\hat{k}$	Our Progam	BFGS	NM
GG53	-444.963	0.886	2.127	3.238	1	17	1
GG54	-430.419	1.144	1.889	1.681	2	12	1
GG55	32.625	0.385	0.200	5.900	3	3	1
GG56	-486.705	1.204	2.736	20.949	3	23	1
GG57	-340.917	1.345	1.011	0.530	1	2	1
GG58	-509.434	0.310	2.658	1.055	2	14	1
GG59	-635.137	-1.700	4.896	0.935	3	14	1
GG60	-127.789	-0.842	0.405	1.308	2	0	1
GG61	-278.946	0.407	0.934	3.807	2	3	1
GG62	-634.879	1.724	3.164	0.176	1	10	1
GG63	-642.303	-0.523	4.512	0.506	3	10	1
GG64	-490.342	-0.756	2.528	1.545	1	4	1
GG65	-503.059	-1.367	2.779	2.212	1	9	1
GG66	-332.103	-1.211	1.119	1.249	1	1	1
GG67	-632.941	-0.755	4.535	0.643	2	18	1
GG68	-447.170	0.201	2.061	1.747	1	13	1
GG69	-472.879	-0.504	2.244	1.168	1	5	1
GG70	-324.897	0.047	1.078	1.237	0	1	1
GG71	-551.800	-0.208	1.327	0.058	4	2	4
GG72	-442.804	1.803	2.092	2.903	1	18	1
GG73	-402.315	-0.788	1.721	3.345	1	4	1
GG74	-569.875	-0.267	3.830	1.870	1	20	1
GG75	-246.196	-0.771	0.818	13.048	3	1	1
GG76	-606.253	-1.852	4.215	0.903	3	4	1
GG77	-219.974	0.236	0.667	1.898	2	4	1
GG78	-509.954	-0.062	2.805	1.643	1	10	1

Table 2.3 Continued

Data	Likemax	$\hat{\mu}$	$\hat{\sigma}$	$\hat{k}$	Our Progam	BFGS	NM
GG79	-367.747	0.084	1.450	3.442	2	10	1
GG80	-276.091	0.718	0.856	1.379	1	2	1
GG81	-316.033	-1.522	1.163	15.676	1	3	1
GG82	-579.212	-0.262	3.654	0.861	1	15	1
GG83	-414.687	1.316	1.810	2.698	2	13	1
GG84	-311.659	-0.093	1.106	4.312	1	5	1
GG85	-312.400	0.012	1.002	1.131	1	2	1
GG86	19.988	-0.826	0.192	1.228	2	1	0
GG87	-431.315	0.067	1.842	1.270	0	8	1
GG88	-427.902	0.837	1.890	1.948	1	14	1
GG89	-408.170	-1.757	1.738	2.380	1	3	1
GG90	-538.403	0.821	3.501	8.373	1	22	2
GG91	-344.514	-0.745	1.270	2.562	3	4	1
GG92	-354.591	1.989	1.280	1.517	1	14	1
GG93	-429.623	-0.731	1.983	3.738	1	10	1
GG94	-552.050	-1.356	3.562	2.317	1	11	1
GG95	-719.875	0.476	7.738	1.194	1	34	4
GG96	-640.854	1.901	5.054	0.954	1	29	3
GG97	-390.833	1.277	1.511	1.317	1	5	1
GG98	-113.550	-0.075	0.341	0.671	1	1	1
GG99	19.087	-1.218	0.164	0.489	1	1	0
GG100	-345.398	0.069	1.008	0.476	0	4	1
SUM	-	-	-	-	135	915	116

Table 2.4: Summary of results in Tables 2.3 and A.3.

	Our Program	BFGS	Nelder-Mead
$n = 200$			
Total number of failures	135	915	116
Total number of "non-global maxima"	134	827	23
Total number of reported errors	1	23	14
Total number of -Inf	0	65	79
$n = 500$			
Total number of failures	129	799	127
Total number of "non-global maxima"	129	711	32
Total number of reported errors	0	24	21
Total number of -Inf	0	64	74

2.5.2 Efficiency of the Iterative Algorithm

In this further preliminary simulation study, we test the efficiency of the algorithm in Subsection 2.4.2 by comparing the estimated parameters from our program to the true ones from which data sets are simulated. For the purpose of illustration, we simulate 10 data sets (each of size $n = 500$) from a $GG(\theta = 0.75, k, \beta = 2)$ distribution while varying k for each data set. Equivalently, we compute μ and σ of the GG pdf Version 4 via the expressions $\mu = \log\theta + \log k/\beta$ and $\sigma = 1/(\beta\sqrt{k})$. The values of the parameters from which the 10 data sets (GG1,...,GG10) were simulated, rounded to three decimal digits, are displayed in Table 2.5. Table 2.6 reports summary results of the computed estimates — of the 10 simulations in Table 2.5 — obtained using our suggested **R** program. By fixing θ and β and varying k from reasonably small to high values, we test whether the algorithm can account for different k values, knowing already that the shape of the GG pdf Version 4 changes only slightly as k gets larger.

The results are encouraging since they show reasonable closeness between the true values of the parameters in Table 2.5 and their estimates in Table 2.6. Estimates of smaller k values are quite accurate. Even when k is very large, we still manage to obtain a reasonable model. We note that these are point estimates obtained from our algorithm while using a set of 10 initial parameters. Taking into consideration more simulations will give a better idea on how good the estimates are.

Table 2.5: Parameter values for the 10 generalised gamma simulations.

Data	k	θ	β	μ	σ
GG1	0.05	0.75	2	-1.786	2.236
GG2	0.1	0.75	2	-1.439	1.581
GG3	0.5	0.75	2	-0.634	0.707
GG4	1	0.75	2	-0.288	0.500
GG5	1.5	0.75	2	-0.085	0.408
GG6	3	0.75	2	0.262	0.289
GG7	6	0.75	2	0.608	0.204
GG8	50	0.75	2	1.668	0.071
GG9	150	0.75	2	2.218	0.041
GG10	250	0.75	2	2.473	0.032

To check how close the estimates are to the true values, we then consider 50 simulations (each of size $n = 500$) from the particular $GG(\theta = 0.75, k = 0.5, \beta = 2)$ — or equivalently from the $GG(\mu = -0.634, \sigma = 0.707, k = 0.5)$ — we estimate the parameters for each data set using our program (repeated with 10 initial parameters where the results of the one with highest likelihood are displayed) and calculate the sample means of the estimates. Lastly, we obtain confidence intervals for the true estimates using the usual asymptotic normality assumptions. Table 2.7 presents a summary of the results. The values are rounded to three decimal digits.

Obviously, the true parameters lie within the 95% CIs. This shows that for reasonable parameter values, the GG model is quite accurate. Increasing the sample size allows, of course,

for more exactness of the estimates. Even for small samples, the algorithm still works reasonably well. A clearer idea on the accuracy of the estimates will be explored in later simulation studies. For example, in Section 5.1 we look at different case-scenarios of three-parameter GG simulations and we analyse the difference between the true and estimated parameters while taking 100 trials of initial parameters to obtain the estimates. Also, in Section 7.5 we look again at the estimates while comparing those of a GG to the estimates of another alternative distribution. Furthermore, in the next section, a study of the three-parameter GG asymptotics allows us to analyse the asymptotic correlations between the parameters, a fact that usually has a great effect on the quality of estimation.

Table 2.6: Parameter estimates of the 10 simulations in Table 2.5.

Data	$\widehat{k}$	$\widehat{\theta}$	$\widehat{\beta}$	$\widehat{\mu}$	$\widehat{\sigma}$
GG1	0.056	0.607	1.928	-1.990	2.183
GG2	0.160	0.526	1.301	-2.053	1.924
GG3	0.437	0.809	2.228	-0.583	0.678
GG4	1.162	0.659	1.800	-0.333	0.515
GG5	1.476	0.757	1.983	-0.082	0.415
GG6	3.541	0.657	1.835	0.268	0.290
GG7	8.661	0.482	1.634	0.592	0.208
GG8	10.404	3.143	4.449	1.672	0.070
GG9	212.330	0.345	1.632	2.217	0.042
GG10	168.762	1.533	2.502	2.477	0.031

Table 2.7: Confidence Intervals for the parameters of the 50 simulations.

$GG5$	$\widehat{k}$	$\widehat{\theta}$	$\widehat{\beta}$	$\widehat{\mu}$	$\widehat{\sigma}$
True	0.5	0.75	2	-0.634	0.707
Mean	0.5	0.75	2.074	-0.634	0.7
StDev	0.111	0.087	0.316	0.064	0.039
95% CI	[0.469, 0.530]	[0.726, 0.774]	[1.986, 2.161]	[−0.651, −0.616]	[0.689, 0.710]

2.6 Asymptotic Properties of the GG ML Parameter Estimates

Having shown the validity of our suggested algorithm for MLE of the three-parameter GG using pdf Version 4, we now present some asymptotic results for the parameter estimates and we deduce asymptotic correlations between $\hat{\mu}$, $\hat{\sigma}$ and $\hat{k}$. We also explore the asymptotics of the GG pdf Version 1. The correlation values between $\hat{\theta}$, $\hat{\alpha}$ and $\hat{\beta}$ are compared later in Subsection 7.4.2 to those of an alternative distribution to the GG. As will be explained in Chapter 7, the alternative distribution, which has the same parameters as the GG, has a special property that somehow specialises the roles of each of α and β to separate regions of the positive real line. It is indeed important to see how θ, α and β affect each distribution and to check whether in the alternative distribution their estimates are less correlated.

2.6.1 Asymptotics of the GG PDF Version 4

As is known, for large n (sample size), the GG ML parameter estimates satisfy the following:

$$\begin{pmatrix} \widehat{\mu} \\ \widehat{\sigma} \\ \widehat{k} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \sigma \\ k \end{pmatrix}, \mathbf{I}^{-1} \right)$$

where $\mathbf{I} = E(-\mathbf{Hessian})$ is the Fisher Information Matrix and

$$\mathbf{Hessian} = \begin{pmatrix} \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial^2 l}{\partial \mu \partial \sigma} & \frac{\partial^2 l}{\partial \mu \partial k} \\ \frac{\partial^2 l}{\partial \mu \partial \sigma} & \frac{\partial^2 l}{\partial \sigma^2} & \frac{\partial^2 l}{\partial \sigma \partial k} \\ \frac{\partial^2 l}{\partial \mu \partial k} & \frac{\partial^2 l}{\partial \sigma \partial k} & \frac{\partial^2 l}{\partial k^2} \end{pmatrix}.$$

Let

$$T_j \equiv \sum_{i=1}^{n} Y_i^j \exp\left(\frac{Y_i - \mu}{\sqrt{k}\sigma}\right); \quad j = 0, 1, 2.$$

Then, the second derivatives in the **Hessian** matrix are

$$\begin{aligned}
\frac{\partial^2 l}{\partial \mu^2} &= -\frac{1}{\sigma^2} T_0; \\
\frac{\partial^2 l}{\partial \mu \partial \sigma} &= \frac{\sqrt{k}}{\sigma^2}\left\{\left(\frac{\mu}{\sqrt{k}\sigma} - 1\right) T_0 - \left(\frac{1}{\sqrt{k}\sigma}\right) T_1 + n\right\}; \\
\frac{\partial^2 l}{\partial \mu \partial k} &= \frac{1}{2k\sigma^2}\left\{\left(\sqrt{k}\sigma + \mu\right) T_0 - T_1 - n\sqrt{k}\sigma\right\}; \\
\frac{\partial^2 l}{\partial \sigma^2} &= \frac{1}{\sigma^4}\left\{\left(2\mu\sqrt{k}\sigma - \mu^2\right) T_0 - \left(2\sqrt{k}\sigma - 2\mu\right) T_1 - T_2 + n\left(\sigma^2 + 2\sqrt{k}\sigma\left(\overline{Y} - \mu\right)\right)\right\}; \\
\frac{\partial^2 l}{\partial \sigma \partial k} &= \frac{1}{2k\sigma^3}\left\{-\left(\mu\sqrt{k}\sigma + \mu^2\right) T_0 + \left(\sqrt{k}\sigma + 2\mu\right) T_1 - T_2 - n\sqrt{k}\sigma\left(\overline{Y} - \mu\right)\right\}; \\
\frac{\partial^2 l}{\partial k^2} &= \left(\frac{1}{k\sqrt{k}\sigma}\right)\left\{\left(\frac{-\mu}{4} - \frac{\mu^2}{4\sqrt{k}\sigma}\right) T_0 + \left(\frac{\mu}{2\sqrt{k}\sigma} + \frac{1}{4}\right) T_1 - \frac{1}{4\sqrt{k}\sigma} T_2\right\} \\
&+ n\left\{\frac{1}{k} + \frac{1}{2k^2} - \psi'(k) - \frac{\overline{Y} - \mu}{4k\sqrt{k}\sigma}\right\}.
\end{aligned}$$

We are interested in finding the **Hessian** matrix at the ML estimates. Therefore, we first refer to the score equations (2.31) and (2.32) to find expressions for T_0 and T_1 respectively, as:

$$T_0 = n;$$

$$T_1 = n\left\{\frac{\sigma}{\sqrt{k}} + \overline{Y}\right\}.$$

Consequently, the second derivatives in the **Hessian** matrix at the ML estimates simplify to

$$\begin{aligned}
\frac{\partial^2 l}{\partial \mu^2} &= -\frac{n}{\sigma^2}; \\
\frac{\partial^2 l}{\partial \mu \partial \sigma} &= \frac{n}{\sigma^3}\left\{\mu - \frac{\sigma}{\sqrt{k}} - \overline{Y}\right\}; \\
\frac{\partial^2 l}{\partial \mu \partial k} &= \frac{n}{2k\sigma^2}\left\{\mu - \frac{\sigma}{\sqrt{k}} - \overline{Y}\right\}; \\
\frac{\partial^2 l}{\partial \sigma^2} &= \frac{n}{\sigma^4}\left\{2\mu\left(\frac{\sigma}{\sqrt{k}} + \overline{Y} - \frac{\mu}{2}\right) - \sigma^2 - \frac{1}{n} T_2\right\}; \\
\frac{\partial^2 l}{\partial \sigma \partial k} &= \frac{n}{2k\sigma^3}\left\{2\mu\left(\frac{\sigma}{\sqrt{k}} + \overline{Y} - \frac{\mu}{2}\right) + \sigma^2 - \frac{1}{n} T_2\right\}; \\
\frac{\partial^2 l}{\partial k^2} &= \frac{n}{k\sqrt{k}\sigma}\left\{\left(\frac{-\mu}{4} - \frac{\mu^2}{4\sqrt{k}\sigma}\right) + \left(\frac{\mu}{2\sqrt{k}\sigma} + \frac{1}{4}\right)\left(\frac{\sigma}{\sqrt{k}} + \overline{Y}\right) - \frac{1}{n4\sqrt{k}\sigma} T_2\right\} \\
&+ n\left\{\frac{1}{k} + \frac{1}{2k^2} - \psi'(k) - \frac{\overline{Y} - \mu}{4k\sqrt{k}\sigma}\right\}.
\end{aligned}$$

Knowing that:

$$\begin{aligned}
\Gamma(k) &= \int_0^\infty y^{k-1}e^{-y}dy;\\
\Gamma'(k) &= \int_0^\infty \log(y)y^{k-1}e^{-y}dy;\\
\Gamma''(k) &= \int_0^\infty (\log y)^2 y^{k-1}e^{-y}dy;\\
\Gamma'(k+1) &= \Gamma(k)+k\Gamma'(k) = \int_0^\infty \log(y)y^{k}e^{-y}dy;\\
\Gamma''(k+1) &= 2\Gamma'(k)+k\Gamma''(k) = \int_0^\infty (\log y)^2 y^{k}e^{-y}dy;\\
\psi(k) &= \frac{\Gamma'(k)}{\Gamma(k)};\\
\psi'(k) &= \frac{\Gamma''(k)}{\Gamma(k)} - \psi(k)^2,
\end{aligned}$$

and setting $w_i = (y_i - \mu)/\sigma$, we find the expectation of $\overline{Y}$ and T_2 to be

$$E(\overline{Y}) = \frac{1}{n}\frac{k^{k-\frac{1}{2}}}{\sigma\Gamma(k)}\sum_{i=1}^{n}\int_0^\infty y_i \exp\left\{\sqrt{k}w_i - ke^{\frac{w_i}{\sqrt{k}}}\right\}dy_i = \sqrt{k}\sigma\{\psi(k) - \log k\} + \mu$$

and

$$\begin{aligned}
E(T_2) &= \frac{1}{n}\frac{k^{k-\frac{1}{2}}}{\sigma\Gamma(k)}\sum_{i=1}^{n}\int_0^\infty y_i^2 e^{\frac{w_i}{\sqrt{k}}}\exp\left\{\sqrt{k}w_i - ke^{\frac{w_i}{\sqrt{k}}}\right\}dy_i\\
&= n\left\{\sigma\log k\left(\sigma k\log k - 2\sigma - 2\mu\sqrt{k}\right) + 2\sigma\psi(k)\left(\mu\sqrt{k}+\sigma-\sigma k\log k\right)\right.\\
&+ \left. k\sigma^2\frac{\Gamma''(k)}{\Gamma(k)} + \mu^2 + \frac{2\sigma\mu}{\sqrt{k}}\right\},
\end{aligned}$$

where $\Gamma''(k)/\Gamma(k)$ can be replaced by $\psi'(k) + \psi(k)^2$.

Let

$$\begin{aligned}
f_1(k) &= \sqrt{k}\left\{\psi(k) - \log k + \frac{1}{k}\right\};\\
f_2(k) &= \frac{1}{2\sqrt{k}}\left\{\psi(k) - \log k + \frac{1}{k}\right\};\\
g_1(k) &= k\frac{\Gamma''(k)}{\Gamma(k)} + 2\psi(k)(1 - k\log k) + \log k(k\log k - 2) + 1;\\
g_2(k) &= \frac{1}{2k}\left\{k\frac{\Gamma''(k)}{\Gamma(k)} + 2\psi(k)(1-k\log k) + \log k(k\log k - 2) - 1\right\};\\
g_3(k) &= \frac{1}{4k^2}\left\{k\frac{\Gamma''(k)}{\Gamma(k)} + 4k^2\psi'(k) + 2\psi(k)(1-k\log k) + \log k(k\log k - 2) - 4k - 3\right\}.
\end{aligned}$$

Given the above — and using that the score function has expectation zero — we deduce the components of the matrix $\mathbf{I}$ to be

$$\begin{aligned}
E\left(-\frac{\partial^2 l}{\partial \mu^2}\right) &= \frac{n}{\sigma^2};\\
E\left(-\frac{\partial^2 l}{\partial \mu \partial \sigma}\right) &= \frac{n}{\sigma^2} f_1(k);\\
E\left(-\frac{\partial^2 l}{\partial \mu \partial k}\right) &= \frac{n}{\sigma} f_2(k);\\
E\left(-\frac{\partial^2 l}{\partial \sigma^2}\right) &= \frac{n}{\sigma^2} g_1(k);\\
E\left(-\frac{\partial^2 l}{\partial \sigma \partial k}\right) &= \frac{n}{\sigma} g_2(k);\\
E\left(-\frac{\partial^2 l}{\partial k^2}\right) &= n g_3(k).
\end{aligned}$$

Therefore, the matrix $\mathbf{I}$ and its determinant are asymptotically respectively

$$\mathbf{I} = n \begin{pmatrix} \frac{1}{\sigma^2} & \frac{1}{\sigma^2} f_1(k) & \frac{1}{\sigma} f_2(k) \\ \frac{1}{\sigma^2} f_1(k) & \frac{1}{\sigma^2} g_1(k) & \frac{1}{\sigma} g_2(k) \\ \frac{1}{\sigma} f_2(k) & \frac{1}{\sigma} g_2(k) & g_3(k) \end{pmatrix}$$

and

$$\begin{aligned}
\det(\mathbf{I}) = \frac{n^3}{\sigma^4} \{ & g_1(k) g_3(k) - g_2(k)^2 \; - \; f_1(k) \left[f_1(k) g_3(k) - f_2(k) g_2(k) \right] \\
& + \; f_2(k) \left[f_1(k) g_2(k) - g_1(k) f_2(k) \right] \}.
\end{aligned}$$

Hence, the asymptotic variance covariance matrix is

$$\mathbf{V} = \mathbf{I}^{-1} = (v_{ij}) \; ; \quad i = 1,2,3 \text{ and } j = 1,2,3 :$$

$$\begin{aligned}
v_{11} &= \frac{1}{\det(I)\sigma^2 n} \{ g_1(k) g_3(k) - g_2(k)^2 \};\\
v_{12} &= \frac{1}{\det(I)\sigma^2 n} \{ f_2(k) g_2(k) - f_1(k) g_3(k) \};\\
v_{13} &= \frac{1}{\det(I)\sigma^3 n} \{ f_1(k) g_2(k) - f_2(k) g_1(k) \};\\
v_{22} &= \frac{1}{\det(I)\sigma^2 n} \{ g_3(k) - f_2(k)^2 \};\\
v_{23} &= \frac{1}{\det(I)\sigma^3 n} \{ f_1(k) f_2(k) - g_2(k) \};\\
v_{33} &= \frac{1}{\det(I)\sigma^4 n} \{ g_1(k) - f_1(k)^2 \}.
\end{aligned}$$

Let

$$p = \psi(k) - \log k + \frac{1}{k};$$

$$q = k\frac{\Gamma''(k)}{\Gamma(k)} + 2\psi(k)\{1 - k\log k\} + \log k\{k\log k - 2\};$$

$$r = k^2\psi'(k) - k.$$

We can now express the asymptotic correlations between $\hat{\mu}$, $\hat{\sigma}$ and $\hat{k}$ as the following:

- The correlation between the parameters μ and σ is

$$\begin{aligned}\text{Corr}(\hat{\mu},\hat{\sigma}) = \frac{v_{12}}{\sqrt{v_{11}v_{22}}} &= \frac{\{f_2(k)g_2(k) - f_1(k)g_3(k)\}}{\sqrt{\{g_1(k)g_3(k) - g_2(k)^2\}\{g_3(k) - f_2(k)^2\}}} \\ &= \frac{\{2p(1/2 - r)\}}{\sqrt{\{qr + r - 1\}\{(1/k)(q + 4r - 3) - p^2\}}}.\end{aligned}$$

- The correlation between the parameters μ and k is

$$\begin{aligned}\text{Corr}(\hat{\mu},\hat{k}) = \frac{v_{13}}{\sqrt{v_{11}v_{33}}} &= \frac{\{f_1(k)g_2(k) - f_2(k)g_1(k)\}}{\sqrt{\{g_1(k)g_3(k) - g_2(k)^2\}\{g_1(k) - f_1(k)^2\}}} \\ &= \frac{\{-\sqrt{k}p\}}{\sqrt{\{qr + r - 1\}\{q + 1 - kp^2\}}}.\end{aligned}$$

- The correlation between the parameters k and σ is

$$\begin{aligned}\text{Corr}(\hat{k},\hat{\sigma}) = \frac{v_{23}}{\sqrt{v_{22}v_{33}}} &= \frac{\{f_1(k)f_2(k) - g_2(k)\}}{\sqrt{\{g_3(k) - f_2(k)^2\}\{g_1(k) - f_1(k)^2\}}} \\ &= \frac{\{p^2 - (1/k)(q - 1)\}}{\sqrt{(1/k)\{(1/k)(q + 4r - 3) - p^2\}\{q + 1 - kp^2\}}}.\end{aligned}$$

It is obvious that the correlation matrix is independent of σ, μ and n (the sample size); it is only dependent on k. Let us look at the values of the correlations between $\hat{\mu}$, $\hat{\sigma}$, and $\hat{k}$ as k tends to infinity and as it tends to zero.

- We will start with the case where $k \to 0$. For small k, $\psi(k) \sim -1/k$ and $\psi'(k) \sim 1/k^2$; therefore, $p \sim -\log(k)$, $q \sim k\log^2 k$ and $r \sim 1-k$ leading to

$$\lim_{k\to 0} \text{Corr}(\hat{\mu},\hat{\sigma}) \sim \lim_{k\to 0} \frac{\log(k)(1-2k)}{|\log(k)|\sqrt{k^2-2k+1+(k-1)/\log(k)^2}} = -1;$$

$$\lim_{k\to 0} \text{Corr}(\hat{\mu},\hat{k}) \sim \lim_{k\to 0} \frac{\log(k)}{|\log(k)|\sqrt{1-k-1/\log(k)^2}} = -1;$$

$$\lim_{k\to 0} \text{Corr}(\hat{\sigma},\hat{k}) \sim \lim_{k\to 0} \frac{1}{\sqrt{1-4k}} = 1.$$

- Consider, now, the case where $k \to \infty$. For large k, $\psi(k) \sim \log(k) - 1/2k - 1/12k^2$ and $\psi'(k) \sim 1/k+1/2k^2+1/6k^3$; therefore, $p \sim 1/2k-1/12k^2$, $q \sim 1-1/4k+1/12k^2+1/144k^3$ and $r \sim 1/6k + 1/2$ leading to

$$\begin{aligned}
\lim_{k\to\infty} \text{Corr}(\hat{\mu},\hat{\sigma}) &\sim \lim_{k\to\infty} \frac{1/36k^3 - 1/6k^2}{\sqrt{\{5/24k + 5/288k^3 + 1/864k^4\}\{3/2k - 1/3k^2 + 1/6k^3\}}} \\
&\sim \lim_{k\to\infty} \frac{-1/6k^2}{\sqrt{5/16k}} = 0;
\end{aligned}$$

$$\begin{aligned}
\lim_{k\to\infty} \text{Corr}(\hat{\mu},\hat{k}) &\sim \lim_{k\to\infty} \frac{1/2\sqrt{k}(1/6k-1)}{\sqrt{\{5/24k + 5/288k^3 + 1/864k^4\}\{2 - 1/2k + 1/6k^2\}}} \\
&\sim \lim_{k\to\infty} \frac{-1/2\sqrt{k}}{\sqrt{5/12k}} \\
&= -\sqrt{\frac{3}{5}} \\
&\sim -0.7745967;
\end{aligned}$$

$$\begin{aligned}
\lim_{k\to\infty} \text{Corr}(\hat{\sigma},\hat{k}) &\sim \lim_{k\to\infty} \frac{1/2k^2 - 1/6k^3}{\sqrt{\{1/6k^3 + 1/6k^4\}\{2 - 1/2k + 1/6k^2\}}} \\
&\sim \lim_{k\to\infty} \frac{1/2k^2}{\sqrt{1/3k^{3/2}}} \\
&= 0.
\end{aligned}$$

It is not clear yet what sample size ensures that the asymptotics are reached. These asymptotic results can be checked by a simulation study that provides insight about sample dimension adequacy. Such a study is not given in this thesis, however our aim is to investigate it

in future work. The previous results of the correlations are confirmed and illustrated in Tables 2.8, 2.9 and 2.10. The correlation values are rounded to three decimal digits. Notice that in all three Tables 2.8, 2.9 and 2.10, as k increases, the correlations unsurprisingly decrease. In Table 2.8, a negative correlation exists between $\hat{\mu}$ and $\hat{\sigma}$. At $k = 0.01$, $\text{Corr}(\hat{\mu}, \hat{\sigma}) = -0.978$ which confirms the previous results that the correlation tends to minus one for very small k. Also, for $k = 1000$, $\text{Corr}(\hat{\mu}, \hat{\sigma}) = -0.028$ showing that as k increases, $\text{Corr}(\hat{\mu}, \hat{\sigma})$ gets closer to zero. Note that for $k = 2$, $\text{Corr}(\hat{\mu}, \hat{\sigma}) = -0.5$ midway between 0 and 1. Table 2.9 shows that as k increases from 0.01 to 1000, $\text{Corr}(\hat{\mu}, \hat{k})$ decreases (in absolute value) from -0.978 (very close to -1) to -0.775 (approximately $-\sqrt{3/5}$). Table 2.10 shows that $\hat{\sigma}$ and $\hat{k}$ are positively correlated with the correlation decreasing from 0.987 to 0.027 as k goes from 0.01 to 1000. Also, for $k = 2$, $\text{Corr}(\hat{\sigma}, \hat{k}) = 0.476$ which is about midway between 0 and 1. These results are also confirmed in the above limits. Overall, we can say that the results are quite reasonable. The correlations are small or moderate except when k is small.

Table 2.8: Correlation between $\hat{\mu}$ and $\hat{\sigma}$ for eight k values.

k	0.01	0.1	1	2	5	9	100	1000
$\text{Corr}(\hat{\mu}, \hat{\sigma})$	-0.978	-0.907	-0.616	-0.5	-0.357	-0.279	-0.089	-0.028

Table 2.9: Correlation between $\hat{\mu}$ and $\hat{k}$ for eight k values.

k	0.01	0.1	1	2	5	9	100	1000
$\text{Corr}(\hat{\mu}, \hat{k})$	-0.978	-0.916	-0.785	-0.772	-0.77	-0.772	-0.774	-0.775

Table 2.10: Correlation between $\hat{\sigma}$ and $\hat{k}$ for eight k values.

k	0.01	0.1	1	2	5	9	100	1000
$\text{Corr}(\hat{\sigma}, \hat{k})$	0.987	0.895	0.584	0.476	0.343	0.269	0.086	0.027

2.6.2 Asymptotics of the GG PDF Version 1

The aim in this subsection is to study the asymptotics of the GG pdf Version 1 ML parameter estimates. The purpose is to obtain asymptotic correlations between $\hat{\theta}$, $\hat{\alpha}$ and $\hat{\beta}$. As already mentioned, we would like to compare the correlations between $\hat{\theta}$, $\hat{\alpha}$ and $\hat{\beta}$ of the GG distribution with those of its alternative distribution later. Let $T_1, ..., T_n$ be n independent identically distributed random variables from a GG distribution with pdf Version 1. The loglikelihood is

$$n\log\beta - n\alpha\log\theta - n\log\Gamma(\alpha/\beta) + (\alpha-1)\sum_{i=1}^{n}\log t_i - \frac{1}{\theta^\beta}\sum_{i=1}^{n} t_i^\beta.$$

We find that the expressions of the score equations are

$$0 = \frac{n}{\theta}\left\{-\alpha + \frac{\beta}{n}\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^\beta\right\}, \tag{2.43}$$

$$0 = n\left\{-\log\theta + \frac{1}{n}\sum_{i=1}^{n}\log t_i - \frac{1}{\beta}\psi(\alpha/\beta)\right\} \tag{2.44}$$

and

$$0 = n\left\{\frac{1}{\beta} - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^\beta \log\left(\frac{t_i}{\theta}\right) + \frac{\alpha}{\beta^2}\psi(\alpha/\beta)\right\}. \tag{2.45}$$

The second derivatives in the **Hessian** matrix turn out to be

$$\begin{aligned}
\frac{\partial^2 l}{\partial\theta^2} &= \frac{n}{\theta^2}\left\{\alpha - \frac{\beta(\beta+1)}{n}\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^\beta\right\};\\
\frac{\partial^2 l}{\partial\theta\partial\alpha} &= -\frac{n}{\theta};\\
\frac{\partial^2 l}{\partial\theta\partial\beta} &= \frac{n}{\theta}\left\{\frac{\beta}{n}\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^\beta\log\left(\frac{t_i}{\theta}\right) + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^\beta\right\};\\
\frac{\partial^2 l}{\partial\alpha^2} &= -\frac{n}{\beta^2}\psi'(\alpha/\beta);\\
\frac{\partial^2 l}{\partial\alpha\partial\beta} &= n\left\{\frac{\alpha}{\beta^3}\psi'(\alpha/\beta) + \frac{1}{\beta^2}\psi(\alpha/\beta)\right\};\\
\frac{\partial^2 l}{\partial\beta^2} &= n\left\{-\frac{1}{\beta^2} - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^\beta\log\left(\frac{t_i}{\theta}\right)^2 - \frac{\alpha^2}{\beta^4}\psi'(\alpha/\beta) - \frac{2\alpha}{\beta^3}\psi(\alpha/\beta)\right\}.
\end{aligned}$$

From the score equations (2.43) and (2.45) respectively, we obtain

$$\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta} = n\frac{\alpha}{\beta}$$

and

$$\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta}\log\left(\frac{t_i}{\theta}\right) = \frac{n}{\beta} + n\frac{\alpha}{\beta^2}\psi(\alpha/\beta).$$

Therefore, at the ML estimates, the second derivatives in the **Hessian** simplify to

$$\begin{aligned}
\frac{\partial^2 l}{\partial\theta^2} &= -\frac{n}{\theta^2}\{\alpha\beta\};\\
\frac{\partial^2 l}{\partial\theta\partial\alpha} &= -\frac{n}{\theta};\\
\frac{\partial^2 l}{\partial\theta\partial\beta} &= -\frac{n}{\theta}\left\{-1-\frac{\alpha}{\beta}-\frac{\alpha}{\beta}\psi(\alpha/\beta)\right\};\\
\frac{\partial^2 l}{\partial\alpha^2} &= -\frac{n}{\beta^2}\psi'(\alpha/\beta);\\
\frac{\partial^2 l}{\partial\alpha\partial\beta} &= -n\left\{-\frac{\alpha}{\beta^3}\psi'(\alpha/\beta)-\frac{1}{\beta^2}\psi(\alpha/\beta)\right\};\\
\frac{\partial^2 l}{\partial\beta^2} &= -n\left\{\frac{1}{\beta^2}+\frac{1}{n}\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta}\log\left(\frac{t_i}{\theta}\right)^2+\frac{\alpha^2}{\beta^4}\psi'(\alpha/\beta)+\frac{2\alpha}{\beta^3}\psi(\alpha/\beta)\right\}.
\end{aligned}$$

Given the above — and using that the score function has expectation zero — we deduce the components of the matrix **I** to be

$$\begin{aligned}
E\left(-\frac{\partial^2 l}{\partial\theta^2}\right) &= \frac{n}{\theta^2}\alpha\beta;\\
E\left(-\frac{\partial^2 l}{\partial\theta\partial\alpha}\right) &= \frac{n}{\theta};\\
E\left(-\frac{\partial^2 l}{\partial\theta\partial\beta}\right) &= \frac{n}{\theta}\left\{-1-\frac{\alpha}{\beta}-\frac{\alpha}{\beta}\psi(\alpha/\beta)\right\};\\
E\left(-\frac{\partial^2 l}{\partial\alpha^2}\right) &= n\left\{\frac{1}{\beta^2}\psi'(\alpha/\beta)\right\};\\
E\left(-\frac{\partial^2 l}{\partial\alpha\partial\beta}\right) &= n\left\{-\frac{\alpha}{\beta^3}\psi'(\alpha/\beta)-\frac{1}{\beta^2}\psi(\alpha/\beta)\right\};\\
E\left(-\frac{\partial^2 l}{\partial\beta^2}\right) &= \left\{\frac{1}{\beta^2}+\left(\frac{\alpha}{\beta^3}+\frac{\alpha^2}{\beta^4}\right)\psi'(\alpha/\beta)+\left(\frac{2}{\beta^2}+2\frac{\alpha}{\beta^3}\right)\psi(\alpha/\beta)+\frac{\alpha}{\beta^3}\psi(\alpha/\beta)^2\right\}.
\end{aligned}$$

Knowing the form of the **Hessian** matrix, similarly to the previous section, we then deduce the components of the asymptotic covariance and correlation matrices. The correlations turn out to be, unsurprisingly, independent of θ. Tables 2.11, 2.12, 2.13, 2.14, 2.15 and 2.16 display values of the correlations between $\hat{\theta}$, $\hat{\alpha}$ and $\hat{\beta}$ for variable α and β. The values are rounded to three decimal digits. We observe that $\hat{\theta}$ and $\hat{\alpha}$ are negatively correlated as well as $\hat{\alpha}$ and $\hat{\beta}$. On the other hand, $\hat{\theta}$ and $\hat{\beta}$ are positively correlated. For fixed β, as α increases, $\text{Corr}(\hat{\theta}, \hat{\alpha})$ increases (in absolute value) giving the impression that it converges to zero for very small α values and rises up towards -1 at higher α values. Now as we fix α and we increase β, $\text{Corr}(\hat{\theta}, \hat{\alpha})$ decreases (in absolute value) from values around -1 to values around zero. A similar pattern happens for $\text{Corr}(\hat{\alpha}, \hat{\beta})$. As for $\text{Corr}(\hat{\theta}, \hat{\beta})$, when we fix β and increase α from 0.001 to 5, the correlation increases to about 1 at 5. Finally, if we fix α and increase β, $\text{Corr}(\hat{\theta}, \hat{\beta})$ decreases from values around 1 at 0.1. Overall, we can say that the correlations are reasonably small or moderate for reasonable values of α and β. Generally, the correlations are smaller for smaller α values and larger β values. In Subsection 7.4.2, we will see how these values compare with those of an alternative distribution to the GG.

Table 2.11: Correlations between $\hat{\theta}$ and $\hat{\alpha}$ for $\beta = 2$ and variable α.

α	0.001	0.01	0.1	0.5	1	2	5
$\text{Corr}(\hat{\theta}, \hat{\alpha})$	-0.033	-0.104	-0.332	-0.680	-0.817	-0.904	-0.961

Table 2.12: Correlations between $\hat{\theta}$ and $\hat{\alpha}$ for $\alpha = 3$ and variable β.

β	0.1	1	2	5	10	100	1000
$\text{Corr}(\hat{\theta}, \hat{\alpha})$	-0.996	-0.967	-0.935	-0.845	-0.720	-0.257	-0.080

Table 2.13: Correlation between $\hat{\theta}$ and $\hat{\beta}$ for $\beta = 2$ and variable α.

α	0.001	0.1	0.2	0.5	1	2	5
$\text{Corr}(\hat{\theta}, \hat{\beta})$	0.467	0.581	0.670	0.826	0.921	0.970	0.993

Table 2.14: Correlation between $\hat{\theta}$ and $\hat{\beta}$ for $\alpha = 3$ and variable β.

β	0.1	1	5	10	10^2	10^3	10^4
$\text{Corr}(\hat{\theta}, \hat{\beta})$	1	0.995	0.938	0.855	0.539	0.474	0.466

Table 2.15: Correlation between $\hat{\alpha}$ and $\hat{\beta}$ for $\beta = 2$ and variable α.

α	0.01	0.1	0.5	1	3	5	10
$\text{Corr}(\hat{\alpha}, \hat{\beta})$	-0.088	-0.271	-0.547	-0.693	-0.877	-0.927	-0.964

Table 2.16: Correlation between $\hat{\alpha}$ and $\hat{\beta}$ for $\alpha = 3$ and variable β.

β	0.1	1	2	5	10	100	1000
$\text{Corr}(\hat{\alpha}, \hat{\beta})$	-0.994	−0.939	-0.877	−0.730	-0.586	-0.212	-0.068

This chapter presented an introduction to and a deep analysis of the three-parameter GG distribution. Aiming to go beyond the limitations of the three-parameter model when a covariate is present, we explore, in the next two chapters, a wider family of the GG which we make available through using QR.

Chapter 3

Parametric Quantile Regression with Generalised Gamma I

Most commonly, researchers try to explain regression data through one curve fitted to the whole data set. However, not all data sets are the same everywhere and observations might behave differently at different percentiles. From here comes the idea of QR aimed at modeling data at every quantile.

QR has recently seen a great increase in practical applications especially in modeling life time data where researchers need to find upper and lower limits (quantiles) within the data beyond which data points are considered outliers (or uncommon). The main objective of this study is using QR to obtain reference charts that researchers can refer to in positive-valued data sets to make decisions. Reference charts are used especially in medicine and pharmacology to help specialists decide whether patients lie within or outside the most common range or the so-called "normal range".

We consider a promising parametric approach to QR using the GG distribution. We extend the work done on the three-parameter GG to a wider model in the context of QR. This involves estimating the four-, five- and six-parameter GG obtained by making the parameters dependent on a univariate covariate. The quantiles, which are functions of the parameters, are

hence functions of the covariate as well. In Section 3.1, a general overview on QR is presented. Section 3.2 offers a literature review of QR. In this framework, the GG model is explored in Section 3.3. Shapes of its quantiles are studied in different cases in Section 3.4.

3.1 Parametric Quantile Regression

As in Koenker (2005), for a real-valued random variable, Z, and for $0 < q < 1$, the q^{th} quantile of Z is defined as

$$F^{-1}(q) = \inf\{z : F(z) \geq q\}, \tag{3.1}$$

where $F(z)$ is the (right continuous) distribution function of Z and the median, $F^{-1}(1/2)$, plays the central role.

Given a random variable with pdf f, making one (or more) of the parameters dependent on a random variable, say X, extends the model to QR. For example, if μ is a location parameter of the given density function and σ is a scale parameter, the distribution function can be written as

$$F\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sigma}\int_{-\infty}^{y} f\left(\frac{z-\mu}{\sigma}\right) dz = q.$$

Using the definition in (3.1), the q^{th} quantile of Y then satisfies

$$F^{-1}(q) = \frac{y-\mu}{\sigma}.$$

Rearranging this yields the equation

$$y_q = \mu + \sigma F^{-1}(q), \tag{3.2}$$

an expression of the quantile function y_q in terms of the location and scale parameters and the location- and scale-free quantile function. Making one (or more) of the parameters dependent

on X, e.g. $\mu = a + bx$, where a and b are appropriate constants, we obtain

$$y_q = a + bx + \sigma F^{-1}(q),$$

a linear regression model. By letting μ (or similarly σ) be dependent on X, and introducing the q^{th} quantile, we obtain the idea of QR. For every quantile q, and for every set of points (X_i, Y_i), $i = 1, ..., n$, we have $y = a + bx + \sigma F^{-1}(q)$, the curve representing the q^{th} quantile when μ is a linear function of X. Similarly, we can make any other parameter dependent on X allowing more curvature in the quantile function. For a distribution with additional shape parameters, it is possible to make the shape parameter dependent on X as well. This idea will be elaborated and explained more in Section 3.3 while using the GG distribution. The next section summarises the work done by previous researchers on QR. We highlight the main contributions and we set the scene for Section 3.3 where we explain our approach to QR using the GG particularly.

3.2 Quantile Regression Literature Review

Parametric quantile regression (PQR) is the grand title under which our study takes place. In this section, we present a quick review of QR and some of the main work done in this area. We start by reporting some of the theory done, then we move to the most recent computational contributions involving software that can implement most of the theory. Finally, we list some of the research done to plot quantiles for reference charts.

3.2.1 Review of the Theory

Sir Francis Galton (1875) introduced QR. His approach used simple statistical elements to analyse data, such as sorting and ranking. It started with simple ideas such as placing objects side by side, gathering their descriptive properties, and telling which one has the larger share

of the quality involved. One of his first examples targeted analysing the height of individuals. He lined up men in order of height and measured the middlemost one which is now known as the median. Hence, this special case of QR is often referred to as "median regression". After measuring the median, the quartiles were obtained and those measurements were used to compare between populations. To obtain a continuous measure, objects were marshalled in order of their magnitude along a level base at equal distances apart. A line was drawn freely through the tops representing their magnitudes. The line, logically, rises up vertically at zero, then becomes nearly horizontal over a long space in the middle and rises up at the other end until there also it is vertical. Such a curve was called an ogive and it is what became later known as the distribution quantile function.

In his measurements, Galton plotted a variable t_q against its probability q i.e. $t_q = Q(q)$, which is the quantile function, whereas, as the years passed statisticians switched the axes around and plotted q against t_q i.e. $q = F(t_q)$, which is the known distribution function. Hence, if we denote by F the distribution function of a random variable T, the quantile function Q satisfies

$$Q(q) = F^{-1}(q) = t_q.$$

More research was done later in Galton (1883) and Galton (1889). For example, in the latter, Galton presented the median regression model for the diameters of sweet peas in successive generations. His model for the conditional median was $M(y|x) = 15.5 + (1/3)(x - 15.5)$, where the dependent variable y is an observation on the new generation, the covariate x on the parent and 15.5 is the median of x. Through history, the topic gradually emerged as a unified statistical methodology for estimating models of conditional quantile functions.

Koenker (2005) is the most popular book on QR. This book offers a first comprehensive treatment of the subject encompassing models that are linear and nonlinear, parametric and nonparametric, illustrated with a variety of applications from economics, biology, ecology and finance. The main idea in this book was then referred to by some researchers as being semi-parametric in the sense that it is based on a linear regression model but without taking into consideration any distributional assumptions.

Gilchrist (2000) is another book that offered a way to statistical modeling using the quantile function in a parametric framework. If $Q_0(q)$ denoted the quantile function of the basic form of a distribution, this book uses the fact that the general quantile function of a distribution satisfies

$$Q(p) = M + SQ_0(q), \tag{3.3}$$

where the median M is the location parameter and S is the scale parameter of the distribution. In this scenario, M represents the deterministic part of the regression equation and $SQ_0(q)$ the stochastic one. The author claims that Galton was the first to present this form for the normal distribution and that this representation of the quantile function, in its generalised form for any distribution, was systematically first given by Parzen (1979). Equation (3.3) allows for regression by introducing to the first term in the right hand side a covariate, say X, and considering the second term to be similar to a random variate scaled error. Gilchrist (2000) attributes to the error term (which is of course a quantile function here) discrete distributions such as the binomial and geometric. Other distributions were also considered such as the logistic, three-, four-, and five-parameter lambda, the extreme value, and the Burr family of distributions. It is claimed that this approach hasn't been explored systematically in the literature and that what has been referred to as PQR is assigning a distributional model to the deterministic part of the regression equation only, rather than the stochastic one.

These are the two most significant books that discussed QR, the former follows more or less a nonparametric approach while the latter follows a parametric one. For a brief review of the (chiefly nonparametric) QR technique and its applications, the paper by Yu *et al.* (2003) is also a good reference. We now present a summary of the most significant research that led to or emerged from the ideas in these books where the QR world was divided into two schools of thought: parametric and nonparametric. In the parametric framework, distributional assumptions are assigned to the model, whereas, in the nonparametric approach no distributional assumptions are made and the model is based totally on the individual data points of the sample. In both schools, regression is attained by making the model dependent on a covariate either linearly or through a kernel function that allows more curvature in the fitted model. Both Gilchrist (2000) and Koenker (2005) considered a linear dependence on a covariate X. Eventually, any study mentioned in the literature emerges either directly from or at least revolves closely around one of four approaches: PQR based on a linear model, PQR based on a kernel regression function, nonparametric QR based on a linear model and nonparametric QR based on a kernel regression function.

Parzen (1979) describes his approach as being simultaneously parametric and nonparametric by taking four stages for modeling. In the first stage, a parametric model is considered assuming that Q_0 has a known distribution and correspondingly the parameters are estimated. In the next stages, goodness-of-fit tests are done and robustness of the model is tested for different distributions symmetric and nonsymmetric. If these tests fail, then a nonparametric model is suggested by trying to find suitable estimates of Q_0 through the sample data. The paper emphasises the last stage that considers a nonparametric model.

In fact, nonparametric quantile regression was widely explored. Koenker and Bassett (1978) is one of the leading papers that considered this framework while taking into consid-

eration a linear model. This paper along with other extensions constituted the basis of the book by Koenker (2005). The Koenker and Bassett (1978) approach starts from the very basic regression equation $Y = \beta X^T + \varepsilon$. The least squares method is the most popular way to estimate β through solving the minimisation $\sum_{i=1}^{n} (y_i - \beta x_i)^2$. It turns out, of course, that the mean is the solution and hence regression was referred to as "mean regression". Koenker and Bassett (1978) introduce a natural generalisation to this linear model using the concept of quantiles. The problem then became a more general one that involved solving for the q^{th} regression quantile, $0 < q < 1$, in the minimisation

$$\min_{\beta \in \mathbb{R}^k} \left\{ \sum_{i \in \{i: y_i \geq x_i \beta\}} q|y_i - \beta x_i| + \sum_{i \in \{i: y_i \leq x_i \beta\}} (1-q)|y_i - \beta x_i| \right\}. \tag{3.4}$$

In fact, for a given density function f with distribution function F and using the basic minimisation procedure of setting the derivative equal to zero, the solution to (3.4) turns out to be unsurprisingly $F^{-1}(q)$, the quantile function defined in (3.1).

Other papers in the same framework are by Koenker and Bassett (1982), Bassett (1986), Koenker and d'Orey (1993), Koenker and Machado (1999) who introduce a goodness-of-fit process for QR, Bassett *et al.* (2002), and most recently the work by Portnoy (2003) where a recursively reweighted estimator of the regression quantile process is developed. Kocherginsky *et al.* (2005) develop a time-saving method to construct confidence intervals in QR and lastly Neocleous and Portnoy (2008) present some theory behind the monotonicity of the increasing quantiles.

As a matter of fact, most of the contributions in quantile regression are nonparametric. In the previous paragraph, we mentioned a few nonparametric contributions that used a linear model. We move, now, to discussing a few papers that used kernel estimation instead. Yu and Jones (1998) suggest two approaches for kernel weighted local linear fitting. One considers the minimisation of (3.4) to obtain the quantiles, while the other inverts a local linear distribution

function coming from the data. Both approaches were applied to skinfold triceps data sets from three Gambian villages and double-smoothed quantiles were plotted. Again, Jones and Yu (2007) improve on their earlier work by introducing a third approach which they claim to be of better performance (and more straightforward) than the two earlier ones. Kernel estimation was novel in both papers.

Following up on kernel estimation, we address this issue now from a parametric point of view and we start with a leading paper by Cole and Green (1992) where the quantile regression model is summarised by three curves representing the median, coefficient of variation and skewness (also referred to as the Box-Cox power transformation) fitted as cubic splines using penalised likelihood. This method was introduced originally by Cole (1988) and was termed the "LMS" method. Other significant papers in this context are by Yu and Stander (2007) who follow a Bayesian approach where the Laplace distribution plays a central role and Thompson *et al.* (2010) where Bayesian inference is based on the posterior density of a spline with a smoothing parameter. In a Bayesian framework, we also mention Yu and Moyeed (2001) and the paper by Cai and Stander (2008) where the latter considers a Bayesian approach to quantile self-exciting threshold autoregressive time series models. Is is also worth mentioning Serfling (2004) who studies multivariate descriptive measures such as the multivariate location, spread, skewness and kurtosis.

Very few papers addressed the parametric framework where a linear model is taken. The only recent contributions known to us are by Warren Gilchrist of which we mention the most recent ones Gilchrist (1997) and Gilchrist (2008). In the former, the quantile function was given for particular distributions such as the uniform, pareto, exponential, Weibull and the logistic. In addition to the location and scale parameters' effect on the distribution and its quantile as shown in (3.3), a small discussion is given on how a shape parameter controlling skewness can

be integrated in the model. Issues like estimation, hypothesis testing and goodness-of fit were discussed very briefly. In the latter paper, the author brought up the fact that nonparametric QR "ignores the actual distribution of the "error" term". His aim was to show how both the deterministic and stochastic elements of regression can be modeled and to highlight the importance of modeling the variation in the data. A wider range of distributions is suggested especially those involving skewness. The fact that QR can focus on the tails of a distribution is discussed and further ideas for model fitting and diagnostics are given. The author emphasises the fact that little theory is available in this context and mentions the following: "Clearly, a well-fitting Normal-based model has many advantages in terms of properties. However, if there is, say, a long-tailed model that is noticeably better, then one is going to be misled by the application of Normal assumptions."

3.2.2 Computational Contributions

QR has a wide history and some of the related theory was explored in the literature. With the rise of computational ability nowadays and the immense power of computers to run algorithms via software such as **R** and SAS used for statistical computing, computational algorithms were introduced to implement a big part of the theory. The best known procedure that provides computational access to the QR theory is the QUANTREG procedure provided in SAS 9.1 and also in **R**. This package computes estimates and related quantities for QR by solving a modification of the least squares criterion as explained in Koenker and Bassett (1978). It is explained and described through its SAS and **R** manuals and through some publications such as the report by Chen (2005). The **R** package `rqmcmb2`, by Maria Kocherginsky and Xuming He, also computes QR-related statistics. It is used basically for Markov chain marginal bootstrapping in QR. Among many other computational procedures, we mention lastly the **R** package `lmsqreg` which implements the "LMS" method by Cole and Green.

3.2.3 Recent Contributions to Reference Charts

The rise in computational ability and the availability of software for computing QR measures made it possible and easier for researchers to consider applying QR to real data sets. In reality, researchers are often asked to construct reference charts and limiting thresholds that specify the common and uncommon range in data. This comes from the need in different fields such as pharmacology, economy and health to have ranges within data that they can refer to in order to tell in which percentile of the data individuals lie. The issue of plotting reference charts has been addressed through history using different approaches some of which plot centiles using linear or additive linear models while others, most recently, use QR.

We now mention some of the most significant researchers and papers that have modeled reference charts. Wei *et al.* (2006) offer a comparison between the LMS method using the penalised likelihood of Cole and Green (1992) and a nonparametric approach based on B-spline expansions following Koenker and Bassett (1978). Quantile curves were plotted for "height-vs-age-data" to detect thresholds of their growth. Similar results were obtained from both approaches, the former reported as being more stable while the latter more flexible. Other models were considered and the effect of more variables was studied as well. Some papers considered using generalised additive models to relate variables with each other. For example Cole *et al.* (2009) highlight the importance of age-related reference ranges to assess growth in children. Data is obtained from four different countries: U.S.A. (NHANES III), Belgium, England and Canada. The computational work was implemented using the package `GAMLSS` in **R**. This package is discussed and explained in Stasinopoulos and Rigby (2007). Van Buuren *et al.* (2009) construct regional centiles for "weight-vs-age-data" sets from different countries. They use a generalised additive model for location, scale and shape parameters and suggest either a Box-Cox t or a Box-Cox power exponential distribution. The package `GAMLSS` was also

used for computational purposes. Heagerty and Pepe (1999) propose a method for reference charts where quantiles are modeled as a smooth function of covariates in a nonparametric framework. The reference chart is done for female infants under the age of three to study the implication of weight for both height and age.

A report by Chen (2005) studied BMI charts of individuals from all ages. BMI is defined as the ratio of weight to the square of the height (kg/m^2). QR is proposed as an alternative to other procedures such as fitting smoothing curves on sample quantiles of segmented age groups and the LMS method. It is also claimed that as BMI across all age groups is skewed to the right, normality assumptions fall down and preference is given to QR procedures such as the Koenker and Bassett (1978) approach. The QUANTREG procedure in SAS is introduced and the three different algorithms therein are applied in a QR context to compare BMI growth charts of two data sets. A model involving six powers of age and the logarithm of BMI is fitted.

It is almost an impossible task to mention all the researchers that have addressed the topic of QR. Through this literature review, we have tried to display the most significant contributions hoping to form a clear picture in this regard and introducing the reader to the next sections of the thesis, in particular to our contribution to PQR and further to applying it to the GG distribution.

3.3 The GG Distribution in the Context of QR

The GG distribution encompasses both the gamma ($\beta = 1$) and Weibull ($\beta = \alpha$) distributions — and hence also the exponential distribution — as special cases and the lognormal and normal distributions as limiting cases. Similar to other life distributions it can be used in modeling regression data. Given an observed set of regression data (X_i, Y_i), $i = 1, ..., n$, satisfying

$$Y_i = g(X_i; \rho; \epsilon_i),$$

where $g(X_i; \rho, \epsilon_i)$ is some function of X_i, ρ is a vector of parameters and ϵ_i is the error term, the aim is to fit a GG distribution to the error while taking into account the quantile to be modeled. Let $T_1, ..., T_n$ be a random sample taken from the GG distribution with pdf Version 2. The cumulative distribution functions of T_i, $i = 1, ..., n$, and of $Y_i = \log T_i$, parametrised as in (2.19), are respectively

$$F(t) = \frac{\Gamma_{(t/\theta)^\beta}(k)}{\Gamma(k)} \tag{3.5}$$

and

$$F(y) = \frac{\Gamma_{k\exp\left(\frac{y-\mu}{\sigma\sqrt{k}}\right)}(k)}{\Gamma(k)}, \tag{3.6}$$

both taking the form of a regularised gamma function which is defined in equation (2.1). Let $r(k, q)$ be the location- and scale-free q^{th} quantile of a gamma distribution with shape parameter k. From (3.5) and , we deduce respectively that

$$r(k, q) = (t/\theta)^\beta$$

and

$$r(k, q) = k \exp\left(\frac{y - \mu}{\sigma\sqrt{k}}\right).$$

Consequently, the q^{th} quantile functions (in terms of μ, σ and k) of T and Y are respectively

$$t_q = \exp(\mu)\left(\frac{r(k, q)}{k}\right)^{\sigma\sqrt{k}} \tag{3.7}$$

and

$$y_q = \mu + \sigma\sqrt{k}\log\left(\frac{r(k, q)}{k}\right). \tag{3.8}$$

In the expressions of the quantile curves (3.7) and (4.14), as explained in Section 3.1, making one of the parameters dependent on a covariate, e.g. X, allows t_q, and y_q in turn, to be conditional on $X = x$, extending the model to quantile regression where t_q (and y_q) is the dependent variable and X the independent variable.

In this thesis, we consider three cases of the conditional GG distribution:

1. The four-parameter GG, where we set μ to be a linear function of a covariate X through $\mu = a + bX$ ($a, b \in \mathbb{R}$), and σ and k positive constants.

2. The five-parameter GG, where we make μ and σ dependent on a covariate X through $\mu = a + bX$ and $\sigma = \exp(c + dX)$ ($a, b, c, d \in \mathbb{R}$), and k a positive constant.

3. The six-parameter GG, where all three parameters μ, σ and k are dependent on a covariate X through $\mu = a + bX$, $\sigma = \exp(c + dX)$ and $k = \exp(f + gX)$ ($a, b, c, d, f, g \in \mathbb{R}$).

The aim from this study is to find the best GG model fit to every quantile of given regression data. Is it the three-, four-, five- or six-parameter GG? For that, we explain in Section 4.3 how we conduct LRTs to check whether a full model (six-parameter model) is required to explain the data or whether a smaller one is enough. In the next section, the shapes of the quantiles for the GG models given in items (1), (2) and (3) are studied.

3.4 Shapes of the Quantile Curves for the Suggested Four-, Five- and Six-Parameter GG

Let X be a positive random variable and T a GG-distributed random variable conditional on the covariate X through the parameters. The shapes of the quantile curves in the three different parameter case-scenarios are explored in what follows.

1. The four-parameter GG quantile curve is given by

$$t_q = \exp(a + bx)\left(\frac{r(k,q)}{k}\right)^{\sigma\sqrt{k}}, \qquad x > 0,$$

where a, $b \in \mathbb{R}$ and σ, $k > 0$. Table 3.1 presents the possible shapes that t_q can take.

For the purpose of illustration, we consider a particular case of the four-parameter conditional GG with $k = 4$, $\sigma = 1.25$, and we make μ conditional on random variable X via the equation $\mu = -1 - 2X$ where $X \sim U(0,1)$. We compute the corresponding θ, α and β from which we generate a data set T_i, $i = 1, ..., 500$, conditional on the random variables X_i, $i = 1, ..., 500$. The simulated regression data (X_i, T_i), $i = 1, ..., 500$, and the corresponding quantile curves t_q for $q = \{0.1, 0.25, 0.5, 0.75, 0.9\}$ are shown in Figure 3.1. As $b = -2 < 0$, we observe that the quantiles are decreasing.

Table 3.1: Shapes of the four-parameter GG quantile curves.

cases	modes	shape
$b > 0$	monotone increasing	exponential increase from $e^a(r/k)^{\sigma\sqrt{k}}$ to $+\infty$ (at ∞)
$b < 0$	monotone decreasing	exponential decrease from $e^a(r/k)^{\sigma\sqrt{k}}$ to 0 (at ∞)

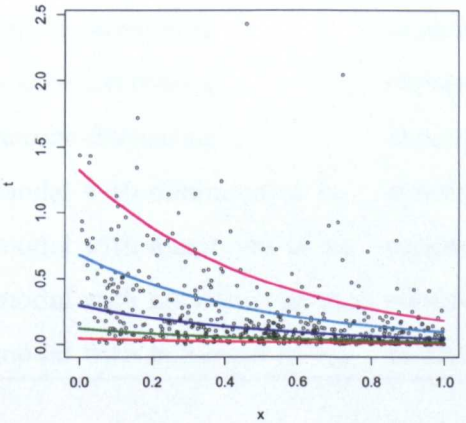

Figure 3.1: 10%, 25%, 50%, 75% & 90% quantile curves of a GG($\mu = -1-2X, \sigma = 1.25, k = 4$).

2. The five-parameter GG quantile curve is given by

$$t_q = \exp(a + bx)\left(\frac{r(k,q)}{k}\right)^{\sqrt{k}\exp(c+dx)}, \qquad x > 0,$$

where a, b, c, $d \in \mathbb{R}$ and $k > 0$.

Let $M = (r(k,q)/k)^{e^c\sqrt{k}}$ and $x_m = (1/d)\log(-b/(d\log M))$ when the definition makes sense. Note that $M > 0$,

$$\frac{\partial \log t_q}{\partial x} = b + d\log M \exp(dx)$$

and

$$\frac{\partial^2 \log t_q}{\partial x^2} = d^2 \log M \exp(dx).$$

Therefore, the shape of t_q depends on whether b and d are positive or negative and whether M is greater or less than one. Table 3.2 presents the possible shapes that t_q can take.

Table 3.2: Shapes of the five-parameter GG quantile curves.

cases	modes	shape
$b > 0$; $d > 0$; $M > 1$	monotone increasing	exponential increase from Me^a to $+\infty$
$b > 0$; $d < 0$; $M < 1$	monotone increasing	exponential increase from Me^a to $+\infty$
$b < 0$; $d < 0$; $M > 1$	monotone decreasing	exponential decrease from Me^a to 0
$b < 0$; $d > 0$; $M < 1$	monotone decreasing	exponential decrease from Me^a to 0
$b > 0$; $d < 0$; $M > 1$	unimodal with minimum at x_m	concave up starting from Me^a to $+\infty$
$b < 0$; $d < 0$; $M < 1$	unimodal with maximum at x_m	concave down starting from Me^a to 0
$b > 0$; $d > 0$; $M < 1$	unimodal with maximum at x_m	concave down starting from Me^a to 0
$b < 0$; $d > 0$; $M > 1$	unimodal with minimum at x_m	concave up starting from Me^a to $+\infty$

For example, setting $k = 0.75$ we make σ and μ conditional on a random variable $X \sim U(0,1)$ via the equations $\mu = 1 - 0.1X$ and $\sigma = \exp(-1.5 - 2X)$. We simulate

a data set of size 500 conditional on random variables X_i, $i = 1, ..., 500$. The simulated regression data (X_i, T_i), $i = 1, ..., 500$, and the corresponding quantile curves t_q for $q = \{0.1, 0.25, 0.5, 0.75, 0.9\}$, are shown in Figure 3.2. In this case, $b = -0.1 < 0$ and $d = -2 < 0$. Of course, $M > 1$ when $r(k, q) > k$ and analogously $M < 1$ when $r(k, q) < k$. Looking at Figure 3.2, for $q = \{0.75, 0.9\}$, $M > 1$, therefore the quantiles are monotone decreasing. For $q = \{0.1, 0.25, 0.5\}$, $M < 1$, hence the quantiles are unimodal with maximum calculated at $x_m = \{1.203, 0.907, 0.331\}$ rounded to three decimal digits. Note that $x_m = 1.203$ is not shown in Figure 3.2 as the graph is restricted to region where $x \in [0, 1]$.

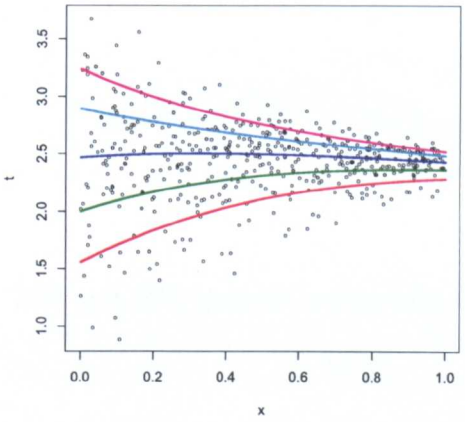

Figure 3.2: 10%, 25%, 50%, 75% and 90% quantile curves of a GG($\mu = 1 - 0.1X, \sigma = \exp(-1.5 - 2X), k = 0.75$).

3. The six-parameter GG quantile curve is given by

$$t_q = \exp(a + bx) \left(\frac{r(f, g, q)}{e^{f+gx}} \right)^{\exp(c+dx+\frac{1}{2}(f+gx))}, \qquad x > 0,$$

where a, b, c, d, f, $g \in \mathbb{R}$.

The form of the six-parameter GG quantile curve is much more complicated than the four- and five-parameter cases since, here, $r(k, q)$ is dependent on k and therefore on X.

A proper summary table is not straightforward. However, it was noticed that bimodal quantiles can exist in this case. We demonstrate this fact by a particular example where all three parameters are dependent on random variable $X \sim U(0,1)$ via the equations $\mu = 5 + 3X$, $\sigma = \exp(0.5 - 2X)$ and $k = \exp(9 - 15X)$. A simulated data set of size 500 conditional on X_i, $i = 1, ..., 500$, with the corresponding quantile curves t_q for $q = \{0.1, 0.25, 0.5, 0.75, 0.9\}$, are shown in Figure 3.3.

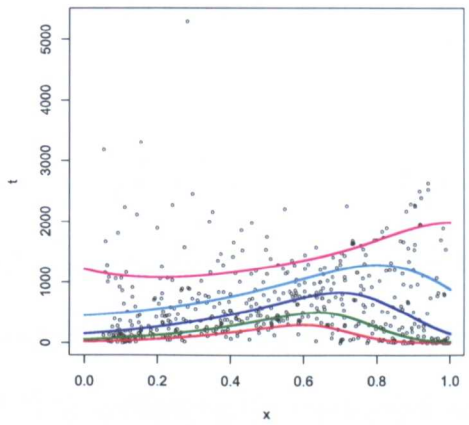

Figure 3.3: 10%, 25%, 50%, 75% and 90% quantile curves of a GG($\mu = 5 + 3X, \sigma = \exp(0.5 - 2X), k = \exp(9 - 15X)$).

We have, now, formed an idea of the GG quantile shapes in the three-, four-, five- and six-parameters cases. The three-parameter quantiles are horizontal and parallel to each other. The four-parameter quantiles are exponentially monotone increasing or decreasing. In the five-parameter case, they exhibit either monotone exponential shapes or a unimodal structure. Finally, the six-parameter case shows all different monotone exponential, unimodal and bimodal shapes. It is important to note that, perhaps surprisingly, despite the very simple linear and loglinear forms used for the dependence on x, the resulting quantiles can exhibit a very wide range of very useful shapes.

Chapter 4

Parametric Quantile Regression with Generalised Gamma II – Estimation

Following up on Chapter 3 where we set the scene for the PQR approach that we use to study our suggested four-, five- and six-parameter GG models (introduced respectively in items 1, 2 and 3 of Section 3.3), we move now to estimating these models. In this chapter, we complete the picture for MLE of the GG by estimating the conditional distribution when the number of parameters $p = 4, 5, 6$. We also demonstrate various steps of the modeling package introduced in Chapter 1 such as performing LRTs for model selection, finding the confidence bands around the estimated quantiles and introducing our goodness-of-fit test.

MLE of the four-parameter GG conditional distribution is performed in Section 4.1 and an iterative algorithm is developed. The five- and six-parameter GG cases are also discussed. In Section 4.2, a simulation study validates the proposed iterative algorithm. Both MLE and the simulation study were also discussed in Noufaily and Jones (2010). In Section 4.3, we explain how LRTs can be conducted to test whether a full-parameter (six-parameter GG) model is required or a smaller one is enough. Section 4.4 displays the asymptotics for the four-, five- and six-parameter GG. These are used in Section 4.5 to obtain expressions for CIs around the estimated quantiles. Finally, a goodness-of-fit test is suggested in Section 4.6.

4.1 MLE for the Conditional GG Distribution

4.1.1 MLE for the Four-Parameter Distribution of the Log of a GG Random Variable

In this subsection, we revisit the work of Subsection 2.4.1 for the simple linear regression case in which the data are (X_i, T_i), $i = 1, ..., n$, $Y_i = \log T_i$ and the X's are a univariate covariate. We thus model μ not as a constant but in terms of the covariate and two parameters to be estimated, through $\mu = a + bX$. Note that both $a, b \in \mathbb{R}$. On the original scale, this means that the scale parameter θ is modeled as a loglinear function of the covariate.

We find it convenient to reparametrise from a and b to $a' = a/\sigma$, $b' = b/\sigma$. Also, write

$$S_{jl} = \frac{1}{n}\sum_{i=1}^{n} Y_i^j X_i^l \exp\left(\frac{Y_i}{\sigma\sqrt{k}} - \frac{b'}{\sqrt{k}}X_i\right); \quad j, l = 0, 1, 2.$$

Given the new parametrisation, the loglikelihood is

$$n\left\{-\log(\sigma) + \left(k - \frac{1}{2}\right)\log(k) - \log\Gamma(k) + \sqrt{k}\left(\frac{\overline{Y}}{\sigma} - a' - b'\overline{X}\right)\right. \tag{4.1}$$

$$\left. - \frac{k}{n}\exp\left(\frac{-a'}{\sqrt{k}}\right)\sum_{i=1}^{n}\exp\left(\frac{Y_i}{\sigma\sqrt{k}} - \frac{b'}{\sqrt{k}}X_i\right)\right\}. \tag{4.2}$$

Differentiating w.r.t a', b', σ and k in turn yields the score equations:

$$0 = n\sqrt{k}\left\{\exp\left(\frac{-a'}{\sqrt{k}}\right)S_{00} - 1\right\}; \tag{4.3}$$

$$0 = n\sqrt{k}\left\{\exp\left(\frac{-a'}{\sqrt{k}}\right)S_{01} - \overline{X}\right\}; \tag{4.4}$$

$$0 = \frac{n\sqrt{k}}{\sigma^2}\left\{\exp\left(\frac{-a'}{\sqrt{k}}\right)S_{10} - \frac{\sigma}{\sqrt{k}} - \overline{Y}\right\}; \tag{4.5}$$

$$0 = n\left\{\exp\left(\frac{-a'}{\sqrt{k}}\right)\left(\frac{S_{10}/\sigma - a'S_{00} - b'S_{01}}{2\sqrt{k}} - S_{00}\right) + \log(k) + 1 - \frac{1}{2k} - \psi(k)\right. \tag{4.6}$$

$$\left. + \frac{\overline{Y}/\sigma - (a' + b'\overline{X})}{2\sqrt{k}}\right\}. \tag{4.7}$$

As in Lawless (1980), (4.3) yields

$$\exp(a') = S_{00}^{\sqrt{k}}, \tag{4.8}$$

an expression for a' in terms of b', σ and k.

Rearranging (4.4) yields

$$B_R(b') \equiv \overline{X} \exp\left(\frac{a'}{\sqrt{k}}\right) - S_{01} = 0. \tag{4.9}$$

We think of (4.9) as an equation in b' for any a', k and σ and solve it numerically. This is easy because we can show that B_R is monotone increasing in b':

$$\frac{\partial B_R(b')}{\partial b'} = \frac{1}{\sqrt{k}} S_{02} > 0.$$

Also, $\lim_{b' \to -\infty} B_R(b') = -\infty$ except for the case where $x_i < 0 \; \forall \, i = 1, ..., n$, then $\lim_{b' \to -\infty} B_R(b') = \overline{X} \exp\left(\frac{a'}{\sqrt{k}}\right)$ (negative horizontal asymptote). Similarly, $\lim_{b' \to \infty} B_R(b') = \infty$ except for the case where $x_i > 0 \; \forall \, i = 1, ..., n$, then $\lim_{b' \to \infty} B_R(b') = \overline{X} \exp\left(\frac{a'}{\sqrt{k}}\right)$ (positive horizontal asymptote). Hence, there must be precisely one value of b' (for any a', k and σ) for which $B_R(b') = 0$. Using (4.8), we can reduce (4.9) to

$$B_R(b') = \frac{S_{01}}{S_{00}} - \overline{X} = 0 \tag{4.10}$$

an equation in b' for any k and σ. We already showed that there exists precisely one value of b' for which $B_R(b') = 0$. Therefore, we solve $B_R(b') = 0$ numerically to find the root b' for any k and σ.

Using (4.8), we can reduce (4.5) to

$$R_R(\sigma) \equiv \frac{S_{10}}{S_{00}} - \overline{Y} - \frac{\sigma}{\sqrt{k}} = 0 \tag{4.11}$$

Again following Lawless (1980), we think of $R_R(\sigma)$ as an equation in σ for any k and b' and solve it numerically. The same argument that applied to (2.35) shows that there must be precisely one value of $\sigma > 0$ (for any b and k) for which $R_R(\sigma) = 0$.

Using (4.8), (4.10) and (4.11) to remove the quantities involving $\exp\left(-a'/(\sigma\sqrt{k})\right)$, S_{00}, S_{01} and S_{10}, we find that (4.6) reduces to

$$T_R(k) \equiv \frac{(\overline{Y}/\sigma) - (a' + b'\overline{X})}{\sqrt{k}} + \log(k) - \psi(k) = 0.$$

Finally, the analogue of (2.38) is

$$T_{L_R}(k) \equiv \log(k) - \psi(k) - \frac{L_R}{\sqrt{k}}, \tag{4.12}$$

a function of k for any L_R where

$$L_R = a' + b'\overline{X} - \frac{\overline{Y}}{\sigma}.$$

Again, Jensen's inequality can be used to show that $L_R > 0$ provided a', b' and σ satisfy (4.8), and hence (2.38) also has a single root in k in those circumstances.

4.1.2 Iterative Algorithm and MLE for the Five- and Six- Parameter Cases

We have now proved that each of the equations (4.8), (4.11), (4.10) and (4.12) has one root $\hat{a}$, $\hat{\sigma}$, $\hat{b}$ and $\hat{k}$ respectively. We will solve them simultaneously and iteratively to obtain the ML estimates of the four-parameter GG. Work is done twice, each time using one of either the Newton Raphson algorithm or the bisection method to solve for b. Our algorithm for estimating the four-parameter GG becomes:

1. Set the iteration number i to 0 and obtain an initial guess for $L_R = L_{R,0} > 0$ and $b' = b'_0$.

2. Set $i = i + 1$.

3. For given $L_{R,i-1}$, compute $\widehat{k}_i$ by solving (4.4) using either the bisection method or the Newton Raphson algorithm (we have used the former).

4. Replace the obtained $\widehat{k}_i$ in $R_R(\sigma)$ to compute $\widehat{\sigma}_i$ by solving (4.3) using the bisection method or the Newton Raphson algorithm (we have used the latter).

5. Replace the obtained $\widehat{k}_i$ and $\widehat{\sigma}_i$ in $B_R(b')$ to compute $\widehat{b'}_i$ by solving (4.2) using the bisection method or the Newton Raphson algorithm (we used the former after experiments with the latter led to too many program failures).

6. Substitute $\widehat{k}_i$, $\widehat{\sigma}_i$ and $\widehat{b'}_i$ into (4.1) to obtain the corresponding $\widehat{a'}_i$.

7. Use these estimates to obtain $L_{R,i}$ and to compute the value of the loglikelihood function.

8. Repeat steps (2), (3), (4), (5), (6) and (7) until desired accuracy of the likelihood is achieved.

9. Set $\widehat{a} = \widehat{\sigma}\widehat{a'}$ and $\widehat{b} = \widehat{\sigma}\widehat{b'}$.

This iterative algorithm is tested via simulations in Section 4.2. It is shown that, computationally, MLE of the GG distribution is perfectly reasonable for p-parametric distributions where p is small to moderate (p =4, 5 or 6). Our simulation study also shows that the general-purpose optimisation procedures, Nelder Mead and BFGS, are competitive with our algorithm and behave even slightly better in some cases. We will use the general-purpose Nelder Mead (which proved to be generally the best) optimisation algorithm to maximise the five- and six- parameter versions of the likelihood in (4.1) and hence to estimate the parameters. To demonstrate the accuracy of the GG in this respect, a grand simulation study that includes estimation of all three-, four-, five- and six-parameter GG distributions will be given in Section 5.3.

4.2 Simulation Study Using the Iterative Algorithm in Subsection 4.1.2

4.2.1 Efficiency of Iterative Algorithm and Comparison with Other Optimisation Procedures

The aim in this section is to show how reliable our suggested algorithm (in its 2 versions as explained in Subsection 4.1.2) is for MLE of the four-parameter GG regarding its well-behaviour for different initial values and relative to other general-purpose optimisation procedures. For this purpose, we compare it with other optimisation methods such as the Nelder-Mead, BFGS

and another method introduced by Prentice (1974). This is done via simulations. We repeated the experiment of Subsection 2.5.1 with the addition of the extra parameter b being generated, like a, from the standard normal distribution. Note that, in this case, the "real parameters" and "initial parameters" are given in Tables A.4 and A.5 of the Appendix. For the regression situation, we also tested an **R** implementation of the Prentice (1974) approach to MLE provided in the `VGAM` package of T.W. Yee. However, we were unable to make this program work anything like as well as any of the others and so we have removed it once more from our comparisons. Tables 4.1 and A.6 present summaries of the results for the $n = 200$ and $n = 500$ cases respectively. "$\hat{a}$", "$\hat{b}$", "$\hat{\sigma}$" and "$\hat{k}$" are values of the parameter estimates (of a, b, σ and k respectively) leading to "Likemax". "Progbis", "Prognr" "BFGS", and "NM" represent the number of times (out of 100) each of our program's bisection version, our program's Newton Raphson version, BFGS, and Nelder-Mead respectively fails to reach "Likmax" up to an error equal to 0.01. Similarly to the three-parameter case, the estimated parameters still give reasonable results. We also observe that our program, in the bisection method version, and the Nelder-Mead procedure report the least number of failures.

As before, Table 4.2 reports a summary of the total number of failures of each method for the $n = 200$ and $n = 500$ cases. The values in Tables 4.1, A.6 and 4.2 are rounded to three decimal digits. Compared with Table 2.4, each method, unsurprisingly, has an increased number of reported errors, but the situation is still very good. For example, when $n = 200$, our program still has a 96.8% success rate, Nelder-Mead a 97.8% success rate, and BFGS an 89.7% success rate. It should be admitted that it was the less successful version of our program that uses the Newton-Raphson algorithm at Step 5 of the iterative algorithm in Subsection 4.1.2 that was matched with the other programs for speed in the regression case, the purely bisection version reported on in Tables 4.1 and A.6 typically taking several times longer.

Table 4.1: Number of times each method fails to reach the maximal likelihood in the four-parameter case for $n = 200$.

Data	Likemax	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$	$\hat{k}$	Progbis	Prognr	BFGS	NM
GGQ1	-442.906	-0.695	1.563	2.013	1.706	1	6	15	1
GGQ2	-184.868	1.076	-1.713	0.544	1.425	1	13	3	1
GGQ3	-367.967	0.178	0.557	1.435	2.754	0	2	14	1
GGQ4	-331.580	0.200	0.551	1.251	11.358	0	4	13	1
GGQ5	-577.164	-0.839	-3.383	4.225	6.428	2	5	21	3
GGQ6	-373.025	1.972	0.099	1.474	2.857	1	13	20	1
GGQ7	-392.633	-0.720	0.973	1.171	0.343	0	12	3	0
GGQ8	-398.024	0.194	1.122	1.652	2.377	0	2	8	1
GGQ9	-664.888	-1.573	-0.362	5.484	0.755	7	23	15	3
GGQ10	-255.322	1.296	-0.297	0.544	0.263	0	5	4	1
GGQ11	-424.828	0.643	-1.541	1.906	2.734	0	2	14	1
GGQ12	-426.109	-0.918	0.785	1.929	3.039	2	13	10	1
GGQ13	-323.809	-0.249	-1.961	1.001	0.774	1	25	2	0
GGQ14	-406.563	1.181	-0.235	1.515	0.777	0	0	8	1
GGQ15	-404.495	-0.969	0.105	1.443	0.635	1	3	2	0
GGQ16	-539.365	0.463	1.515	3.529	9.818	0	3	29	3
GGQ17	-614.401	0.364	-0.312	5.050	4.950	1	6	22	3
GGQ18	-563.332	-0.449	0.068	3.698	1.820	1	10	11	2
GGQ19	-154.092	-2.016	1.005	0.467	1.421	0	14	0	0
GGQ20	-587.273	-0.914	-0.390	4.133	1.656	3	14	12	2
GGQ21	-596.564	-0.736	0.678	4.605	4.524	0	3	25	3
GGQ22	-424.361	-0.379	1.913	1.250	0.253	0	5	3	1
GGQ23	-268.639	-1.763	0.971	0.823	1.364	0	16	2	0
GGQ24	-11.923	0.930	1.550	0.244	3.310	6	33	8	1

Table 4.1 Continued

Data	Likemax	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$	$\hat{k}$	Progbis	Prognr	BFGS	NM
GGQ25	-412.259	-1.062	0.999	1.566	0.799	1	9	8	0
GGQ26	-164.104	0.049	-0.846	0.507	2.020	0	1	0	0
GGQ27	-420.395	-0.055	1.790	1.952	11.788	0	3	20	2
GGQ28	-410.758	-0.458	1.163	1.394	0.471	1	9	5	1
GGQ29	-207.683	0.822	0.866	0.557	0.748	0	11	6	1
GGQ30	-335.375	-0.008	0.863	1.162	1.508	0	2	7	1
GGQ31	-571.334	0.524	-0.988	3.178	0.516	4	13	10	1
GGQ32	-346.282	0.182	-0.591	1.169	1.013	0	3	4	1
GGQ33	-118.025	1.125	0.044	0.411	2.759	1	5	10	0
GGQ34	-493.246	-0.089	-0.013	2.355	0.813	0	8	5	1
GGQ35	-238.472	-1.442	0.445	0.784	10.322	0	10	2	0
GGQ36	-413.775	-0.857	1.874	1.589	0.831	0	8	5	1
GGQ37	33.285	2.759	0.437	0.183	1.406	3	35	8	1
GGQ38	-514.620	-0.989	0.038	2.150	0.340	3	8	1	0
GGQ39	-101.314	0.578	1.103	0.399	23.888	1	11	13	1
GGQ40	-124.867	1.078	-0.725	0.375	0.833	0	7	3	1
GGQ41	-492.900	0.385	0.235	2.447	1.056	0	4	7	1
GGQ42	-353.617	-0.852	0.249	1.238	1.179	0	9	2	0
GGQ43	-596.422	0.740	-1.061	4.167	1.179	1	15	23	3
GGQ44	-453.328	0.224	2.324	2.064	1.309	0	4	20	1
GGQ45	-607.338	-0.745	1.007	3.858	0.548	3	14	12	2
GGQ46	-321.427	0.520	0.799	1.205	94.460	99	99	83	100
GGQ47	-247.047	1.658	-0.554	0.672	0.712	1	9	7	1
GGQ48	-277.319	1.174	0.920	0.715	0.470	0	8	8	1
GGQ49	-153.138	0.690	0.282	0.483	2.204	0	6	4	0
GGQ50	-502.474	1.656	1.198	1.974	0.313	0	3	4	1

Table 4.1 Continued

Data	Likemax	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$	$\hat{k}$	Progbis	Prognr	BFGS	NM
GGQ51	-254.243	-0.079	-0.011	0.783	1.692	2	4	3	0
GGQ52	-142.850	-1.869	1.666	0.459	2.241	0	22	1	0
GGQ53	-225.844	0.457	-0.513	0.680	1.689	0	2	2	0
GGQ54	-383.357	-1.688	-1.045	1.619	10.492	2	13	3	0
GGQ55	-459.992	0.412	0.432	1.742	0.429	1	10	7	1
GGQ56	-468.659	0.619	-0.159	2.453	6.186	0	0	20	2
GGQ57	-285.011	0.366	0.660	0.961	3.573	1	2	9	1
GGQ58	-402.213	-0.393	-0.116	1.510	0.866	0	11	4	1
GGQ59	-470.161	1.394	-0.659	2.331	1.918	1	1	13	2
GGQ60	-531.635	1.120	-0.651	3.340	4.985	0	6	25	3
GGQ61	-265.784	-0.503	-0.657	0.877	4.033	1	10	3	0
GGQ62	-369.778	-0.865	-2.013	1.514	11.011	1	8	8	1
GGQ63	-408.642	1.070	0.341	1.724	2.066	0	3	11	1
GGQ64	-215.231	-0.488	-0.402	0.676	3.348	0	2	2	0
GGQ65	-502.374	-1.469	-0.330	2.750	2.022	2	5	11	1
GGQ66	-460.777	0.903	-0.796	1.664	0.356	0	7	14	1
GGQ67	-401.359	-1.134	1.672	1.746	5.481	1	12	12	1
GGQ68	-485.569	1.317	0.448	2.480	1.622	0	2	22	2
GGQ69	-395.804	0.264	-0.150	1.410	0.703	1	7	2	1
GGQ70	-287.722	-0.256	0.087	1.000	8.511	0	1	7	1
GGQ71	-169.578	0.197	0.947	0.496	1.232	0	0	0	0
GGQ72	-329.366	0.701	-0.204	1.185	2.852	0	0	6	1
GGQ73	-648.818	-2.102	0.832	4.930	0.657	9	21	5	1
GGQ74	-316.697	-0.353	0.565	1.121	3.315	4	4	5	0
GGQ75	-519.385	-0.682	1.684	3.175	7.354	0	0	18	3
GGQ76	-553.142	0.312	0.548	3.682	3.836	3	7	17	3

Table 4.1 Continued

Data	Likemax	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$	$\hat{k}$	Progbis	Prognr	BFGS	NM
GGQ77	105.427	-0.880	-1.502	0.111	0.596	55	99	1	2
GGQ78	-374.305	-0.327	1.236	1.408	1.465	0	6	10	1
GGQ79	-505.434	0.494	1.690	2.963	7.550	1	3	19	3
GGQ80	-489.499	-0.775	0.563	2.246	0.694	2	13	7	1
GGQ81	-254.088	-0.767	2.125	0.803	2.319	0	2	5	1
GGQ82	-500.152	-0.575	-0.819	2.708	1.916	2	10	10	1
GGQ83	-392.419	0.176	-1.224	1.531	1.380	1	16	1	1
GGQ84	-523.335	0.729	-0.574	2.550	0.563	3	7	5	1
GGQ85	-563.943	0.413	0.775	3.782	2.333	0	1	17	3
GGQ86	-646.923	0.984	-0.487	5.739	2.414	0	6	21	5
GGQ87	-378.393	-0.448	-1.883	1.573	8.211	0	7	6	1
GGQ88	-655.215	0.341	-1.495	5.525	1.077	3	19	17	3
GGQ89	227.843	1.675	0.719	0.072	2.310	69	99	15	2
GGQ90	-452.535	-1.217	-1.600	2.019	1.132	2	16	1	0
GGQ91	-560.585	1.658	0.803	3.671	1.967	0	2	22	3
GGQ92	-527.449	-1.738	-0.342	2.743	0.732	3	38	0	1
GGQ93	-517.305	1.715	-1.005	2.848	1.334	0	0	15	2
GGQ94	-524.773	-1.284	-2.068	3.191	3.714	2	16	9	1
GGQ95	-641.756	0.897	-0.657	5.582	2.344	0	5	26	4
GGQ96	-378.337	-0.190	0.488	1.491	2.239	0	4	11	1
GGQ97	-357.932	2.117	0.130	1.388	3.887	1	5	25	1
GGQ98	-324.076	-0.583	-0.503	0.966	0.637	0	9	2	0
GGQ99	-309.508	0.153	1.475	1.096	4.522	0	5	14	1
GGQ100	-512.432	-0.120	-0.797	2.641	0.912	2	13	5	1
SUM	-	-	-	-	-	320	1107	1035	220

Table 4.2: Summary of the results in Tables 4.1 and A.6.

	Progbis	Prognr	BFGS	Nelder-Mead
$n = 200$				
Total number of failures	320	1107	1035	220
Total number of "non-global maxima"	290	210	1035	179
Total number of reported errors	30	897	0	0
Total number of -Inf	0	0	0	41
$n = 500$				
Total number of failures	168	949	766	136
Total number of "non-global maxima"	140	115	766	86
Total number of reported errors	28	834	0	0
Total number of -Inf	0	0	0	50

4.2.2 Interpretation of the Results

Our underlying thesis is that, computationally, MLE for p-parameter parametric distributions, where p is small to moderate, say $p = 3, 4, 5, 6$, is generally straightforward. Since Nelder-Mead performance was the best among the optimisation algorithms for estimating the three- and four-parameter cases, we will use it to estimate the five- and six-parameter cases as well, later on in the thesis. One should take care that one is working with a sensible parameterisation in which the parameters are clearly identifiable (which often corresponds to being clearly interpretable). And yes, away from the very special cases of exponential families and simple location-scale models, likelihoods are not concave and may have local maxima. The existence of more than one substantial local maximum and/or of local maxima close to the global maximum can cause problems. But, in our substantial simulation experience, the GG distribution is one for which

neither of these circumstances arises — GG likelihoods appear to have a clear global maximum with any local maxima being much smaller and distant. Reasonable maximisation algorithms therefore find the global maximum 90% of the time or more, typically, and much more of the time in some cases. This leads to the simple strategy of running, say, our program or the BFGS method (or even a mix of the two) from a small number, m, of randomly chosen sets of starting values resulting in almost guaranteed location of the global ML estimate. You can think of running the algorithms as independent Bernoulli trials with the probability of success $p = 0.9$. For example, $m = 5$ independent Bernoulli trials with $p = 0.9$ give $p(0) = 0.00001$, $m = 10$ such give $p(0) = 0.0000000001$, where $p(0)$ is the probability that the ML value obtained in the runs corresponding to the 'trials' is not the global maximum. Alternatively, instead of fixing m, the number of intial values to a small number, such as 4 or 5, one could simply run the algorithm from different initial points until one gets j, say 3, equal values of the maximised likelihood. Hence our claim that, despite other assertions in the literature to the contrary, the GG distribution is actually one for which ML estimation is — in a computational sense — quite straightforward and reliable.

The previous concluding results, concerning the reliability of MLE of the GG distribution, allow us to move further in our study into model selection where the best GG model is chosen. After identifying a convenient distribution for given data, a crucial task is then to find the best subset of the model that fits the data. The LRT, which we will use to compare the likelihoods of the different GG models, is appropriate in this case. A detailed demonstration of the procedure we follow for model selection is given in the next section.

4.3 Likelihood Ratio Test for Choosing the Best Subset of the GG Family

Knowing that MLE of the three-, four-, five- and six-parameter GG is manageable and quite straightforward, we wish to choose the appropriate subset of the GG family that best fits given data. Obviously, a more accurate fit is obtained as the number of parameters increases. What remains to check is whether a model with more parameters is significantly more accurate. Otherwise, it is better to retreat to a simpler one with fewer parameters.

To test for this significance, we propose the well-known LRT which, using the usual asymptotic normality assumptions, compares the p-parametric GG models (where $p \geq 3$). For given data, we denote the loglikelihoods of the three-, four-, five- and six-parameter GG by L_3, L_4, L_5 and L_6 respectively. We start by computing the test statistic $D_1 = 2(L_6 - L_5)$. The probability distribution of this test statistic can be approximated by a chi-square distribution with $(df1 - df2)$ degrees of freedom, where $df1$ and $df2$ are the degrees of freedom of the six-parameter and five-parameter models respectively. In this case, this will be the χ^2_1 whose 95% quantile is approximately 3.84. Therefore, with a 5% level of significance, if $D_1 > 3.84$ we conclude that a six-parameter model is required, otherwise, we compute $D_2 = 2(L_5 - L_4)$. Similarly to what we did before, if $D_2 > 3.84$, we conclude that a five-parameter model is required, otherwise, we compute $D_3 = 2(L_4 - L_3)$. If $D_3 > 3.84$, a four-parameter model is suggested, otherwise, a three-parameter one is considered enough.

Knowing the appropriate subset of the GG family for given data, we can then move to plotting the corresponding estimated quantiles and later on the confidence bands around them as will be shown in Section 4.5.

4.4 Asymptotics of the Four-, Five- and Six-Parameter GG ML Estimates

In the four-parameter GG, as explained in item 1 of Section 3.4, the parameters to be estimated are a, b, σ and k. The model is dependent on a covariate X through $\mu = a + bX$. Referring to the four-parameter GG loglikelihood equation (4.1) and the corresponding score equations (4.8), (4.10), (4.11) and (4.12) and similarly to the three-parameter case, we find that the components of the Fisher information matrix at the ML estimates are

$$\begin{aligned}
E\left(-\frac{\partial^2 l}{\partial a^2}\right) &= \frac{n}{\sigma^2};\\
E\left(-\frac{\partial^2 l}{\partial a \partial b}\right) &= \frac{n\overline{X}}{\sigma^2};\\
E\left(-\frac{\partial^2 l}{\partial a \partial \sigma}\right) &= \frac{n}{\sigma^2} f_1(k);\\
E\left(-\frac{\partial^2 l}{\partial a \partial k}\right) &= \frac{n}{\sigma} f_2(k);\\
E\left(-\frac{\partial^2 l}{\partial b^2}\right) &= \frac{n\overline{X^2}}{\sigma^2};\\
E\left(-\frac{\partial^2 l}{\partial b \partial \sigma}\right) &= \frac{n\overline{X}}{\sigma^2} f_1(k);\\
E\left(-\frac{\partial^2 l}{\partial b \partial k}\right) &= \frac{n\overline{X}}{\sigma} f_2(k);\\
E\left(-\frac{\partial^2 l}{\partial \sigma^2}\right) &= \frac{n}{\sigma^2} g_1(k);\\
E\left(-\frac{\partial^2 l}{\partial \sigma \partial k}\right) &= \frac{n}{\sigma} g_2(k);\\
E\left(-\frac{\partial^2 l}{\partial k^2}\right) &= n g_3(k).
\end{aligned}$$

Being the inverse of the information matrix, it turns out that the covariance matrix is independent of a and b.

We now move to the five-parameter case where σ is also dependent on X through $\sigma = \exp(c + dX)$. The parameters are now a, b, c, d and k, leading to a (5×5) information matrix whose components at the ML estimates turn out to be

$$
\begin{aligned}
E\left(-\frac{\partial^2 l}{\partial a^2}\right) &= \sum_{i=1}^{n}\left(\frac{1}{\exp(c+dX_i)}\right)^2;\\
E\left(-\frac{\partial^2 l}{\partial a\partial b}\right) &= \sum_{i=1}^{n} X_i\left(\frac{1}{\exp(c+dX_i)}\right)^2;\\
E\left(-\frac{\partial^2 l}{\partial a\partial c}\right) &= f_1(k)\sum_{i=1}^{n}\frac{1}{\exp(c+dX_i)};\\
E\left(-\frac{\partial^2 l}{\partial a\partial d}\right) &= f_1(k)\sum_{i=1}^{n}\frac{X_i}{\exp(c+dX_i)};\\
E\left(-\frac{\partial^2 l}{\partial a\partial k}\right) &= f_2(k)\sum_{i=1}^{n}\frac{1}{\exp(c+dX_i)};\\
E\left(-\frac{\partial^2 l}{\partial b^2}\right) &= \sum_{i=1}^{n}\left(\frac{X_i}{\exp(c+dX_i)}\right)^2;\\
E\left(-\frac{\partial^2 l}{\partial b\partial c}\right) &= f_1(k)\sum_{i=1}^{n}\frac{X_i}{\exp(c+dX_i)};\\
E\left(-\frac{\partial^2 l}{\partial b\partial d}\right) &= f_1(k)\sum_{i=1}^{n}\frac{X_i^2}{\exp(c+dX_i)};\\
E\left(-\frac{\partial^2 l}{\partial b\partial k}\right) &= f_2(k)\sum_{i=1}^{n}\frac{X_i}{\exp(c+dX_i)};\\
E\left(-\frac{\partial^2 l}{\partial c^2}\right) &= ng_1(k);\\
E\left(-\frac{\partial^2 l}{\partial c\partial d}\right) &= n\overline{X}g_1(k);\\
E\left(-\frac{\partial^2 l}{\partial c\partial k}\right) &= ng_2(k)\\
E\left(-\frac{\partial^2 l}{\partial d^2}\right) &= n\overline{X^2}g_1(k);\\
E\left(-\frac{\partial^2 l}{\partial d\partial k}\right) &= n\overline{X}g_2(k);\\
E\left(-\frac{\partial^2 l}{\partial k^2}\right) &= ng_3(k).
\end{aligned}
$$

Finally, in the six-parameter GG, in additional to the other parameters, the shape parameter k is dependent on X through $k = \exp(f + gX)$. By transitivity, $r(k, q)$ is also dependent on X. The six parameters to be estimated are a, b, c, d, f and g and the components of the Fisher information matrix at the ML estimates are

$$E\left(-\frac{\partial^2 l}{\partial a^2}\right) = \sum_{i=1}^{n}\left(\frac{1}{\exp(c+dX_i)}\right)^2;$$

$$E\left(-\frac{\partial^2 l}{\partial a \partial b}\right) = \sum_{i=1}^{n} X_i\left(\frac{1}{\exp(c+dX_i)}\right)^2;$$

$$E\left(-\frac{\partial^2 l}{\partial a \partial c}\right) = \sum_{i=1}^{n}\left\{\frac{1}{\exp(c+dX_i)} f_1(\exp(f+gX_i))\right\};$$

$$E\left(-\frac{\partial^2 l}{\partial a \partial d}\right) = \sum_{i=1}^{n}\left\{\frac{X_i}{\exp(c+dX_i)} f_1(\exp(f+gX_i))\right\};$$

$$E\left(-\frac{\partial^2 l}{\partial a \partial f}\right) = \frac{1}{2}\sum_{i=1}^{n}\left\{\frac{1}{\exp(c+dX_i)} f_1(\exp(f+gX_i))\right\};$$

$$E\left(-\frac{\partial^2 l}{\partial a \partial g}\right) = \frac{1}{2}\sum_{i=1}^{n}\left\{\frac{X_i}{\exp(c+dX_i)} f_1(\exp(f+gX_i))\right\};$$

$$E\left(-\frac{\partial^2 l}{\partial b^2}\right) = \sum_{i=1}^{n}\left(\frac{X_i}{\exp(c+dX_i)}\right)^2;$$

$$E\left(-\frac{\partial^2 l}{\partial b \partial c}\right) = \sum_{i=1}^{n}\left\{\frac{X_i}{\exp(c+dX_i)} f_1(\exp(f+gX_i))\right\};$$

$$E\left(-\frac{\partial^2 l}{\partial b \partial d}\right) = \sum_{i=1}^{n}\left\{\frac{X_i^2}{\exp(c+dX_i)} f_1(\exp(f+gX_i))\right\};$$

$$E\left(-\frac{\partial^2 l}{\partial b \partial f}\right) = \frac{1}{2}\sum_{i=1}^{n}\left\{\frac{X_i}{\exp(c+dX_i)} f_1(\exp(f+gX_i))\right\};$$

$$E\left(-\frac{\partial^2 l}{\partial b \partial g}\right) = \frac{1}{2}\sum_{i=1}^{n}\left\{\frac{X_i^2}{\exp(c+dX_i)} f_1(\exp(f+gX_i))\right\};$$

$$E\left(-\frac{\partial^2 l}{\partial c^2}\right) = \sum_{i=1}^{n} g_1(\exp(f+gX_i));$$

$$E\left(-\frac{\partial^2 l}{\partial c \partial d}\right) = \sum_{i=1}^{n}\{X_i g_1(\exp(f+gX_i))\};$$

$$E\left(-\frac{\partial^2 l}{\partial c \partial f}\right) = \sum_{i=1}^{n}\{\exp(f+gX_i) g_2(\exp(f+gX_i))\};$$

$$E\left(-\frac{\partial^2 l}{\partial c \partial g}\right) = \sum_{i=1}^{n}\{X_i \exp(f+gX_i) g_2(\exp(f+gX_i))\};$$

$$E\left(-\frac{\partial^2 l}{\partial d^2}\right) = \sum_{i=1}^{n}\{X_i^2 g_1(\exp(f+gX_i))\};$$

$$E\left(-\frac{\partial^2 l}{\partial d \partial f}\right) = \sum_{i=1}^{n}\{X_i \exp(f+gX_i) g_2(\exp(f+gX_i))\};$$

$$
\begin{aligned}
E\left(-\frac{\partial^2 l}{\partial d \partial g}\right) &= \sum_{i=1}^{n}\left\{X_i^2 \exp(f+gX_i) g_2(\exp(f+gX_i))\right\}; \\
E\left(-\frac{\partial^2 l}{\partial f^2}\right) &= \sum_{i=1}^{n}\left\{\exp(2f+2gX_i) g_3(\exp(f+gX_i))\right\}; \\
E\left(-\frac{\partial^2 l}{\partial f \partial g}\right) &= \sum_{i=1}^{n}\left\{X_i \exp(2f+2gX_i) g_3(\exp(f+gX_i))\right\}; \\
E\left(-\frac{\partial^2 l}{\partial g^2}\right) &= \sum_{i=1}^{n}\left\{X_i^2 \exp(2f+2gX_i) g_3(\exp(f+gX_i))\right\};
\end{aligned}
$$

Given the form of the information matrix in all of the four-, five- and six-parameter cases we deduce the asymptotic covariance and correlation matrices. We use the covariance matrix to compute confidence bands around the quantiles, numerically.

4.5 Confidence Intervals for the Quantile Curves

Knowing the expressions of the quantile functions t_q (and y_q) in the three-, four-, five- and six-parameter case-scenarios, we aim to fit confidence bands for them around the estimated quantile functions. This is done using the asymptotic results of the GG ML parameter estimates (from Sections 2.6 and 4.4). We note that these are pointwise confidence bands calculated at every X_i, $i = 1, ..., n$.

Recall that

$$t_q = \exp(\mu)\left(\frac{r(k,q)}{k}\right)^{\sigma\sqrt{k}} \tag{4.13}$$

and

$$y_q = \mu + \sigma\sqrt{k}\log\left(\frac{r(k,q)}{k}\right), \tag{4.14}$$

where μ, σ and k are the location, scale and shape parameters respectively that will be linked to the covariate.

Let $\mathbf{\Lambda}$ be the vector of first derivatives of t_q with respect to the GG parameters. Denote by $\hat{\mathbf{\Lambda}}$ and $\hat{t_p}$ the estimates of $\mathbf{\Lambda}$ and t_q respectively, obtained when replacing the GG parameters

by their ML estimates. It is known, using the delta method (explained in Davison (2003)), that the asymptotic expectation and variance of $\hat{t_p}$ are respectively t_p and $\mathbf{\Lambda}^T\mathbf{I}^{-1}\mathbf{\Lambda}$, where $\mathbf{I}$ is the Fisher information matrix. The variance is usually estimated by $\hat{\mathbf{\Lambda}}^T\mathbf{I}^{-1}\hat{\mathbf{\Lambda}}$. Denote the variance of t_q by Σ and its estimate by $\widehat{\Sigma}$. Then, using the Central Limit Theorem and the asymptotic normality assumptions, expressions for the 95% and 99% confidence intervals for t_q are respectively $\left[\hat{t_q} - 1.96\sqrt{\hat{\Sigma}}, \hat{t_q} + 1.96\sqrt{\hat{\Sigma}}\right]$ and $\left[\hat{t_q} - 2.575\sqrt{\hat{\Sigma}}, \hat{t_q} + 2.575\sqrt{\hat{\Sigma}}\right]$. A similar argument holds for the confidence intervals of y_q. $\mathbf{I}^{-1}$ can be found by inverting the Fisher information matrix whose expressions for the three-, four-, five- and six-parameter GG are given in Sections 2.6 and 4.4. It remains to find $\mathbf{\Lambda}$, the vector of derivatives.

As stated in Section 3.3, the q^{th} quantile of the gamma distribution, $r(k,q)$, is the inverse of the distribution function at q which turns out to be the inverse of the regularised gamma function satisfying

$$\frac{\Gamma_{r(k,q)}(k)}{\Gamma(k)} = q.$$

Notation:

- $P(k, r(k,q)) = \frac{\Gamma_{r(k,q)}(k)}{\Gamma(k)}$.
- $r_k = \frac{\partial r(k,q)}{\partial k}$.
- $P^{10}(k, r(k,q)) = \frac{\partial P(k,r(k,q))}{\partial k}$ and $P^{01}(k, r(k,q)) = \frac{\partial P(k,r(k,q))}{\partial r(k,q)}$.
- $r_f = \frac{\partial r(f,g,q)}{\partial f}$ and $r_g = \frac{\partial r(f,g,q)}{\partial g}$ (in the six-parameter case when $k = \exp(f + gX)$).

We know that for known k, $P(k, r(k,q)) = q$. Differentiating this equation with respect to k on both sides, we obtain

$$P^{10}(k, r(k,q)) + r_k P^{01}(k, r(k,q)) = 0,$$

where

$$P^{10}(k, r(k,q)) = \frac{\int_0^{r(k,q)} \log(t) t^{k-1} e^{-t} dt - \psi(k)\Gamma_{r(k,q)}(k)}{\Gamma(k)}$$

and

$$P^{01}(k, r(k,q)) = \frac{r(k,q)^{k-1} e^{-r(k,q)}}{\Gamma(k)}.$$

It therefore turns out that

$$\begin{aligned} r_k &= -\frac{P^{10}(k, r(k,q))}{P^{01}(k, r(k,q))} \\ &= \frac{\psi(k)\Gamma_{r(k,q)}(k) - \int_0^{r(k,q)} \log(t) t^{k-1} e^{-t} dt}{r(k,q)^{k-1} e^{-r(k,q)}}. \end{aligned}$$

In the six-parameter case when $k = \exp(f + gx)$,

$$r_f = \exp(f + gx) \frac{\psi(\exp(f + gx))\Gamma_{r(f,g,q)}(\exp(f + gx)) - \int_0^{r(f,g,q)} \log(t) t^{\exp(f+gx)-1} e^{-t} dt}{r(f,g,q)^{\exp(f+gx)-1} e^{-r(f,g,q)}}.$$

and

$$r_g = x r_f.$$

We deduce that expressions for $\mathbf{\Lambda}$, the vector of first derivatives of t_q, for the three-, four-, five- and six-parameter GG are

- The three-parameter case:

$$\mathbf{\Lambda} = \begin{pmatrix} \exp(\mu)\left(\frac{r(k,q)}{k}\right)^{\sigma\sqrt{k}} \\ \sqrt{k}\exp(\mu)\left(\frac{r(k,q)}{k}\right)^{\sigma\sqrt{k}} \log\left(\frac{r(k,q)}{k}\right) \\ \sigma\exp(\mu)\left(\frac{r(k,q)}{k}\right)^{\sigma\sqrt{k}} \left\{\frac{1}{2\sqrt{k}} \log\left(\frac{r(k,q)}{k}\right) + \frac{k r_k - r(k,q)}{r(k,q)\sqrt{k}}\right\} \end{pmatrix}$$

- The four-parameter case:

$$\Lambda = \begin{pmatrix} \exp(a+bx)\left(\frac{r(k,q)}{k}\right)^{\sigma\sqrt{k}} \\ x\exp(a+bx)\left(\frac{r(k,q)}{k}\right)^{\sigma\sqrt{k}} \\ \sqrt{k}\exp(a+bx)\left(\frac{r(k,q)}{k}\right)^{\sigma\sqrt{k}}\log\left(\frac{r(k,q)}{k}\right) \\ \sigma\exp(a+bx)\left(\frac{r(k,q)}{k}\right)^{\sigma\sqrt{k}}\left\{\frac{1}{2\sqrt{k}}\log\left(\frac{r(k,q)}{k}\right)+\frac{kr_k-r(k,q)}{r(k,q)\sqrt{k}}\right\} \end{pmatrix}$$

- The five-parameter case:

$$\Lambda = \begin{pmatrix} \exp(a+bx)\left(\frac{r(k,q)}{k}\right)^{\sqrt{k}\exp(c+dx)} \\ x\exp(a+bx)\left(\frac{r(k,q)}{k}\right)^{\sqrt{k}\exp(c+dx)} \\ \sqrt{k}\exp(a+bx)\left(\frac{r(k,q)}{k}\right)^{\sqrt{k}\exp(c+dx)}\log\left(\frac{r(k,q)}{k}\right)\exp(c+dx) \\ x\sqrt{k}\exp(a+bx)\left(\frac{r(k,q)}{k}\right)^{\sqrt{k}\exp(c+dx)}\log\left(\frac{r(k,q)}{k}\right)\exp(c+dx) \\ \exp(a+bx+c+dx)\left(\frac{r(k,q)}{k}\right)^{\sqrt{k}\exp(c+dx)}\left\{\frac{1}{2\sqrt{k}}\log\left(\frac{r(k,q)}{k}\right)+\frac{kr_k-r(k,q)}{r(k,q)\sqrt{k}}\right\} \end{pmatrix}$$

- The six-parameter case:

$$\Lambda = \begin{pmatrix} \exp(a+bx)\left(\frac{r(f,g,q)}{\exp(f+gx)}\right)^{E} \\ x\exp(a+bx)\left(\frac{r(f,g,q)}{\exp(f+gx)}\right)^{E} \\ E\exp(a+bx)\left(\frac{r(f,g,q)}{\exp(f+gx)}\right)^{E}\log\left(\frac{r(f,g,q)}{\exp(f+gx)}\right) \\ xE\exp(a+bx)\left(\frac{r(f,g,q)}{\exp(f+gx)}\right)^{E}\log\left(\frac{r(f,g,q)}{\exp(f+gx)}\right) \\ E\exp(a+bx)\left(\frac{r(f,g,q)}{\exp(f+gx)}\right)^{E}\left\{\frac{1}{2}\left(\log(r(f,g,q))-f-gx\right)+\frac{r_f-r(f,g,q)}{r(f,g,q)}\right\} \\ xE\exp(a+bx)\left(\frac{r(f,g,q)}{\exp(f+gx)}\right)^{E}\left\{\frac{1}{2}\left(\log(r(f,g,q))-f-gx\right)+\frac{r_g-r(f,g,q)}{r(f,g,q)}\right\} \end{pmatrix},$$

where $E = \exp(c+dx+1/2(f+gx))$.

We now have all the components for plotting the confidence bands for the quantiles of a three-, four-, five- and six-parameter GG.

4.6 Goodness-of-Fit Test for the GG Model

Following model identification, parameter estimation and model selection, the final stage in our overall package is to test the goodness of the fitted model. In this case, the main question is: How good a fit is our GG model?

To answer this question, we will benefit from the fact that QR allows for the fitting of quantiles at every percentile covering the whole range of the data. These curves represent, ideally, the best fit at every percentile. Ideally, $100j\%$ of the data should lie between the q^{th} and $(q+j)^{th}$ quantile. For our goodness-of-fit test, we consider five main quantiles: $q = \{0.1, 0.25, 0.5, 0.75, 0.9\}$. These quantiles divide the data into six regions. Denote by $\hat{m}_l$ and m_l, $l = 1, ..., 6$, the observed and expected number of data points in each of the six regions. If n is the sample size, obviously $m_1 = 0.1n$, $m_2 = 0.15n$, $m_3 = 0.25n$, $m_4 = 0.25n$, $m_5 = 0.15n$ and $m_6 = 0.1n$.

Based on the χ^2 goodness-of-fit test, we propose the following test statistic

$$\tau \equiv \sum_{l=1}^{6} \frac{(\hat{m}_l - m_l)^2}{m_l}. \tag{4.15}$$

In a simulation study we present later in Section 5.2, we show that, approximately,

$$\tau = \frac{3}{4}\varrho, \quad \text{where } \varrho \sim \chi^2_4 \text{ or } \tau \sim G(\theta = 1.5, k = 2).$$

Note that we were not able to prove this result theoretically. It is based merely on computational evidence obtained from the simulations. We use this result to test the goodness-of-fit of the GG model. The 95% quantile of a $G(\theta = 1.5, k = 2)$ is approximately 7.12. Therefore, at a 0.05 level of significance, if $\tau > 7.12$, we reject the fact that the GG is a good fit, otherwise, we conclude that the GG is a good model. Similarly, at a 0.01 level of significance the rejection criterion is $\tau > 9.96$. A possible improvement to the test, that we may explore in the future (but haven't done in this thesis), may be achieved by splitting up the region considered by the covariate as

well as the quantiles. Dividing the abscissa axis into regions and computing residuals might allow a more accurate test and a better scope for modifying the null hypothesis.

Chapter 5

Simulation Study

In this chapter, we validate, via simulations, some of the theoretical and methodological work proposed throughout the thesis. Two simulation studies are presented and summary results are given. In the first study we apply the different steps of the modeling package suggested in Chapter 1 to an overall collection of 1600 data sets simulated from the three-, four-, five- and six-parameter GG (400 from each). For each data set, LRTs are conducted to choose the best GG model from within the family. On that basis, quantile curves and confidence bands are plotted. It is shown that the quantile curves lie within their CIs most of the time and that the relative biases and squared errors of the estimated parameters and quantiles are reasonably small. In the second study, using the same three-, four-, five- and six-parameter data sets of the first study, we validate via simulations the goodness-of-fit test suggested in Section 4.6.

5.1 LRTs, Quantiles, Confidence Bands, Biases and Mean Squared Errors for GG Simulations

In this simulation study, for each of the three-, four-, five- and six-parameter GG distributions, we addressed four different situations (coming from four different sets of parameters) by generating 100 data sets from each and analysing the results in the context of QR. In total we explored 16 case-scenarios using 1600 data sets. The parameters of the models estimated from

the data sets were estimated using four different sets of initial parameters each including ten groups of parameters used as initial values. Among the ten repetitions, we consider the highest likelihood attained as being the ML and we find the corresponding estimates.

On the one hand, pretending that we don't know the actual number of parameters from which each data set was simulated, we conduct LRTs to test for the number of parameters and we find the proportion of times the obtained results match with the reality. Based on the results from the LRTs concerning the number of parameters to be used, we estimate the corresponding GG parameters using MLE. Then, accordingly, we compute the estimated quantile curves. We call this the "LRT Version". For each estimated quantile curve corresponding to each of the 100 data sets coming from one of the case-scenarios, we plot CIs and find what proportion of the true quantile lies within the CIs, then we average over the hundred data sets. Ideally, for 95% CIs, we should obviously expect the true quantiles to lie within the CIs 95% of the time.

On the other hand, we repeat the same mechanism while using, this time, the actual (true) number of parameters from which each data set was simulated originally, rather than the one obtained from the LRT. We refer to this as the "Known p Version".

Finally, we compare between both approaches the proportion of time the true quantiles lie within the CIs. We would like the results obtained from the former to be as close as possible to the latter. We divide this study into four subsections. In the first one, we present the 16 case-scenarios of parameters used to simulate the data sets. Secondly, we display and discuss the LRT results. Thirdly, we analyse the closeness of the estimated values of the parameters to the known ones. Lastly, we study how the estimates affect the shapes of the quantiles, reporting the percentage of times these quantiles lie within their CIs. Table 5.1 displays all the notations used in the tables of this section along with their explanation.

Table 5.1: Explanation of the notations in the tables of Section 5.3.

Notation	Explanation
D31, D32, D33, D34	Data sets of the four case-scenarios from the three-parameter GG. Each one comprises 100 simulated data sets.
D41, D42, D43, D44	Data sets of the four case-scenarios from the four-parameter GG. Each one comprises 100 simulated data sets.
D51, D52, D53, D54	Data sets of the four case-scenarios from the five-parameter GG. Each one comprises 100 simulated data sets.
D61, D62, D63, D64	Data sets of the four case-scenarios from the six-parameter GG. Each one comprises 100 simulated data sets.
N3, N4, N5, N6	Number of times the LRT predicts a three-, four-, five- or six-parameter GG respectively
Known p Version	Reported results taking into consideration the real number of parameter from which data was simulated.
LRT Version	Reported results taking into consideration the number of parameter predicted by the LRT.
q0.1, q0.25, q0.5, q0.75, q0.9	10%, 25%, 50%, 75% and 90% quantiles respectively
biasa, biasb, biasc, biasd, biasf, biasg, biasmu, biassigma, biask	Mean relative bias of a, b, c, d, f, g, μ, σ, and k respectively
msea, mseb, msec, msed, msef, mseg, msemu, msesigma, msek	Square root of the mean relative squared error of a, b, c, d, f, g, μ, σ, and k respectively
mrbias , mrse	Pointwise mean relative bias and root mean relative squared error of the quantiles, respectively
CI95prop, CI99prop	Proportion of times the estimated quantiles lie within their 95% and 99% CIs respectively

5.1.1 Data Sets

The 16 cases of parameters were chosen in a way to cover most scenarios a GG can portray. Tables 5.2, 5.3, 5.4 and 5.5 present respectively the four scenarios of parameters from which each of the three-, four-, five- and six-parameter GG data sets were generated. Note that 100 data sets were simulated from each of the 16 parameter cases. The four-, five- and six-parameter cases depend on a covariate, say X, which we make uniformly distributed over the interval (0,1). A different covariate was simulated for each data set of the four-, five- and six-parameter case-scenarios. For each of those case-scenarios, we make accordingly (one or more of) the parameters dependent on the corresponding covariate and we calculate the corresponding θ, α and β. Using those, we then simulate the data sets as explained in Section 2.2 by considering the $1/\beta^{th}$ power of a gamma distribution. In the three-parameter case, we tackle situations where $\alpha < 1$ (D32) as well as $\alpha > 1$ (D31, D33 and D34) which reflect respectively one- and two-tailed GG pdfs. The parameters are also chosen in a way so as to deal with data sets that have fairly large values (D32) as well as ones that take only smaller values (D31, D33 and D34). In this context, we deal with highly-dispersed and non-highly-dispersed data as well. In the four-parameter case scenarios, data is dependent on a covariate. We deal with various cases such as when the quantiles are increasing ($b > 0$ in D42 and D43) or decreasing ($b < 0$ in D42 and D44). We also make sure that the parameters cover a large range of values from relatively small (e.g. $b = -0.1$ in D42) to relatively large (e.g. $k = 5$ in D43). Similarly, in the five-parameter case-scenarios, we cover a reasonably wide range of values for each parameter making sure at the same time that most different quantile shapes are covered. We take different combinations of positive and negative parameters b and d which result in monotone and unimodal quantiles. Likewise, in the six-parameter case, we make sure that the four case-scenarios address situations with monotone, unimodal and bimodal quantiles.

Table 5.2: Parameters used to generate the four case-scenarios of the three-parameter GG.

	μ	σ	k	θ	α	β
D31	-1.6	0.50	0.5	0.258	1.414	2.828
D32	4.5	2.80	2.0	5.785	0.505	0.253
D33	-1.0	0.35	5.0	0.104	6.389	1.278
D34	0.5	1.00	3.0	0.246	1.732	0.577

Table 5.3: Parameters used to generate the four case-scenarios of the four-parameter GG.

	a	b	σ	k
D41	1.5	0.5	0.75	0.2
D42	5.0	-0.1	1.00	1.5
D43	0.2	7.0	1.50	5.0
D44	-1.0	-2.0	2.00	4.0

Table 5.4: Parameters used to generate the four case-scenarios of the five-parameter GG.

	a	b	c	d	k
D51	6.0	-0.4	-2.5	0.8	1.50
D52	3.0	0.5	-0.1	-0.5	3.00
D53	1.0	-0.1	-1.5	-2.0	0.75
D54	-1.5	-6.0	0.5	1.5	4.00

Table 5.5: Parameters used to generate the four case-scenarios of the six-parameter GG.

	a	b	c	d	f	g
D61	0.5	5.00	1.00	-0.75	4.0	1.5
D62	-2.0	-0.75	-0.50	-4.00	-0.2	-1.0
D63	2.0	-1.50	0.25	0.10	-3.0	5.0
D64	3.0	0.10	-5.00	1.50	2.0	-4.0

5.1.2 Performance of the Likelihood Ratio Tests

As already mentioned, LRTs are applied to each of the 1600 data sets. Table 5.6 displays the number of parameters reported by the LRTs for each of the 16 case-scenarios. Each of the 16 columns presents the proportion of times each of the hundred data sets (coming from one of the 16 case scenarios) are reported to come from a three-, four-, five- or six-parameter GG by the LRT. All four case-scenarios of the three-parameter GG were reported as having three parameters at an average of 86% of the time for $n = 200$ and 84.5% for $n = 500$. The reason for that is D32 where the true number of parameters is reported more times in the $n = 200$ than in the $n = 500$. We suspect that this is due to the randomness of the simulation process. The four-parameter GG case-scenarios were reported as having four parameters most of the time in D41, D43 and D44. The only exception is D42 which is reported as being a three-parameter GG 88% of the time for $n = 200$ and 81% of the time for $n = 500$. Looking back at the parameters used to generate D42, we find that $b = -0.1$ which is relatively close to zero. This difference from zero could not be picked up in LRT terms, hence the result. This suggests that the three- and four-parameter GG are very similar in this case and should give very similar results, so the LRT performance provides us with no practical problem. The LRT reports five parameters most of the time for the five-parameter GG (89.5% on average for $n = 200$ and 95.24% on average for $n = 500$). Very rarely were three or four parameters reported. Finally, in the six-parameter GG case-scenarios, D63 and D64 were reported as having six parameters most of the time, whereas D61 and D62 were reported as having five parameters mostly. Looking at the quantile shapes in the latter two cases, we don't observe any bimodalities in (0,1). They can be obtained using a five-parameter GG which made the LRT predict this model. Choosing a reasonable five-parameter model by the LRT when, in reality a six-parameter model is the true one, is shown later to have not affected the quality of estimation.

Table 5.6: Predicted number of parameters by the LRT.

	D31	D32	D33	D34	D41	D42	D43	D44	D51	D52	D53	D54	D61	D62	D63	D64
$n = 200$																
N3	84	90	83	87	16	88	0	0	0	9	0	0	0	0	0	0
N4	6	4	5	3	70	7	90	91	1	11	0	0	1	0	0	0
N5	2	3	4	5	6	5	8	5	91	77	94	96	99	75	0	6
N6	8	3	8	5	8	0	2	4	8	3	6	4	0	25	100	94
$n = 500$																
N3	86	81	84	87	2	81	0	0	0	0	0	0	0	0	0	0
N4	3	8	3	4	85	8	90	91	0	1	0	0	0	0	0	0
N5	7	3	7	6	6	4	5	5	98	92	95	96	98	40	0	0
N6	4	8	6	3	7	7	5	4	2	7	5	4	2	58	100	100

In all, we did not find any noticeable differences between the $n = 200$ and $n = 500$ cases. However, from this analysis, questions especially concerning the situations where the LRT did not report the expected number of parameters should be addressed: How does that affect the estimated quantiles which is reflected in the first place by the estimated parameters? Generally, in both the "Known p Version" and the "LRT Version", how good are the results?

5.1.3 Analysis of the "LRT Version" and the "Known p Version"

In this analysis, we compare the estimated parameters to the true ones (from which data was simulated) for each of the "LRT Version" and the "Known p Version". For each data set of the 16 case-scenarios, we compute two statistics: the component of the relative bias given by

$$\frac{(Estimated\ Parameter - Real\ Parameter)}{Real\ Parameter}$$

and the component of the relative squared error

$$\left(\frac{Estimated\ Parameter - Real\ Parameter}{Real\ Parameter}\right)^2$$

and we average over the 100 data sets from every case-scenario. We then take the square root of the mean relative squared error. This is straightforward in the "Known p Version" because the number of estimated parameters is the same as the true number of parameters. A few alterations are done in the "LRT Version" when the predicted number of estimated parameters differs from the true one. If the number of predicted parameters is greater than the true one, the actual bias and the squared error of the extra estimated parameter are given by the averages of $(Estimated\ Parameter)$ and $(Estimated\ Parameter)^2$, the latter square-rooted, respectively. If the true number of parameters is greater than the predicted one, the relative bias and the relative squared error for the extra real parameter are given by the averages of the relative quantities where 0 is entered in the place of the $(Estimated\ Parameter)$. It remains to say

that, always, the estimated parameters are compared to the real ones, the latter ones taken as reference, hence the order of parameters in the "LRT Version" of the tables of results.

For $n = 200$, Tables 5.7, 5.8, 5.9 and 5.10 report the mean relative biases and the square root of the mean relative squared errors of the parameters for each of the three-, four-, five- and six-parameter GG simulations respectively. This is given for both the "LRT Version" and the "Known p Version". The analogues of those table for the $n = 500$ case are given in Tables A.7, A.8, A.9 and A.10 of the Appendix.

Analysing those results, we realise that generally the mean relative biases and the root mean relative squared errors are quite small. For most of the cases, the "Known p Version" is slightly better than the "LRT Version", though the difference is very small. Also, there is very little difference between the $n = 200$ and $n = 500$ cases. Although the $n = 500$ case has slightly smaller biases and errors, our method still performs well in the $n = 200$ case inferring that the approach still works reasonably well for smaller-size data sets. Of particular interest is that D42, D61 and D62 behave fairly well. Even though the LRT failed to predict the right number of parameters for those data sets, it still managed to predict a reasonable model. In rare cases, we notice large values in the tables. For example, "msek" and "biask" are quite large sometimes in Table 5.7 for D33 and D34. This reflects the fairly rare failure of the GG to accurately estimate k when k is relatively large, as the shape of the distribution only slightly changes when k approaches infinity. We do not observe a noticeable greater/less number of positive or negative biases in general which means that there is no systematic over/under-estimation.

Having compared the estimated and true parameters for both versions, we move now to find out, most importantly, the effect of those results on the quantile shapes and hence on the overall model. This was done for all data sets of the four case-scenarios of which we display the results of D32, D42, D54 and D62 in the next section.

Table 5.7: Mean relative bias (biasmu, biassigma, biask, biasb, biasd, biasg) and root mean relative squared error (msemu, msesigma, msek, mseb, msed, mseg) of the estimated parameters of the three-parameter GG data sets for $n = 200$. In the case of b, d and g in the "LRT Version" 'relative' is replaced by 'actual'.

Known p Version				
	D31	D32	D33	D34
biasmu	-0.008	0.000	-0.003	-0.004
msemu	0.039	0.070	0.037	0.212
biassigma	-0.024	-0.012	-0.005	-0.003
msesigma	0.087	0.069	0.055	0.061
biask	0.036	0.209	1.251	0.844
msek	0.340	0.915	4.812	2.358
LRT Version				
biasmu	-0.009	-0.008	-0.010	-0.020
msemu	0.068	0.088	0.063	0.290
biassigma	-0.029	-0.013	-0.010	0.005
msesigma	0.125	0.076	0.093	0.096
biask	0.058	1.301	1.882	1.790
msek	0.532	8.434	6.407	7.171
biasb	-0.003	0.068	-0.014	0.015
mseb	0.176	0.493	0.100	0.194
biasd	0.009	-0.001	0.006	-0.020
msed	0.210	0.082	0.135	0.143
biasg	0.093	-0.065	0.118	-0.040
mseg	0.847	0.932	1.622	1.129

Table 5.8: Mean relative bias (biasa, biasb, biassigma, biask, biasd, biasg) and root mean relative squared error (msea, mseb, msesigma, msek, msed, mseg) of the estimated parameters of the four-parameter GG data sets for $n = 200$. In the case of d and g in the "LRT Version" 'relative' is replaced by 'actual'.

Known p Version				
	D41	D42	D43	D44
biasa	0.010	0.000	0.034	-0.002
msea	0.104	0.036	1.178	0.271
biasb	0.070	0.085	0.005	0.007
mseb	0.381	2.664	0.049	0.227
biassigma	-0.058	-0.009	-0.022	-0.013
msesigma	0.136	0.049	0.058	0.060
biask	-0.070	0.161	1.505	1.000
msek	0.314	0.607	5.554	4.004
LRT Version				
biasa	0.022	-0.006	0.071	0.009
msea	0.125	0.029	1.184	0.294
biasb	-0.024	-0.548	0.003	-0.006
mseb	0.618	1.837	0.051	0.276
biassigma	-0.043	0.001	-0.027	-0.016
msesigma	0.179	0.068	0.080	0.080
biask	-0.023	0.162	1.503	1.223
msek	0.455	0.591	5.567	4.250
biasd	-0.012	-0.013	0.009	0.004
msed	0.246	0.101	0.113	0.113
biasg	0.012	0.000	0.108	-0.087
mseg	0.573	0.000	0.765	0.908

Table 5.9: Mean relative bias (biasa, biasb, biasc, biasd, biask, biasg) and root mean relative squared error (msea, mseb, msec, msed, msek, mseg) of the estimated parameters of the five-parameter GG data sets for $n = 200$. In the case of g in the "LRT Version" 'relative' is replaced by 'actual'.

Known p Version				
	D51	D52	D53	D54
biasa	0.000	0.002	0.001	-0.012
msea	0.003	0.049	0.021	0.256
biasb	0.004	-0.028	-0.003	0.029
mseb	0.078	0.387	0.225	0.182
biasc	0.000	0.206	0.017	-0.079
msec	0.048	1.126	0.084	0.232
biasd	0.002	0.001	0.003	0.025
msed	0.250	0.382	0.091	0.127
biask	0.346	1.022	0.036	1.117
msek	1.210	5.170	0.353	3.915
LRT Version				
biasa	-0.001	0.006	0.000	-0.004
msea	0.003	0.055	0.023	0.264
biasb	-0.008	-0.076	-0.020	0.024
mseb	0.097	0.473	0.263	0.187
biasc	-0.002	0.451	0.013	-0.079
msec	0.053	1.514	0.085	0.234
biasd	-0.019	-0.094	0.015	0.023
msed	0.286	0.532	0.106	0.129
biask	2.541	1.450	0.180	1.917
msek	12.239	6.653	1.030	6.368
biasg	-0.255	-0.032	-0.114	-0.114
mseg	1.274	0.883	0.702	0.918

Table 5.10: Mean relative bias (biasa, biasb, biasc, biasd, biasf, biasg) and root mean relative squared error (msea, mseb, msec, msed, msef, mseg) of the six-parameter GG data sets for $n = 200$.

Known p Version				
	D61	D62	D63	D64
biasa	0.293	-0.003	0.040	0.000
msea	0.833	0.014	0.237	0.001
biasb	-0.002	0.009	0.069	0.000
mseb	0.109	0.039	0.455	0.061
biasc	-0.036	0.027	-0.438	0.004
msec	0.108	0.265	1.256	0.026
biasd	-0.023	0.013	0.901	-0.027
msed	0.244	0.074	4.201	0.187
biasf	-0.179	0.296	0.082	0.046
msef	0.450	2.799	0.263	0.570
biasg	-0.764	0.025	0.093	0.052
mseg	1.846	1.003	0.315	0.355
LRT Version				
biasa	0.175	-0.002	0.040	0.000
msea	0.711	0.014	0.237	0.001
biasb	0.009	0.008	0.069	-0.004
mseb	0.090	0.038	0.455	0.063
biasc	-0.033	0.057	-0.438	0.006
msec	0.106	0.260	1.256	0.026
biasd	-0.024	-0.006	0.901	-0.010
msed	0.248	0.071	4.201	0.181
biasf	-0.047	1.153	0.082	0.001
msef	0.314	2.831	0.263	0.608
biasg	-1.000	-0.448	0.093	0.009
mseg	1.000	1.111	0.315	0.403

5.1.4 Analysis of D32, D42, D54 and D62

In this subsection, we select four of the 16 simulated case-scenarios to discuss more closely. These are D32, D42, D54 and D62 for which we display respectively in Figures 5.1, 5.2, 5.3 and 5.4 a plot of the true underlying 10%, 25%, 50%, 75%, and 90% quantile curves. To each one of these figures (which in fact represents a set of quantiles that fit corresponding data) we use the predicted results of the LRT to estimate the parameters of every simulation.

The LRT results and the estimated parameters were discussed in the previous two subsections. It remains to see how these affect the estimated quantiles. For that, we calculate the mean relative biases and the square root of the mean relative squared errors between the estimated and the real quantiles as follows.

Consider first a certain data set from one of the case-scenarios, say the first data set from D31. Also, consider a certain quantile of the data set, say the 10% quantile, and a sequence of points $n_i = n_{i-1} + 0.001$, $i = 1, ..., 1000$ where $n_0 = 0$. For each point of abscissa n_i, $i = 1, ..., 1000$, we calculate the corresponding true and estimated quantiles for each of the "LRT Version" and the "Known p Version", then we deduce the component of the relative bias given by

$$\frac{(Estimated\ Quantile - Real\ Quantile)}{Real\ Quantile}$$

and the component of the relative squared error

$$\left(\frac{Estimated\ Quantile - Real\ Quantile}{Real\ Quantile}\right)^2$$

and we average over the 1000 points. We repeat the same procedure for the 100 data sets of D31 and we average again the obtained 100 mean relative biases and mean relative squared errors over the 100 data sets to get respectively "mrbias" and "mrsc", the latter square rooted.

We then obtain the confidence bands around the estimated quantiles and check what proportion of the true quantiles lies within the corresponding CIs. As the CIs are pointwise, we consider the sequence of points n_i, $i = 1, ..., 1000$, on the true quantiles and find out what proportion of those 1000 points are within the points (of same abscissa) on the confidence bands. This is also done for each of the 100 data sets of the given case-scenario. An average of the 100 proportions is then obtained. This procedure is then repeated for all case-scenarios D31, D32, ..., D64. The results for each of D32, D42, D54 and D62 when $n = 500$ are displayed in Tables 5.11, 5.12, 5.13 and 5.14 respectively.

Figure 5.1 of D32 shows highly dispersed quantiles. In its corresponding Table 5.11 of results, "mrbias" and "mrse" are quite small. Also, about 95% of the time the real quantiles lie within their 95% confidence bands on average and similarly for the 99% confidence bands. This pattern almost repeats itself in the other tables. In Table 5.12 which corresponds to the decreasing quantiles of D42, the "Known p Version" slightly over-estimates "CI95prop" and "CI99prop" compared with the "LRT Version" which shows slight under-estimation. In Table 5.13, "mrbias" and "mrse" are slightly higher than in other tables. This is due to the quantile shapes where the 90% quantile has a strong minimum while the 10%, 25%, 50% and 75% quantiles are monotone decreasing. "CI95prop" and "CI99prop" are slightly (though reasonably) under-estimated. Table 5.14 corresponds to the monotone and unimodal quantiles of D62. Similarly to D42, the 'Known p Version" slightly over-estimates "CI95prop" and "CI99prop", whereas in the "LRT Version" we observe a slight under-estimation. Noticeably, "mrbias" and "mrse" are generally very small. Also, any biases or large variances in parameter estimation seem not to have a great effect on the quality of quantile estimation. Particularly, large relative mean squared errors of $\hat{k}$ don't seem to have a noticeable negative influence on the results since this corresponds to distributions, and hence quantiles, that are similar over a broad range of k.

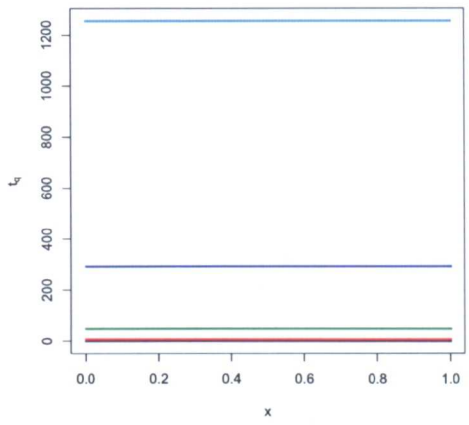

Figure 5.1: 10%, 25%, 50%, 75%, and 90% quantile curves underlying D32.

Table 5.11: Mean relative bias (mrbias) and root mean relative squared error (mrse) of quantiles and proportions of the real quantiles inside the confidence bands for D32 when $n = 500$.

Data32				
	mrbias	mrse	CI95prop	CI99prop
Known p Version				
q0.1	0.065	0.079	0.95	0.99
q0.25	0.030	0.034	0.96	1.00
q0.5	0.013	0.021	0.94	1.00
q0.75	0.005	0.018	0.94	0.98
q0.9	0.004	0.022	0.92	0.99
LRT version				
q0.1	0.078	0.337	0.958	0.979
q0.25	0.039	0.224	0.946	0.990
q0.5	0.018	0.183	0.932	0.990
q0.75	0.007	0.165	0.926	0.980
q0.9	0.006	0.177	0.897	0.976

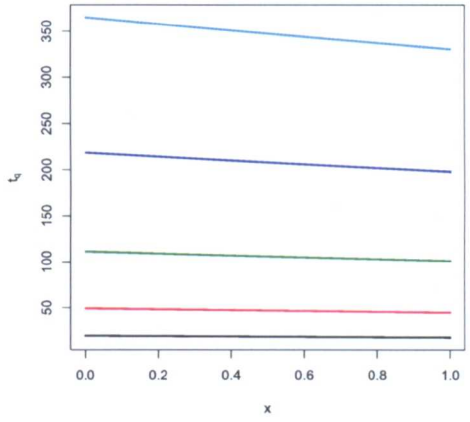

Figure 5.2: 10%, 25%, 50%, 75%, and 90% quantile curves underlying D42.

Table 5.12: Mean relative bias (mrbias) and root mean relative squared error (mrse) of quantiles and proportions of the real quantiles inside the confidence bands for D42 when $n = 500$.

Data42				
	mrbias	mrse	CI95prop	CI99prop
Known p Version				
q0.1	0.018	0.011	0.973	0.999
q0.25	0.011	0.006	0.969	0.992
q0.5	0.004	0.004	0.965	0.991
q0.75	-0.001	0.004	0.964	0.988
q0.9	-0.004	0.004	0.928	0.984
LRT version				
q0.1	0.019	0.116	0.950	0.997
q0.25	0.010	0.080	0.934	0.985
q0.5	0.004	0.066	0.924	0.980
q0.75	-0.001	0.059	0.921	0.975
q0.9	-0.004	0.063	0.886	0.963

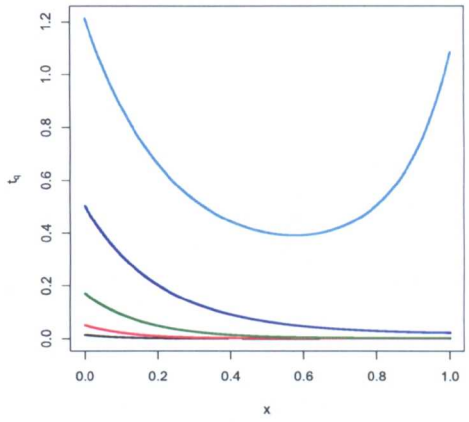

Figure 5.3: 10%, 25%, 50%, 75%, and 90% quantile curves underlying D54.

Table 5.13: Mean relative bias (mrbias) and root mean relative squared error (mrse) of quantiles and proportions of the real quantiles inside the confidence bands for D54 when $n = 500$.

Data54				
	mrbias	mrse	CI95prop	CI99prop
Known p Version				
q0.1	0.198	0.784	0.911	0.967
q0.25	0.100	0.457	0.932	0.975
q0.5	0.046	0.311	0.928	0.967
q0.75	0.013	0.264	0.925	0.964
q0.9	0.001	0.299	0.903	0.956
LRT version				
q0.1	0.199	0.784	0.911	0.967
q0.25	0.100	0.457	0.938	0.975
q0.5	0.046	0.311	0.932	0.967
q0.75	0.014	0.265	0.925	0.964
q0.9	0.002	0.304	0.910	0.957

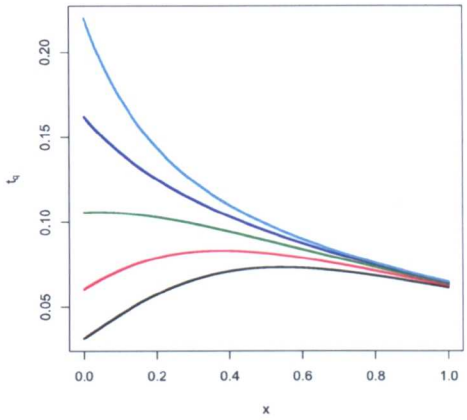

Figure 5.4: 10%, 25%, 50%, 75%, and 90% quantile curves underlying D62.

Table 5.14: Mean relative bias (mrbias) and root mean relative squared error (mrse) of quantiles and proportions of the real quantiles inside the confidence bands D62 when $n = 500$.

Data62				
	mrbias	mrse	CI95prop	CI99prop
Known p Version				
q0.1	0.010	0.003	0.977	0.991
q0.25	0.005	0.001	0.983	0.992
q0.5	0.002	0.000	0.987	0.993
q0.75	0.000	0.000	0.989	0.999
q0.9	0.000	0.000	0.992	1.000
LRT version				
q0.1	0.005	0.057	0.919	0.969
q0.25	0.002	0.031	0.958	0.987
q0.5	0.000	0.016	0.930	0.982
q0.75	-0.001	0.013	0.905	0.954
q0.9	-0.002	0.017	0.927	0.972

Overall, we deduce that the difference between the estimated and true GG models (which is reflected through the difference between the estimated and true quantile curves) is quite small. This is observed from the relatively small "mrbias" and "mrse" of Tables 5.11, 5.12, 5.13 and 5.14 in both the "Known p Version" and "LRT Version". We also note the possibility that any positive and negative biases in some parameter estimates (mentioned in the previous subsection) might cancel when combined in formulae to make quantiles. This might have been a factor that allowed the relative biases and squared errors of the quantiles to be relatively small. Also, we observe that the true quantiles lie within the CIs around the estimated quantiles most of the time with about the right percentage. The (practical) "LRT Version" has performance almost and very acceptably as good as the (practically unavailable) "Known p Version". These are very encouraging results for the practical application of the methodology.

5.2 Computational Evidence Via Simulations of the Suggested Goodness-of-Fit Test for the GG Model

In this study, the aim is to show, via simulations, that the test statistic (and its approximate distribution) we suggested for our goodness-of-fit test in Section 4.6 is a suitable one. Knowing that the 10%, 25%, 50%, 75%, and 90% quantile curves divide data into six regions, we recall that this test statistic is

$$\tau \equiv \sum_{l=1}^{6} \frac{(\widehat{m_l} - m_l)^2}{m_l},$$

where $\widehat{m_l}$ and m_l, $l = 1, ..., 6$, are respectively the observed and expected number of data points in each of the six regions.

For each of the 100 data sets in each of D31, D32, D33, ..., D64, we compute τ resulting in S31, S32, S33, ..., S64 respectively (each of size 100 consisting of 100 summations, of course). To validate our claim in Section 4.6 that $\tau \sim G(\theta = 1.5, k = 2)$, we fit by ML a gamma G$(\theta, k)$

distribution to each of S31, S32, S33, ..., S64. We plot a default **R** histogram for each one of those and using the bins from the histograms, we also perform a χ^2 goodness-of-fit test to compare them with a $G(\theta = 1.5, k = 2)$ and we obtain the corresponding p-values which we denote by "Gamp-values". The ML estimates $\hat{\theta}$ and $\hat{k}$ for θ and k respectively, for each of S31, S32, S33, ..., S64, along with the "Gamp-values" are given in Table 5.15.

It is very obvious that the $\hat{\theta}$ and $\hat{k}$ values are respectively approximately 1.5 and 2 for each of S31, S32, S33, ..., S64. These results are remarkably consistent throughout the table. This was the initial fact that lead us to think that the $G(\theta = 1.5, k = 2)$ distribution is a good approximation to the sampling distribution for τ.

To support this fact and to get a better idea of how well the $G(\theta = 1.5, k = 2)$ fits, we also plot the histograms of S32, S42, S54 and S62 along with the $G(\theta = 1.5, k = 2)$ pdfs in Figures 5.5, 5.6, 5.7 and 5.8 respectively. We observe that although the size of each of S31, S32, S33, ..., S64 is not very large ($n = 100$), the $G(\theta = 1.5, k = 2)$ pdfs still fit the histograms reasonably well. Looking back at the "Gamp-values" of S31, S32, S33, ..., S64 in Table 5.15, we notice that for a 0.05 level of significance, 10 out of 16 (62.5%) are well-estimated by a $G(\theta = 1.5, k = 2)$. For a 0.01 level of significance 81.25% are well-estimated. We failed to make the test work for S62 as the number of bins (in the corresponding histogram) was very small. The probabilities are good enough to support our initial guess that the $G(\theta = 1.5, k = 2)$ is a good fit.

For S31, S32, S33, ..., S64, the consistency of the values of the ML estimates (approximately 1.5 for $\hat{\theta}$ and 2 for $\hat{k}$), the relatively high "Gamp-values" and the observable fact that $G(\theta = 1.5, k = 2)$ fits well the histograms are three factors that lead us to conclude that $G(\theta = 1.5, k = 2)$ is a good approximation to the sampling distribution for τ. Note also that the approximate sampling distribution is independent of p.

Table 5.15: ML estimates of the summations along with the goodness-of-fit test p-values.

	$\hat{\theta}$	$\hat{k}$	Gamp-values
S31	1.439	2.006	0.631
S32	1.404	2.049	0.077
S33	1.579	2.165	0.033
S34	1.728	1.722	0.004
S41	1.609	1.772	0.292
S42	1.534	2.004	0.689
S43	1.419	1.900	0.253
S44	1.374	2.287	0.578
S51	1.837	1.694	0.003
S52	1.579	1.935	0.447
S53	1.825	1.866	0.014
S54	1.516	1.940	0.156
S61	1.468	2.093	0.644
S62	2.218	1.551	-
S63	1.710	1.868	0.028
S64	1.867	1.500	0.108

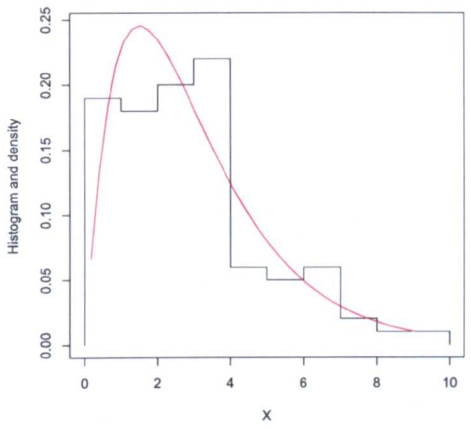

Figure 5.5: Histogram of S32 & G($\theta = 1.5, k =$ 2) pdf.

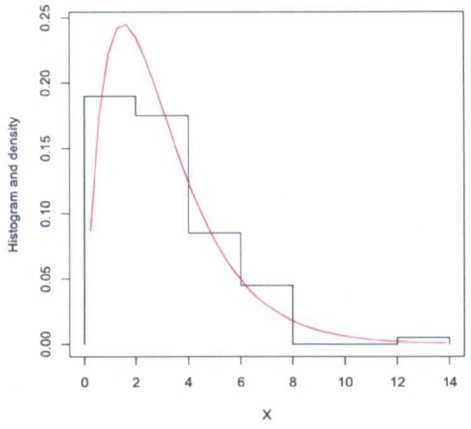

Figure 5.6: Histogram of S42 & G($\theta = 1.5, k =$ 2) pdf.

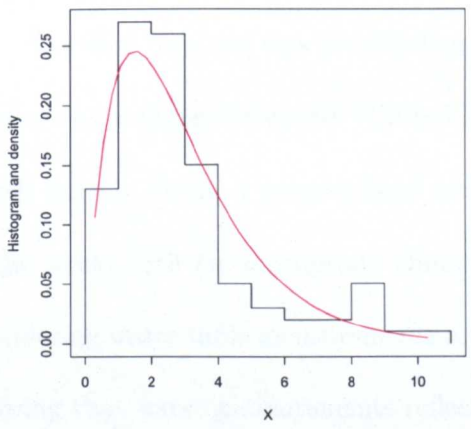

Figure 5.7: Histogram of S54 & G($\theta = 1.5, k =$ 2) pdf.

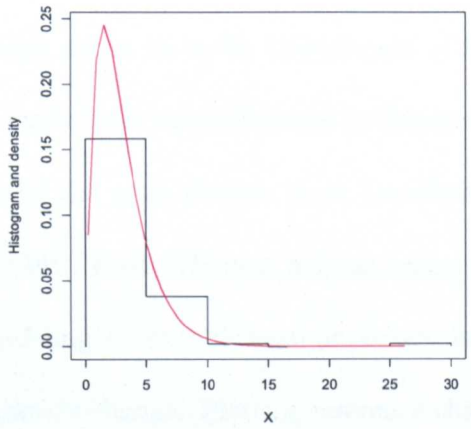

Figure 5.8: Histogram of S62 & G($\theta = 1.5, k =$ 2) pdf.

Chapter 6

Application to Real Data

In this chapter, we apply the theoretical work underlying this thesis, that was validated via simulations, to real-life data. We apply all five stages of the modeling package we introduced in Chapter 1 to two data sets:

1. Water Table Depth vs Flux
2. Weight vs Height

The first data set was provided to us by Dr. Yoseph Araya from the Department of Life Science at the Open University (United Kingdom) who made these measurements in Capetown, South Africa. Being a geographical area that has one of the most diverse plant populations in the world with an appropriate climate for rainfall study, South Africa is a great target for considering water table measurements and research. Modeling water table is an important issue knowing that water measurements reflect the world's climate change. Plotting reference charts for such data allows scientists to class geographical areas as having water-rich soil, dry soil, floods etc., which has a direct effect on the plant population.

The second data set was obtained from The National Health and Nutrition Examination Survey (NHANES) in 2007-2008 which is a program of studies designed to assess the health and nutritional status of adults and children in the United States. For more information about

NHANES and how to obtain the data, refer to NHANES (2007). The survey is unique in that it combines interviews and physical examinations. From this survey, we model the weight vs the height of males. The issue of individuals' fitness is of great interest to society nowadays, where there is a great need to set limiting thresholds that specify when individuals are considered as lying in the "common range" or not. The percentiles of weight for specified height is of particular interest in public health where upper percentiles are referenced for overweight or obesity and the lower percentiles for underweight.

The two mentioned data sets, the former being environment-related and the latter health-related, are studied in the light of QR in Sections 6.1 and 6.2 respectively.

6.1 Water Table Depth vs Flux

Flux is the measure of rainfall that goes into the ground. Knowing that some of this water evaporates as it touches the ground, flux can take negative values. As water infiltrates through pore spaces in the soil, it first passes through a zone where the soil is unsaturated. At increasing depths water fills in more spaces, until the zone of saturation is reached. The relatively horizontal plane atop this zone constitutes the water table. Water regimes dictate different vegetation types. The Cape Floristic Region (Capetown, South Africa) has about 9000 plant species most of which are unique to this area. According to the changing climate scenarios, the annual rainfall in the Cape is likely to decrease. A better understanding of the water distribution will enhance the way this biodiversity spot is managed. Researchers from the Open University, working in collaboration with a South African team and other environmental institutions, established around ten sites in the Cape to monitor the hydrology of the area.

For our QR study, we obtain one of the "water table depth (WTD) vs flux" data sets collected at 307 different times by Open University researchers in the Cape and study it by

plotting its reference charts. A scatter plot of the data set is given in Figure 6.1. Note that negative values of flux arise when the ground is dried by the sun's action. The data points are quite condensed near zero, therefore we aim to take the logarithm of the Flux. To avoid taking the log of negative numbers and zero, we shift the flux to the positive real line. As the minimum value for flux is -6.8 in this case, we consider log(flux + 6.9). We then rescale the resulting values to the interval (0,1) and we fit a GG distribution to the obtained data. A scatter plot of the "rescaled version" is given in Figure 6.2. It appears that WTD decreases as Flux increases.

LRTs show that $L_6 = -228.7693$ and $L_5 = -240.717$, hence $D_1 = 23.894$. The significant difference between L_6 and L_5 means that a six-parameter GG (as given in item 3 of Section 3.3) is required. The values of the estimated parameters are $\hat{a} = 5.278$, $\hat{b} = -4.013$, $\hat{c} = -2.424$, $\hat{d} = 3.106$, $\hat{f} = -3.886$ and $\hat{g} = 11.159$. We recall that the expression of the quantile function corresponding to this model, given in item 3 of Section 3.4 is

$$t_q = \exp(a + bx)\left(\frac{r(f, g, q)}{e^{f+gx}}\right)^{\exp(c+dx+\frac{1}{2}(f+gx))}, \qquad x > 0,$$

which is estimated by

$$t_q = \exp(5.278 - 4.013x)\left(\frac{r(-3.886, 11.159, q)}{e^{-3.886+11.159x}}\right)^{\exp(-2.424+3.106x+\frac{1}{2}(-3.886+11.159x))}, \qquad x > 0.$$

Figure 6.3 presents a plot of the 10%, 25%, 50%, 75% and 90% estimated quantiles. In the interval $(0, 1)$, the quantiles show interesting shapes. The 90% quantile is obviously unimodal with minimum in (0.6,1). It increases towards 1 to account for the data points that have higher values although the degree of upturn might be being unduly influenced by a single point. The other quantiles are monotone decreasing. Also, interestingly, the 10% quantile decreases very slowly (almost horizontally) between 0 and 0.1 in contrast with the other quantiles which display a sharp decrease in this region. This might also be due to the influence of a single point, in this region, on the 10% quantile. The confidence bands around the 10%, 25%, 50%,

75% and 90% estimated quantiles are shown in Figure 6.4, which seems to be a bit obscured by overplotting. For more clarity, we plot separately the confidence bands around the 10%, 50% and 90% estimated quantiles in Figure 6.5 and the confidence bands around the 25% and 75% estimated quantiles in Figure 6.6. We observe that near the center (where data is condensed), the confidence bands are quite narrow around the quantiles showing more certainty in the results. They are wide at the edges where there are very few data points allowing more variation and less certainty in the model. Finally, taking into consideration the 10% and 90% quantiles, a reference chart is given in Figure 6.7. A corresponding reference chart using the original covariate scale is given in Figure 6.8. A closer view of the latter figure is given in Figure 6.9. Data within the 10% and 90% quantiles are considered as lying in the common range. Data below the 10% quantile indicate over-saturation of water in the soil or equivalently excessive rainfall. Above the 90% quantile are data points taken at times where the WTD is quite low.

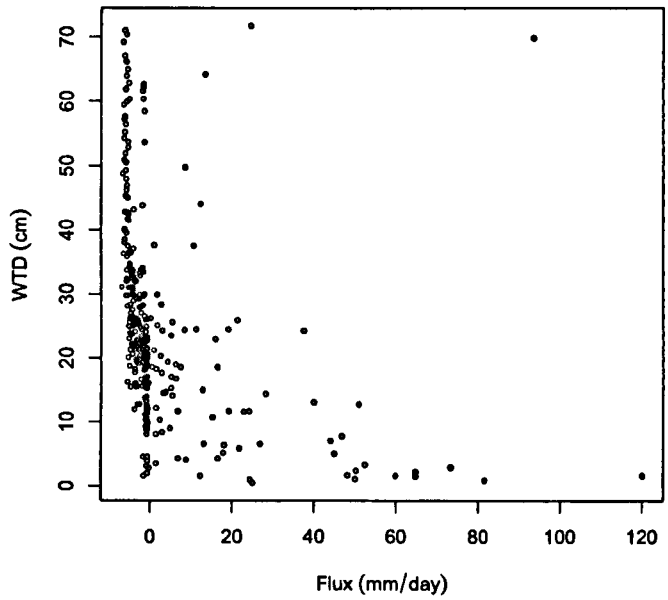

Figure 6.1: WTD vs Flux data of size 307 at various times in Capetown.

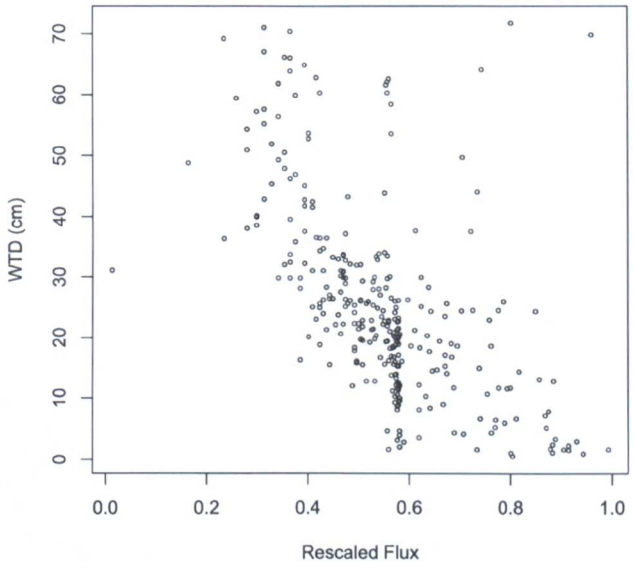

Figure 6.2: WTD vs rescaled logarithm of Flux in Capetown at 307 various times.

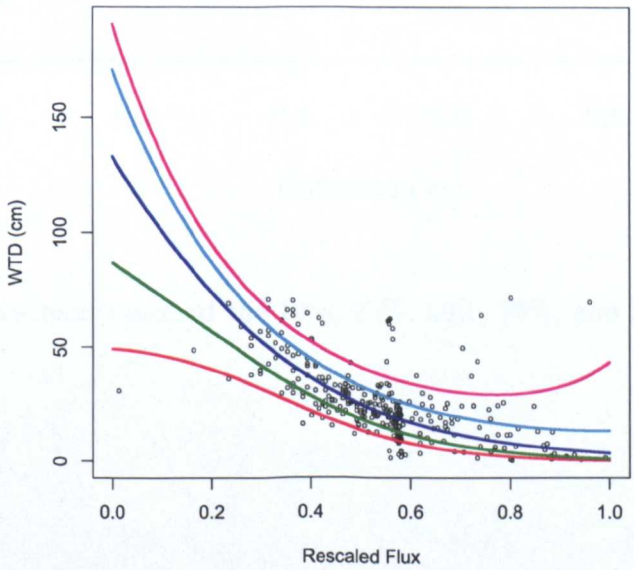

Figure 6.3: 10%, 25%, 50%, 75%, and 90% estimated quantiles of WTD vs rescaled logarithm of Flux data in Figure 6.2.

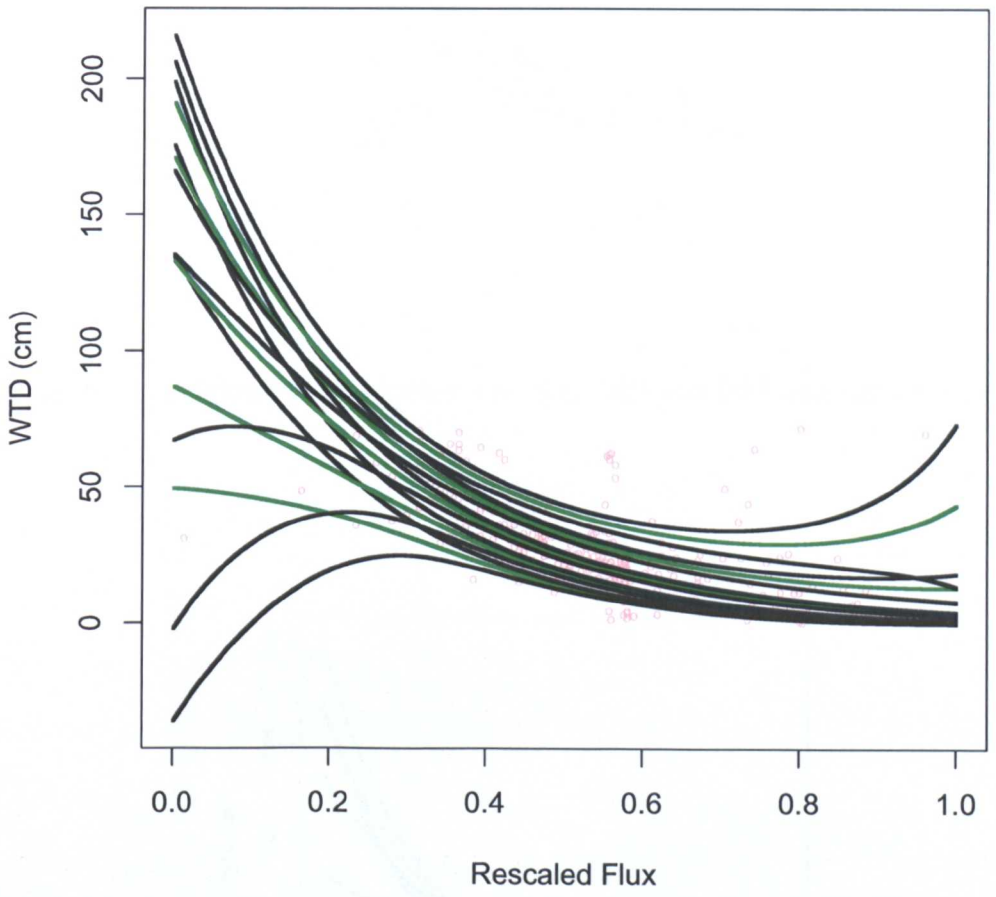

Figure 6.4: Confidence bands around the 10%, 25%, 50%, 75%, and 90% quantiles in Figure 6.3.

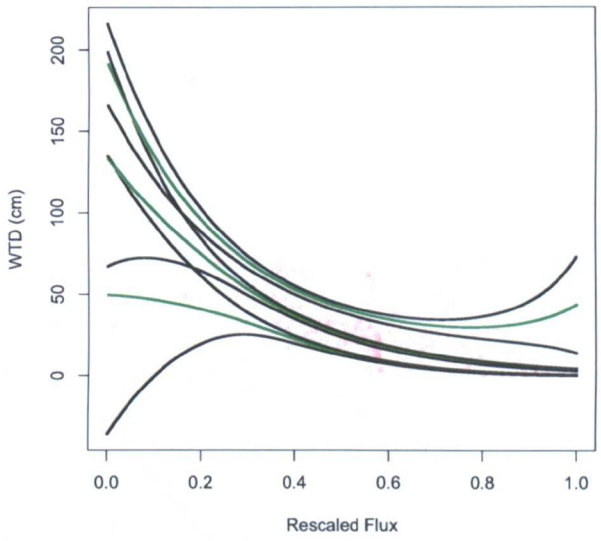

Figure 6.5: Confidence bands around the 10%, 50% and 90% quantiles in Figure 6.3.

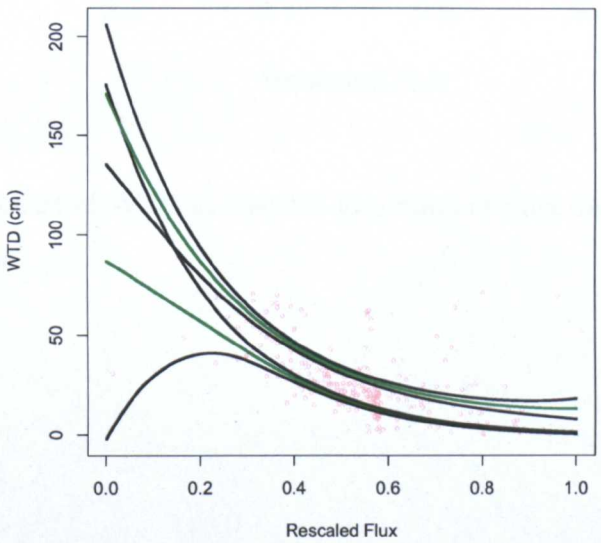

Figure 6.6: Confidence bands around the 25% and 75% quantiles in Figure 6.3.

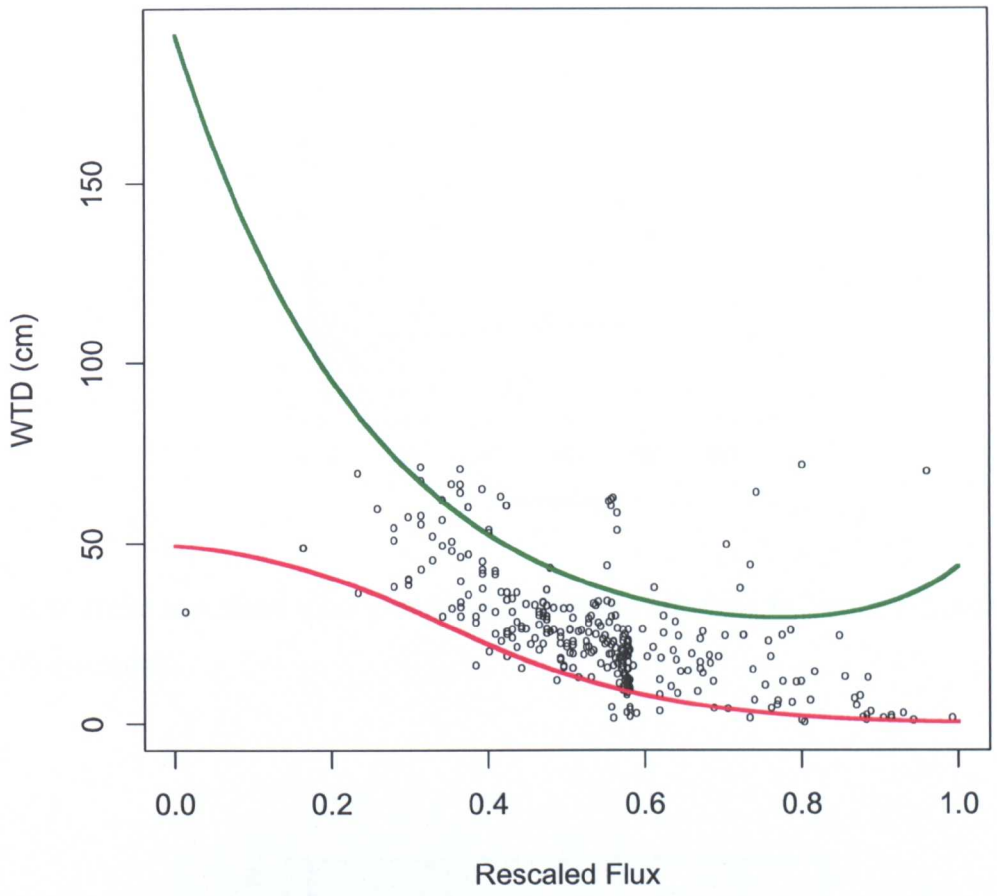

Figure 6.7: Reference chart of WTD vs rescaled logarithm of Flux data in Figure 6.2 using the 10% and 90% quantiles.

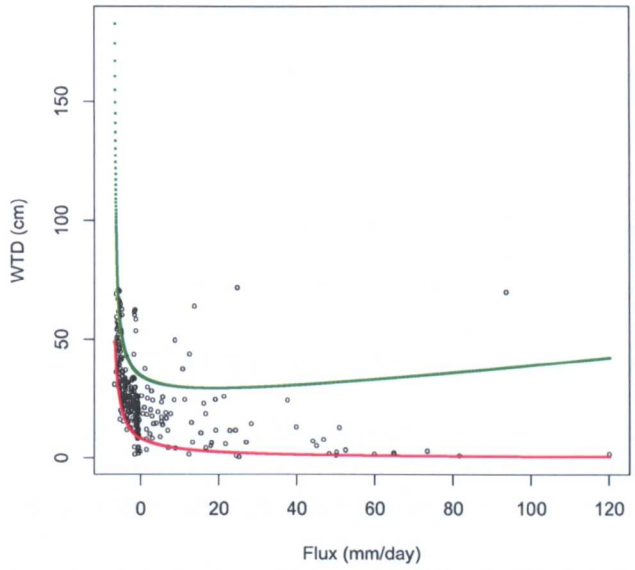

Figure 6.8: Reference chart (closer view) of WTD vs Flux data in Figure 6.1 using the 10% and 90% quantiles.

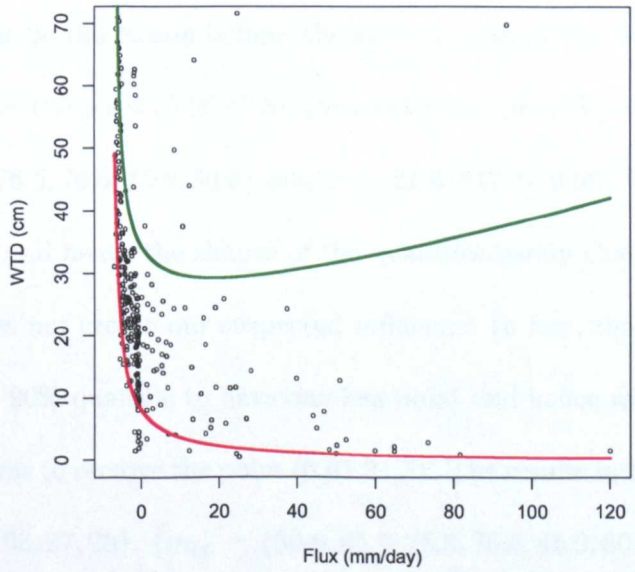

Figure 6.9: Reference chart of WTD vs Flux data in Figure 6.1 using the 10% and 90% quantiles.

We now move to the last step of our modeling procedure where we check the goodness-of-fit of our estimated model. As explained in Sections 4.6 and 5.2, we compute $\widehat{m_l}$ and m_l, $l = 1, ..., 6$, the observed and expected number of data points in each of the six regions formed by the 10%, 25%, 50%, 75%, and 90% estimated quantiles and we deduce the test statistic τ. It turns out that $\{\widehat{m_l}\} = \{21, 54, 75, 102, 26, 29\}$, $\{m_l\} = \{30.7, 46.05, 76.75, 76.75, 46.05, 30.7\}$ and $\tau = 21.60803 > 9.96$. Based on those results, we reject the fact the the GG model is a good fit in this case at 0.01 level of significance, although our observation of the data and its estimated quantiles shows that the latter fit quite well and take the general shape of the data. This reasoning leads us to think that the high value of τ is caused by the influence on the quantiles of single data points that are outliers. Therefore, we study the same problem again while removing this time the influence of some data points. We first remove the influence of the point of coordinates (0.96,69.8), with abscissa rounded to the nearest hundredth, which we suspect causes the upturn of the 90% quantile towards 1. In another attempt, we remove the influence of the point of coordinates (0.01,31.3), with abscissa rounded to the nearest hundredth, which we think might be the reason behind the slow decrease of the 10% quantile towards 0.

When we remove the point (0.96,69.8), the results become $\{\widehat{m_l}\} = \{21, 54, 75, 102, 26, 28\}$, $\{m_l\} = \{30.6, 45.9, 76.5, 76.5, 45.9, 30.6\}$ and $\tau = 21.81917 > 9.96$. We also realise that the parameter estimates and hence the shapes of the quantiles barely change. We deduce that the point (0.96,69.8) does not create our suspected influence. In fact, removing this point caused the region above the 90% quantile to have one less point and hence worse results.

The next step was to remove the point (0.01,31.3). The results in this case turned out to be $\{\widehat{m_l}\} = \{22, 51, 75, 103, 27, 28\}$, $\{m_l\} = \{30.6, 45.9, 76.5, 76.5, 45.9, 30.6\}$ and $\tau = 20.19608 > 9.96$. Plots of the data, the estimated quantiles, the confidence bands around them and the reference charts are given respectively in Figures 6.10, 6.11, 6.12 and 6.13. A close view of the

corresponding reference chart using the original covariate scale is given in Figure 6.14. Removing point (0.01,31.3) causes the shape of the 10% quantile to become similar (about parallel) to the other ones and to lose the horizontal behavior near zero. Looking at the confidence bands, we notice that they are still narrow in the center. Around the edges, they become far narrower especially around zero where we don't observe any more the wide deviation away from the quantiles. In fact, they take a very similar shape to that of the quantiles. This fact supports our initial guess that the general shape of the data is decreasing from relatively high values (around 130) at 0 to smaller ones (around 20) near 1. Our goodness-of-fit test still rejects the fact that the GG model is a good fit where the resulting value of τ is only slightly better.

Finally, removing both points yields very similar results to the previous case. We conclude that the general decrease of WTD with the increase in Flux is well-reflected by the decreasing quantiles, a fact that becomes clearer when the influence of point (0.01,31.3) is removed. Despite that, the goodness-of-fit test still rejects the GG model as a good fit. Looking closely at the values of $\widehat{m_l}$ and m_l in each of the mentioned scenarios, we realise that the biggest difference between the observed and expected results happens between the fourth and fifth regions ($l = 4$ and $l = 5$ respectively). When (0.01,31.3) is removed, $\widehat{m_4} = 103$ which is considerably bigger than $m_4 = 76.5$. Also, $\widehat{m_5} = 27$ which is quite smaller than $m_5 = 45.9$. Removing this point changed the behaviour of the quantiles around zero and shifted downwards the 75% quantile only very slightly, though not to the extent that allowed the observed and estimated count of points in this area to match more. We observe many points in the upper part of the fourth region very close to the 75% quantile. Although our goodness-of-fit test rejects the GG model as a good fit, it shows through the count of data points in each region that the observed and expected counts match quite well most of the time except around the fourth and fifth regions where shifting downwards the 75% quantile only slightly will enable far much better results.

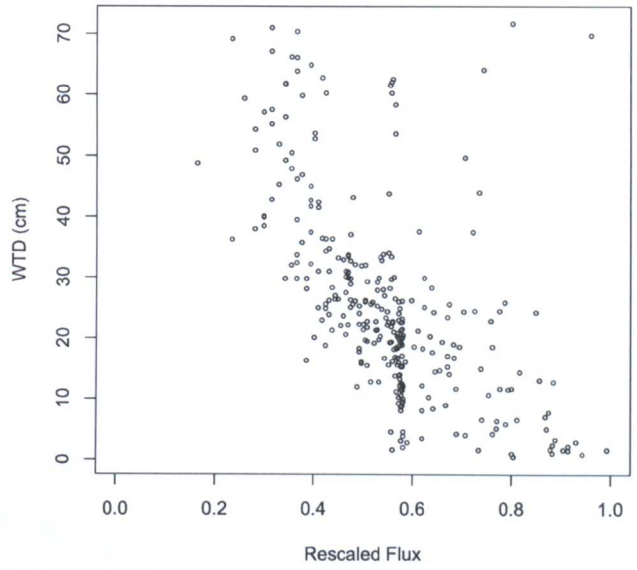

Figure 6.10: WTD vs rescaled logarithm of Flux in Capetown at 306 various times where point of coordinates (0.01,31.3) is removed from the original data set.

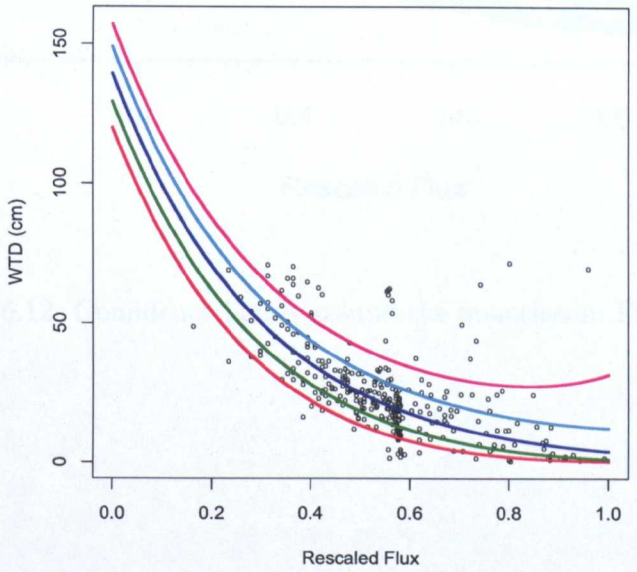

Figure 6.11: 10%, 25%, 50%, 75%, and 90% estimated quantiles of WTD vs rescaled logarithm of Flux data in Figure 6.10.

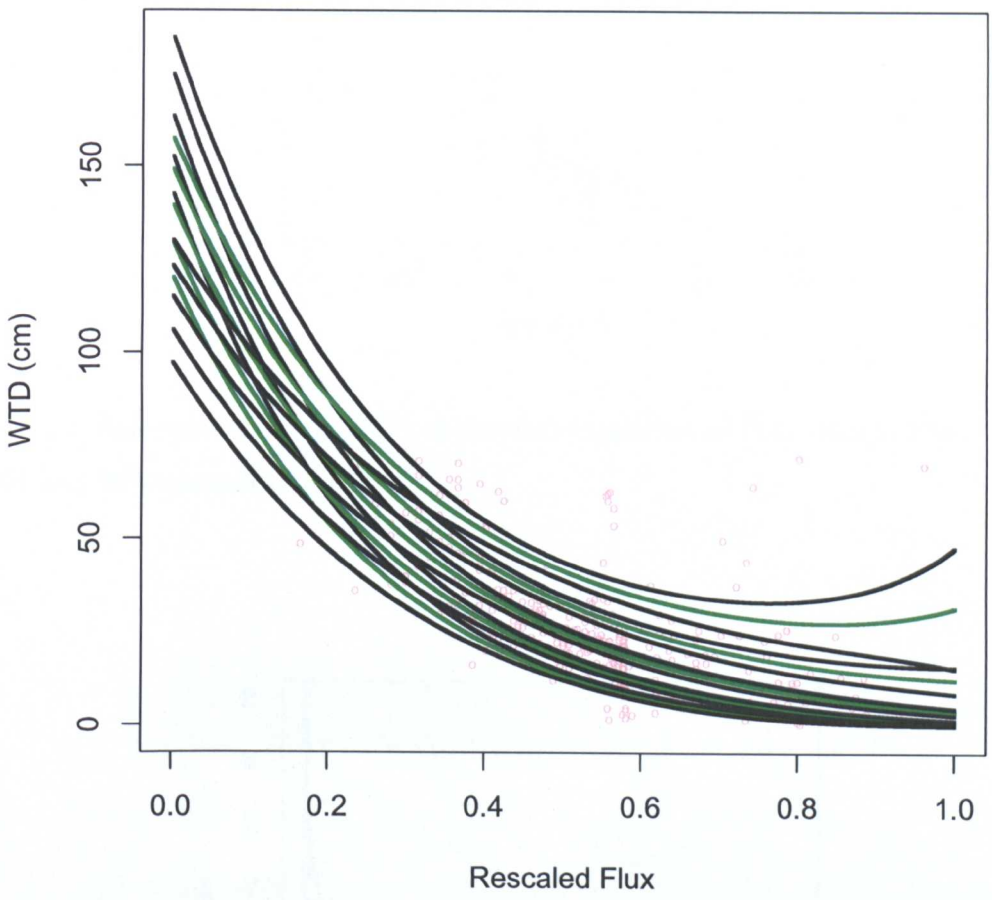

Figure 6.12: Confidence bands around the quantiles in Figure 6.11.

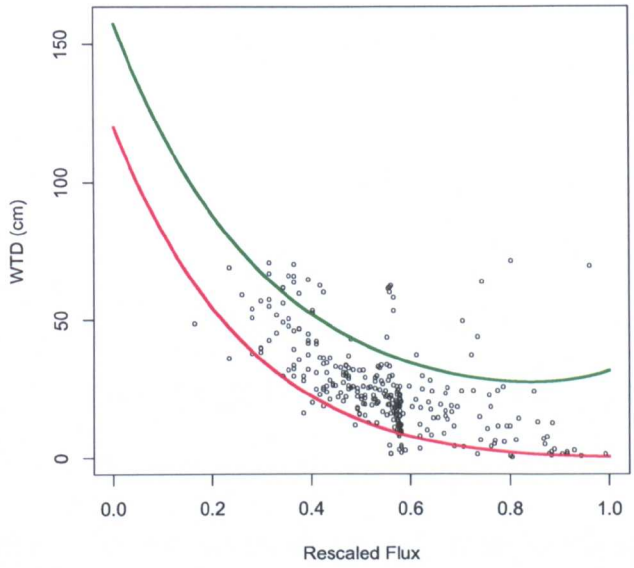

Figure 6.13: Reference chart of WTD vs rescaled logarithm of Flux data in Figure 6.10 using the 10% and 90% quantiles.

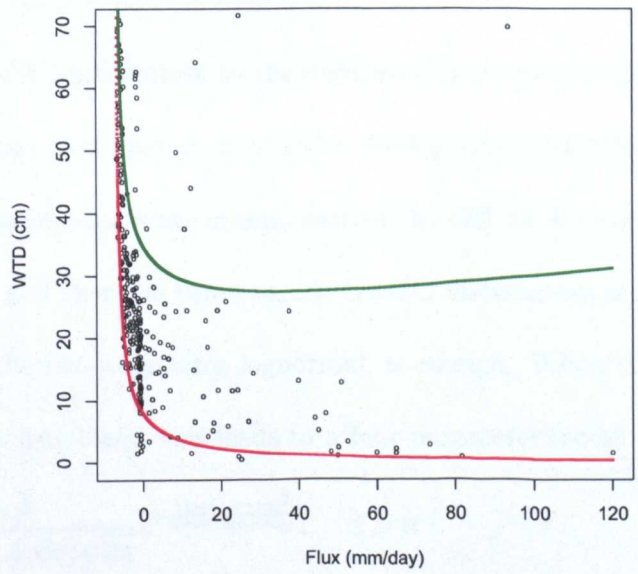

Figure 6.14: Reference chart of WTD vs Flux data (where point of coordinates (-6.8,31.1) is removed from the original data set) using the 10% and 90% quantiles.

6.2 Weight vs Height

Weight problems are becoming more and more common so that some of the world's leaders are considering this issue a threat to societies and are therefore seeking solutions. Since 1999, the National Center for Health Statistics in the United States has conducted the NHANES survey annually. The latest survey released is the NHANES 2007-2008. Each release includes "demographic", "examination" and "laboratory" measurements. From the "examination" data file we considered two variables BMXWT and BMXHT reflecting respectively weights (in kg) and heights (in cm) of individuals across all age ranges. After filtering out the individuals that have missing gender, weight and height as well as all the females, we ended-up with a total of 4448 males with known weight and height. A scatter plot of the data set is given in Figure 6.15. As expected, generally, weight is increasing with height. For smaller heights, data is quite narrow and it widens as height increases. This is because shorter individuals have less weight variability. After rescaling the height to the interval (0,1) we present another scatter plot of the new "rescaled version" in Figure 6.16.

As we fitted a GG distribution to the data, results showed that the estimated k is taking very high values, a fact that lead us to consider the lognormal distribution as an alternative fit to the GG. As the lognormal is the special case of the GG for $k \to \infty$, we concluded that it is a convenient model and that the three-parameter GG distribution is not required in this case. A simpler version, the two-parameter lognormal, is enough. When the parameters depend on the covariates in the usual way, this leads to a four-parameter model with density

$$f(t) = \frac{1}{t \exp(c+dx)\sqrt{2\pi}} e^{-\frac{(\log t - a - bx)^2}{2\exp\{2(c+dx)\}}}; \quad t > 0.$$

The LRT showed that $L_4 = 870.757$ and $L_3 = 703.184$, hence $D_1 = 335.146$. This significant difference means that a four-parameter lognormal is required. The values of the estimated parameters are $\hat{a} = 2.378$, $\hat{b} = 2.685$, $\hat{c} = -2.325$, and $\hat{d} = 1.101$. The lognormal quantile

function is given by the exponential of the quantile function of a normal distribution

$$t_q = \exp\{2.378 + 2.685x + e^{-2.325+1.101x}\Phi^{-1}(q)\}$$

where $\Phi^{-1}(q)$ is the quantile function of the standard normal distribution. Figure 6.17 presents a plot of the 10%, 25%, 50%, 75% and 90% estimated quantiles.

As expected, the quantiles are increasing with the increase in height (and weight). Also, they spread out from each other as height increases reflecting the change in the data quite well. The confidence bands around the estimated quantiles, shown in Figure 6.18, are very narrow. This is because the data set is quite large and covering all height ranges. The large number of data points allows more certainty in the model (i.e. less variation) and hence narrower CIs. Taking into consideration the 10% and 90% quantiles, an example of how a reference chart might appear is given in Figure 6.19. To obtain thresholds for "normal", "obese" or "underweight" individuals, it is sometimes advised to use the 5% and 95% quantiles as reference. A reference chart with the 5% and 95% quantiles is given in Figure 6.20. Note that no further complicated calculations are needed to obtain Figure 6.20. It is a simple matter of replacing q by the new values 5% and 95%. A corresponding reference chart using the original covariate scale is given in Figure 6.21.

Finally, we check the goodness-of-fit of the GG model. The observed and expected counts are $\{\widehat{m_l}\} = \{427, 763, 1086, 1118, 606, 448\}$ and $\{m_l\} = \{444.8, 667.2, 1112, 1112, 667.2, 444.8\}$ respectively, while $\tau = 20.74475 > 9.96$. Based on the results, we reject the fact that the GG model is a good fit to our data, though, if we look at $\widehat{m_l}$ and m_l, we find that they follow a similar pattern. Also, the general shape of the quantiles reflect the exponential increase in the data and they widen as the normalised height increases towards one, as the data widens.

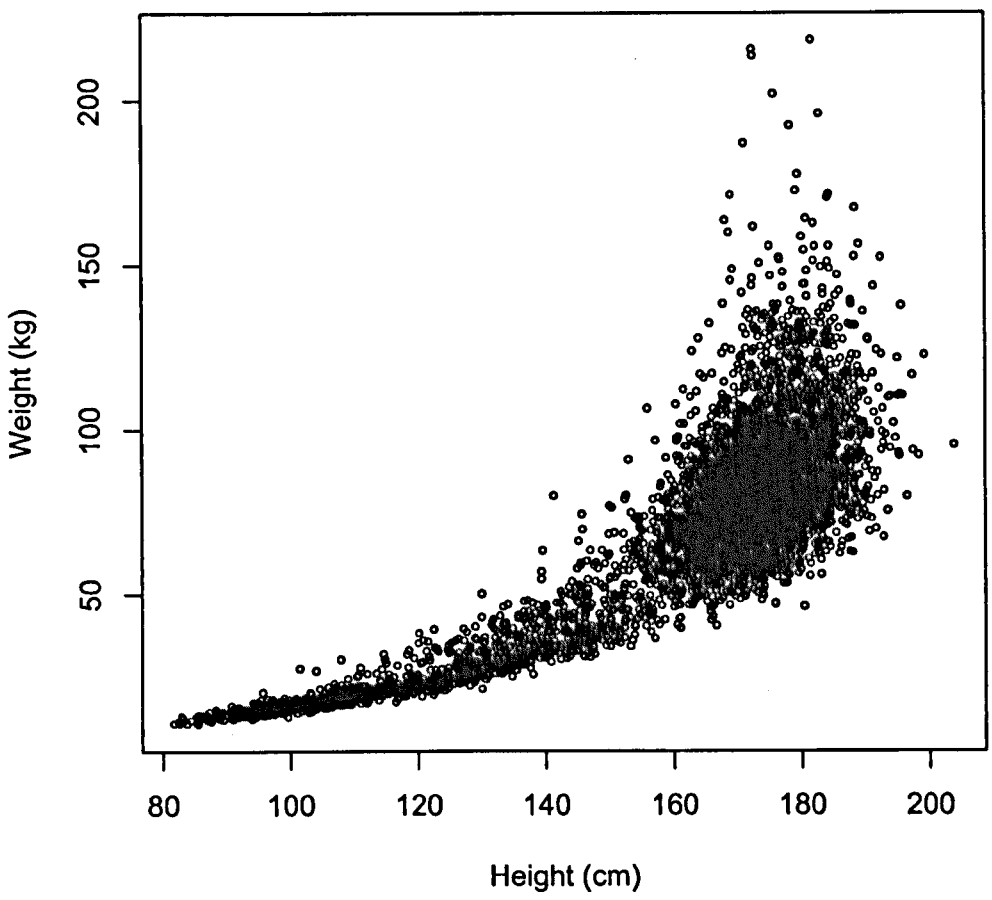

Figure 6.15: NHANES 2007-2008 weight vs height data of 4448 males.

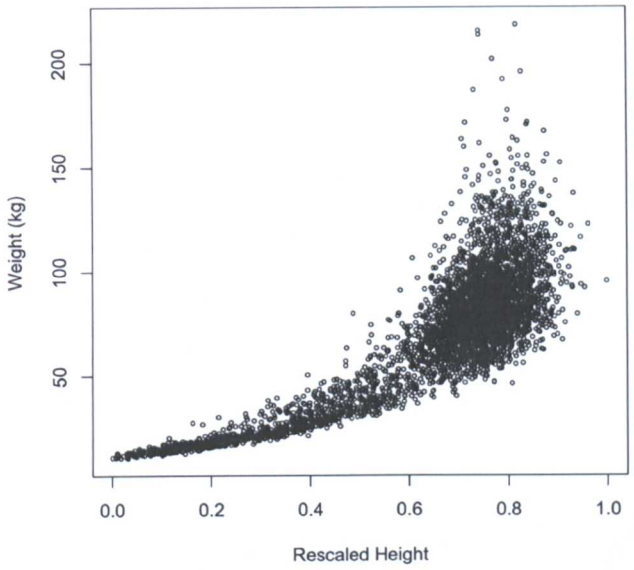

Figure 6.16: NHANES 2007-2008 weight vs rescaled height data of 4448 males.

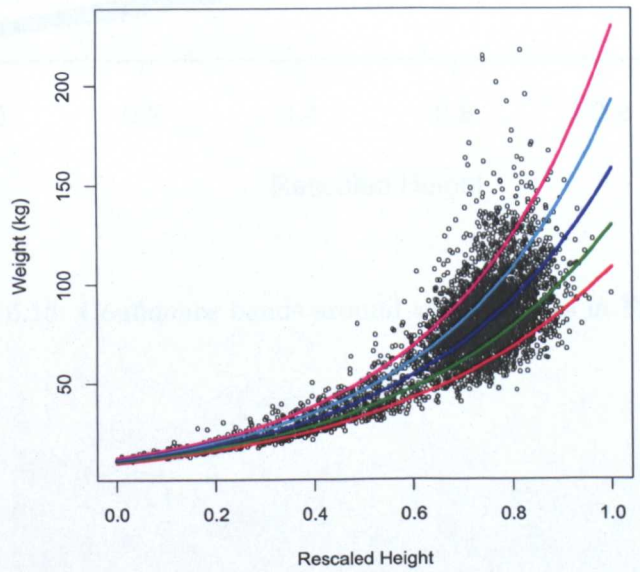

Figure 6.17: 10%, 25%, 50%, 75%, and 90% of weight vs rescaled height data in Figure 6.16.

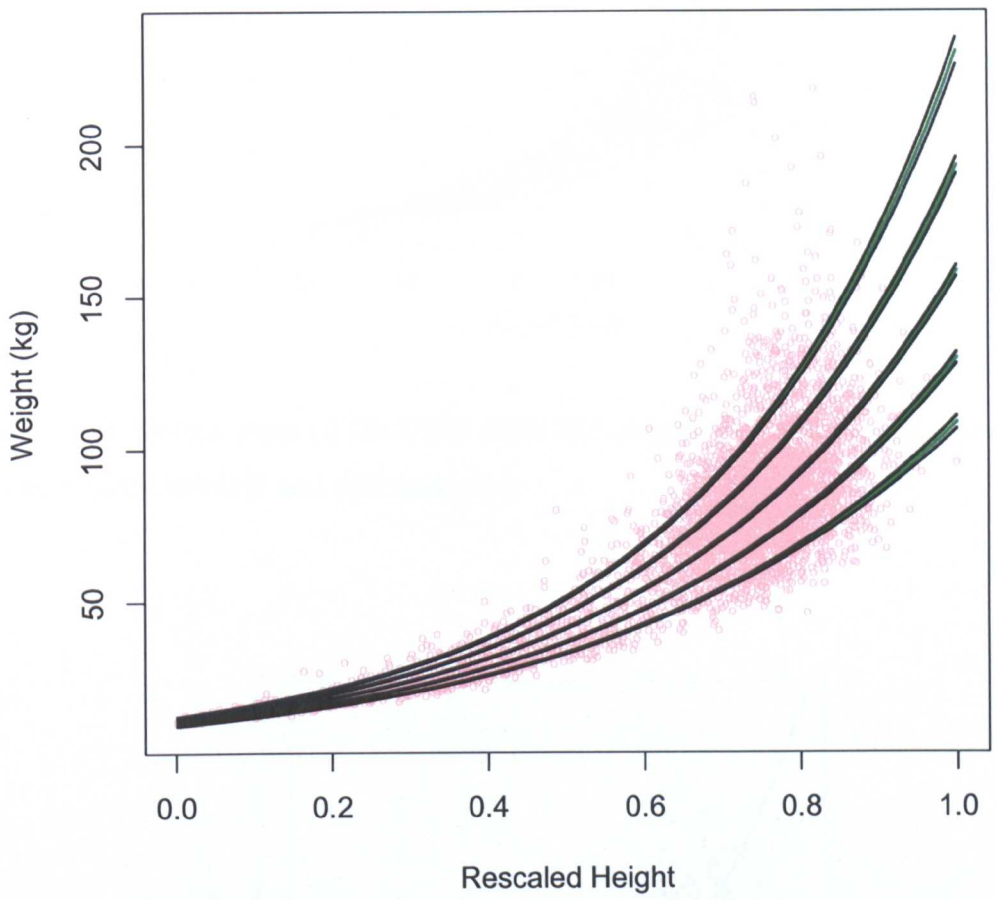

Figure 6.18: Confidence bands around the quantiles in Figure 6.17.

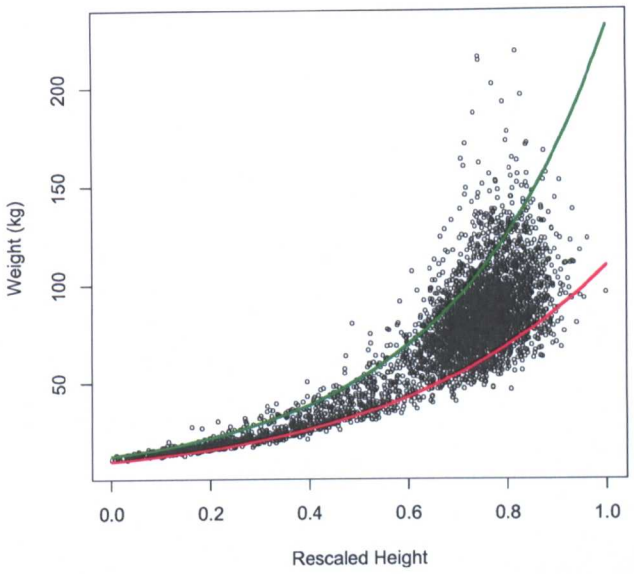

Figure 6.19: Reference chart of NHANES 2007-2008 weight vs rescaled height males' data in Figure 6.16 using the 10% and 90% quantiles.

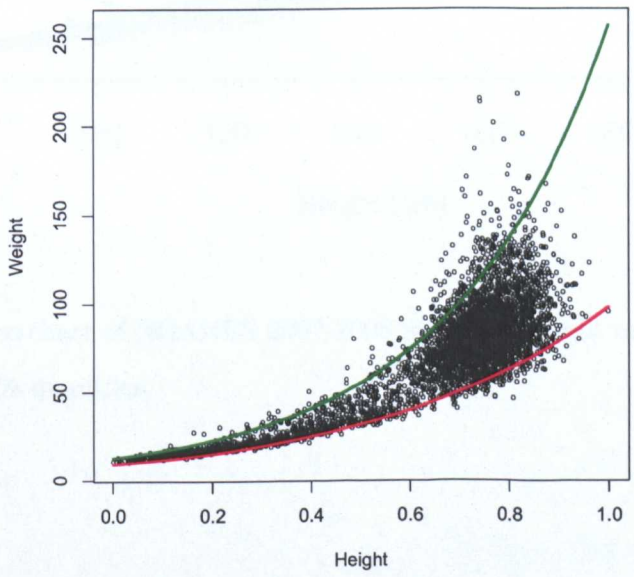

Figure 6.20: Reference chart of NHANES 2007-2008 weight vs rescaled height males' data in Figure 6.16 using the 5% and 95% quantiles.

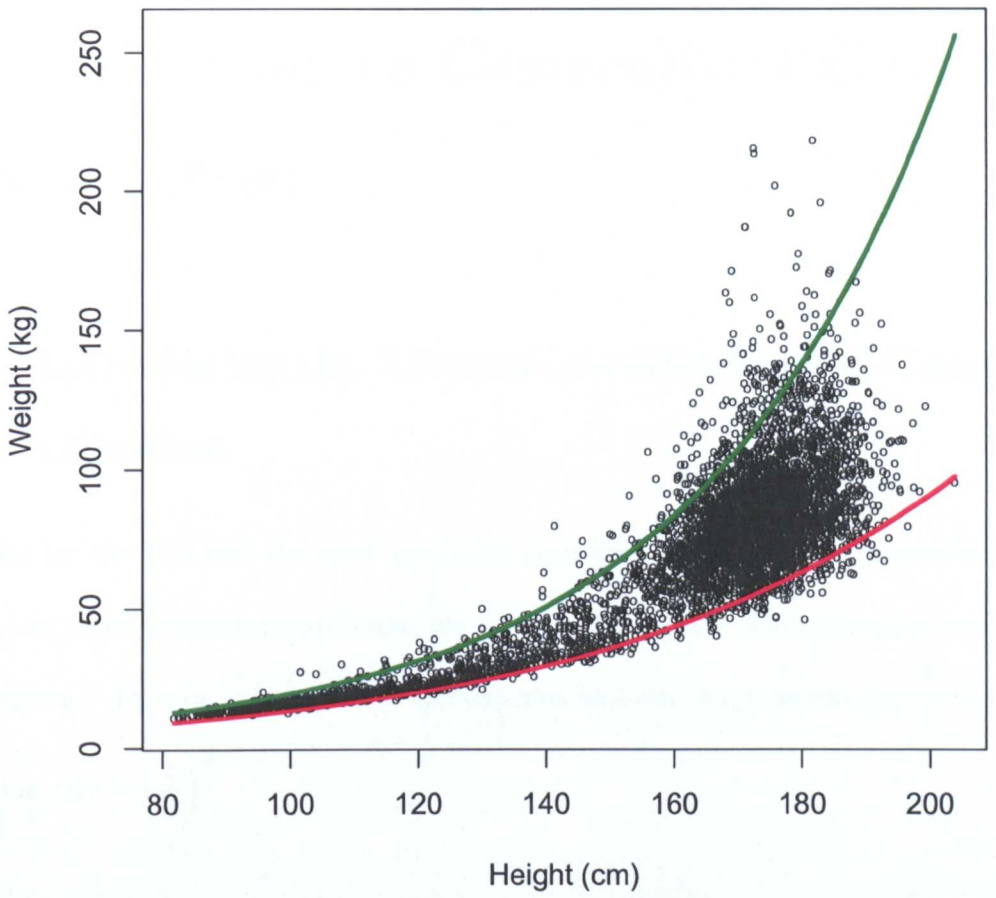

Figure 6.21: Reference chart of NHANES 2007-2008 weight vs height males' data in Figure 6.15 using the 5% and 95% quantiles.

Chapter 7

The Alternative Generalised Gamma Distribution

7.1 Introducing the Alternative Generalised Gamma Distribution

Inspired by the GG and the need to model skewness, we propose an alternative unimodal univariate three-parameter continuous life distribution. While behaviour near zero of the GG pdf Version 1 depends only on α, the tail depends on both shape parameters α and β:

$$\lim_{t\to 0} f(t) \sim \left(\frac{t}{\theta}\right)^{\alpha-1}$$

and

$$\lim_{t\to \infty} f(t) \sim \left(\frac{t}{\theta}\right)^{\alpha-1} e^{-\left(\frac{t}{\theta}\right)^{\beta}} \to 0,$$

where generically '$\sim$' means that if $\lim_{t\to a} f(t) \sim g(t)$, then $\lim_{t\to a} f(t)/g(t) = K$, $0 < K < \infty$.

Analogously, the basic density function of the alternative generalised gamma (AGG) distribution consists of two shape parameters; one controls the part next to zero and the other controls the tail. Apart from the normalising constant, the product $(t/\theta)^{\alpha-1}\exp\{-(t/\theta)^{\beta}\}$ is the main expression affecting the shape of the GG pdf Version 1. To construct the AGG distribution, and building upon the GG, we multiply this expression by the factor $1/(1+t/\theta)^{\alpha-1}$.

This ensures that behavior near zero of the AGG pdf depends only on α while the tail depends only on β. Let

$$f(t) = \frac{K_{\alpha,\beta}}{\theta} \frac{\left(\frac{t}{\theta}\right)^{\alpha-1}}{\left(1+\frac{t}{\theta}\right)^{\alpha-1}} e^{-\left(\frac{t}{\theta}\right)^{\beta}}, \qquad t > 0$$

be the basic form of the AGG pdf, where

$$K_{\alpha,\beta} = \frac{1}{\int_0^{\infty} \frac{t^{\alpha-1}}{(1+t)^{\alpha-1}} e^{-t^{\beta}} dt}.$$

Obviously,

$$\lim_{t\to 0} f(t) \sim \frac{\left(\frac{t}{\theta}\right)^{\alpha-1}}{\left(1+\frac{t}{\theta}\right)^{\alpha-1}} \sim \left(\frac{t}{\theta}\right)^{\alpha-1}$$

and

$$\lim_{t\to\infty} f(t) \sim e^{-\left(\frac{t}{\theta}\right)^{\beta}} \to 0.$$

The behavior of the AGG distribution is the same as that of the GG (and hence gamma and Weibull) distribution as $t \to 0$. However, as $t \to \infty$, the dependence on α that afflicts the GG distribution is removed, allowing the upper tail dependence of the AGG distribution to be controlled by β only.

To draw a little more comparison between the GG and the AGG densities, we look at the difference between t and $t/(1+t)$. Of course, $t/(1+t) < t$ for all $t > 0$ and they both converge to zero at zero. However, at ∞, t converges to ∞ while $t/(1+t)$ converges to 1. We notice an interesting case when we plot the functions $c_1(t) = t/2$ and $c_2(t) = t/(1+t)$ in Figure 7.1. Noticeably, $c_1(t)$ and $c_2(t)$ behave similarly for $t < 1$.

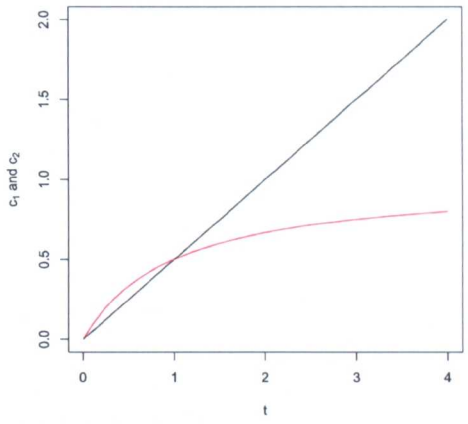

Figure 7.1: Plots of c_1 and c_2.

7.2 Some Basic Properties of the AGG Distribution

Set

$$\zeta(n,d,p) = p\int_0^\infty \frac{z^{n-1}}{(1+z)^{d-1}} e^{-z^p} dz.$$

In contrast with the GG, the different pdf versions of the AGG turn out to be

- Version 1, $AGG(\theta,\alpha,\beta)$:

$$f(t) = \frac{\beta}{\theta^\alpha \zeta(\alpha,\alpha,\beta)} \left(\frac{t}{1+\frac{t}{\theta}}\right)^{\alpha-1} e^{-\left(\frac{t}{\theta}\right)^\beta}, \qquad t>0,$$

where θ is a scale parameter and α and β are shape parameters each of which we take to be positive and

$$\begin{aligned}\zeta(\alpha,\alpha,\beta) &= \beta\int_0^\infty \left(\frac{z}{1+z}\right)^{\alpha-1} e^{-z^\beta} dz \\ &= \int_0^\infty \frac{z^{\alpha/\beta-1}}{(1+z^{1/\beta})^{\alpha-1}} e^{-z} dz.\end{aligned}$$

Setting $\alpha = k\beta$, we get:

- Version 2, $AGG(\theta, k, \beta)$:

$$f(t) = \frac{\beta}{\theta^{k\beta}\zeta(k\beta, k\beta, \beta)} \left(\frac{t}{1+\frac{t}{\theta}}\right)^{k\beta-1} e^{-\left(\frac{t}{\theta}\right)^{\beta}}, \qquad t > 0,$$

where θ is a scale parameter and k and β shape parameters all being positive and

$$\begin{aligned}\zeta(k\beta, k\beta, \beta) &= \beta \int_0^{\infty} \left(\frac{z}{1+z}\right)^{k\beta-1} e^{-z^{\beta}} dz \\ &= \int_0^{\infty} \frac{z^{k-1}}{(1+z^{1/\beta})^{k\beta-1}} e^{-z} dz.\end{aligned}$$

- Version 3, $AGG(\gamma, \sigma, q)$ for $q \neq 0$ only:

$$f(y) = \frac{|q|}{\sigma\zeta\left(\frac{1}{\sigma q}, \frac{1}{\sigma q}, \frac{q}{\sigma}\right)} \exp\left\{w_1 q^{-2} - e^{w_1}\right\} (1 + qe^{\sigma w_1})^{1-\frac{1}{\sigma q}}$$

where $w_1 = (y-\gamma)\sigma^{-1}q + \mu^*$, $y = \log t \in \mathbb{R}$, $\mu^* = \psi(q^{-2})$, ψ being the first derivative of the logarithm of the gamma function and $q = k^{-\frac{1}{2}}$ (c being some positive constant). γ is a location parameter related to the parameters in Version 2 by $\gamma = \log\theta + \mu^*/\beta$, σ is a scale parameter $\sigma = q/\beta$, and q is a shape parameter. w_1 and $\gamma \in \mathbb{R}$. σ, $q > 0$ and

$$\begin{aligned}\zeta\left(\frac{1}{\sigma q}, \frac{1}{\sigma q}, \frac{q}{\sigma}\right) &= \int_0^{\infty} \left(\frac{z}{1+z}\right)^{\frac{1}{\sigma q}-1} e^{-z^{\frac{q}{\sigma}}} dz \\ &= \int_0^{\infty} \frac{z^{q^{-2}-1}}{(1+z^{\frac{\sigma}{q}})^{\frac{1}{\sigma q}-1}} e^{-z} dz.\end{aligned}$$

- Version 4, $AGG(\mu, \sigma, k)$:

$$f(y) = \frac{k^{k-\frac{1}{2}}}{\sigma\zeta\left(\frac{\sqrt{k}}{\sigma}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right)} \exp\left\{\sqrt{k}w - ke^{\frac{w}{\sqrt{k}}}\right\} \left(1 + k^{\sigma\sqrt{k}} e^{\sigma w}\right)^{1-\frac{\sqrt{k}}{\sigma}}, \tag{7.1}$$

where $w = (y-\mu)/\sigma$ and $y = \log t \in \mathbb{R}$. μ is a location parameter related to the parameters in Version 2 by $\mu = \log\theta + \log k/\beta$, σ is a scale parameter $\sigma = 1/(\beta\sqrt{k})$ and k is a shape parameter. w and $\mu \in \mathbb{R}$ while σ and k are positive and

$$\zeta\left(\frac{\sqrt{k}}{\sigma}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right) = \frac{1}{\sigma\sqrt{k}} \int_0^{\infty} \left(\frac{z}{1+z}\right)^{\frac{\sqrt{k}}{\sigma}-1} e^{-z^{1/\sigma\sqrt{k}}} dz \tag{7.2}$$

$$= \int_0^{\infty} \frac{z^{k-1}}{\left(1+z^{\sigma\sqrt{k}}\right)^{\frac{\sqrt{k}}{\sigma}-1}} e^{-z} dz. \tag{7.3}$$

It is now time to illustrate how the GG pdf versions actually look and how they behave for different values of the parameters by presenting some pdf plots. Figures 7.2 and 7.3 present plots of the AGG pdf Version 1. Similarly to the GG plots in Section 2.1, the former figure plots the pdfs for fixed values of θ and β and variable α where the particular case of an AGG($\theta = 2, \alpha, \beta = 1.5$) for $\alpha \in \{0.1, 0.25, 0.5, 1, 2, 5, 25\}$ is taken. In the latter figure, we fix θ and α and we vary β, this time considering the particular AGG($\theta = 2, \alpha = 3, \beta$) for $\beta \in \{0.75, 0.9, 1, 1.5, 2, 3, 4\}$. We observe that, generally, the AGG($\theta = 2, \alpha, \beta = 1.5$) pdf shapes of Figure 7.2 (but not the scales) are very similar to the GG($\theta = 2, \alpha, \beta = 1.5$) ones of Figure 2.1. One noticeable difference is that the AGG($\theta = 2, \alpha, \beta = 1.5$) pdfs are more shifted to the left hand side of the positive real line and this is more obvious when α is larger. In Figure 7.3, we observe that the AGG($\theta = 2, \alpha = 3, \beta$) pdfs are closer to zero than the GG($\theta = 2, \alpha = 3, \beta$) ones in Figure 2.2. While the pdf of AGG($\theta = 2, \alpha = 3, \beta = 0.75$) at $x = 2$ is approximately 0.131, the GG($\theta = 2, \alpha = 3, \beta = 0.75$) pdf at $x = 2$ is approximately 0.023. We note that, similarly to the GG, the AGG densities are, mathematically, decreasing or unimodal. However, we are not able to provide an explicit form of the mode here.

Referring to the AGG pdf Version 2, the r^{th} moments and the cumulative distribution function of the AGG distribution are respectively

$$E(t^r) = \theta^r \frac{\zeta(k\beta + r, k\beta, \beta)}{\zeta(k\beta, k\beta, \beta)}$$

and

$$F(t) = \frac{\int_0^t \left(\frac{z}{1+z}\right)^{k\beta-1} e^{-z^\beta} dz}{\zeta(k\beta, k\beta, \beta)}. \tag{7.4}$$

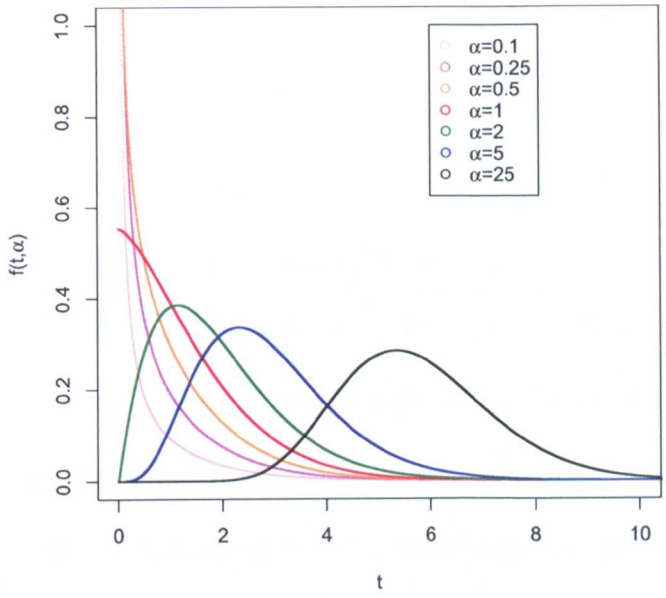

Figure 7.2: AGG pdf Version 1 plotted at seven α values for fixed θ and β.

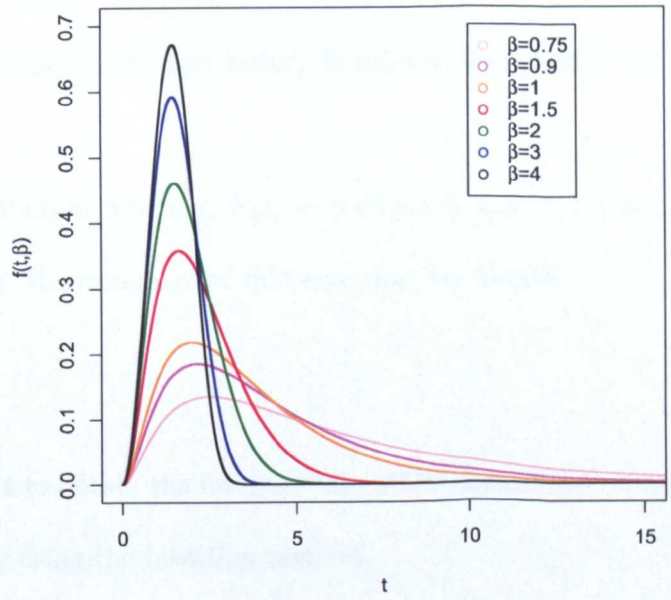

Figure 7.3: AGG pdf Version 1 plotted at seven β values for fixed θ and α.

7.3 Generating Random Variables from the Alternative Generalised Gamma Distribution

There are many known ways to generate random variables from known density functions. In our attempt to generate data from the AGG distribution, we visit two of the most known procedures: the inversion and the rejection methods. In the inversion method, the main idea involves finding the inverse of the distribution function, whereas in the rejection method the main challenge lies in finding an appropriate known distribution that the AGG can use as reference for comparison to accept or reject random variables. Using these two methods, we present two attempts at generating AGG random variables along with the drawbacks and advantages of each.

7.3.1 AGG Data Generation Using the Inversion Method

In the inversion method, by inverting the distribution function of uniformly distributed random variables in the interval (0,1), we generate f-distributed random variables. Refer to Devroye (1986) for details. Let $T \sim AGG(\theta, \alpha, \beta)$, $F(t)$ its distribution function and $U \sim U(0,1)$. Then, $F^{-1}(U)$ is $AGG(\theta, \alpha, \beta)$ distributed. It remains to obtain $F^{-1}(U)$ for the $AGG(\theta, \alpha, \beta)$ distribution.

Being a distribution function, $F(t) = p$ where $0 < p < 1$ and consequently, its inverse satisfies $F^{-1}(p) = t$. Rearranging the first equation, we obtain

$$F(t) - p = 0$$

which we solve for t to obtain the inverse of the AGG distribution function at $0 < p < 1$. This is done numerically using the bisection method.

The aim now is to calculate $F(t)$ which is the ratio of two integrals as explained in equation 7.4. This process is not easy as we did not manage to calculate the integrals explicitly.

Let us look closely at the integral $\int_0^t \left(\frac{z}{1+z}\right)^{\alpha-1} e^{-z^\beta} dz$ for both finite and infinite $t > 0$ and, in particular, at the integrand. Let

$$\eta(z) = \left(\frac{z}{1+z}\right)^{\alpha-1} e^{-z^\beta},$$

then $\lim_{z\to\infty} \eta(z) = 0$ (converges very slowly if $\beta < 1$). The limit at zero depends on the value of α such as

$$\lim_{z\to 0} \eta(z) = \begin{cases} 0 & \text{if } \alpha > 1 \\ \infty & \text{if } \alpha < 1 \\ 1 & \text{if } \alpha = 1. \end{cases}$$

Numerical computation of $F(t)$ is not always accurate especially when $\alpha < 1$ since $\eta(z)$ goes to ∞ at zero in this case. To overcome this problem, we consider the transformation $v = \log z$. This transformation extends the integral of the function η to the negative real line avoiding any "infinity" at zero. The integral then takes the form

$$\beta \int_{-\infty}^{\log t} \frac{\exp\{\alpha v - e^{\beta v}\}}{(1+e^v)^{\alpha-1}} dv \qquad t > 0$$

and the limit of the function inside the integral, at $\pm\infty$, is zero. We now have a bell-shaped function that is easier to integrate . This was done numerically using either the trapezoidal rule or the `int` function of the `rmutil` package in **R**, which can be downloaded from *http://www.commanster.eu/rcode.html.*

The main advantage of using the inversion method for $AGG(\theta, \alpha, \beta)$ data generation is that it is not restrictive. It works for all θ, α and β. Its disadvantage is that the bisection method is not very quick and requires the specification of upper and lower limits. Also, it requires calculating $\zeta(\alpha, \alpha, \beta)$ which is not always straightforward as already explained.

7.3.2 AGG Data Generation Using the Rejection Method

According to Devroye (1986), to generate n data points from a pdf f using the rejection method, start by finding an appropriate known density g and let $h = f/g$. Find the value t_m that maximises $h(t)$. The iterative procedure for generating the random variables is then as follows

1. Set the iteration number i to 0.
2. Set $i = i + 1$.
3. Simulate a random variable $T_i \sim g$.
4. Simulate another random variable $U_i \sim U(0, 1)$.
5. If

$$U_i \leq h(T_i)/h(t_m), \tag{7.5}$$

then T_i is AGG distributed, otherwise T_i is rejected.

6. Repeat steps (3), (4) and (5) until we obtain a T_i that satisfies (7.5).
7. Repeat steps (2), (3), (4), (5) and (6) until we obtain n AGG distributed random variables.

By comparing the AGG distribution to the Weibull distribution, we used the rejection method to develop an algorithm for random number generation from the AGG with pdf Version 1. Let $f \sim AGG(\theta, \alpha, \beta)$ and $g \sim Weibull(\theta, \beta)$ (the Weibull distribution with scale parameter θ and shape parameter β) with well-known pdf

$$g(t) = \frac{\beta}{\theta^\beta} t^{\beta-1} \exp\left\{-\left(\frac{t}{\theta}\right)^\beta\right\}, \qquad t > 0,$$

where the scale parameter θ and the shape parameter β are both positive. Let

$$h(t) = \frac{f(t)}{g(t)} = \frac{\theta^{\beta-\alpha}}{\zeta(\alpha, \alpha, \beta)} \frac{t^{\alpha-\beta}}{(1 + t/\theta)^{\alpha-1}}.$$

We would like to maximise $h(t)$ over $t > 0$. This is done by solving $h'(t) = (\alpha - \beta)t^{-1}(1 + t/\theta) - (\alpha - 1)/\theta = 0$ for t. The solution to this equation turns out to be $t_m = (\beta - \alpha)\theta/(1 - \beta)$ leading to four cases:

1. $(\alpha - \beta) > 0$ and $(1 - \beta) < 0 \Rightarrow t_m > 0$ is maximum.

2. $(\alpha - \beta) < 0$ and $(1 - \beta) > 0 \Rightarrow t_m > 0$ is minimum.

3. $(\alpha - \beta) > 0$ and $(1 - \beta) > 0 \Rightarrow t_m < 0$.

4. $(\alpha - \beta) < 0$ and $(1 - \beta) < 0 \Rightarrow t_m < 0$.

The first case is the only acceptable one; all the rest are rejected since our aim is to find a maximum for a positive t. Therefore, using this procedure, we can generate random variables from the AGG distribution only when $\alpha > \beta > 1$. Conditional on the previous inequality, we can then proceed with generating random variables from the AGG where it turns out that

$$t_m = (\beta - \alpha)\theta/(1 - \beta) > 0$$

and the acceptance criterion for simulated $t_i \sim Weibull(\theta, \beta)$ when $u_i \sim U(0, 1)$ is that

$$u_i \leq \frac{(1 + t_m/\theta)^{\alpha-1}}{t_m^{\alpha-\beta}} \frac{t_i^{\alpha-\beta}}{(1 + t_i/\theta)^{\alpha-1}}.$$

The main advantage for AGG data generation using the rejection method is that the rejection criterion is independent of $\zeta(\alpha, \alpha, \beta)$, a fact that makes computation more accurate and far easier and quicker. The main drawback is that it is restrictive to the case $\alpha > \beta > 1$.

7.4 Asymptotics for the Maximum Likelihood Estimates of the Alternative Generalised Gamma

Similarly to the GG distribution, the aim in this section is to find the asymptotics of the ML parameter estimates of an AGG distribution. Let us consider first the following notations:

Notations:

- $\zeta_1(n,d,p) = p^2 \int_0^\infty \log z \frac{z^{n-1}}{(1+z)^{d-1}} e^{-z^p} dz$
- $\zeta_2(n,d,p) = p^3 \int_0^\infty (\log z)^2 \frac{z^{n-1}}{(1+z)^{d-1}} e^{-z^p} dz$
- $\zeta_3(n,d,p) = p \int_0^\infty \log\left(\frac{z}{1+z}\right) \frac{z^{n-1}}{(1+z)^{d-1}} e^{-z^p} dz$
- $\zeta_4(n,d,p) = p \int_0^\infty \left(\log\left(\frac{z}{1+z}\right)\right)^2 \frac{z^{n-1}}{(1+z)^{d-1}} e^{-z^p} dz$
- $\zeta_5(n,d,p) = p^2 \int_0^\infty \log z \log\left(\frac{z}{1+z}\right) \frac{z^{n-1}}{(1+z)^{d-1}} e^{-z^p} dz$
- $\zeta_6(n,d,p) = p \int_0^\infty \log(1+z) \frac{z^{n-1}}{(1+z)^{d-1}} e^{-z^p} dz$

7.4.1 Asymptotics of the AGG PDF Version 4

We approach the problem through the distribution of $Y_i = \log T_i$ parametrised as in (7.1). The loglikelihood is

$$n\{-\log\sigma + \left(k-\frac{1}{2}\right)\log(k) - \log\zeta\left(\frac{\sqrt{k}}{\sigma}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right) + \sqrt{k}\frac{\overline{Y}-\mu}{\sigma} - \frac{k}{n}\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)\sum_{i=1}^{n}\exp\left(\frac{Y_i}{\sigma\sqrt{k}}\right) + \left(1-\frac{\sqrt{k}}{\sigma}\right)\sum_{i=1}^{n}\log\left(1+k^{\sigma\sqrt{k}}e^{Y_i-\mu}\right)\Big\}.$$

Let

$$M_j \equiv \frac{1}{n}\sum_{i=1}^{n}\left(\frac{k^{\sigma\sqrt{k}}\exp(Y_i-\mu)}{1+k^{\sigma\sqrt{k}}\exp(Y_i-\mu)}\right)^j; \quad j=1,2,$$

$$N \equiv \frac{1}{n}\sum_{i=1}^{n}\log\left(1+k^{\sigma\sqrt{k}}\exp(Y_i-\mu)\right)$$

Note that $0 < M_j < 1$ and $N > 0$ and recall that

$$S_j \equiv \frac{1}{n}\sum_{i=1}^{n} Y_i^j \exp\left(\frac{Y_i}{\sigma\sqrt{k}}\right); \quad j=0,1,2.$$

Notations:

- $\zeta = \zeta\left(\frac{\sqrt{k}}{\sigma}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right)$
- $\zeta_\sigma = \frac{\partial\zeta}{\partial\sigma} = -\frac{1}{\sigma}\zeta + \frac{1}{\sigma}\zeta_1\left(\frac{\sqrt{k}}{\sigma} + \frac{1}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right) - \frac{\sqrt{k}}{\sigma^2}\zeta_3\left(\frac{\sqrt{k}}{\sigma}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right)$
- $\zeta_{\sigma,\sigma} = \frac{\partial^2\zeta}{\partial\sigma^2} = \frac{2}{\sigma^2}\zeta - \frac{4}{\sigma^2}\zeta_1\left(\frac{\sqrt{k}}{\sigma} + \frac{1}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right)$
  $- \frac{1}{\sigma^2}\left\{\zeta_2\left(\frac{\sqrt{k}}{\sigma} + \frac{1}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right) - \zeta_2\left(\frac{\sqrt{k}}{\sigma} + \frac{2}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right)\right\}$
  $+ \frac{4\sqrt{k}}{\sigma^3}\zeta_3\left(\frac{\sqrt{k}}{\sigma}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right) + \frac{k}{\sigma^4}\zeta_4\left(\frac{\sqrt{k}}{\sigma}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right) - \frac{2\sqrt{k}}{\sigma^3}\zeta_5\left(\frac{\sqrt{k}}{\sigma} + \frac{1}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right)$
- $\zeta_k = \frac{\partial\zeta}{\partial k} = -\frac{1}{2k}\zeta + \frac{1}{2k}\zeta_1\left(\frac{\sqrt{k}}{\sigma} + \frac{1}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right) + \frac{1}{2\sigma\sqrt{k}}\zeta_3\left(\frac{\sqrt{k}}{\sigma}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right)$
- $\zeta_{k,k} = \frac{\partial^2\zeta}{\partial k^2} = \frac{3}{4k^2}\zeta - \frac{5}{4k^2}\zeta_1\left(\frac{\sqrt{k}}{\sigma} + \frac{1}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right)$
  $- \frac{1}{4k^2}\left\{\zeta_2\left(\frac{\sqrt{k}}{\sigma} + \frac{1}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right) - \zeta_2\left(\frac{\sqrt{k}}{\sigma} + \frac{2}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right)\right\}$
  $- \frac{3}{4\sigma k\sqrt{k}}\zeta_3\left(\frac{\sqrt{k}}{\sigma}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right) + \frac{1}{4\sigma^2 k}\zeta_4\left(\frac{\sqrt{k}}{\sigma}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right)$
  $+ \frac{1}{2\sigma k\sqrt{k}}\zeta_5\left(\frac{\sqrt{k}}{\sigma} + \frac{1}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right)$
- $\zeta_{\sigma,k} = \frac{\partial^2\zeta}{\partial\sigma\partial k} = \frac{1}{2\sigma k}\zeta - \frac{3}{2\sigma k}\zeta_1\left(\frac{\sqrt{k}}{\sigma} + \frac{1}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right)$
  $- \frac{1}{2\sigma k}\left\{\zeta_2\left(\frac{\sqrt{k}}{\sigma} + \frac{1}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right) - \zeta_2\left(\frac{\sqrt{k}}{\sigma} + \frac{2}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right)\right\}$
  $- \frac{1}{2\sigma^2\sqrt{k}}\zeta_3\left(\frac{\sqrt{k}}{\sigma}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right) - \frac{1}{2\sigma^3}\zeta_4\left(\frac{\sqrt{k}}{\sigma}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}}\right)$

Differentiating the loglikelihood with respect to μ, σ and k in turn yields the score equations

$$0 = \frac{n\sqrt{k}}{\sigma}\left\{\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)S_0 - 1 - \frac{\sigma}{\sqrt{k}}\left(1 - \frac{\sqrt{k}}{\sigma}\right)M_1\right\}, \tag{7.6}$$

$$\begin{aligned} 0 &= \frac{n\sqrt{k}}{\sigma^2}\left\{\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)(S_1 - \mu S_0) - \frac{\sigma}{\sqrt{k}} - \overline{Y}\right. \\ &+ \left.\mu + \sigma^2\log k\left(1 - \frac{\sqrt{k}}{\sigma}\right)M_1 + N - \frac{\sigma^2}{\sqrt{k}}\frac{\zeta_\sigma}{\zeta}\right\} \end{aligned} \tag{7.7}$$

and

$$\begin{aligned} 0 &= n\left\{\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)\left(\frac{S_1 - \mu S_0}{2\sigma\sqrt{k}} - S_0\right) + \log k + 1 - \frac{1}{2k} - \frac{\zeta_k}{\zeta} + \frac{\overline{Y} - \mu}{2\sigma\sqrt{k}}\right. \\ &+ \left.\frac{\sigma}{2\sqrt{k}}(\log k + 2)\left(1 - \frac{\sqrt{k}}{\sigma}\right)M_1 - \frac{1}{2\sigma\sqrt{k}}N\right\}. \end{aligned} \tag{7.8}$$

Equation (7.6) yields

$$\exp(\mu) = \left(\frac{S_0}{1 + \frac{\sigma}{\sqrt{k}} \left(1 - \frac{\sqrt{k}}{\sigma}\right) M_1} \right)^{\sigma\sqrt{k}}. \tag{7.9}$$

Using this, we can reduce (7.7) and (7.8) to

$$AR(\sigma) \equiv \frac{S_1}{S_0} - \overline{Y} - \frac{\sigma}{\sqrt{k}} + \frac{\sigma}{\sqrt{k}} \left(1 - \frac{\sqrt{k}}{\sigma}\right) \left\{ \frac{S_1}{S_0} - \mu \right\} M_1 + \sigma^2 \log k \left(1 - \frac{\sqrt{k}}{\sigma}\right) M_1 + N - \frac{\sigma^2}{\sqrt{k}} \frac{\zeta_\sigma}{\zeta} = 0.$$

and

$$AT_L(k) \equiv \log(k) - \frac{\zeta_k}{\zeta} + \frac{\overline{Y} - \mu}{\sigma\sqrt{k}} - \frac{1}{\sigma\sqrt{k}} N + \frac{\sigma}{2k} \frac{\zeta_\sigma}{\zeta} = 0 \tag{7.10}$$

Recall that

$$T_j \equiv \sum_{i=1}^{n} Y_i^j \exp\left(\frac{Y_i - \mu}{\sqrt{k}\sigma} \right); \quad j = 0, 1, 2;$$

Using these expressions, we find the components of the **Hessian** matrix at the ML estimates.

Similarly to the GG, we find that $E(T_1)$, $E(T_2)$, $E(M_1)$, $E(M_2)$ and $E(\overline{Y})$ of the AGG are

$$\begin{aligned}
E(T_1) &= \frac{1}{\zeta} \left\{ \left(\frac{\mu}{k} - \frac{\sigma}{\sqrt{k}} \log k \right) \zeta \left(\frac{\sqrt{k}}{\sigma} + \frac{1}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}} \right) \right. \\
&+ \left. \frac{\sigma}{\sqrt{k}} \zeta_1 \left(\frac{\sqrt{k}}{\sigma} + \frac{1}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}} \right) \right\}; \\
E(T_2) &= \frac{1}{\zeta} \left\{ \left(\frac{\mu^2}{k} + \sigma^2 (\log k)^2 - \frac{2\mu\sigma}{\sqrt{k}} \log k \right) \zeta \left(\frac{\sqrt{k}}{\sigma} + \frac{1}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}} \right) \right. \\
&+ \left. \left(\frac{2\mu\sigma}{\sqrt{k}} - 2\sigma^2 \log k \right) \zeta_1 \left(\frac{\sqrt{k}}{\sigma} + \frac{1}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}} \right) + \sigma^2 \zeta_2 \left(\frac{\sqrt{k}}{\sigma} + \frac{1}{\sigma\sqrt{k}}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}} \right) \right\}; \\
E(M_1) &= \frac{1}{\zeta} \left\{ \zeta \left(\frac{\sqrt{k}}{\sigma} + 1, \frac{\sqrt{k}}{\sigma} + 1, \frac{1}{\sigma\sqrt{k}} \right) \right\}; \\
E(M_2) &= \frac{1}{\zeta} \left\{ \zeta \left(\frac{\sqrt{k}}{\sigma} + 2, \frac{\sqrt{k}}{\sigma} + 2, \frac{1}{\sigma\sqrt{k}} \right) \right\}; \\
E(N) &= \frac{1}{\zeta} \left\{ \zeta_6 \left(\frac{\sqrt{k}}{\sigma}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}} \right) \right\}; \\
E(\bar{Y}) &= \frac{1}{\zeta} \left\{ \left(\mu - \sigma\sqrt{k} \log k \right) \zeta + \sigma\sqrt{k} \zeta_1 \left(\frac{\sqrt{k}}{\sigma}, \frac{\sqrt{k}}{\sigma}, \frac{1}{\sigma\sqrt{k}} \right) \right\}.
\end{aligned}$$

Given the above — and using that the score function has expectation zero — we deduce the components of the matrix $\mathbf{I}$ to be

$$
\begin{aligned}
E\left(-\frac{\partial^2 l}{\partial \mu^2}\right) &= \frac{n}{\sigma^2} - \left(1-\frac{\sqrt{k}}{\sigma}\right)\left\{\left(1-\frac{1}{\sigma\sqrt{k}}\right)E(M_1) - E(M_2)\right\};\\
E\left(-\frac{\partial^2 l}{\partial \mu \partial \sigma}\right) &= \frac{n}{\sigma^2\sqrt{k}} - \frac{n\mu}{\sigma^3} + \sqrt{k}\log k\left(1-\frac{\sqrt{k}}{\sigma}\right)\left\{\left(1-\frac{1}{\sigma\sqrt{k}}\right)E(M_1) - E(M_2)\right\}\\
&\quad + \frac{1}{\sigma}E(M_1) + \frac{1}{\sigma^3}\left\{E(\overline{Y}) - E(N)\right\} + \frac{N}{\sigma\sqrt{k}}\frac{\zeta_\sigma}{\zeta};\\
E\left(-\frac{\partial^2 l}{\partial \mu \partial k}\right) &= \frac{n}{2\sigma k\sqrt{k}} - \frac{n\mu}{2\sigma^2 k} + \frac{\sigma}{2\sqrt{k}}\left(1-\frac{\sqrt{k}}{\sigma}\right)\left\{-\frac{\log k}{\sigma\sqrt{k}}E(M_1)\right.\\
&\quad \left. - (\log k + 2)\left(E(M_1) - E(M_2)\right)\right\} - \frac{1}{2k}E(M_1) + \frac{1}{2\sigma^2 k}\left\{E(\overline{Y}) - E(N)\right\} + \frac{n}{2k\sqrt{k}}\frac{\zeta_\sigma}{\zeta};\\
E\left(-\frac{\partial^2 l}{\partial \sigma^2}\right) &= \frac{n}{\sigma^2} + \frac{n\mu^2}{\sigma^4} - k(\log k)^2\left(1-\frac{\sqrt{k}}{\sigma}\right)\left\{E(M_1) - E(M_2)\right\}\\
&\quad - \left\{\frac{2\sqrt{k}\log k}{\sigma} + \left(\frac{2k\log k}{\sigma^2} - \frac{\mu^2}{\sigma^3\sqrt{k}}\right)\left(1-\frac{\sqrt{k}}{\sigma}\right)\right\}E(M_1)\\
&\quad + \frac{1}{\sigma^4}\left\{E(T_2) - 2\mu E(T_1)\right\} + n\left\{\left(\frac{2}{\sigma} + \frac{2\sigma}{\sqrt{k}}\right)\frac{\zeta_\sigma}{\zeta} + \frac{\sigma^2}{\sqrt{k}}\left(\frac{\zeta_{\sigma,\sigma}}{\zeta} - \left(\frac{\zeta_\sigma}{\zeta}\right)^2\right)\right\};\\
E\left(-\frac{\partial^2 l}{\partial \sigma \partial k}\right) &= -\frac{n}{2\sigma k} + \frac{n\mu^2}{2\sigma^3 k} - \left(\frac{\sigma(\log k)^2}{2} + \sigma\log k\right)\left(1-\frac{\sqrt{k}}{\sigma}\right)\left\{E(M_1) - E(M_2)\right\}\\
&\quad - \left(1-\frac{\sqrt{k}}{\sigma}\right)\left\{\frac{1}{\sqrt{k}} - \frac{\mu^2}{2\sigma^2 k\sqrt{k}}\right\}E(M_1) - \frac{1}{\sigma}E(M_1)\\
&\quad + \frac{1}{2\sigma^3 k}\left\{E(T_2) - 2\mu E(T_1)\right\} - \frac{n\sigma^2}{\sqrt{k}}\left\{\left(\frac{1}{2\sigma^2\sqrt{k}} - \frac{1}{2k}\right)\frac{\zeta_\sigma}{\zeta} - \frac{\zeta_{\sigma,k}}{\zeta} + \frac{\zeta_k\zeta_\sigma}{\zeta^2}\right\};\\
E\left(-\frac{\partial^2 l}{\partial k^2}\right) &= -\frac{n}{k} - \frac{3n}{4k^2} + \frac{n\mu^2}{4\sigma^2 k^2} - \left(\frac{\sigma\log k}{2\sqrt{k}} + \frac{\sigma}{\sqrt{k}}\right)^2\left(1-\frac{\sqrt{k}}{\sigma}\right)\left\{E(M_1) - E(M_2)\right\}\\
&\quad + \left(1-\frac{\sqrt{k}}{\sigma}\right)\left\{\frac{\sigma\log k}{2k\sqrt{k}} + \frac{\mu^2}{4\sigma k^2\sqrt{k}}\right\}E(M_1) + \left\{\frac{\log k}{2k} + \frac{1}{k}\right\}E(M_1)\\
&\quad + \frac{1}{4\sigma^2 k^2}\left\{E(T_2) - 2\mu E(T_1)\right\} + n\left\{-\frac{\sigma}{4k^2}\frac{\zeta_\sigma}{\zeta} + \frac{\zeta_{k,k}}{\zeta} - \left(\frac{\zeta_k}{\zeta}\right)^2\right\}.
\end{aligned}
$$

Inverting the information matrix, we deduce the asymptotic covariance and correlation matrices.

7.4.2 Asymptotics of the AGG PDF Version 1

The loglikelihood of the AGG pdf Version 1 is

$$l = n\{\log\beta - \alpha\log\theta - \log\zeta(\alpha,\alpha,\beta) + \frac{(\alpha-1)}{n}\sum_{i=1}^{n}\log t_i - \frac{(\alpha-1)}{n}\sum_{i=1}^{n}\log\left(1+\frac{t_i}{\theta}\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta}\right\}.$$

Differentiating the loglikelihood with respect to θ, α and β, we find the score equations to be

$$0 = \frac{n}{\theta}\left\{-\alpha + \frac{\alpha-1}{n}\sum_{i=1}^{n}\frac{t_i/\theta}{1+t_i/\theta} + \frac{\beta}{n}\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta}\right\},$$

$$0 = n\left\{-\log\theta + \frac{1}{n}\sum_{i=1}^{n}\log t_i - \frac{1}{n}\sum_{i=1}^{n}\log\left(1+\frac{t_i}{\theta}\right) - \frac{\zeta_3(\alpha,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)}\right\}$$

and

$$0 = n\left\{-\frac{1}{n}\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta}\log\left(\frac{t_i}{\theta}\right) + \frac{1}{\beta}\frac{\zeta_1(\alpha+\beta,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)}\right\}.$$

Therefore, at the ML estimates, the second derivatives in the **Hessian** simplify to

$$\frac{\partial^2 l}{\partial\theta^2} = \frac{n}{\theta^2}\left\{\alpha - (\alpha-1)\left(\frac{2}{n}\sum_{i=1}^{n}\frac{t_i/\theta}{1+t_i/\theta} - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{t_i/\theta}{1+t_i/\theta}\right)^2\right) - \frac{\beta(\beta+1)}{n}\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta}\right\};$$

$$\frac{\partial^2 l}{\partial\theta\partial\alpha} = \frac{n}{\theta}\left\{-1 + \frac{1}{n}\sum_{i=1}^{n}\frac{t_i/\theta}{1+t_i/\theta}\right\};$$

$$\frac{\partial^2 l}{\partial\theta\partial\beta} = \frac{n}{\theta}\left\{\frac{\beta}{n}\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta}\log\left(\frac{t_i}{\beta}\right) + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta}\right\};$$

$$\frac{\partial^2 l}{\partial\alpha^2} = n\left\{\left(\frac{\zeta_3(\alpha,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)}\right)^2 - \frac{\zeta_4(\alpha,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)}\right\};$$

$$\frac{\partial^2 l}{\partial\alpha\partial\beta} = \frac{n}{\beta}\left\{-\frac{\zeta_3(\alpha,\alpha,\beta)\zeta_1(\alpha+\beta,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)^2} + \frac{\zeta_5(\alpha+\beta,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)}\right\};$$

$$\frac{\partial^2 l}{\partial\beta^2} = \frac{n}{\beta^2}\left\{-\frac{\beta^2}{n}\sum_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta}\left(\log\left(\frac{t_i}{\beta}\right)\right)^2 + \left(\frac{\zeta_1(\alpha+\beta,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)}\right)^2 + \frac{\zeta_2(\alpha+\beta,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)} - \frac{\zeta_2(\alpha+2\beta,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)}\right\}.$$

Given the above — and using that the score function has expectation zero — we deduce the components of the matrix **I** to be

$$
\begin{aligned}
E\left(-\frac{\partial^2 l}{\partial\theta^2}\right) &= \frac{n}{\theta^2}\left\{-\alpha+(\alpha-1)\left(2\frac{\zeta(\alpha+1,\alpha+1,\beta)}{\zeta(\alpha,\alpha,\beta)}-\frac{\zeta(\alpha+2,\alpha+2,\beta)}{\zeta(\alpha,\alpha,\beta)}\right)\right. \\
&+ \left.\beta(\beta+1)\frac{\zeta(\alpha+\beta,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)}\right\}; \\
E\left(-\frac{\partial^2 l}{\partial\theta\partial\alpha}\right) &= \frac{n}{\theta}\left\{1-\frac{\zeta(\alpha+1,\alpha+1,\beta)}{\zeta(\alpha,\alpha,\beta)}\right\}; \\
E\left(-\frac{\partial^2 l}{\partial\theta\partial\beta}\right) &= \frac{n}{\theta}\left\{-\frac{\zeta_1(\alpha+\beta,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)}-\frac{\zeta(\alpha+\beta,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)}\right\}; \\
E\left(-\frac{\partial^2 l}{\partial\alpha^2}\right) &= n\left\{\frac{\zeta_4(\alpha,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)}-\left(\frac{\zeta_3(\alpha,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)}\right)^2\right\}; \\
E\left(-\frac{\partial^2 l}{\partial\alpha\partial\beta}\right) &= \frac{n}{\beta}\left\{\frac{\zeta_3(\alpha,\alpha,\beta)\zeta_1(\alpha+\beta,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)^2}-\frac{\zeta_5(\alpha+\beta,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)}\right\}; \\
E\left(-\frac{\partial^2 l}{\partial\beta^2}\right) &= \frac{n}{\beta^2}\left\{-\left(\frac{\zeta_1(\alpha+\beta,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)}\right)^2+\frac{\zeta_2(\alpha+2\beta,\alpha,\beta)}{\zeta(\alpha,\alpha,\beta)}\right\}.
\end{aligned}
$$

Finally, we invert the information matrix to obtain the asymptotic covariance and correlation matrices. The correlations turn out to be, unsurprisingly, independent of θ. Tables 7.1, 7.2, 7.3, 7.4, 7.5 and 7.6 display values of the correlations between $\hat{\theta}$, $\hat{\alpha}$ and $\hat{\beta}$. The values are rounded to three decimal digits. As these computations required the numerical calculation of some complicated integrals, we have used the software MAPLE for this purpose. In principal, MAPLE provides a good accuracy when computing such numerical integrals, though it takes time to do that. As previously mentioned, the main aim of this analysis is to compare the AGG correlations to the GG ones hoping that the former ones are smaller. The reason some correlation values were given for the GG distribution case but not for the AGG, is that for certain values of the parameters, say for $\beta < 1$ and for β very large, we weren't able to obtain reasonable AGG correlations, therefore we skipped those results and we did not include them. Despite that, we could still observe the general pattern of the AGG correlations. We notice that the results are very similar to the GG ones given in Tables 2.11 to 2.16. The same pattern

occurs with generally (most of the time) slightly smaller correlations for the AGG distribution. When the values of the AGG correlations are less than the GG ones, the differences are small reaching to a maximum of around 0.07. In a very few cases, the values of the GG correlations are smaller than the AGG ones. The noticeable ones are the correlations between $\hat{\theta}$ and $\hat{\beta}$. When α is small, the correlation difference gets to around 0.3. Overall, the AGG correlations are disappointingly only slightly better given that the primary motivation underlying the AGG distribution was to separate the tail influences of α and β with the hope that they could be estimated 'more separately' and hence better in practice. To explore the usefulness of the AGG distribution more, we look at its performance with MLE using simulations in the next section.

Table 7.1: Correlations between $\hat{\theta}$ and $\hat{\alpha}$ for $\beta = 2$ and variable α.

α	0.001	0.01	0.1	0.5	1	2	5
Corr$(\hat{\theta}, \hat{\alpha})$	-0.035	-0.109	-0.320	-0.607	-0.752	-0.866	-0.948

Table 7.2: Correlations between $\hat{\theta}$ and $\hat{\alpha}$ for $\alpha = 3$ and variable β.

β	1	2	5	10	100
Corr$(\hat{\theta}, \hat{\alpha})$	-0.940	-0.911	-0.797	-0.648	$-.216$

Table 7.3: Correlation between $\hat{\theta}$ and $\hat{\beta}$ for $\beta = 2$ and variable α.

α	0.001	0.1	0.2	0.5	1	2	5
Corr$(\hat{\theta}, \hat{\beta})$	0.730	0.772	0.803	0.862	0.912	0.952	0.981

Table 7.4: Correlation between $\hat{\theta}$ and $\hat{\beta}$ for $\alpha = 3$ and variable β.

β	1	5	10	100
Corr$(\hat{\theta}, \hat{\beta})$	0.982	0.899	0.792	0.516

Table 7.5: Correlation between $\hat{\alpha}$ and $\hat{\beta}$ for $\beta = 2$ and variable α.

α	0.01	0.1	0.5	1	3	5	10
$\text{Corr}(\hat{\alpha}, \hat{\beta})$	-0.097	-0.277	-0.506	-0.637	-0.829	-0.890	-0.943

Table 7.6: Correlation between $\hat{\alpha}$ and $\hat{\beta}$ for $\alpha = 3$ and variable β.

β	1	2	5	10	100
$\text{Corr}(\hat{\alpha}, \hat{\beta})$	−0.881	-0.829	−0.669	-0.518	-0.176

7.5 Simulation Study Comparing AGG with GG when $\alpha > \beta > 1$

In this section, we perform a simulation study that compares the AGG distribution to the GG distribution in terms of MLE. As generating AGG random variables, for any θ, α and β, is quite complicated (as explained in Section 7.3), we will restrict this study to data sets that satisfy $\alpha > \beta > 1$ only since in this case generating data proved to be quite straightforward using the rejection method. We note that the case $\alpha > 1$ defines bell-shaped GG and AGG density functions which are of more interest to us because they are very common and we would like to see how each of α and β affect both tails. When $\beta < 1$ the AGG pdf converges very slowly to zero at infinity and estimating it has many complications. For these reason, in this simulation study, we will only address the cases where $\alpha > \beta > 1$.

To compare both distributions, we simulate a set of "real parameters" that consists of 100 triplets α, θ and β (all satisfying $\alpha > \beta > 1$). θ is simulated from a $G(2, 1)$ distribution. We take $\beta \sim G(2, 1)$ and $\alpha \sim G(2.5, 1.5)$ while filtering out all cases that don't satisfy the inequality. This ensures that α and β take a reasonably wide range of values in the domain.

In a similar way, we simulate 100 triplets of "initial parameters" that will be used as initial values for our MLE. The set of "real parameters" is given in Table A.11 of the Appendix. From each of the 100 triplets of "real parameters", we simulate 100 AGG and GG data sets of size $n = 500$. To simulate the AGG data sets, we use the rejection method, of course. Using MLE via the Nelder-Mead procedure, we estimate each of the 100 AGG and GG data sets using the correct form of model each time, that is, the AGG model fit to AGG data sets, the GG model to GG data sets. For each data, we repeat estimation 100 times, each time using a triplet of the "initial parameters" and we choose the maximal likelihood attained. Corresponding to this maximal likelihood, we compute the estimates $\widehat{\theta}$, $\widehat{\alpha}$ and $\widehat{\beta}$ (of θ, α and β respectively) for each of AGG and GG. The maximal likelihoods of the data sets and their corresponding estimates for both the AGG and GG models are given in Table A.12 of the Appendix.

To check how well each distribution estimates reality and to compare between both of them, we compute the relative bias and the squared error for the parameter estimates of both distribution. The mean relative biases and the root mean relative squared errors for each of the AGG and GG estimates (of the 100 data sets) are given in Table 7.7. The values are rounded to three decimal digits.

The square root of the mean relative squared error of the AGG θ estimate (which is 0.216) is smaller than the GG one (which is 0.381). The square root of the mean relative squared error of the β estimate is also smaller in the AGG case (0.162 for the AGG vs 0.184 for the GG). As for the α case, the GG performs better. The reason for that is the 30^{th} of the 100 data sets where the AGG highly over-estimates α. If we exclude this data set, the GG still performs better, but this time the AGG mean squared error becomes 0.263 instead of 1.690. In general, the difference between the performance of the two distributions is very small. Among the hundred data sets the AGG relative squared error for θ is less than the GG one 62% of the time. For α

and β it is respectively 48% and 48% of the time less.

The results do not support the hypothesis that the AGG distribution will perform much better because each of α and β influence separately the respective regions close to zero and close to infinity. This is disappointing. Any practical benefits that might be forthcoming by using the AGG distribution in place of the GG distribution are minuscule; moreover, the AGG distribution adds a little complication relative to the GG distribution in some respects. For these reasons, it seems inappropriate to pursue the AGG distribution further.

Table 7.7: Mean relative bias and root mean relative squared error of the estimated parameters of the AGG and GG data sets.

Distribution	AGG			GG		
Parameter	θ	α	β	θ	α	β
mean relative bias	-0.02	0.222	0.002	0.013	0.041	-0.005
root mean relative squared error	0.216	1.690	0.162	0.381	0.222	0.184

Chapter 8

Conclusion

In this dissertation, we presented a thorough methodological statistical study on QR using the GG distribution. The theory was complemented by supporting computational implementations in the statistical software **R**. This is an original unique study that targeted PQR by making the parameters of the GG distribution dependent on a univariate covariate. The study followed a logical chronology that combined researching QR simultaneously with the GG distribution. We presented a detailed literature review that explained and clarified all the previous historical arguments in each of these areas and set the scene for our further research. Once the basic definitions and history were given, we presented our contributions to these fields through a rounded study that included all stages of statistical modeling. We validated our work via simulations. We applied our modeling package to two real-life examples and plotted the corresponding reference charts. Finally, we proposed an alternative distribution to the GG and explored its properties and usefulness.

8.1 Summary and Main Contributions

Following the study we described in the chapters of this thesis, we present a list of the main contributions offered:

- We promoted the GG distribution as a suitable distribution for regression analysis, a field

that has been dominated by normality assumptions. In the cases where the unexplained variation in the dependent variable is not symmetric, we proposed the GG distribution whose shape parameters control skewness, applicable to response data that take positive values only.

- We helped rehabilitate MLE of the GG while improving on earlier work by being the first to formulate the iterative approach to the solution of the likelihood score equations in such a way that the individual equations involved are uniquely solvable and far from being problematic as a number of authors have suggested. All attempts mentioned in the literature review to estimate the GG distribution showed computational difficulties and complications. Procedures based on MLE often assumed that the shape parameter (or one of the parameters) is considered known. We extended the work done by Lawless (1980) to solve the score equations in a simple and yet efficient way taking all parameters to be unknown and none of them fixed.

- We provided bounds on the values of the ML estimates. Those bounds can be used to obtain initial values for MLE.

- We explored the GG distribution in a QR framework, a study that no researcher known to us has addressed before. As explained in the literature review, very little work was done on PQR let alone using the GG distribution for that. The aim was to go beyond the limitations of regression, which models the mean only, into modeling every quantile of the data using QR.

- We integrated the idea of QR by making the GG parameters dependent on a univariate covariate. The main idea lies in studying how the covariate influences the location, scale and shape of the entire distribution. The closest research to us that addressed PQR is

by Warren Gilchrist, yet he never used the idea of a covariate in the same way. In a QR framework, we extended the work done on the three-parameter GG to a wider model that involved the four-, five- and six-parameter GG obtained by making the parameters dependent on the covariate. The quantiles, which are functions of the parameters, hence became functions of the covariate as well. We considered three cases of the conditional GG distribution: the four-parameter GG, where we set μ to be a linear function of a covariate, the five-parameter GG, where we made μ and σ dependent respectively linearly and loglinearly on a covariate and the six-parameter GG, where all three parameters μ, σ and k are dependent on a covariate, μ linearly, and σ and k loglinearly.

- We offered an overall statistical study that started with model identification and ended with suggesting a goodness-of-fit test. Along the way, we passed by MLE of the three-, four-, five- and six-parameter GG and model selection using LRTs. We gave expressions for the quantiles and confidence bands and presented some plots. This overall modeling package is unique in the sense that it was done particularly for the GG in a PQR framework, an area that hasn't been explored in a similar way before.

- We presented the asymptotics of the three-, four-, five- and six-parameter GG ML parameter estimates. We deduced the covariance and correlation matrices and analysed the correlations between the different parameter estimates.

- We proposed a goodness-of-fit test statistic for the GG distribution based on the χ^2 goodness-of-fit test.

- We validated the theoretical work computationally via simulations. The simulation study in Section 2.5 validated our iterative algorithm for MLE of the three-parameter GG distribution. Similarly, the simulation study in Section 4.2 validated our iterative algorithm for

MLE of the four-parameter GG distribution. Section 5.1 offered a simulation study that applied all the different steps of our suggested modeling package to three-, four-, five- and six-parameter GG simulations. Section 5.2 validated our proposed goodness-of-fit test.

- We applied our modeling package to two real-life examples: one studied an environmental issue that analysed the effect of flux on water table depth and the other a health issue that studied the effect of height on weights of individuals. We plotted reference charts for both data.

- We introduced the AGG distribution. This alternative distribution specialises the effect of each of the parameters α and β respectively to the region near zero of the real line and the region near infinity. We proposed two approaches for random variate generation from the AGG. We compared it to the GG by analysing the correlations between its parameters and by studying its performance for MLE.

8.2 Main Results Obtained

To each of the above contributions, we display the main results obtained from the computations we made and the simulation studies we implemented. We summarise them as follows:

- The simulation study in Subsection 2.5.1 showed that, computationally, MLE of the three-parameter GG is generally straightforward. Overall, our program and the Nelder-Mead algorithm attained the global maximum likelihood the highest proportion of times. Local maxima of the likelihood, on the rare occasions they were observed, are both far from the global maximum and have much smaller values of the likelihood. We note that our theoretical results are very much still partial in the sense that they do not guarantee convergence of our algorithm nor uniqueness of the ML estimates. The simulation study in Subsection 2.5.2 checks how close the estimates are from the true values. Obviously,

the true parameters lie in 95% CIs of their estimates' averages. Increasing the sample size allows, of course, for more exactness of the estimates.

- We found the asymptotic correlations between the GG parameters $\hat{\mu}$, $\hat{\sigma}$ and $\hat{k}$ as k tends to zero to be $\text{Corr}(\hat{\mu}, \hat{\sigma}) = -1$, $\text{Corr}(\hat{\mu}, \hat{k}) = -1$, $\text{Corr}(\hat{\sigma}, \hat{k}) = 1$ and as k tends to infinity $\text{Corr}(\hat{\mu}, \hat{\sigma}) = 0$, $\text{Corr}(\hat{\mu}, \hat{k}) = -\sqrt{3/5}$ and $\text{Corr}(\hat{\sigma}, \hat{k}) = 0$. We noticed that in all cases, as k increases, the correlations decrease. There is a negative correlation between $\hat{\mu}$ and $\hat{\sigma}$ and $\hat{\mu}$ and $\hat{k}$, whereas $\hat{\sigma}$ and $\hat{k}$ are positively correlated.

- We will now analyse the correlation results between the GG parameters $\hat{\theta}$, $\hat{\alpha}$ and $\hat{\beta}$. While $\hat{\theta}$ and $\hat{\alpha}$ are negatively correlated as well as $\hat{\alpha}$ and $\hat{\beta}$, $\hat{\theta}$ and $\hat{\beta}$ are positively correlated. For fixed β, it appears that as $\text{Corr}(\hat{\theta}, \hat{\alpha})$ increases (in absolute value) from zero to -1 as α increases. For fixed α, $\text{Corr}(\hat{\theta}, \hat{\alpha})$ decreases (in absolute value) from -1 to values around zero as β increases. A similar pattern happens for $\text{Corr}(\hat{\alpha}, \hat{\beta})$. As for $\text{Corr}(\hat{\theta}, \hat{\beta})$, for fixed β the correlation increases to about 1 while as we fix α, $\text{Corr}(\hat{\theta}, \hat{\beta})$ decreases from values around 1 as β increases. Overall, we can say that the correlations are reasonably small or moderate for reasonable values of α and β.

- In the simulation study of Subsection 4.2.1 we showed that, computationally, MLE of the four-parameter GG is generally straightforward. We compared it with the three-parameter GG case. Each method, unsurprisingly, has an increased number of reported errors, but the situation is still very good. When $n = 200$, our program still has a 96.8% success rate, Nelder-Mead a 97.8% success rate, and BFGS an 89.7% success rate.

- In the simulation study of Subsection 5.1, we addressed four different situations (coming from four different sets of parameters) of each of the four-, five- and six-parameter GG. In total we explored 16 case-scenarios using 1600 data sets. As already mentioned, LRTs are

applied to each of the 1600 data sets. Our LRT approach predicted the right number of parameters most of the time. Generally, when the LRT did not pick up the true number of parameters, one of the parameters was very small so that it wouldn't affect the model much: this case happened when data could be modeled by a distribution with fewer parameters. In all, we did not find any noticeable differences between the $n = 200$ and $n = 500$ cases. They both behave reasonably well.

- In the simulation study of Subsection 5.1, generally the mean relative biases and the root mean relative squared errors are quite small. For most of the cases, the "Known p Version" is slightly better than the "LRT Version", though the difference is very small. Also, we observe that the true quantiles lie within the CIs around the estimated quantiles most of the time with about the right percentage.

- From the simulation study in Section 5.2, it appears to us that the $\hat{\theta}$ and $\hat{k}$ (ML estimates of θ and k respectively) values are respectively approximately 1.5 and 2 for each of S31, S32, S33, ..., S64. These results were remarkably consistent, a fact that lead us to think that the $G(\theta = 1.5, k = 2)$ distribution is a good approximation to our test statistic τ. We also observe that although the size of each of S31, S32, S33, ..., S64 is not very large ($n = 100$), the $G(\theta = 1.5, k = 2)$ pdf still fits the histograms reasonably well. From the ML estimates for each of S31, S32, S33, ..., S64, along with the p-values of the χ^2 test, we notice that for a 0.05 level of significance, 10 out of 16 (62.5%) are well-estimated by a $G(\theta = 1.5, k = 2)$. For a 0.01 level of significance 81.25% are well-estimated.

- In the "WTD vs Flux" data, the LRT predicts a six-parameter GG. The values of the estimated parameters are $\hat{a} = 5.278$, $\hat{b} = -4.013$, $\hat{c} = -2.424$, $\hat{d} = 3.106$, $\hat{f} = -3.886$ and $\hat{g} = 11.159$. The quantiles have monotone and unimodal shapes. The confidence bands around the center are quite narrow. They are wide at the edges where there are very

few data points allowing more variation and less certainty in the model. Applying our suggested goodness-of-fit test, it turns out that $\{\widehat{m_l}\} = \{21, 54, 75, 102, 26, 29\}$, $\{m_l\} = \{30.7, 46.05, 76.75, 76.75, 46.05, 30.7\}$ and $\tau = 21.60803 > 9.96$. Based on those results, we reject the fact the the GG model is a good fit in this case at 0.01 level of significance, although our observation of the data and its estimated quantiles shows that the latter fit quite well and take the general shape of the data.

When we remove the point of coordinates (0.96,69.8), which we suspect is an outlier, the results barely change, whereas when the point of coordinates (0.01,31.3) is removed, the 10% quantile loses its horizontal behaviour near zero and maintains a monotone exponentially decreasing pattern in this region. Still, in this case, $\tau = 21.81917 > 9.96$. The results are only slightly better. Removing this point changed the behaviour of the quantiles around zero and managed to shift downwards the 75% quantile (which we think is causing the high value of τ) only very slightly, though not to the extent that allowed the observed and estimated count of points in this area to match more. We observe many points in the upper part of the fourth region very close to the 75% quantile. Although the goodness-of-fit test rejects the GG model as a good fit, the counts of observed and expected data points in each region match quite well except around the fourth and fifth regions where shifting downwards the 75% quantile only slightly would enable much better results.

- In the "Weight vs Height" data, the estimated k was taking very high values. As the lognormal is the special case of the GG for $k \to \infty$, we concluded that it is a convenient model and that the three-parameter GG distribution is not required in this case. The LRT predicts a four-parameter lognormal. The values of the estimated parameters are $\hat{a} = 2.377747$, $\hat{b} = -2.325$, $\hat{c} = 2.685$, and $\hat{d} = 1.101$. The quantiles are increasing with

the increase in height (and weight). They spread out from each other as height increases reflecting the change in the data quite well. The confidence bands are very narrow around the estimated quantiles. The large number of data points allows more certainty in the model (i.e. less variation) and hence narrower CIs. The goodness-of-fit test shows that $\{\widehat{m_l}\} = \{427, 763, 1086, 1118, 606, 448\}$, $\{m_l\} = \{444.8667.2, 1112, 1112, 667.2, 444.8\}$ and $\tau = 20.74475 > 9.96$. Based on the results, we reject the fact that the GG model is a good fit to our data, though, if we look at $\widehat{m_l}$ and m_l, we find that they follow a similar pattern. Also, the general shape of the quantiles reflect the exponential increase in the data and they widen as normalised height goes to one, as the data widens.

- The AGG distribution is a useful distribution that we have attempted to explore because it separates the tail influences of α and β where α influences separately the region close to zero while β influences the region close to infinity. We haven't managed to develop an iterative algorithm for its MLE. Yet, using the general-purpose optimisation procedures to estimate its parameters proved to be successful. The distribution has great similarities with the GG. We notice that the distributions are similar close to zero. Pdf plots show that they have similar density functions, the AGG being more shifted to the left than the GG. Random variate generation from the AGG was not fully successful. The inversion method is not very quick and requires calculating $\zeta(\theta, \alpha, \beta)$ which is not always straightforward. The rejection method is more accurate and far easier and quicker. Its drawback is that it is restrictive to the case $\alpha > \beta > 1$.

- We now briefly analyse the correlation results between the AGG parameters $\hat{\theta}$, $\hat{\alpha}$ and $\hat{\beta}$ and compare them with the GG ones. A similar pattern to that of the GG occurs with generally (most of the time) slightly smaller correlations for the AGG distribution. In a very few cases, the values of the GG correlations are smaller than the AGG ones.

- From the simulation study in Section 7.5 performed to evaluate the AGG distribution in terms of MLE, we find that the square root of the mean relative squared errors of the AGG θ and β estimates are smaller than the GG ones. As for the α case, the GG performs better. The reason for that is the 30^{th} of the 100 data sets where the AGG highly over-estimates α. Generally, the difference between the performance of the two distributions is very small.

8.3 Overall Thesis

The GG is a multi-task distribution that can model small and large data. It can easily and accurately model skewness in data. It encompasses a variety of known and useful special cases. When a special case is enough to model given data, our modeling package was able to detect that and a special case was used instead. The correlations between its parameter estimates in the different versions were quite reasonable. This fact was reflected through its accuracy in MLE of its parameters. More importantly, we put an end to all previous arguments that mentioned the complications of the implementation of the GG MLE. We can confidently say the GG distribution is one where MLE is perfectly manageable.

Using the GG distribution for PQR proved to be of great usefulness due to the various shapes the quantiles offered. Linear, monotone, unimodal and bimodal quantile shapes were observed targeting different skewness aspects in data sets. This was done by studying the three-, four-, five- and six-parameter GG obtained by making the parameters dependent on a covariate. This problem was approached in an original way that only very few researchers have considered before, yet with little similarity and far less detailness. The overall modeling package linked the different parts of the thesis together, making it more coherent. This made modeling data sets using the GG very approachable with a clear starting point and a concluding

evaluation of the goodness-of-fit of the model.

Our suggested alternative distribution to the GG, that was analysed computationally and compared with the GG, showed interesting results. The results did not perfectly support the hypothesis that the AGG distribution will perform better because each of α and β influence separately the respective regions close to zero and close to infinity. Any practical benefits that might be forthcoming by using the AGG distribution in place of the GG distribution are minor. This issue stressed the fact that the GG is a leader distribution in this respect. The AGG was the simplest three-parameter distribution we could think of that had very similar properties to the GG, yet associated separate tasks to its parameters. Even this well-developed least complicated version did not manage to overcome the advantages of using the GG distribution.

8.4 Further Work

A doctoral thesis is a study that never really ends. In fact, research generally is a continuous process where one idea brings forth another. Hence, every conclusion is the beginning of another new research. Our study is, therefore, an inspirational work to further ideas of which we mention

- Applying our methodology to more data sets for further evaluation of our procedure.
- Conducting a simulation study that provides insight about sample dimension adequacy for the asymptotics to be reached.
- Attempting improvements on our suggested GG goodness-of-fit test by splitting up the region considered by the covariate as well as the quantiles.
- Considering nonparametric approaches to QR using the GG distribution, such as the methods available in the QUANTREG **R** package and comparing with our approach.
- Exploring more distributions that can model skewness.

- Using QR to the benefit of other fields such as modeling outbreaks of diseases and developing risk thresholds for finance and banking.

- Analysing QR problems when more than one covariate is involved.

The mentioned points are a few of many further directions to our study. We look back at the first question in the introduction of the thesis in Chapter 1: What best shape or model can be attributed to the data?

We may confidently say that our study proved the GG distribution in a QR context to be an efficient and flexible model that offers a variety of nice shapes which address different aspects of data, especially skewness. It is a powerful combination of simpler life distributions with an additional property that allows it to model both distributional tails more accurately. Knowing that there is no perfect model, we aspire for more research on further models in the future.

Appendix A

More Tables

A.1 Tables of Simulation Study in Section 2.5

Table A.1: Table of "real parameters" from Section 2.5.

Data	μ	σ	k	θ	α	β
GG1	-0.170	4.869	1.787	0.019	0.275	0.154
GG2	0.237	1.920	0.379	3.990	0.321	0.847
GG3	1.966	2.091	1.243	4.304	0.533	0.429
GG4	0.384	6.046	0.705	8.641	0.139	0.197
GG5	0.598	1.077	0.557	2.911	0.693	1.244
GG6	-0.286	0.604	2.560	0.303	2.648	1.034
GG7	0.500	0.924	0.808	1.967	0.973	1.204
GG8	-0.413	1.814	0.371	1.979	0.336	0.905
GG9	-0.235	1.749	2.610	0.053	0.924	0.354
GG10	0.911	1.466	5.329	0.009	1.574	0.295
GG11	0.333	2.281	3.239	0.011	0.789	0.244
GG12	-1.089	3.188	3.374	0.000	0.576	0.171
GG13	-0.731	1.146	1.781	0.199	1.165	0.654
GG14	0.627	1.680	1.774	0.519	0.793	0.447
GG15	-1.159	1.802	1.623	0.103	0.707	0.435

Table A.1 Continued

Data	μ	σ	k	θ	α	β
GG16	-0.234	2.415	1.902	0.093	0.571	0.300
GG17	-0.410	1.718	0.743	1.030	0.502	0.675
GG18	0.822	2.290	2.249	0.141	0.655	0.291
GG19	0.260	2.284	0.617	3.087	0.344	0.558
GG20	-2.617	2.217	0.169	0.369	0.185	1.097
GG21	-0.093	0.610	2.356	0.408	2.516	1.068
GG22	1.656	1.195	1.699	2.296	1.091	0.642
GG23	0.536	1.373	3.741	0.051	1.409	0.377
GG24	0.959	1.021	3.166	0.321	1.742	0.550
GG25	-0.083	4.340	2.575	0.001	0.370	0.144
GG26	0.355	0.614	1.356	1.148	1.895	1.398
GG27	0.145	4.967	1.519	0.090	0.248	0.163
GG28	-0.077	1.755	0.487	2.234	0.398	0.817
GG29	-1.505	3.498	2.685	0.001	0.468	0.174
GG30	-1.804	1.757	2.062	0.027	0.817	0.396
GG31	-1.455	0.955	0.886	0.260	0.986	1.113
GG32	-0.280	1.293	0.355	1.678	0.461	1.297
GG33	-2.776	0.863	1.030	0.061	1.176	1.142
GG34	0.747	6.230	1.662	0.036	0.207	0.125
GG35	-1.118	1.307	3.112	0.024	1.350	0.434
GG36	0.365	3.094	2.702	0.009	0.531	0.197
GG37	-0.556	2.646	4.860	0.000	0.833	0.171
GG38	0.399	2.770	1.649	0.252	0.464	0.281
GG39	0.660	0.646	0.875	2.097	1.447	1.654
GG40	1.042	2.239	1.729	0.566	0.587	0.340
GG41	-1.316	0.400	0.217	0.357	1.166	5.360

Table A.1 Continued

Data	μ	σ	k	θ	α	β
GG42	-0.534	3.646	3.046	0.000	0.479	0.157
GG43	0.180	1.541	1.852	0.329	0.883	0.477
GG44	-0.550	3.128	1.023	0.538	0.323	0.316
GG45	1.112	3.516	1.611	0.362	0.361	0.224
GG46	-0.402	1.589	3.633	0.013	1.199	0.330
GG47	0.201	0.811	1.753	0.669	1.632	0.931
GG48	-0.817	1.705	0.733	0.696	0.502	0.685
GG49	0.592	2.544	4.685	0.000	0.851	0.182
GG50	-0.533	1.677	0.819	0.795	0.540	0.659
GG51	-1.534	0.427	1.875	0.149	3.209	1.711
GG52	0.777	2.306	3.542	0.009	0.816	0.230
GG53	0.686	2.310	2.179	0.139	0.639	0.293
GG54	1.255	2.028	1.197	2.355	0.540	0.451
GG55	0.405	0.191	5.178	0.734	11.912	2.301
GG56	1.385	2.544	5.593	0.000	0.930	0.166
GG57	1.356	0.981	0.622	5.600	0.804	1.292
GG58	0.537	2.598	1.116	1.265	0.407	0.364
GG59	-1.126	4.790	0.878	0.581	0.196	0.223
GG60	-0.804	0.448	1.376	0.378	2.615	1.901
GG61	0.521	0.872	2.792	0.377	1.916	0.686
GG62	1.188	3.292	0.202	34.950	0.137	0.675
GG63	-1.456	5.582	0.769	0.842	0.157	0.204
GG64	-0.573	2.486	1.381	0.220	0.473	0.342
GG65	-1.248	3.174	4.791	0.000	0.690	0.144
GG66	-1.219	1.209	1.363	0.191	0.965	0.708
GG67	-0.436	4.688	0.543	5.325	0.157	0.289

Table A.1 Continued

Data	μ	σ	k	θ	α	β
GG68	-0.098	2.238	2.337	0.050	0.683	0.292
GG69	-0.471	2.281	1.302	0.314	0.500	0.384
GG70	0.028	1.193	1.507	0.564	1.029	0.683
GG71	-0.585	1.730	0.130	1.989	0.208	1.604
GG72	1.622	2.342	1.836	0.735	0.579	0.315
GG73	-0.720	1.692	2.864	0.024	1.000	0.349
GG74	-0.808	4.285	3.583	0.000	0.442	0.123
GG75	-0.705	0.812	4.405	0.039	2.585	0.587
GG76	-2.077	3.993	0.872	0.209	0.234	0.268
GG77	0.264	0.645	1.229	1.124	1.720	1.399
GG78	-0.636	3.057	3.119	0.001	0.578	0.185
GG79	-0.072	1.500	3.294	0.036	1.210	0.367
GG80	0.638	0.930	1.839	0.878	1.459	0.793
GG81	-1.073	1.169	1.641	0.163	1.096	0.668
GG82	-0.024	3.607	0.526	5.242	0.201	0.382
GG83	1.376	1.830	3.114	0.101	0.964	0.310
GG84	-0.105	1.137	2.750	0.134	1.459	0.530
GG85	-0.035	1.020	1.204	0.784	1.076	0.893
GG86	-0.807	0.193	0.980	0.448	5.139	5.246
GG87	-0.397	1.952	2.852	0.021	0.865	0.303
GG88	0.474	1.801	1.910	0.321	0.768	0.402
GG89	-1.942	1.913	2.400	0.011	0.810	0.337
GG90	1.085	3.398	4.958	0.000	0.655	0.132
GG91	-0.891	1.339	7.197	0.000	2.004	0.278
GG92	2.091	1.281	0.925	8.904	0.751	0.812
GG93	-0.730	1.966	2.513	0.027	0.806	0.321

Table A.1 Continued

Data	μ	σ	k	θ	α	β
GG94	-0.746	3.413	1.090	0.349	0.306	0.281
GG95	0.641	8.109	1.692	0.007	0.160	0.095
GG96	1.662	4.863	0.865	10.137	0.191	0.221
GG97	1.464	1.648	1.247	2.881	0.677	0.543
GG98	-0.041	0.326	0.527	1.117	2.225	4.225
GG99	-1.231	0.173	0.512	0.317	4.125	8.058
GG100	0.186	0.888	0.337	2.109	0.654	1.939

Table A.2: Table of "initial parameters" from Section 2.5.

Initial value	μ_0	σ_0	k_0	θ_0	α_0	β_0
1	-2.400	2.408	0.929	0.108	0.400	0.431
2	1.158	0.509	1.729	2.206	2.581	1.493
3	-0.923	1.268	1.615	0.183	1.003	0.621
4	0.869	3.298	0.767	5.125	0.266	0.346
5	-1.244	2.811	0.864	0.422	0.331	0.383
6	-0.898	2.409	1.347	0.177	0.482	0.358
7	-0.876	1.633	1.308	0.252	0.700	0.535
8	1.630	4.524	1.299	1.324	0.252	0.194
9	1.258	1.212	4.351	0.085	1.721	0.396
10	-1.193	6.121	0.604	3.343	0.127	0.210
11	-0.805	1.157	0.337	0.929	0.502	1.489
12	-0.608	1.177	2.613	0.088	1.374	0.526
13	-1.549	1.341	2.198	0.044	1.105	0.503
14	0.958	5.356	1.246	0.699	0.208	0.167
15	-0.074	2.510	4.465	0.000	0.842	0.189
16	-0.663	0.014	0.490	0.519	49.110	100.140
17	-0.030	1.143	1.501	0.550	1.072	0.714
18	-0.238	3.953	2.937	0.001	0.433	0.148
19	-0.532	3.314	2.998	0.001	0.522	0.174
20	-0.303	1.033	1.084	0.678	1.008	0.930
21	1.671	0.340	0.259	6.720	1.498	5.772
22	0.216	1.558	2.797	0.085	1.073	0.384
23	-1.043	1.275	1.428	0.205	0.938	0.657
24	1.248	5.694	1.977	0.015	0.247	0.125
25	-0.511	1.948	1.725	0.149	0.674	0.391
26	1.653	1.036	4.682	0.164	2.088	0.446

Table A.2 Continued

Initial value	μ_0	σ_0	k_0	θ_0	α_0	β_0
27	-0.385	1.762	1.021	0.656	0.573	0.562
28	-0.863	0.846	0.236	0.764	0.574	2.433
29	2.481	2.698	1.533	2.866	0.459	0.299
30	0.762	0.997	0.433	3.710	0.660	1.524
31	-0.521	2.100	5.218	0.000	1.088	0.208
32	-0.242	1.351	0.861	0.946	0.687	0.7983
33	-1.001	0.753	1.642	0.228	1.702	1.0379
34	-1.341	2.153	3.437	0.002	0.861	0.2500
35	-0.848	2.566	1.449	0.136	0.469	0.3248
36	1.450	0.470	0.703	4.895	1.785	2.5381
37	0.832	1.885	3.610	0.023	1.008	0.2796
38	1.290	1.132	0.628	5.512	0.700	1.1156
39	0.328	2.325	0.300	6.427	0.236	0.7851
40	0.056	1.130	10.231	0.000	2.830	0.2776
41	-0.243	1.202	4.654	0.015	1.795	0.3865
42	0.277	0.976	1.037	1.272	1.043	1.0064
43	-0.320	1.982	3.059	0.015	0.882	0.2889
44	1.264	3.561	3.916	0.000	0.556	0.1426
45	-0.207	1.111	6.798	0.003	2.346	0.3457
46	-1.250	2.354	1.345	0.128	0.493	0.3665
47	-1.604	0.471	2.114	0.120	3.089	1.4617
48	1.523	4.830	5.311	0.000	0.477	0.0909
49	0.436	2.096	5.360	0.000	1.105	0.2064
50	0.731	2.889	2.497	0.032	0.547	0.2191
51	-1.074	7.303	2.135	0.000	0.200	0.0944
52	-1.136	1.718	2.516	0.026	0.923	0.3673

Table A.2 Continued

Initial value	μ_0	σ_0	k_0	θ_0	α_0	β_0
53	-1.496	0.831	1.792	0.117	1.612	0.8991
54	-0.041	1.855	1.153	0.723	0.579	0.5029
55	0.153	0.843	2.941	0.245	2.035	0.6920
56	1.092	1.865	3.451	0.041	0.996	0.2898
57	0.368	0.651	0.549	1.930	1.138	2.0732
58	0.603	3.898	1.289	0.593	0.291	0.2267
59	-2.047	3.270	5.066	0.000	0.688	0.1362
60	0.010	0.620	1.555	0.718	2.011	1.2937
61	-1.496	2.696	1.386	0.079	0.437	0.3159
62	-0.623	2.709	1.495	0.142	0.451	0.3021
63	-0.461	4.592	2.268	0.002	0.328	0.1457
64	-0.209	4.289	0.583	4.747	0.178	0.3059
65	1.585	1.377	0.907	5.546	0.691	0.7629
66	0.993	0.479	2.018	1.674	2.966	1.4705
67	-0.706	0.542	2.311	0.247	2.803	1.213
68	-1.573	1.973	0.289	0.774	0.272	0.943
69	-1.167	2.681	0.671	0.748	0.305	0.456
70	-0.506	1.172	4.541	0.014	1.819	0.400
71	-0.632	1.292	1.586	0.251	0.975	0.614
72	-1.061	3.103	0.836	0.575	0.295	0.352
73	-0.073	0.494	5.837	0.113	4.894	0.838
74	-0.150	1.387	4.465	0.011	1.523	0.341
75	-0.632	0.375	1.756	0.401	3.531	2.010
76	1.983	2.311	3.861	0.016	0.850	0.220
77	-0.118	2.110	6.308	0.000	1.191	0.189
78	1.337	0.649	1.704	2.426	2.013	1.181

Table A.2 Continued

Initial value	μ_0	σ_0	k_0	θ_0	α_0	β_0
79	1.926	0.636	2.470	2.779	2.471	1.000
80	-0.005	3.364	3.449	0.000	0.552	0.160
81	-0.635	1.789	0.299	1.728	0.306	1.023
82	-1.905	2.484	3.203	0.001	0.720	0.225
83	-0.832	2.957	1.665	0.062	0.436	0.262
84	-0.118	1.658	5.414	0.001	1.403	0.259
85	-0.543	2.438	3.035	0.005	0.715	0.235
86	-0.277	3.532	2.253	0.010	0.425	0.189
87	-1.328	1.106	3.203	0.026	1.618	0.505
88	-0.297	1.053	1.785	0.329	1.269	0.711
89	-1.528	0.453	1.334	0.187	2.551	1.913
90	-0.656	1.705	1.008	0.512	0.589	0.584
91	-1.351	0.419	1.257	0.233	2.674	2.127
92	-0.674	3.373	2.759	0.002	0.493	0.179
93	0.490	0.471	3.962	0.448	4.222	1.066
94	0.845	1.116	0.983	2.370	0.889	0.904
95	-0.940	0.460	0.609	0.467	1.696	2.784
96	-1.952	0.804	1.071	0.134	1.288	1.202
97	0.643	0.642	0.787	2.180	1.381	1.754
98	0.347	3.732	6.222	0.000	0.668	0.107
99	-0.163	0.243	2.154	0.646	6.048	2.808
100	-0.251	1.916	3.041	0.019	0.910	0.299

Table A.3: Number of times each method fails to reach the maximal likelihood for $n = 500$.

Data	Likemax	$\hat{\mu}$	$\hat{\sigma}$	$\hat{k}$	Our Progam	BFGS	NM
GG1	-1536.441	-0.397	4.906	2.603	1	15	2
GG2	-1181.214	0.118	1.904	0.478	1	1	1
GG3	-1121.897	2.240	1.888	0.820	1	5	1
GG4	-1737.515	-0.439	6.323	0.722	3	24	3
GG5	-880.550	0.614	1.104	0.615	1	2	1
GG6	-480.523	-0.293	0.601	3.223	1	2	1
GG7	-760.787	0.585	0.902	0.746	2	7	1
GG8	-1157.147	-0.154	1.645	0.330	1	1	1
GG9	-996.975	0.009	1.571	1.305	0	3	1
GG10	-902.135	0.927	1.403	3.546	1	12	1
GG11	-1142.043	0.204	2.316	6.579	3	18	1
GG12	-1274.876	-1.001	2.908	2.612	1	8	1
GG13	-811.702	-0.657	1.088	1.342	2	2	1
GG14	-1000.820	0.528	1.667	2.289	2	4	1
GG15	-1036.757	-1.200	1.782	2.135	1	3	1
GG16	-1156.168	-0.050	2.211	1.630	2	12	1
GG17	-1070.604	-0.125	1.543	0.501	1	2	1
GG18	-1164.577	0.973	2.292	2.035	1	21	1
GG19	-1236.591	-0.162	2.483	1.102	1	4	1
GG20	-1436.673	-3.071	2.518	0.216	3	1	3
GG21	-504.875	-0.078	0.593	1.429	1	0	1
GG22	-822.199	1.604	1.150	1.916	1	7	1
GG23	-889.256	0.514	1.400	7.104	1	9	1
GG24	-746.157	0.933	1.040	4.889	1	5	1
GG25	-1450.189	0.102	3.984	1.645	2	11	1
GG26	-551.443	0.330	0.642	1.263	1	2	1

Table A.3 Continued

Data	Likemax	$\hat{\mu}$	$\hat{\sigma}$	$\hat{k}$	Our Progam	BFGS	NM
GG27	-1546.648	0.060	5.014	2.660	2	20	2
GG28	-1133.804	-0.330	1.811	0.582	1	7	1
GG29	-1335.270	-1.121	3.282	2.614	1	17	1
GG30	-1058.496	-1.751	1.821	1.657	1	4	1
GG31	-775.005	-1.476	0.961	0.919	1	0	1
GG32	-1023.554	-0.279	1.303	0.372	1	3	1
GG33	-712.502	-2.830	0.850	0.932	1	1	1
GG34	-1669.130	-0.009	6.416	2.733	1	34	8
GG35	-863.488	-1.238	1.310	4.371	2	5	1
GG36	-1302.941	0.156	3.121	3.402	0	9	1
GG37	-1194.321	-0.229	2.489	2.854	1	8	1
GG38	-1282.329	0.211	2.859	1.718	1	12	1
GG39	-580.381	0.640	0.660	1.004	1	3	1
GG40	-1169.161	0.866	2.345	2.463	1	18	1
GG41	-494.301	-1.254	0.349	0.167	1	0	0
GG42	-1371.153	-0.733	3.580	3.454	1	19	1
GG43	-973.959	0.342	1.461	1.062	0	3	1
GG44	-1322.295	-0.510	3.027	1.370	2	4	1
GG45	-1382.708	0.753	3.581	2.321	2	17	1
GG46	-975.819	-0.463	1.629	3.696	1	3	1
GG47	-650.099	0.200	0.806	1.675	1	4	1
GG48	-1062.999	-0.751	1.589	0.612	1	4	1
GG49	-1176.741	0.556	2.467	5.248	2	18	2
GG50	-1035.735	-0.701	1.628	0.953	1	8	1
GG51	-332.759	-1.528	0.436	2.171	1	0	0
GG52	-1131.288	0.782	2.242	4.572	1	18	1

Table A.3 Continued

Data	Likemax	$\hat{\mu}$	$\hat{\sigma}$	$\hat{k}$	Our Progam	BFGS	NM
GG53	-1167.494	0.660	2.283	1.807	2	13	1
GG54	-1123.347	1.499	1.981	1.109	1	8	1
GG55	109.096	0.386	0.187	4.492	3	6	1
GG56	-1185.208	1.345	6.243	2.521	2	23	2
GG57	-828.955	1.310	1.015	0.678	1	1	1
GG58	-1256.115	0.591	2.578	1.089	0	6	1
GG59	-1544.157	-1.309	4.494	0.945	1	4	1
GG60	-363.085	-0.770	0.432	1.093	1	0	1
GG61	-664.774	0.451	0.888	5.646	1	5	1
GG62	-1582.361	1.556	3.004	0.157	1	3	1
GG63	-1646.145	-1.242	5.197	0.672	3	20	2
GG64	-1208.395	-0.482	2.375	1.209	1	2	1
GG65	-1294.619	-1.244	3.100	4.275	2	14	1
GG66	-866.334	-1.201	1.196	1.192	1	1	1
GG67	-1615.649	-0.205	4.415	0.427	2	16	1
GG68	-1154.335	0.055	2.184	1.498	1	8	1
GG69	-1174.977	-0.502	2.264	1.421	1	4	1
GG70	-863.170	-0.018	1.195	1.247	1	1	1
GG71	-1342.239	-0.193	1.368	0.074	4	2	3
GG72	-1188.295	1.727	2.344	1.536	1	5	1
GG73	-1008.780	-0.701	1.701	2.456	1	3	1
GG74	-1454.763	-1.009	4.246	3.717	1	15	2
GG75	-619.711	-0.718	0.807	4.700	0	2	1
GG76	-1493.220	-2.386	4.137	1.080	3	7	1
GG77	-571.003	0.220	0.678	1.455	1	2	1
GG78	-1272.677	-0.739	2.939	3.424	2	9	1

Table A.3 Continued

Data	Likemax	$\hat{\mu}$	$\hat{\sigma}$	$\hat{k}$	Our Progam	BFGS	NM
GG79	-967.506	-0.222	1.560	2.315	0	3	1
GG80	-723.384	0.607	0.934	1.708	2	5	1
GG81	-830.841	-1.114	1.211	3.205	1	5	1
GG82	-1476.033	0.144	3.513	0.526	1	7	1
GG83	-1031.432	1.191	1.866	8.202	2	12	1
GG84	-800.151	0.009	1.078	1.525	0	3	1
GG85	-801.581	-0.023	1.077	1.478	2	4	1
GG86	31.591	-0.806	0.192	0.941	1	0	0
GG87	-1080.230	-0.531	1.988	3.043	2	10	1
GG88	-1053.781	0.672	1.792	1.543	1	5	1
GG89	-1069.627	-2.005	1.910	2.244	1	5	1
GG90	-1322.735	1.061	3.291	4.700	2	27	2
GG91	-863.172	-0.899	1.336	9.281	2	7	1
GG92	-921.673	2.099	1.293	0.940	1	7	1
GG93	-1100.892	-0.679	2.033	2.250	1	9	1
GG94	-1402.308	-0.594	3.406	0.988	1	10	1
GG95	-1811.848	1.427	8.143	1.510	1	46	9
GG96	-1616.439	1.588	5.195	0.947	1	26	4
GG97	-996.703	1.327	1.605	1.609	1	7	1
GG98	-290.158	-0.036	0.332	0.560	1	0	1
GG99	7.368	-1.254	0.187	0.618	1	0	0
GG100	-843.262	0.344	0.804	0.248	1	1	1
SUM	-	-	-	-	129	799	127

A.2 Tables of Simulation Study in Section 4.2

Table A.4: Table of "real parameters" from Section 4.2.

Data	a	b	σ	k
GGQ1	-0.671	1.195	2.133	2.453
GGQ2	1.128	-1.716	0.532	1.036
GGQ3	0.063	0.758	1.507	2.572
GGQ4	0.223	0.747	1.318	6.317
GGQ5	-1.680	-0.047	4.036	5.666
GGQ6	1.519	0.639	1.621	2.850
GGQ7	-0.797	0.800	1.357	0.465
GGQ8	0.268	0.957	1.699	2.842
GGQ9	-2.073	-0.129	6.439	0.933
GGQ10	1.294	-0.358	0.549	0.275
GGQ11	0.980	-1.852	1.834	1.203
GGQ12	-0.417	-0.237	2.068	2.236
GGQ13	-0.165	-1.859	0.860	0.711
GGQ14	1.127	-0.120	1.504	1.124
GGQ15	-1.347	0.235	1.525	1.453
GGQ16	0.605	1.913	3.302	4.096
GGQ17	0.703	-0.306	4.979	5.363
GGQ18	-0.776	0.338	3.914	2.097
GGQ19	-2.047	0.950	0.474	2.630
GGQ20	-0.880	-0.954	4.026	1.879
GGQ21	-0.356	0.897	4.268	4.656
GGQ22	-1.016	2.418	1.662	0.521
GGQ23	-1.784	0.918	0.880	1.536
GGQ24	0.949	1.505	0.262	3.794

Table A.4 Continued

Data	a	b	σ	k
GGQ25	-0.534	-0.032	1.614	0.904
GGQ26	-0.013	-0.680	0.539	2.302
GGQ27	0.571	1.763	1.710	1.805
GGQ28	-0.613	0.876	1.369	1.093
GGQ29	0.837	0.685	0.634	1.074
GGQ30	-0.107	0.489	1.202	2.392
GGQ31	0.169	-1.430	3.278	0.729
GGQ32	0.113	-0.897	1.395	1.736
GGQ33	1.066	0.166	0.473	3.211
GGQ34	-0.738	0.482	2.503	0.974
GGQ35	-1.314	0.390	0.864	3.202
GGQ36	-0.566	1.130	1.673	1.096
GGQ37	2.794	0.412	0.184	0.787
GGQ38	-0.814	-0.343	2.132	0.378
GGQ39	0.592	1.166	0.410	2.876
GGQ40	1.137	-0.929	0.383	0.823
GGQ41	0.271	0.196	2.578	1.739
GGQ42	-0.711	0.236	1.167	0.905
GGQ43	0.325	-0.546	3.816	1.248
GGQ44	0.445	1.314	2.161	2.561
GGQ45	-0.878	1.762	4.104	0.490
GGQ46	0.671	0.830	1.173	4.093
GGQ47	1.366	-0.276	0.756	1.530
GGQ48	0.982	0.839	0.790	0.687
GGQ49	0.651	0.249	0.459	3.394
GGQ50	1.767	0.832	2.104	0.444

Table A.4 Continued

Data	a	b	σ	k
GGQ51	-0.095	0.139	0.803	2.448
GGQ52	-1.859	1.716	0.437	1.372
GGQ53	0.238	-0.370	0.757	2.269
GGQ54	-1.555	-0.990	1.543	5.086
GGQ55	-0.199	0.717	2.056	0.799
GGQ56	0.781	-1.438	2.666	5.002
GGQ57	0.466	0.756	0.912	2.491
GGQ58	-0.434	-0.807	1.677	1.422
GGQ59	1.097	-0.025	2.668	2.113
GGQ60	0.699	0.233	3.365	4.446
GGQ61	-0.596	-0.465	0.953	4.488
GGQ62	-0.758	-1.840	1.491	5.788
GGQ63	1.219	-0.271	1.719	2.471
GGQ64	-0.488	-0.528	0.756	3.914
GGQ65	-1.733	0.359	2.762	2.134
GGQ66	1.035	-0.746	1.559	0.352
GGQ67	-1.217	1.683	1.739	3.436
GGQ68	1.530	-0.416	2.490	1.485
GGQ69	0.559	-0.103	1.279	0.519
GGQ70	-0.139	-0.253	0.986	5.160
GGQ71	0.044	1.111	0.525	1.914
GGQ72	0.735	-0.548	1.228	2.965
GGQ73	-1.565	-1.223	5.415	0.799
GGQ74	-0.150	0.344	1.097	2.469
GGQ75	0.043	0.460	3.268	3.717
GGQ76	0.341	0.377	3.813	3.185

Table A.4 Continued

Data	a	b	σ	k
GGQ77	-0.889	-1.536	0.129	1.002
GGQ78	0.051	0.266	1.601	2.478
GGQ79	1.075	1.545	2.839	2.886
GGQ80	-0.584	-0.002	2.452	1.020
GGQ81	-0.826	2.011	0.854	1.974
GGQ82	-0.447	-1.037	2.571	2.530
GGQ83	-0.023	-1.398	1.572	2.559
GGQ84	0.073	-0.118	2.885	0.967
GGQ85	1.057	-0.905	4.085	2.666
GGQ86	0.917	1.249	5.563	2.416
GGQ87	-0.403	-1.447	1.637	3.180
GGQ88	-0.183	-1.392	6.037	1.286
GGQ89	1.666	0.713	0.072	2.070
GGQ90	-1.541	-1.280	2.076	1.372
GGQ91	1.635	0.160	3.785	3.057
GGQ92	-1.499	-1.084	2.634	0.601
GGQ93	1.797	-1.228	3.095	1.613
GGQ94	-1.179	-1.149	3.123	2.101
GGQ95	1.009	0.313	5.900	2.806
GGQ96	-0.322	0.439	1.628	2.631
GGQ97	2.231	0.036	1.362	2.374
GGQ98	-0.878	-0.075	0.993	1.026
GGQ99	0.352	1.116	1.096	5.293
GGQ100	0.207	-0.772	2.723	0.984

Table A.5: Table of "initial parameters" from Section 4.2.

Initial value	a_0	b_0	σ_0	k_0
1	0.213	-0.076	0.176	0.985
2	-1.093	0.293	0.849	1.777
3	0.190	-0.062	3.179	0.817
4	0.751	1.820	2.222	0.950
5	1.801	0.320	4.820	3.213
6	-0.483	-0.356	0.743	0.520
7	-1.030	1.040	0.407	4.977
8	-0.477	-0.142	1.773	6.200
9	0.387	0.491	0.413	1.584
10	0.449	-0.502	2.965	3.861
11	0.998	0.482	4.041	2.383
12	-0.009	-1.547	2.606	0.652
13	0.452	1.001	3.364	2.250
14	0.644	0.930	1.875	1.985
15	0.663	-2.854	1.626	0.854
16	-0.074	0.327	3.925	2.444
17	-1.675	1.827	4.009	4.555
18	-1.914	-0.180	1.428	0.745
19	0.089	-0.721	0.883	2.437
20	1.055	-0.707	2.284	3.510
21	0.004	0.734	2.669	2.970
22	-0.500	0.485	0.558	2.387
23	-1.042	-1.514	2.167	0.959
24	-0.660	0.511	1.330	1.737
25	0.581	-0.363	1.948	2.028
26	2.307	-0.920	9.884	0.981

Table A.5 Continued

Initial value	a_0	b_0	σ_0	k_0
27	0.581	-0.148	2.302	3.408
28	0.404	0.843	3.520	1.437
29	0.273	2.197	2.342	3.331
30	0.071	1.255	1.037	1.131
31	-1.013	-1.003	0.850	7.203
32	-0.351	-0.137	1.570	1.218
33	2.202	1.922	2.674	3.040
34	0.603	-2.119	2.403	1.663
35	-0.699	0.334	1.205	1.846
36	-0.365	-0.848	0.692	1.716
37	0.662	0.793	4.423	2.378
38	-0.122	-0.483	0.802	0.276
39	-2.199	-0.945	3.902	3.763
40	0.093	-0.145	0.552	4.222
41	-0.638	-0.636	2.546	2.105
42	1.153	-2.038	0.416	2.079
43	1.203	-1.484	0.890	2.118
44	-1.049	-0.001	2.982	2.536
45	-0.776	2.052	2.611	4.354
46	1.695	0.572	1.665	0.385
47	-0.455	0.622	0.117	0.329
48	0.495	1.293	1.939	2.167
49	2.182	0.387	1.668	3.078
50	0.695	0.238	1.575	1.126
51	0.237	-2.103	2.957	1.615
52	-0.055	0.103	4.193	2.234

Table A.5 Continued

Initial value	a_0	b_0	σ_0	k_0
53	0.482	0.291	1.275	1.678
54	-0.605	0.410	1.803	0.909
55	-0.200	0.012	0.510	1.414
56	-1.044	-0.382	1.146	1.784
57	1.734	-1.011	2.967	1.731
58	-2.140	0.572	6.465	3.318
59	0.464	1.671	2.791	1.873
60	0.273	0.568	1.938	0.464
61	0.648	0.334	3.910	1.373
62	-1.884	0.506	0.650	1.541
63	-1.501	-0.491	2.539	1.507
64	-1.767	0.236	0.789	1.467
65	-1.194	0.129	7.150	4.792
66	-0.406	0.619	0.863	2.281
67	-1.637	0.844	1.641	6.222
68	-0.080	-0.525	1.216	0.319
69	-0.050	-0.307	3.653	2.639
70	1.661	-1.007	2.211	1.141
71	-2.144	-0.116	1.592	3.595
72	0.916	-0.245	0.587	0.579
73	2.548	1.116	3.364	1.656
74	-1.903	0.062	2.524	4.459
75	0.087	-0.269	0.259	2.458
76	0.631	-0.553	1.299	5.144
77	-1.455	0.383	0.889	1.130
78	-0.712	0.780	2.369	0.798

Table A.5 Continued

Initial value	a_0	b_0	σ_0	k_0
79	-1.781	-1.075	0.552	3.110
80	-0.619	-1.004	2.125	2.855
81	1.497	1.409	2.319	0.799
82	2.035	-1.167	1.110	1.323
83	0.015	-0.939	1.420	1.663
84	1.857	1.322	1.338	2.706
85	-0.062	1.048	3.108	3.292
86	-1.458	0.653	0.305	0.761
87	-0.159	-0.020	4.690	1.427
88	1.708	0.255	2.657	3.229
89	-0.287	-0.328	1.982	2.546
90	-0.493	-0.221	1.683	1.950
91	1.031	1.316	1.909	1.017
92	-1.227	1.350	1.868	2.026
93	2.061	-0.665	0.499	6.703
94	0.872	1.115	3.671	2.910
95	-0.731	0.006	2.137	0.270
96	0.651	-0.955	4.023	2.635
97	-0.399	-1.594	2.415	0.895
98	1.122	-0.700	0.848	1.524
99	0.655	0.415	0.486	3.753
100	1.134	1.966	2.164	0.767

Table A.6: Number of times each method fails to reach the maximal likelihood in the four-parameter case for $n = 500$.

Data	Likemax	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$	$\hat{k}$	Progbis	Prognr	BFGS	NM
GGQ1	-1136.811	-0.527	0.773	2.226	3.040	0	3	11	1
GGQ2	-453.047	1.169	-1.787	0.518	1.097	0	11	2	1
GGQ3	-956.597	-0.063	0.802	1.575	4.151	0	2	16	1
GGQ4	-859.130	0.039	1.127	1.335	15.713	1	4	16	1
GGQ5	-1402.789	-2.122	0.249	3.913	7.471	1	22	21	3
GGQ6	-958.702	1.393	0.785	1.551	2.786	1	7	18	1
GGQ7	-1035.826	-0.638	0.657	1.309	0.347	0	11	2	0
GGQ8	-1010.320	0.791	0.869	1.536	0.908	0	0	8	1
GGQ9	-1692.466	-3.180	0.706	6.168	1.086	3	40	12	3
GGQ10	-656.312	1.533	-0.480	0.471	0.157	0	2	1	1
GGQ11	-1094.673	0.454	-1.653	2.011	2.291	1	8	9	1
GGQ12	-1106.762	-0.521	-0.460	2.080	2.665	0	4	9	1
GGQ13	-734.360	-0.132	-1.955	0.860	0.769	1	29	0	0
GGQ14	-984.957	1.241	-0.330	1.531	1.288	0	0	6	1
GGQ15	-960.734	-1.294	0.128	1.462	1.317	0	6	2	0
GGQ16	-1341.533	0.404	1.834	3.464	7.708	0	3	30	3
GGQ17	-1502.134	0.790	-0.662	4.790	8.847	2	6	27	5
GGQ18	-1433.885	-1.091	1.086	3.948	2.178	0	13	11	3
GGQ19	-349.802	-1.949	0.830	0.449	1.996	1	22	1	0
GGQ20	-1457.474	-1.310	0.236	4.006	1.502	1	16	9	3
GGQ21	-1463.523	0.277	-0.037	4.268	2.898	0	5	9	3
GGQ22	-1090.882	-0.895	1.980	1.679	0.611	1	4	2	1
GGQ23	-677.044	-1.828	1.031	0.852	1.706	0	12	1	0
GGQ24	-78.992	0.927	1.511	0.268	3.005	2	29	10	1

Table A.6 Continued

Data	Likemax	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$	$\hat{k}$	Progbis	Prognr	BFGS	NM
GGQ25	-1053.134	-0.543	0.032	1.715	1.074	0	5	0	1
GGQ26	-438.993	0.099	-0.817	0.510	1.208	0	6	1	0
GGQ27	-1043.584	0.583	1.854	1.772	1.697	0	3	8	2
GGQ28	-939.365	-0.651	1.098	1.356	1.021	0	12	4	1
GGQ29	-551.041	0.909	0.593	0.614	0.915	1	8	5	1
GGQ30	-852.947	-0.298	0.864	1.236	2.192	2	6	7	1
GGQ31	-1398.997	0.037	-0.773	3.258	0.780	2	6	7	1
GGQ32	-917.035	-0.076	-0.613	1.353	1.435	0	3	3	1
GGQ33	-331.188	1.047	0.183	0.450	3.963	0	3	7	1
GGQ34	-1229.262	-0.856	0.487	2.415	1.002	1	7	4	1
GGQ35	-667.047	-1.172	0.275	0.836	1.728	0	5	0	0
GGQ36	-1026.644	-0.823	1.570	1.677	1.377	0	3	8	1
GGQ37	36.372	2.781	0.423	0.190	0.920	1	34	5	1
GGQ38	-1264.442	-0.750	-0.411	2.243	0.472	2	11	2	1
GGQ39	-300.523	0.584	1.182	0.410	2.250	0	19	8	1
GGQ40	-334.930	1.105	-0.948	0.403	0.986	0	8	2	1
GGQ41	-1232.126	0.135	0.666	2.530	1.383	1	6	3	1
GGQ42	-896.902	-0.596	-0.035	1.242	1.002	0	3	0	0
GGQ43	-1468.081	0.085	-0.604	4.011	1.254	0	8	7	3
GGQ44	-1117.375	0.368	0.923	2.167	3.910	0	0	17	2
GGQ45	-1554.908	-0.052	1.682	3.796	0.381	2	5	4	2
GGQ46	-808.004	0.605	0.951	1.141	2.533	0	0	5	1
GGQ47	-633.382	1.402	-0.307	0.740	1.067	0	4	1	1
GGQ48	-705.661	0.941	0.948	0.755	0.533	0	6	0	1
GGQ49	-349.877	0.624	0.317	0.469	4.320	0	1	4	1
GGQ50	-1257.028	1.829	0.748	2.192	0.457	0	2	0	2

Table A.6 Continued

Data	Likemax	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$	$\hat{k}$	Progbis	Prognr	BFGS	NM
GGQ51	-633.252	-0.085	0.062	0.802	2.428	2	3	3	0
GGQ52	-328.067	-1.849	1.755	0.400	1.041	1	19	1	0
GGQ53	-610.761	0.111	-0.144	0.782	3.446	2	4	3	0
GGQ54	-953.648	-1.618	-1.028	1.580	5.411	1	13	4	1
GGQ55	-1178.795	0.376	0.293	1.977	0.575	0	3	2	1
GGQ56	-1226.492	0.854	-1.549	2.740	6.369	1	3	20	3
GGQ57	-709.575	0.418	0.699	0.940	2.644	1	2	2	1
GGQ58	-1028.758	-0.530	-0.993	1.735	1.874	0	10	2	1
GGQ59	-1214.612	1.130	-0.184	2.548	2.192	0	1	11	2
GGQ60	-1345.589	0.689	0.133	3.419	3.856	1	3	15	3
GGQ61	-708.992	-0.460	-0.465	0.914	1.830	1	7	1	0
GGQ62	-929.627	-0.772	-1.878	1.502	4.943	0	15	3	0
GGQ63	-990.903	1.127	-0.103	1.628	2.177	0	2	16	1
GGQ64	-595.306	-0.499	-0.505	0.775	6.298	0	6	1	0
GGQ65	-1244.420	-1.566	-0.033	2.694	2.080	1	9	9	1
GGQ66	-1114.379	1.185	-1.042	1.505	0.326	0	5	1	1
GGQ67	-1015.339	-1.454	2.094	1.761	3.621	1	13	11	1
GGQ68	-1242.674	1.387	-0.358	2.694	2.183	0	3	9	3
GGQ69	-949.376	0.617	-0.262	1.252	0.581	0	2	2	1
GGQ70	-725.138	-0.052	-0.333	0.991	4.084	0	0	4	1
GGQ71	-427.401	0.073	1.095	0.515	1.623	0	6	0	0
GGQ72	-828.805	0.687	-0.452	1.214	3.688	0	0	13	1
GGQ73	-1647.717	-1.997	-0.908	5.590	1.019	3	22	11	3
GGQ74	-781.319	-0.173	0.305	1.109	4.134	0	1	3	1
GGQ75	-1307.765	0.195	0.271	3.168	3.814	0	3	25	2
GGQ76	-1410.755	0.416	-0.284	3.872	3.397	0	6	21	3

Table A.6 Continued

Data	Likemax	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$	$\hat{k}$	Progbis	Prognr	BFGS	NM
GGQ77	228.422	-0.900	-1.528	0.134	1.165	50	97	0	0
GGQ78	-970.028	-0.013	0.384	1.556	2.083	4	6	10	1
GGQ79	-1254.524	0.894	1.619	2.885	5.423	1	2	24	3
GGQ80	-1250.759	-0.973	0.128	2.577	1.180	0	5	5	1
GGQ81	-638.188	-0.943	2.175	0.822	3.079	1	5	3	1
GGQ82	-1180.910	-0.366	-1.591	2.456	3.748	0	7	9	1
GGQ83	-971.894	0.106	-1.567	1.569	2.220	0	10	7	1
GGQ84	-1348.306	-0.274	0.834	2.896	0.713	0	3	9	1
GGQ85	-1421.700	0.665	-0.926	3.995	4.214	1	4	22	3
GGQ86	-1641.202	1.626	0.392	5.861	1.716	0	2	15	4
GGQ87	-983.079	-0.481	-1.682	1.675	5.341	0	7	6	1
GGQ88	-1640.000	0.097	-2.112	5.829	1.659	2	8	13	5
GGQ89	599.510	1.671	0.709	0.067	1.869	60	97	9	1
GGQ90	-1116.125	-1.878	-0.805	2.084	2.079	1	24	1	1
GGQ91	-1383.553	2.134	-0.239	3.578	2.245	0	3	20	3
GGQ92	-1344.011	-1.346	-0.903	2.656	0.492	2	13	1	1
GGQ93	-1332.145	1.985	-1.864	3.103	1.436	0	5	11	3
GGQ94	-1320.367	-1.016	-1.576	3.171	2.439	1	24	9	2
GGQ95	-1618.645	1.432	0.242	5.617	1.767	0	0	11	4
GGQ96	-978.636	-0.174	0.505	1.553	1.666	0	4	9	1
GGQ97	-860.130	2.248	0.149	1.248	2.064	1	10	16	1
GGQ98	-785.183	-0.791	-0.172	0.998	1.037	0	6	3	0
GGQ99	-787.309	0.390	1.030	1.131	5.089	2	9	7	1
GGQ100	-1302.519	0.014	-1.029	2.947	1.546	0	4	8	1
SUM	-	-	-	-	-	168	949	766	136

A.3 Tables of Simulation Study in Section 5.1

Table A.7: Mean relative bias (biasmu, biassigma, biask, biasb, biasd, biasg) and root mean relative squared error (msemu, msesigma, msek, mseb, msed, mseg) of the estimated parameters of the three-parameter GG data sets for $n = 500$. In the case of b, d and g in the "LRT Version" 'relative' is replaced by 'actual'.

Known p Version				
	D31	D32	D33	D34
biasmu	-0.002	-0.002	0.001	-0.004
msemu	0.026	0.042	0.022	0.146
biassigma	-0.008	-0.004	-0.006	-0.001
msesigma	0.047	0.039	0.038	0.036
biask	0.004	0.097	0.250	0.147
msek	0.197	0.355	0.712	0.533
LRT Version				
biasmu	-0.003	0.000	0.001	-0.011
msemu	0.032	0.070	0.032	0.185
biassigma	-0.011	-0.009	-0.008	-0.004
msesigma	0.062	0.053	0.052	0.053
biask	0.025	0.156	0.878	0.288
msek	0.232	0.718	4.262	1.805
biasb	-0.002	-0.018	0.002	0.007
mseb	0.067	0.501	0.049	0.108
biasd	0.004	0.007	0.003	0.005
msed	0.104	0.079	0.074	0.071
biasg	-0.034	0.048	-0.027	0.021
mseg	0.304	0.726	1.026	0.614

Table A.8: Mean relative bias (biasa, biasb, biassigma, biask, biasd, biasg) and root mean relative squared error (msea, mseb, msesigma, msek, msed, mseg) of the estimated parameters of the four-parameter GG data sets for $n = 500$. In the case of d and g in the "LRT Version" 'relative' is replaced by 'actual'.

Known p Version				
	D41	D42	D43	D44
biasa	0.012	-0.001	-0.092	-0.022
msea	0.062	0.022	0.770	0.204
biasb	-0.020	-0.054	0.002	0.015
mseb	0.242	1.442	0.032	0.153
biassigma	-0.020	-0.009	-0.005	-0.003
msesigma	0.073	0.044	0.039	0.037
biask	-0.038	0.060	0.328	0.247
msek	0.191	0.321	1.016	0.952
LRT Version				
biasa	0.005	-0.006	-0.112	-0.016
msea	0.085	0.026	0.792	0.231
biasb	0.021	-0.620	0.004	0.012
mseb	0.442	1.821	0.037	0.180
biassigma	-0.005	-0.009	-0.001	-0.003
msesigma	0.115	0.064	0.053	0.056
biask	0.013	0.180	0.600	0.816
msek	0.366	0.949	3.030	4.010
biasd	-0.029	0.002	-0.008	0.001
msed	0.189	0.094	0.069	0.077
biasg	-0.056	0.015	-0.012	-0.066
mseg	0.469	0.653	0.895	0.844

Table A.9: Mean relative bias (biasa, biasb, biasc, biasd, biask, biasg) and root mean relative squared error (msea, mseb, msec, msed, msek, mseg) of the estimated parameters of the five-parameter GG data sets for $n = 500$. In the case of g in the "LRT Version" 'relative' is replaced by 'actual'.

Known p Version				
	D51	D52	D53	D54
biasa	0.000	-0.004	0.002	0.003
msea	0.002	0.027	0.015	0.141
biasb	-0.010	0.046	0.020	-0.003
mseb	0.048	0.220	0.171	0.104
biasc	-0.001	0.013	0.007	-0.021
msec	0.031	0.693	0.049	0.126
biasd	-0.022	0.010	-0.001	-0.003
msed	0.166	0.235	0.065	0.071
biask	0.057	0.122	0.016	0.169
msek	0.360	0.560	0.224	0.652
LRT Version				
biasa	0.000	-0.004	0.002	0.002
msea	0.002	0.031	0.015	0.142
biasb	-0.008	0.042	0.017	-0.002
mseb	0.050	0.270	0.178	0.104
biasc	0.000	0.036	0.006	-0.021
msec	0.030	0.741	0.051	0.127
biasd	-0.019	0.004	0.000	-0.003
msed	0.165	0.249	0.069	0.071
biask	0.042	0.282	0.034	0.172
msek	0.367	1.471	0.292	0.687
biasg	0.042	0.024	-0.011	0.013
mseg	0.297	0.804	0.362	0.248

Table A.10: Mean relative bias (biasa, biasb, biasc, biasd, biasf, biasg) and root mean relative squared error (msea, mseb, msec, msed, msef, mseg) of the estimated parameters of the six-parameter GG data sets for $n = 500$.

Known p Version				
	D61	D62	D63	D64
biasa	0.186	0.000	0.008	0.000
msea	0.512	0.008	0.141	0.000
biasb	-0.009	0.000	-0.004	0.002
mseb	0.066	0.023	0.256	0.038
biasc	-0.006	0.034	-0.210	-0.001
msec	0.068	0.154	0.712	0.015
biasd	-0.003	0.001	0.583	-0.011
msed	0.169	0.040	2.402	0.114
biasf	-0.066	-0.420	0.024	0.083
msef	0.333	1.543	0.140	0.334
biasg	-0.869	0.181	0.023	0.054
mseg	1.653	0.606	0.159	0.215
LRT Version				
biasa	0.078	0.000	0.008	0.000
msea	0.465	0.008	0.141	0.000
biasb	0.000	0.000	-0.004	0.002
mseb	0.061	0.023	0.256	0.038
biasc	-0.004	0.054	-0.210	-0.001
msec	0.065	0.164	0.712	0.015
biasd	-0.001	-0.008	0.583	-0.011
msed	0.163	0.045	2.402	0.114
biasf	0.043	0.062	0.024	0.083
msef	0.218	1.811	0.140	0.334
biasg	-0.998	-0.081	0.023	0.054
mseg	1.082	0.833	0.159	0.215

A.4 Tables of Simulation Study in Section 7.5

Table A.11: Table of "real parameters" for Section 7.5.

Data	θ	α	β
1	0.289	4.832	2.801
2	2.330	4.406	2.067
3	1.740	2.116	1.254
4	1.403	8.135	3.121
5	5.174	6.551	2.444
6	0.136	7.697	1.291
7	6.836	4.557	3.614
8	4.681	1.826	1.153
9	0.739	4.469	1.744
10	1.284	5.025	2.414
11	4.449	9.592	5.284
12	8.425	2.577	2.348
13	6.087	5.018	3.270
14	0.496	3.055	1.450
15	10.050	3.317	1.401
16	1.047	9.358	1.650
17	4.452	8.475	5.735
18	1.158	4.674	2.010
19	1.358	5.332	2.075
20	2.239	5.390	4.934
21	0.916	3.980	1.452
22	0.861	4.024	1.058
23	0.804	2.034	1.120
24	1.864	4.417	1.367

Table A.11 Continued

Data	θ	α	β
25	0.971	3.785	2.612
26	2.707	1.768	1.545
27	3.482	7.291	4.509
28	3.470	2.138	1.514
29	2.521	4.537	2.333
30	5.229	2.460	1.057
31	6.872	2.606	2.338
32	0.611	16.626	2.248
33	1.469	2.436	1.233
34	4.560	2.915	2.435
35	4.675	5.137	3.427
36	1.064	2.877	2.671
37	1.351	3.857	1.258
38	1.710	5.077	3.220
39	1.032	5.474	2.494
40	0.520	11.378	1.095
41	1.204	1.552	1.386
42	0.209	7.755	1.381
43	2.098	4.624	1.365
44	0.711	2.539	1.025
45	0.461	3.301	2.280
46	0.047	2.187	1.037
47	0.803	3.119	1.874
48	1.361	7.864	1.573
49	2.567	3.898	1.479
50	6.181	1.990	1.504

Table A.11 Continued

Data	θ	α	β
51	4.254	7.977	1.450
52	3.531	7.281	6.064
53	0.331	1.992	1.126
54	1.779	1.829	1.200
55	0.699	9.394	1.774
56	2.651	4.218	2.635
57	0.895	3.367	2.416
58	1.465	2.166	1.792
59	1.000	4.025	2.689
60	1.224	3.044	1.325
61	2.203	1.579	1.276
62	1.980	1.608	1.576
63	2.307	2.539	1.025
64	0.010	6.376	3.247
65	0.639	9.178	1.739
66	0.742	2.720	1.435
67	12.490	6.076	3.374
68	0.798	2.281	1.490
69	1.168	10.224	2.251
70	0.467	1.503	1.123
71	1.834	11.183	1.859
72	0.566	1.664	1.038
73	0.507	2.443	1.257
74	0.585	8.107	1.158
75	0.614	3.651	1.121
76	6.827	10.440	2.888

Table A.11 Continued

Data	θ	α	β
77	0.613	5.640	3.047
78	1.027	8.346	2.623
79	0.486	4.449	1.165
80	1.268	4.223	2.614
81	10.394	1.902	1.860
82	5.364	5.689	1.423
83	0.153	4.630	2.909
84	5.324	6.894	4.993
85	1.897	6.216	2.072
86	0.878	7.382	2.160
87	0.020	5.438	1.590
88	0.003	5.463	2.302
89	0.917	3.121	1.393
90	0.899	2.944	2.438
91	1.337	2.622	2.084
92	2.532	3.274	1.050
93	4.859	3.874	1.823
94	6.519	3.299	2.031
95	0.793	6.883	1.479
96	1.250	2.930	2.505
97	0.224	6.664	1.101
98	0.919	5.817	2.799
99	3.080	4.359	1.984
100	1.158	3.571	1.944

Table A.12: Table of AGG and GG parameter estimates.

	AGG				GG			
Data	Lik	θ	α	β	Lik	θ	α	β
1	482.054	0.280	4.814	2.664	468.470	0.282	4.955	2.694
2	-681.388	2.462	4.182	2.225	-741.842	3.022	3.571	2.630
3	-777.882	1.680	2.104	1.210	-849.819	2.006	2.011	1.393
4	-216.693	1.287	9.359	2.873	-228.795	1.414	8.173	3.105
5	-1020.756	4.112	7.796	1.907	-1059.850	5.663	5.808	2.525
6	396.457	0.218	5.660	1.733	166.666	0.256	5.619	1.627
7	-1027.212	7.184	4.219	3.876	-975.949	6.190	5.477	3.279
8	-1322.916	5.436	1.730	1.238	-1407.788	5.371	1.611	1.179
9	-190.741	0.717	4.756	1.742	-314.757	0.238	5.833	1.058
10	-304.403	1.101	6.057	2.104	-336.946	1.008	5.930	2.036
11	-634.429	4.495	8.960	5.137	-557.845	4.625	8.881	5.636
12	-1261.754	7.902	2.736	2.229	-1309.709	8.507	2.572	2.358
13	-971.722	6.075	4.926	3.370	-966.244	6.135	5.144	3.365
14	-70.559	0.631	2.759	1.873	-166.609	0.504	3.332	1.539
15	-1630.807	12.210	2.980	1.609	-1727.129	11.695	3.142	1.547
16	-444.419	1.031	9.163	1.616	-565.467	1.123	9.097	1.696
17	-602.898	4.132	9.613	4.803	-557.107	4.055	8.958	4.683
18	-395.599	1.418	3.460	2.280	-401.255	1.007	5.180	1.858
19	-415.543	1.163	6.255	1.815	-471.669	1.199	5.584	1.906
20	-365.135	2.188	5.661	4.520	-323.189	2.040	5.936	4.090
21	-433.405	1.059	3.402	1.561	-551.053	0.937	3.809	1.440
22	-609.191	1.086	3.484	1.225	-875.830	1.160	3.561	1.150
23	-453.038	0.867	2.116	1.201	-570.972	1.079	1.880	1.306
24	-818.409	2.243	4.072	1.563	-927.782	0.576	6.784	0.976
25	-145.761	0.977	3.675	2.697	-200.380	1.082	3.329	2.825

Table A.12 Continued

	AGG				GG			
Data	Lik	θ	α	β	Lik	θ	α	β
26	-856.593	2.838	1.574	1.578	-906.166	2.749	1.643	1.535
27	-577.938	3.821	6.024	5.796	-494.527	3.078	9.193	3.884
28	-969.181	2.607	2.714	1.321	-1054.541	2.620	2.563	1.351
29	-649.968	2.770	4.435	2.777	-726.698	2.750	3.938	2.448
30	-1422.743	0.167	43.518	0.452	-1646.035	4.335	2.525	0.964
31	-1155.521	7.067	2.622	2.586	-1228.272	7.720	2.318	2.554
32	36.916	0.577	17.863	2.138	-17.807	0.218	23.744	1.453
33	-687.264	1.668	2.238	1.407	-844.966	1.029	2.578	1.039
34	-958.448	4.652	2.694	2.487	-976.576	4.943	2.797	2.749
35	-835.875	4.816	4.826	3.613	-818.579	4.285	5.691	3.071
36	-172.758	0.999	3.176	2.600	-224.578	0.942	3.244	2.296
37	-704.081	1.454	3.938	1.335	-886.937	1.018	4.179	1.114
38	-318.967	1.530	6.138	2.923	-348.035	1.404	6.032	2.581
39	-177.542	1.095	5.137	2.796	-222.809	1.071	5.077	2.567
40	-463.337	0.274	19.128	0.866	-796.705	1.480	8.467	1.547
41	-474.889	1.128	1.723	1.375	-534.263	1.170	1.457	1.328
42	232.016	0.230	7.113	1.442	76.925	0.121	9.289	1.172
43	-902.535	2.875	3.578	1.654	-1031.438	2.339	4.503	1.439
44	-507.552	0.903	2.122	1.144	-652.472	0.222	3.550	0.766
45	149.491	0.441	3.216	2.071	122.356	0.531	2.992	2.628
46	862.011	0.042	2.330	0.955	734.244	0.084	1.748	1.330
47	-156.957	0.580	4.257	1.533	-248.458	0.829	2.945	1.935
48	-588.561	1.189	8.898	1.466	-738.977	1.255	7.851	1.503
49	-931.425	1.980	4.514	1.230	-1048.014	3.453	3.214	1.707
50	-1270.494	8.111	1.648	2.148	-1344.523	7.624	1.898	1.847
51	-1215.507	4.846	7.522	1.605	-1390.235	5.593	7.238	1.635

Table A.12 Continued

	AGG				GG			
Data	Lik	θ	α	β	Lik	θ	α	β
52	-508.249	3.328	8.080	4.948	-436.615	3.567	7.186	6.222
53	49.497	0.220	2.460	0.975	-126.148	0.364	1.932	1.164
54	-777.444	1.683	1.943	1.193	-882.954	2.618	1.639	1.530
55	-200.843	0.534	11.535	1.467	-287.748	0.539	10.189	1.574
56	-635.201	2.760	4.111	2.909	-660.437	2.241	4.902	2.281
57	-138.753	0.910	3.400	2.498	-171.294	0.911	3.351	2.475
58	-491.236	1.733	2.093	2.274	-545.429	1.600	2.031	1.951
59	-163.081	0.905	4.163	2.319	-179.216	0.973	4.042	2.587
60	-596.835	1.489	2.639	1.534	-716.794	1.184	2.905	1.298
61	-829.442	2.293	1.604	1.350	-898.591	2.893	1.507	1.580
62	-677.298	1.886	1.603	1.502	-718.939	2.019	1.600	1.618
63	-1103.327	2.700	2.289	1.089	-1280.024	2.050	2.528	0.958
64	2211.414	0.011	6.113	3.305	2247.201	0.009	7.528	2.811
65	-168.619	0.880	6.294	2.204	-237.857	0.576	10.187	1.696
66	-282.344	0.617	3.366	1.296	-384.427	0.452	3.137	1.139
67	-1340.606	14.621	4.509	4.491	-1315.905	13.676	5.508	3.755
68	-302.627	0.588	2.667	1.177	-378.854	1.001	2.047	1.708
69	-284.914	0.703	16.009	1.593	-348.162	0.937	10.897	1.944
70	-151.175	0.364	1.686	0.960	-180.468	0.376	1.672	1.059
71	-643.904	1.938	10.408	1.917	-725.697	2.552	9.531	2.247
72	-313.856	0.751	1.634	1.260	-396.433	0.325	1.868	0.845
73	-159.325	0.574	2.251	1.395	-278.872	0.556	2.367	1.326
74	-455.165	0.152	23.478	0.716	-703.771	0.083	11.930	0.761
75	-388.875	0.592	3.966	1.122	-592.568	1.237	2.862	1.545
76	-1029.579	7.736	8.937	3.533	-1058.728	7.931	8.681	3.267
77	154.101	0.528	7.043	2.517	130.655	0.427	6.782	2.149

Table A.12 Continued

	AGG				GG			
Data	Lik	θ	α	β	Lik	θ	α	β
78	-160.540	0.925	9.195	2.318	-153.739	0.709	10.685	2.065
79	-257.457	0.287	6.713	0.938	-489.961	0.724	3.780	1.338
80	-298.264	1.260	4.130	2.499	-293.294	1.398	3.999	3.007
81	-1469.243	10.602	1.632	1.778	-1486.411	10.145	1.986	1.871
82	-1361.572	5.674	4.891	1.411	-1450.403	9.613	4.815	2.069
83	832.002	0.164	4.269	3.455	808.797	0.090	6.337	1.888
84	-754.107	5.782	6.311	6.525	-699.523	4.817	8.184	4.201
85	-590.789	1.680	7.004	1.848	-616.127	1.668	7.166	1.973
86	-145.603	0.702	10.270	1.889	-248.473	1.186	5.984	2.654
87	1549.743	0.017	6.423	1.465	1441.585	0.001	12.277	0.686
88	2626.101	0.004	5.120	2.487	2571.583	0.003	5.424	2.243
89	-431.351	0.898	3.283	1.367	-539.535	1.065	3.001	1.515
90	-142.556	0.872	3.099	2.337	-157.297	0.860	3.149	2.400
91	-409.741	1.610	2.234	2.644	-451.584	1.374	2.669	2.104
92	-1165.868	1.858	3.822	0.895	-1352.101	3.521	3.037	1.194
93	-1117.898	4.726	4.025	1.773	-1188.987	5.225	3.724	1.913
94	-1217.553	5.726	3.717	1.752	-1240.240	5.448	3.668	1.825
95	-327.088	0.489	10.921	1.206	-502.865	0.039	13.533	0.729
96	-275.495	1.100	3.624	2.240	-309.181	1.344	2.806	2.872
97	47.883	0.199	7.713	1.076	-248.515	0.273	6.266	1.175
98	-99.867	0.809	6.661	2.319	-111.384	0.815	6.213	2.430
99	-844.063	3.333	3.731	2.168	-887.240	3.637	4.076	2.351
100	-353.023	1.174	3.792	2.024	-425.099	1.554	2.938	2.562

Bibliography

Abramowitz, M. and Stegun, I. A., editors (1965). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. New York: Dover.

Achkar, J. A. and Bolfarine, H. (1986). The log-linear model with a generalized gamma distribution for the error: a Bayesian approach. *Statistics and Probability Letters*, **4**, 325–332.

Allgower, E. L. and Georg, K., editors (1990). *Numerical Continuation Methods*. Springer, Berlin.

Amoroso, L. (1925). Ricerche intorno alla curva dei redditi di Amoroso. *Ann. Mat. Pura Appl. Ser.*, **21**, 123–159.

Balakrishnan, N., editor (1995). *Recent Advances in Life-Testing and Reliability*. Boca Raton, CRC Press.

Balakrishnan, N. and Peng, Y. (2006). Generalized gamma frailty model. *Statistics in Medicine*, **25**, 2797–2816.

Bassett, G. W. (1986). Strong consistency of regressian quantile and related empirical processes. *Econometric Theory*, **2**, 191–201.

Bassett, G. W., Tam, M.-Y. S., and Knight, K. (2002). Quantile models and estimators for data analysis. *Metrika*, **55**, 17–26.

Cai, Y. and Stander, J. (2008). Quantile self-exciting threshold autoregressive time series models. *Journal of Time Series Analysis*, **29**, 186–202.

Chen, C. (2005). Growth charts of body mass index (BMI) with quantile regression. *Proceedings of 2005 International Conference on Algorithmic Mathematics and Computer Science*. June 2005, Las Vegas, Nevada.

Cohen, A. C. and Whitten, B. J. (1988). *Parameter Estimation in Reliability and Life Span Models*, volume 96. New York: Marcel Dekker.

Cole, T. J. (1988). Fitting smoothed centile curves to reference data. *Journal of the Royal Statistical Society, Series A*, **151**, 385–418.

Cole, T. J. and Green, P. J. (1992). Smoothing reference centile curves: The LMS method and penalized likelihood. *Statisics in Medicine*, **11**, 1305–1319.

Cole, T. J., Stanojevic, S., Stocks, J., Coates, A. L., Hankinson, J. L., and Wade, A. M. (2009). Age- and size-related reference ranges: A case study of spirometry through childhood and adulthood. *Statistics in Medicine*, **28**, 880–898.

Cooray, K. and Ananda, M. M. A. (2008). A generalization of the half-normal distribution with applications to lifetime data. *Communications in Statistics-Theory and Methods*, **37**, 1323–1337.

Cox, C., Chu, H., Shneider, M. F., and Muñoz, A. (2007). Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in Medicine*, **26**, 4352–4374.

D'Addario, R. (1932). Intorno alla curva dei redditi di Amoroso. *Riv. Italiana Statist. Econ. Fnanza*, **anno 4(1)**.

Dang, H. and Weerakkoda, G. (2000). Bounds for the maximum likelihood estimates in two-parameter gamma distribution. *Journal of Mathematical Analysis and its Applications*, **245**, 1–6.

Davison, A. C. (2003). *Statistical Models*. Cambridge University Press.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag, New York.

DiCiccio, T. J. (1987). Approximate inference for the generalized gamma distribution. *Technometrics*, **29**, 33–40.

Farewell, V. T. and Prentice, R. L. (1977). A study of distributional shape in life testing. *American Statistical Association and American Society for Quality*, **19**, 69–75.

Galton, F. (1875). Statistics by intercomparison, with remarks on the law of frequency of error. *Philosophical Magazine*, **49** (4th series), 33–46.

Galton, F. (1883). *Enquiries into Human Faculty and its Development*. Macmillan, London.

Galton, F. (1889). *Natural Inheritance*. Macmillan, London.

Ghilgaber, G. (2005). The extended generalized gamma model and its special cases: applications to modeling and marriage durations. *Quality and Quantity*, **39**, 71–85.

Gilchrist, W. (1997). Modelling with quantile distribution functions. *Journal of Applied Statistics*, **24**, 113–122.

Gilchrist, W. (2008). Regression revisited. *International Statistical Review*, **76**, 401–418.

Gilchrist, W. G. (2000). *Statistical Modelling with Quantile Functions*. Chapman & Hall/CRC Press, Boca Raton, Florida.

Gomes, O., Combes, C., and Dussauchoy, A. (2008). Parameter estimation of the generalized gamma distribution. *Mathematics and Computers in Simulation*, **79**, 955–963.

Hager, H. W. (1970). *Statistical Models and Methods for Lifetime Data.* PhD dissertation, Department of Statistics, Rolla: University of Missouri.

Hager, H. W. and Bain, L. J. (1970). Inferential procedures for the generalized gamma distribution. *Journal of the American Statistical Association*, **65**, 1601–1609.

Hager, H. W., Bain, L. J., and Antle, C. E. (1971). Reliability estimation for the generalized gamma distribution and robustness of the Weibull model. *Technometrics*, **13**, 547–557.

Harter, H. L. (1967). Maximum likelihood estimation of the parameters of a four-parameter generalized gamma population from complete and censored data. *Technometrics*, **9**, 159–165.

Heagerty, P. J. and Pepe, M. S. (1999). Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. *Applied Statistics*, **48**, 533–551.

Hirose, H. (2000). Maximum likelihood parameter estimation by model augmentation with applications to the extended four-parameter generalized gamma distribution. *Mathematics and Computers in Simulation*, **54**, 81–97.

Hogg, S. A., Ciampi, A., and Lawless, J. F. (1982). Ggdmle-a computer program which finds maximum likelihood estimates for the generalized log gamma distribution. *Computer Programs in Biomedicine*, **15**, 201–215.

Huang, P.-H. and Hwang, T.-Y. (2006). On new moment estimation of parameters of the generalized gamma distribution using its characterization. *Taiwanese Journal of Mathematics*, **10**, 1083–1093.

Johnson, N. L. and Kotz, S. (1972). Power transformation of gamma variables. *Biometrika*, **59**, 226–229.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, volume 1. Toronto: Wiley-Interscience, second edition.

Jones, M. C. and Yu, K. (2007). Improved double kernel local linear quantile regression. *Statistical Modelling*, **7**, 377–389.

Kocherginsky, M., He, X., and Mu, Y. (2005). Partial confidence intervals for regression quantiles. *Journal of Computational and Graphical Statistics*, **14**, 41–55.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

Koenker, R. and Bassett, G. W. (1978). Regression quantiles. *Econometrica*, **46**, 33–50.

Koenker, R. and Bassett, G. W. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, **50**, 43–61.

Koenker, R. and d'Orey, V. (1993). Computing regression quantiles. *Applied Statistics*, **43**, 410–414.

Koenker, R. and Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, **94**, 1296–1310.

Kokkinakis, K. and Nandi, A. K. (2007). Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Signal Processing*, **87**, 1156–1162.

Lawless, J. F. (1980). Inference in the generalized gamma and log gamma distributions. *Technometrics*, **22**, 409–419.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data.* John Wiley and Sons, Inc. Hoboken, New Jersey, 1st edition.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data.* John Wiley and Sons, Inc. Hoboken, New Jersey, 2nd edition.

Lefante, J. J. and Turner, M. E. (1985). The average likelihood and a fiducial approximation: One parameter members of the generalized gamma distributions. *Communications in Statistics - Theory and Methods*, **14**, 419–436.

Nelder, J. A. and Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal*, **7**, 308–313.

Neocleous, T. and Portnoy, S. (2008). On monotonicity of regression quantile functions. *Statistics and Probability Letters*, **78**, 1226–1229.

NHANES (2007). Centers for Disease Control and Prevention. National Center for Health Statistics. Health and Nutrition Examination Surveys. NHANES Questionnaires, Datasets, and Related Documentation. Hyattsville, MD. 20782. *http://www.cdc.gov/nchs/nhanes/nhanes2007-2008/nhanes07_08.htm. [Accessed* 22/11/10*].*

Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization.* Springer.

Noufaily, A. and Jones, M. C. (2010). On maximisation of the likelihood of the generalised gamma distribution. *(submitted for publication).*

Ortega, E. M. M., Bolfarine, H., and Paula, G. A. (2003). Influence diagnostics in generalized log-gamma regression models. *Computational Statistics and Data Analysis*, **42**, 165–186.

Ortega, E. M. M., Cancho, V. G., and Paula, G. A. (2009). Generalized log-gamma regression models with cure fraction. *Lifetime Data Analysis*, **15**, 79–106.

Parr, V. B. and Webster, J. T. (1965). A method for discriminating between failure density functions used in reliability predictions. *Technometrics*, **7**, 1–10.

Parzen, E. (1979). Non-parametric statistical data modelling. *Journal of the American Statistical Association*, **74**, 105–121.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, **98**, 1001–1012.

Prentice, R. L. (1974). Log-gamma model and its maximum likelihood estimation. *Biometrika*, **61**, 539–544.

Rao, A. V., Kantam, R. R. L., and Narasimham, V. L. (1991). Linear estimation of location and scale parameters in the generalized gamma ditribution. *Communications in Statistics-Theory and Methods*, **20**, 3823–3848.

Roberts, C. D. (1971). On the distribution of random variables whose m^{th} absolute power is gamma. *Sankhyā, Series A*, **33**, 229–232.

Serfling, R. (2004). Nonparametric multivariate descriptive measures based on spatial quantiles. *Journal of Statistical Planning and Inference*, **123**, 259–278.

Song, K. (2008). Globally convergent algorithms for estimating generalized gamma distributions in fast signal and image processing. *IEEE Transactions on Image Processing*, **17**, 1233–1250.

Stacy, E. W. (1962). A generalization of the gamma distribution. *Annals of Mathematical Statistics*, **33**, 1187–1192.

Stacy, E. W. (1973). Quasimaximum likelihood estimators for two-parameter gamma distributions. *IBM Journal of Research and Development*, **17**, 115–124.

Stacy, E. W. and Mirham, G. A. (1965). Parameter estimation for a generalized gamma distribution. *Technometrics*, **7**, 349–358.

Stasinopoulos, M. and Rigby, B. (2007). Generalized additive models for location scale and shape (GAMLSS) in **R**. *Journal of Statistical Software*, **23**, 1–46.

R (2009). Development core team. **R**: a language for and environment for statistical computing. **R** Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Thompson, P., Cai, Y., Moyeed, R., Reeve, D., and Stander, J. (2010). Bayesian nonparametric quantile regression using splines. *Computational Statistics and Data Analysis*, **54**, 1138–1150.

Tsionas, E. G. (2001). Exact inference in four-parameter generalized gamma distributions. *Communications in Statistics-Theory and Methods*, **30**, 747–756.

Van Buuren, S., Hayes, D. J., Stasinopoulos, D. M., Rigby, R. A., ter Kuile, F. O., and Terlouw, D. J. (2009). Estimating regional centile curves from mixed data sources and countries. *Statistics in Medicine*, **28**, 2891–2911.

Wei, Y., Pere, A., Koenker, R., and He, X. (2006). Quantile regression methods for reference growth charts. *Statistics in Medicine*, **25**, 1369–1382.

Wingo, D. R. (1987). Computing maxium-likelihood parameter estimates of the generalized gamma distribution by numerical root isolation. *IEEE Transactions on Reliability*, **R-36**, 586–590.

Wong, J. Y. (1993). Readily obtaining the maximum-likelihood estimates of the 3-parameter generalized gamma-distribution of Stacy and Mirham. *Microelectronics and Reliability*, **33**, 2243–2251.

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, **93**, 228–237.

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

Yu, K. and Stander, J. (2007). Bayesian analysis of a Tobit quantile regression model. *Journal of Econometrics*, **137**, 260–276.

Yu, K., Lu, Z., and Stander, J. (2003). Quantile regression: Applications and current research areas. *The Statistician*, **52**, 331–350.