DEPARTMENT OF ECONOMICS AND FINANCE

COLLEGE OF BUSINESS AND ECONOMICS

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

Bayesian Extreme Value Mixture Modelling for Estimating VaR

Xin Zhao, Carl John Scarrott, Marco Reale, Les Oxley

WORKING PAPER

No. 15/2009

Department of Economics and Finance
College of Business and Economics
University of Canterbury
Private Bag 4800, Christchurch
New Zealand

WORKING PAPER No. 15/2009

WORKING PAPER No. 15/2009

Bayesian Extreme Value Mixture Modelling for Estimating VaR

Xin Zhao[1], Carl John Scarrott[1],Marco Reale[1], Les Oxley[2*]

27 October,2009

Abstract:A new extreme value mixture modelling approach for estimating Value-at-Risk (VaR) is proposed, overcoming the key issues of determining the threshold which defines the distribution tail and accounts for uncertainty due to threshold choice. A two-stage approach is adopted: volatility estimation followed by conditional extremal modelling of the independent innovations. Bayesian inference is used to account for all uncertainties and enables inclusion of expert prior information, potentially overcoming the inherent sparsity of extremal data. Simulations show the reliability and flexibility of the proposed mixture model, followed by VaR forecasting for capturing returns during the current financial crisis.

Keywords: Extreme values, Bayesian, Threshold estimation, Value-at-Risk
JEL Classifications: C11, G12, G17

[1]Department of Mathematics and Statistics, University of Canterbury, Christchurch
[2]Department of Economics, University of Canterbury, Christchurch
[*]Corresponding Author: Department of Economics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand. Tel. +64 (0)3 3642134. Fax +64 (0)3 3642635. Email: les.oxley@canterbury.ac.nz

WORKING PAPER No. 15/2009

Bayesian Extreme Value Mixture Modelling for Estimating VaR

1. Introduction

Extreme value models have been widely used to assess financial risk such as risk due to adverse market movements, see for example Embrechts et al. (2003). The current financial crisis, as with those of 1990's, has further stimulated interest in describing the probability of such extreme events (Gençay et al. 2003). Extreme value models describe the stochastic dynamics of a process for states with small chances of realization, and typically beyond the range of observed data (Beirlant et al. 2004). They are therefore suitable for capturing tail related quantities for risk measurement and control. Value-at-Risk (VaR) is one such risk measure, which quantifies the largest possible profit-and-gain of a portfolio over a fixed holding period for a given low probability (see Duffie and Pan 1997 and Jorion 2000 for a variety of definitions and expansions of VaR). In statistical terms, the VaR is estimated as the extreme quantiles of returns due to unexpected market shortfalls. Three common statistical approaches to estimate VaR (McNeil and Frey 2000 and Gençay et al. 2003) are non-parametric historical simulation methods, parametric methods based on econometric models with volatility dynamics and extreme value based models.

This paper develops a new extreme value theory (EVT) based model to estimate the VaR, using extreme quantiles of the return series after accounting for the dependence structure induce by the volatility clustering typically observed in financial returns. In particular, a three component mixture is used to capture the entire innovation distribution. A general distribution (normal used in this paper, but others possibilities are discussed below) is used to describe the main mode of the innovation distribution. However, the flexible generalized Pareto distribution (GPD) is used to simultaneously extrapolate the gain and losses beyond some thresholds, which define the upper and lower tails of the innovation distribution.

1.1. Background

Non-parametric methods for estimating VaR are challenging due to the inherent lack of sample information in the tails of the distribution (Embrechts et al. 2003). Parametric methods such as the autoregressive conditional heteroscedasticity (ARCH) and generalized ARCH (GARCH) model, and their many variants, typically make the assumption of conditional normality for the residuals which is typically unrealistic as it is commonly observed that financial series display heavier tails. Extreme value theory (EVT) based models are based upon an asymptotic approximation for the tail distributions, which are very flexible in terms of the allowable tail shape behaviour. The attraction of the EVT based methods is that they can provide mathematically and statistically justifiable parametric model for the tails of distribution which can give reliable extrapolations beyond the range of the observed data.

There are two issues in applying the classical GPD model for estimating the VaR: the dependence of extremes (financial returns typically show clusters of observations in the tails) and the threshold choice of GPD (i.e. at which level of extremity into the tails of the data is the GPD a good model). The classical extreme value theory used to justify the GPD for capturing the tail of a distribution assumes the observations are independent and identically distributed. Under quite general conditions processes with short range dependence can also lead to the GPD being an appropriate model for the tail of the distribution (Beirlant et al. 2004). However, the latter result does not necessarily describe the impact of the dependence on inferences (e.g. uncertainty estimation) for tail measures like the VaR.

There are three common approaches in accounting for the impact of the dependence on our inferences. One of these approaches is to explicitly model the dependence structure using time series or covariate models for the model parameters, e.g. non-stationary covariate models (e.g. Smith 1989; Davison and Ramesh 2000 and Pauli and Coles 2001) or heteroscedastic time series models, (e.g. Bali and Weinbaum 2007, Zhao et al. 2009 and Zhao et al. 2010). However, there are significant challenges with these approaches associated the threshold choice and model specification. The other common approach is to decluster the dependent extremes and then apply standard extreme models to the non-dependent sequence using statistical declustering algorithms (e.g. Ferro and Segers 2003).

A commonly used two stage approach in the finance literature (due to McNeil and Frey 2000) is to capture the dependence in the returns induced by the volatility clustering using a GARCH type model, followed by extreme value modeling of the independent residuals. We will use this two stage methodology as a basis for the approach taken in this paper. Chan et al. (2007) consider an extension of this approach by applying non-parametric heavy tails to the GARCH innovations.

Threshold selection for the GPD can also be problematic. The threshold selection is a balance between reliability of the asymptotic approximation versus the sample variance of estimators. The threshold must be sufficiently high to ensure the threshold excesses have a corresponding approximate distribution within the domain of attraction of the generalized Pareto family. However, the threshold cannot be too high as this will reduce the sample information for inferences. Traditionally, the threshold was chosen (fixed) using various graphically diagnostics, see Coles (2001), which assess features of the model fit for a range of potential thresholds. Once a suitable value has been determined, the threshold is then treated as a known fixed constant in latter inferences. This approach suffers from concerns over subjectivity about the threshold choice and not accounting for threshold uncertainty in inferences. A recently developed approach due to Dupuis (1998) aimed at reducing the subjectivity and ensure robustness. However, this method still requires some subjective assessment. For some applications, the threshold selection can be critical for the extrapolated tail behaviour, so the extra uncertainty associated with the threshold choice needs to be account for. In estimating VaR uncertainty estimation is also important for risk control.

There has been much recent research in the development of mixture type models, which typically treat the threshold as a model parameter to be estimated, and so also automatically accounts for the uncertainty associated with the threshold selection. Behrens et al. (2004) use a truncated Gamma distribution for the bulk of the distribution, and the GPD above the threshold. Tancredi et al. (2006) propose a less restrictive approach to overcome the lack of a natural model below the threshold, by trying to model all the observations above a threshold which is definitely too low, by an unknown number of uniform distributions up to some more suitable threshold and a GPD above that threshold. Frigessi et al. (2002) take a slightly different approach, using a dynamically weighted mixture model of a single GPD and a light-tailed distribution for the bulk of the distribution, with a smooth weight function to transition between the two distributions. This approach avoids the threshold choice, but replaces this problem with choice of transition function parameters.

We will extend these one sided GPD mixture models to a two tailed GPD to capture both the gains (upper) and losses (lower) tails simultaneously (jointly accounting for uncertainties from both tails), potentially permitting both tails to be heavy (and even asymmetric) which is a well documented feature in finance/economics applications. In this paper, we use a two stage GARCH-GPD mixture model following the two stage approach of McNeil and Frey (2000), but with a two tailed GPD mixture model used in the second stage to overcome the difficulty of threshold choice and to fully account for the uncertainty due to the threshold selection.

The distribution selected for the non-extreme data (main mode of the distribution) can affect estimation of the tail quantities (like VaR) and therefore it is necessary to choose it according to the application. The normal distribution is suggested for the financial applications in this paper, due to their inherent unimodal nature, approximate symmetry and quadratic shape around the mode. In applications where significant asymmetry in the mode is expected, then the Weibull or Gamma may be suitably flexible alternatives.

Bayesian inference is used for fitting the mixture model as it can take advantage of any expert prior information, which can be important in tail estimation due to the inherent sparsity of extremal data. The estimation method for the proposed mixture model is firstly evaluated by a simulation study, followed by application of the two stage GARCH-GPD mixture model to forecasting VaR for a stock market index during the current financial crisis.

The paper is organized as follows: Section 2 defines the model; Section 3 describes the estimation method; Section 4 summarises the results from the simulation studies assessing the model and estimation method performance; Section 5 presents the empirical results for estimating VaR, followed by conclusions in Section 6.

2. The GARCH-GPD mixture Model

In the two stage approach of McNeil and Frey (2000), the dependence in return sequence is captured using a generalised autoregressive conditional heteroscedastic (GARCH) process. The residuals from the GARCH approach are treated as independent and their conditional distribution is then modeled using the two tail GPD mixture model.

2.1. Single GPD Tail Model

The generalized Pareto distribution (GPD) is a model with asymptotic justification when applied to excesses which occur over a sufficiently high threshold. The GPD can equivalently be defined for excesses below a suitably low threshold for capturing the lower tail of a distribution. Let X be an IID random variable with $X \geq u$ following a GPD(σ,ξ) with scale parameter σ (dependent on threshold u) and shape parameter ξ, which has a distribution function given by:

$$G(x|\xi,\sigma,u) = P(X < x|X > u) = \begin{cases} 1 - \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]_{+}^{-1/\xi} & \xi \neq 0 \\ 1 - \exp\left[-\left(\frac{x-u}{\sigma}\right)\right] & \xi = 0 \end{cases}, \qquad (1)$$

where $x \geq u, \sigma > 0$ and $y_+ = \max(y, 0)$. There are three types of tail behaviour determined by the shape parameter: $\xi = 0$ gives an exponential tail, $\xi < 0$ gives a short tail with an upper bound given by $u - \sigma/\xi$ and a heavier tail than an exponential is indicated if $\xi > 0$.

2.2. Two Tail GPD Mixture Model

The two tail GPD mixture model has separate GPD's for the upper and lower tails beyond each threshold, with a suitable distribution between the two thresholds. The thresholds are explicitly specified by model parameters to be estimated. The focus of this paper is on applications in finance and economics, hence a sensible choice for the non-extreme data is the normal distribution as in these fields the time series are generally unimodal, approximately symmetric and quadratic in shape around the mode.

We will denote the two tail GPD mixture model as the GNG model (GPD-Normal-GPD). Let X be an IID random variable from the GNG distribution. The distribution function of the mixture model, $P(X \leq x) = F(x)$ where:

$$F(x|\theta) \quad = \quad \{\Phi(u_l|m,s)[1 - G(-x|\xi_l,\sigma_l,-u_l)]\}\, I_{(-\infty,u_l]}(x) \qquad (2)$$

$$+\Phi(x|m,s)I_{(u_l,u_r)}(x) +$$
$$\{\Phi(u_r|m,s) + [1-\Phi(u_r|m,s)]G(x|\xi_r,\sigma_r,u_r)\}\, I_{[u_r,\infty)},$$

and $\Phi(x)$ is the normal cumulative distribution function with mean m and variance s^2 and $G(x|\xi,\sigma,u)$ is the distribution function of GPD defined by equation 1. The subscript on the GPD parameters l denotes the lower (left) tail and r denotes the upper (right) tail. The parameter vector of the model is $\theta = (m, s, u_r, \xi_r, \sigma_r, u_l, \xi_l, \sigma_l)$.

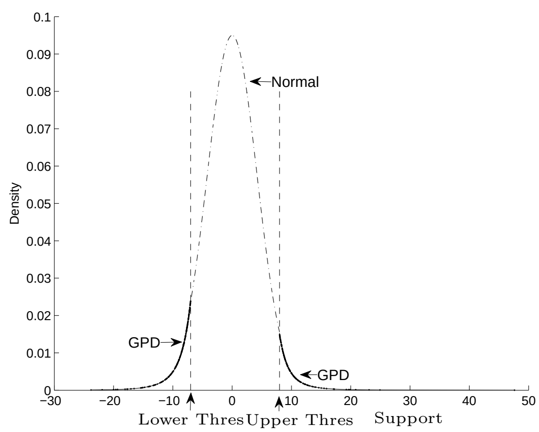

Figure 1: Example of the density for the two tailed GNG (GPD-Normal-GPD) model. The two vertical dash line represent the threshold cut off points for the two GPD distributions.

Figure 1 provides an example of a GNG distribution. The density is deliberately chosen to be smooth at the thresholds, to ensure it is realistic. However, it is worth noting that the density of the mixture distribution may have a discontinuity at the threshold(s). The cumulative distribution function will be continuous. However, we have found in most applications that the estimated density is close to continuous and any lack of continuity is typically of little concern if interest lies in sufficiently high (or low) quantiles away from the thresholds.

The are distinct benefits and potential drawbacks of the mixture modeling approach when compared to the classical fixed threshold method. The principal benefits are that the threshold is estimated (avoiding the often subjective choice in the classical approach) and the uncertainty associated with the estimation is accounted for in inference, which is rather challenging for the fixed threshold method. The automated threshold estimation is a major benefit when trying to automate fitting the GPD to multiple datasets. The principal drawbacks are the added complexity of estimating the additional parameters and the fit in the bulk of the distribution (or the alternate tail) may have an influence on the tail fit. It is clear that different parameters sets could give similar model fits. However, the Bayesian inference approach taken in this paper is shown in the simulation study in Section 4 to provide reliable parameter estimates (including the threshold).

The GNG model is also able to extrapolate two sided tail distributions simultaneously, which is highly relevant in many finance/economics applications. The proposed mixture model has the flexibility in dealing with a variety of distributions, with or without the symmetry, by allowing both tails to follow separate GPD distributions.

2.3. Two Stage Approach

Let $\{R_t\}$ be a strictly stationary daily log return series on a financial asset at time t. The two stage approach to estimate the VaR is as follows:

1. Fit a GARCH volatility model to $\{R_t\}$ and obtain the standardized innovation term x_t as $R_t = E(R_t) + v_t x_t$. Here, the $E(R_t)$ is the expected return at time t and v_t is the volatility estimator from a GARCH model. The form of GARCH can be selected according to the particular application.
2. Fit the proposed GNG mixture model to $\{x_t\}$ (the standardized innovation sequence) as described above. The upper tail of the mixture model represent gains and the lower tail represents the losses.

The first stage GARCH model is fitted using a standard maximum likelihood method, as this stage is less critical for estimating the VaR. However, the GNG mixture model is estimated using Bayesian inference, as the complexity of the likelihood for this model means it would be challenging to maximise directly and Bayesian inference also permits use of prior information which can substantially aid estimation of tail quantities (like VaR) due to the inherent paucity of sample information.

2.4. Estimating the VaR (Extreme Return Quantile)

The GPD cumulative distribution function defined above is defined conditional on being above (or below) the upper (or lower) threshold, i.e. $P(X < x|X > u_r)$ for upper tail. Therefore, if we are interested in estimating the $1-p$ quantile (where p is small so an upper tail quantile) of the entire population distribution (otherwise known as the return level associated with a return period of $1/p$), then we need to scale the conditional GPD distribution function by the probability of being above the threshold $p^* = P(X > u_r)$ giving:

$$p(X < x) = P(X < x|X > u_r)P(X > u_r) = G(x|\xi_r, \sigma_r, u_r)p^*.$$

We can then invert this relation to get the corresponding return level q_p giving:

$$q_p = \begin{cases} u - \frac{\sigma}{\xi}(1 - (p/p^*)^{-\xi}) & \text{for } \xi \neq 0 \\ u - \sigma \log(p/p^*) & \text{for } \xi = 0. \end{cases} \tag{3}$$

In the traditional fixed threshold approach, $p^* = P(X > u_r)$ is estimated using the sample proportion above/below the threshold. Since the proposed GNG model describes the entire sample distribution we use the quantile at each threshold giving, for example, $p^* = 1 - \Phi(u_r|m, s)$ for the upper tail. An equivalent formulation can also be determined for the lower GPD tail quantiles, but is not shown for brevity.

For the fixed threshold GPD, the expected return quantile with a return period of $1/p$ at time t is the sum of the expected return and the expected rise or fall of returns given by:

$$E(R_{p,t}) = E(R_t) + v_t q_p(x).$$

When forecasting, the 1-step ahead prediction of the conditional quantile q_p for the upper tail is defined as:

$$R_{q_p,t(1)} = \inf\{F(R_t) \geq q_p | \varphi_{t-1}\}$$

where φ_{t-1} is the information up to day $t-1$. A similar result is obtainable for the lower tail quantile 1-step ahead forecasts.

3. Bayesian Inference for Mixture Model

Bayesian inference is used to estimate the parameters of the mixture model in order to potentially combine expert prior information along with the sample data. Markov Chain Monte Carlo (MCMC) has been used to obtain the posterior distribution. In this section, we main describe the Bayesian inference.

3.1. Prior Distributions

The parameter vector $\theta = (m, s, u_r, \xi_r, \sigma_r, u_l, \xi_l, \sigma_l)$ can be decomposed into three components $\theta_1 = (m, s)$, $\theta_3 = (\xi_r, \sigma_r, \xi_l, \sigma_l)$, and $\theta_2 = (u_r, u_l)$, associated with the normal, GPD parameters, and the thresholds respectively. In this study we explicitly specify priors with little information, to allow the data to speak for themselves and expose any issues in the estimation method. In specific applications, however, expert information could be included to give more informative priors which could reduce the uncertainty associated with parameter estimates.

3.1.1. Prior for normal parameters

A normal prior is used for the location m and a gamma prior for the scale parameter s of the normal component of the GNG model, under the assumption that m and s are independent such that:

$$\pi(m|m_m, s_m) \propto \exp\left[-\frac{1}{2}\left(\frac{m - m_m}{s_m}\right)^2\right]$$

$$\pi(s|\alpha, \beta) \propto s^{\alpha-1} e^{\frac{-s}{\beta}}$$

with hyperparameters (m_m, s_m) and (α, β) respectively.

3.1.2. Prior for the GPD parameters

Following Coles and Tawn (1996) the GPD priors are specified on the quantile differences since expert prior beliefs are generally easier to elicit on the quantiles themselves, rather than more directly on the parameters. The formulation of the prior elicited on the quantile differences also permits consideration of the known negative dependence between the shape ξ and scale σ parameters of the GPD. A gamma prior distribution is used to describe the quantile differences.

We assume the quantile differences follow a gamma distribution, so that $d_{q_i} \sim Ga(a_i, b_i)$ and $q_{p_0} = 0$ for the excesses above (or below if lower tail) the threshold, the prior for upper tail is defined as:

$$\pi(\xi, \sigma) \propto J \times q_{p_1}^{a_1-1} e^{-b_1 q_{p_1}} (q_{p_2} - q_{p_1})^{a_2-1} e^{-b_2(q_{p_2} - q_{p_1})}$$

where J is the Jacobian transformation followed by the joint distribution of the quantile differences. The prior for the upper tail GPD parameters ξ and σ is then:

$$\begin{aligned}\pi(\sigma, \xi) \quad \propto \quad & \exp\left\{-b_1\left[u + \frac{\sigma}{\xi}(p_1^{-\xi} - 1)\right]\right\}\left[u + \frac{\sigma}{\xi}(p_1^{-\xi} - 1)\right]^{a_1-1} \\ & \times \exp\left\{-b_2\left[\frac{\sigma}{\xi}(p_2^{-\xi} - p_1^{-\xi})\right]\right\}\left[\frac{\sigma}{\xi}(p_2^{-\xi} - p_1^{-\xi})\right]^{a_2-1} \\ & \times \left|\frac{\sigma}{\xi^2}\left[(p_1 p_2)^{-\xi}(\log p_1 - \log p_2) + p_2^{-\xi}\log p_2 - p_1^{-\xi}\log p_1\right]\right|\end{aligned}$$

The prior for the lower tail GPD parameters is similarly defined. In this paper we have used the quantile differences for the conditional tail probabilities $p_1 = 0.1$ and $p_2 = 0.01$ (equivalent to tail probability p/p^* in equation 3 above) following Coles and Tawn (1994). The tail probabilities considered for p_1 and p_2 can be altered according to the application and the available expert information.

3.1.3. Prior for the thresholds

A truncated normal distribution is used as the prior distribution for the thresholds of both tails, which are truncated at the minimum and maximum of the sample data respectively (and thresholds), due to Behrens et al. (2004):

$$\pi(u|m_u, s_u, l_u) \quad \propto \quad \exp\left[-\frac{1}{2}\left(\frac{u - m_u}{s_u}\right)^2\right]$$

for the lower threshold $u = u_l$ and upper threshold $u = u_r$.

3.2. Posterior Distribution

The priors for the normal and GPD components are assumed independent giving the logarithm of the posterior $p(\theta|x) \propto \pi(\theta)l(x|\theta)$ for $\xi \neq 0$:

$$\begin{aligned}
\log p(\theta|x) \quad = \quad & K + \sum_{i=1}^{n} I_{(u_l,u_r)}(x_i)\left[-\log s - \frac{1}{2}\left(\frac{x_i - m}{s}\right)^2\right] \\
& + \sum_{i=1}^{n} I_{[u_r,\infty)}(x_i) \log\left[1 - \Phi(u_r|m, s)\right] \\
& + \sum_{i=1}^{n} I_{[u_r,\infty)}(x_i)\left\{-\log\sigma_r - \frac{1+\xi_r}{\xi_r}\log\left[1 + \xi_r\left(\frac{x_i - u_r}{\sigma_r}\right)\right]\right\} + \\
& + \sum_{i=1}^{n} I_{(-\infty,u_l]}(x_i) \log\left[\Phi(u_l|m, s)\right] \\
& + \sum_{i=1}^{n} I_{(-\infty,u_l]}(x_i)\left\{-\log\sigma_l - \frac{1+\xi_l}{\xi_l} + \log\left[1 + \xi_l\left(\frac{u_l - x_i}{\sigma_l}\right)\right]\right\} \\
- \quad & \left[\frac{1}{2}\left(\frac{m - m_m}{s_m}\right)^2\right]
\end{aligned}$$

$$
\begin{aligned}
&+(\alpha-1)\log(s)-\frac{s}{\beta}-\frac{1}{2}\left(\frac{u_r-m_{u_r}}{s_{u_r}}\right)^2-\frac{1}{2}\left(\frac{u_l-m_{u_l}}{s_{u_l}}\right)^2\\
&-b_{1_r}\left[u_r+\frac{\sigma_r}{\xi_r}(p_{1_r}^{-\xi_r}-1)\right]+(a_{1_r}-1)\log\left[u_r+\frac{\sigma_r}{\xi_r}(p_{1_r}^{-\xi_r}-1)\right]\\
&-b_{2r}\left[\frac{\sigma_r}{\xi_r}(p_{2_r}^{-\xi_r}-p_{1_r}^{-\xi_r})\right]+(a_{2_r}-1)\log\left[\frac{\sigma_r}{\xi_r}(p_{2_r}^{-\xi_r}-p_{1_r}^{-\xi_r})\right]\\
&+\log\left|\frac{\sigma_r}{\xi_r^2}\left[(p_{1_r}p_{2_r})^{-\xi_r}(\log p_{1_r}-\log p_{2_r})+p_{2_r}^{-\xi_r}\log p_{2_r}-p_{1_r}^{-\xi_r}\log p_{1_r}\right]\right|\\
&-b_{1_l}\left[u_l+\frac{\sigma_l}{\xi_l}(p_{1_l}^{-\xi_l}-1)\right]+(a_{1_l}-1)\log\left[u_l+\frac{\sigma_l}{\xi_l}(p_{1_l}^{-\xi_l}-1)\right]\\
&-b_{2_l}\left[\frac{\sigma_l}{\xi_l}(p_{2_l}^{-\xi_l}-p_{1_l}^{-\xi_l})\right]+(a_{2_l}-1)\log\left[\frac{\sigma_l}{\xi_l}(p_{2_l}^{-\xi_l}-p_{1_l}^{-\xi_l})\right]\\
&+\log\left|\frac{\sigma_l}{\xi_l^2}\left[(p_{1_l}p_{2_l})^{-\xi_l}(\log p_{1_l}-\log p_{2_l})+p_{2_l}^{-\xi_l}\log p_{2_l}-p_{1_l}^{-\xi_l}\log p_{1_l}\right]\right|
\end{aligned}
$$

where K is from the normalizing constant. In the case where $\xi = 0$, the posterior can be obtained by replacing within the above function the likelihood and prior of $\xi = 0$ from above, which is not shown for brevity.

The posterior is sampled using Markov Chain Monte Carlo (MCMC) with a random walk Metropolis-Hastings (M-H) algorithm, which is a common approach with the advantage of being free of functional form when the posterior distribution function is not a proper probability function. In the algorithm implementation, each subset of parameters is updated at each iteration step in terms of the importance order of the parameters as (ξ_r, σ_r, u_r), (ξ_l, σ_l, u_l) and finally (m, s).

The convergence of the chain of MCMC is checked by monitoring multiple posterior simulation sequences with over dispersed starting values suggested by Gelman et al. (2004). We can estimate the marginal posterior variance by a weighted average of the between and within sequence variances for each of parameter estimator. We then assess the convergence by monitoring whether the scale of the current posterior distribution for θ might be reduced if the simulation continues in the limit $n \to \infty$. The potential scale reduction is defined as the ratio of the marginal posterior variance and within variance, which should decline to one as the chain length goes to infinity. Only once the chain has converged is it regarded as an approximate sample from the posterior distribution. The second half of the chain is used as the posterior distribution and the estimated parameter is calculated as the mean of the posterior parameters within the highest posterior density (HPD) interval. The full details of the MCMC algorithm are given by Zhao (2009) and are available upon request.

4. Simulation Studies

Various simulation studies were undertaken to assess the performance of the GNG model and estimation method for various applications. The first simulation study was designed to assess the performance of the Bayesian inference approach, via an application to data simulated directly from the model with known parameter values. Some of the key results are discussed in Section 4.1.

Section 4.2 compares the threshold estimate of the proposed GNG model with the threshold choice via the robust estimation approach of Dupuis (1998). The second major simulation study in Section 4.3 was designed to assess the ability of the model to reliably estimate quantiles from various population distributions.

4.1. Simulations from GNG Mixture Model

A single simulated dataset from the GNG model is presented as Figure 2. We will consider the results from fitting the model to this single sample, before considering the full simulation study. The sample is from the GNG model using the parameter value $\theta = (\mu_m = 0, s_m = 4.2, u_r = 6, \xi_r = 0.3, \sigma_r = 2.2, u_l = -5, \xi_l = 0.2, \sigma_l = 2.5)$ with a sample size 3000, with approximately 35 observations in the lower tail below u_l and 230 in the upper tail above u_r. The parameters are chosen to give two reasonably heavy tails, and a near continuous density function at the thresholds. Figure 2 shows the posterior predictive density and corresponding cumulative distribution function (CDF).

Figure 2.1 *Fitted Density* Figure 2.2 *CDF and Tail Return Levels*

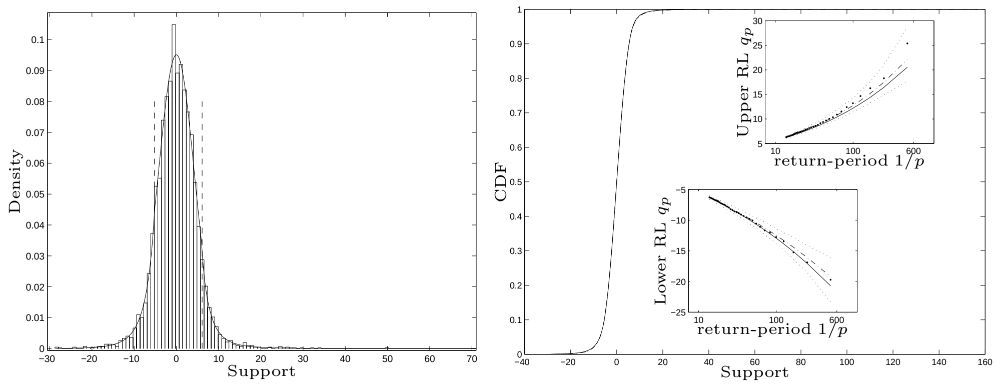

Figure 2: Example dataset from the GNG model with parameter set $\theta = (\mu_m = 0, s_m = 4.2, u_r = 6, \xi_r = 0.3, \sigma_r = 2.2, u_l = -5, \xi_l = 0.2, \sigma_l = 2.5)$. (1) is the density with fitted model and (2) gives the fitted CDF with excerpts showing the fit in the tails in more detail. In (1) the estimated thresholds are shown by vertical dashed lines and the posterior predictive density estimate is shown by the solid line. The true CDF and return level are denoted by solid line, sample values are presented by dots, and posterior predictive estimates are shown by dashed lines in (2).

It is clear from Figure 2 and Table 1 that the Bayesian inferences are reliable, with the fitted model providing a very good fit. Notice that point estimates of the parameters are close to the true values which are well within the 95% credible intervals.

Figure 3 shows the estimated relationship of the GPD parameters for both tails as contours from their joint posterior distribution. As expected, the ξ and σ are negatively correlated, and the shape parameter also appears independent of the threshold which is as expected. The scale parameters are also linearly related to the threshold as we would expect. The shape

Table 1: Results from Bayesian inference for single simulated dataset with sample size n=3000 from from the GNG model. The true parameter values (True) and estimated parameters (Estimated) using the mean of the MCMC samples within the 95% highest posterior credible interval are shown.

Parameters	m	s	u_r	ξ_r	σ_r	u_l	ξ_l	σ_l
True	0	4.2	6	0.3	2.2	-5	0.2	2.5
Estimated	-0.0011	4.1960	6.0986	0.3015	2.4318	-5.2267	0.1532	2.5251
Lower CI	-0.1179	4.1022	5.6346	0.2038	2.0456	-5.7949	0.0583	2.1555
Upper CI	0.1108	4.2827	6.6472	0.4055	2.8273	-4.6149	0.2572	2.9343

parameter posterior density is positively skewed and the scale parameter is slightly positively skewed as we would expect. The threshold posterior densities also appears to be approximately normally distributed.

Posterior predictive checks are important diagnostics for MCMC methods to assess their performance. The basic idea is to compare a specified test quantity and an appropriate predictive distribution from posterior replications. A large discrepancy between them would indicate that the model is not a good fit to the data. The obvious quantities for diagnostics are the quantiles of the simulated sample. The posterior predictive quantile distributions are shown in Figure 4, along with the true quantiles and direct sample estimates of the quantiles. The sample and true quantiles are well within the 95% credible intervals and are located near the mode of the posterior predictive quantile distributions. You will notice that as the tail quantiles get more extreme the posterior predictive distribution becomes more skewed, representing the asymmetry in the information available for estimation.

Multiple samples from the model were then simulated, under different parameters sets and sample sizes, to verify the performance of the Bayesian

Figure 3.1 *Upper Tail* Figure 3.2 *Lower Tail*

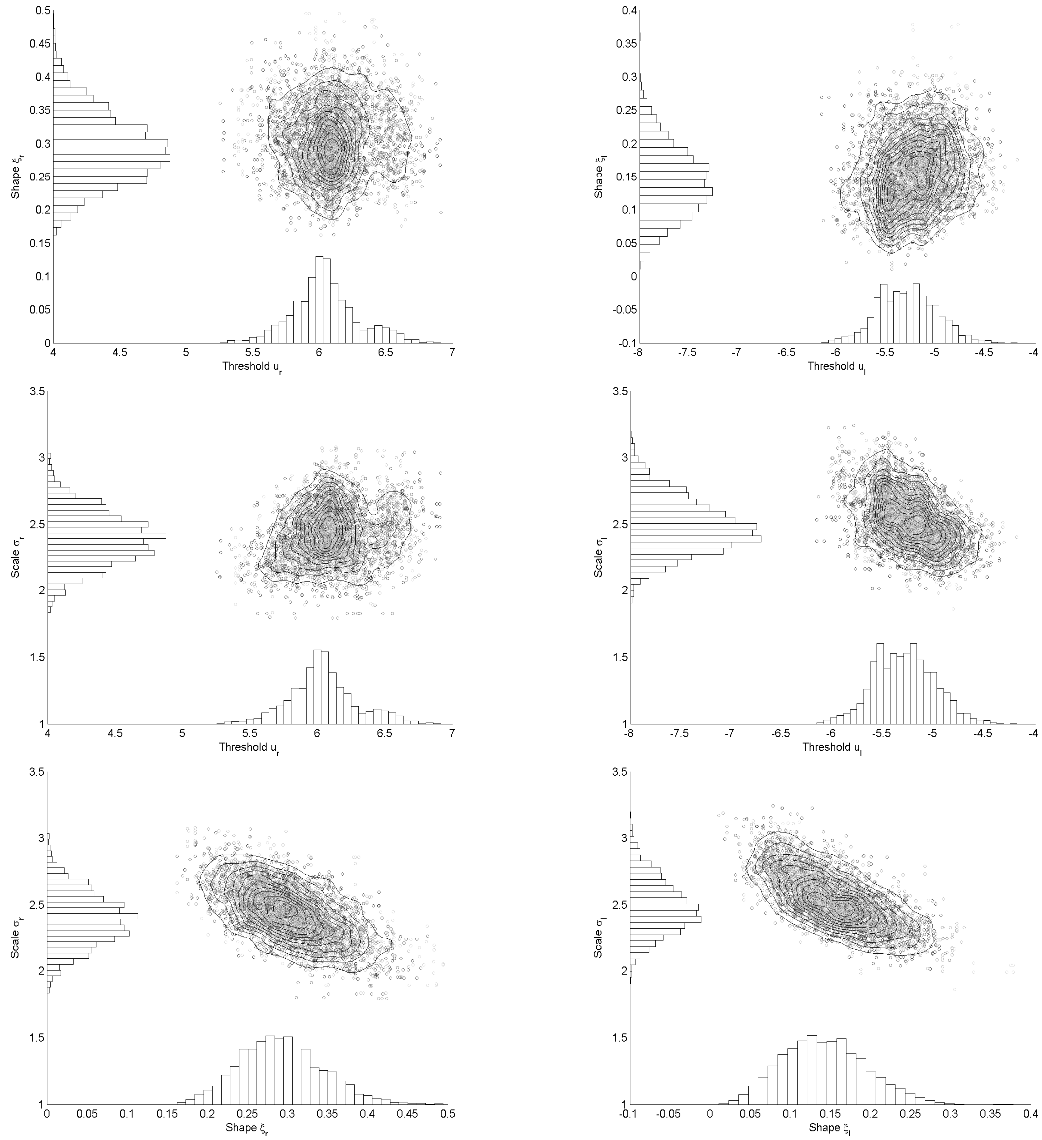

Figure 3: Pairwise contours of the posterior density of the GPD related parameters for the lower tail (u_l, σ_l, ξ_l) and upper tail (u_r, σ_r, ξ_r). the histogram of each posterior distribution is also shown.

inference estimation method. The parameters were chose to represent different combinations of tail behaviors (and all providing asymmetric population distributions, which are more challenging). Tables 2 and 3 lists the six pa-

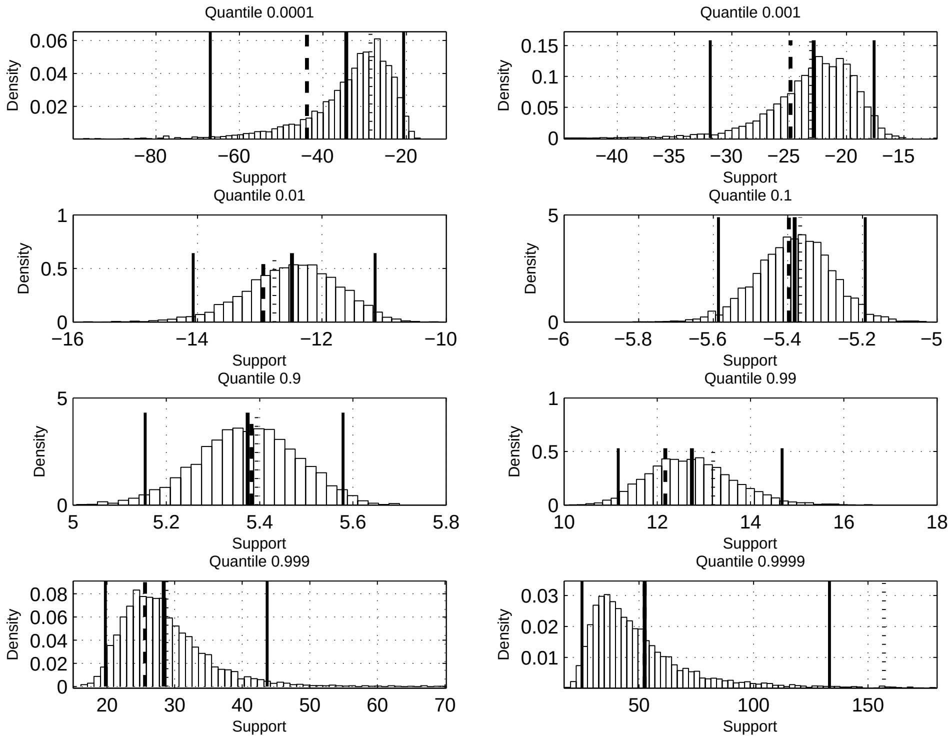

Figure 4: Histogram of the posterior predictive quantiles for unconditional tail probabilities 0.01%, 0.1%, 1% and 10%. The sample quantile is shown by the dotted line and the true quantile by the interior dashed line. The posterior predictive quantile (PPQ) shown by the solid line is the mean of posterior predictive quantiles within the 95% credible intervals, shown by the exterior solid lines.

rameter sets. The parameter set 1 has both Type I (exponential) tails with $\xi = 0$. The parameter set 2 has both Type II tails (heavier tails than exponential) with $\xi > 0$. The third parameter set has both Type III tails (short bounded tails) with $\xi < 0$. The remaining parameter sets 4 to 6 represent the different combinations of the three tail types. For each of the parameter sets and sample sizes, 100 simulation samples were simulated from the GNG distribution. The final parameter values were chosen to ensure the density was near continuous at the upper and lower thresholds, as this is most likely in real world applications.

The MCMC chain was run for a sample of length 10,000, from which the first half were discarded. The estimated parameters are calculated as the mean of the sample posterior values in the 95% highest posterior density (HPD), and the credible interval (CI) of the estimators are the boundary of the associated HPD interval.

All the HPD means of the estimators for these simulated data are close to the real parameters, indicating low bias across the samples. Table 2 also reports the root mean square error (RMSE) of the estimated parameters (a frequentist property useful for summarising variation across simulated datasets). It is clear that the RMSE (and bias) decreases with sample size (approximately in proportion to the sample size) for all the parameters, as is expected. Generally, the RMSE for the threshold is much more uncertain than all the other parameters, indicative of the fact that various thresholds can still give a similar model fit.

The simulation shows the reliability in the model estimation and classification. Even for the smallest sample size 1000, the model can still estimate values close to the true parameters and describe both tails accurately. The simulation also shows the flexibility of the model for various tail behaviors.

4.2. Threshold estimation

As previously discussed, there are many techniques which can be used to estimate the threshold (see for example Coles 2001 and Beirlant et al. 2004). Typically these techniques have been difficult to automate, often requiring manual intervention requiring subjective judgement. Dupuis (1998) suggests a robust threshold selection method which examines the weights applied to the extremes to assess the validity of the GPD under a range of proposed

Table 2: Summary of properties of the Bayesian estimates of the GNG model parameters, for a range of different parameters sets (tail behaviours). There are 100 simulated datasets for parameter set. The true parameters along with the mean and root mean square error (RMSE) of the point estimates across the 100 sample estimations. The point estimates for each sample are the mean of the posterior within the 95% highest posterior density.

1. I-N-I	**True**	**size=1000**		**size=3000**		**size=5000**	
Param	**Value**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
m	0.00	0.0081	0.0416	0.0019	0.0228	-0.0004	0.0182
s	2.00	2.0087	0.0334	2.0055	0.0181	2.0042	0.0132
u_r	2.30	2.2524	0.1839	2.2774	0.1940	2.3084	0.1689
ξ_r	0.00	-0.0339	0.0981	-0.0156	0.0538	-0.0064	0.0426
σ_r	1.20	1.2401	0.1552	1.2189	0.0844	1.2170	0.0799
u_l	-2.50	-2.2895	0.2706	-2.3570	0.2568	-2.3442	0.2564
ξ_l	0.00	-0.0299	0.0877	-0.0065	0.0562	-0.0130	0.0444
σ_l	1.15	1.1915	0.1699	1.1577	0.0948	1.1776	0.0817
2. II-N-II	**True**	**size=1000**		**size=3000**		**size=5000**	
Param	**Value**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
m	0.00	0.0045	0.0342	-0.0005	0.0201	0.0000	0.0150
s	2.00	2.0139	0.0304	2.0055	0.0158	2.0070	0.0138
u_r	2.00	2.0953	0.2418	2.1138	0.2236	2.0577	0.2150
ξ_r	0.20	0.1575	0.0857	0.1868	0.0614	0.1851	0.0518
σ_r	1.30	1.3334	0.1586	1.3354	0.1205	1.3294	0.0913
u_l	-1.80	-1.9485	0.2438	-1.8757	0.1796	-1.8498	0.1635
ξ_l	0.30	0.2715	0.0861	0.2793	0.0605	0.2924	0.0438
σ_l	1.40	1.4496	0.1950	1.4446	0.1211	1.4249	0.0902
3. III-N-III	**True**	**size=1000**		**size=3000**		**size=5000**	
Param	**Value**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
m	0.00	0.0031	0.0493	-0.0004	0.0200	-0.0023	0.0194
s	2.00	2.0046	0.0334	2.0014	0.0186	2.0025	0.0146
u_r	2.30	2.3811	0.2951	2.3721	0.2860	2.3660	0.3185
ξ_r	-0.30	-0.2865	0.0898	-0.2947	0.0592	-0.2971	0.0469
σ_r	1.20	1.1721	0.1951	1.1704	0.1145	1.1770	0.0839
u_l	-2.50	-2.4820	0.3477	-2.4782	0.3267	-2.4580	0.3075
ξ_l	-0.20	-0.1722	0.0827	-0.2078	0.0614	-0.1941	0.0451
σ_l	1.20	1.1671	0.1452	1.2259	0.1277	1.2097	0.1057

Table 3: Summary of properties of the Bayesian estimates of the GNG model parameters, for a range of different parameters sets (tail behaviours). There are 100 simulated datasets for parameter set. The true parameters along with the mean and root mean square error (RMSE) of the point estimates across the 100 sample estimations. The point estimates for each sample are the mean of the posterior within the 95% highest posterior density.

4. III-N-II	**True**	**size=1000**		**size=3000**		**size=5000**	
Param	**Value**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
m	0.00	0.0108	0.0405	0.0004	0.0251	-0.0005	0.0192
s	2.00	2.0088	0.0305	2.0038	0.0167	2.0023	0.0139
u_r	2.50	2.3155	0.1886	2.3534	0.2267	2.4185	0.1978
ξ_r	0.20	0.1747	0.0865	0.1909	0.0457	0.1895	0.0357
σ_r	1.15	1.1291	0.1536	1.1404	0.1204	1.1482	0.0948
u_l	-2.60	-2.4365	0.1637	-2.5101	0.2035	-2.5190	0.2162
ξ_l	-0.15	-0.1476	0.0934	-0.1439	0.0561	-0.1413	0.0390
σ_l	1.10	1.1226	0.1714	1.1146	0.1100	1.1093	0.0841
5. III-N-I	**True**	**size=1000**		**size=3000**		**size=5000**	
Param	**Value**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
m	0.00	-0.0123	0.0418	-0.0035	0.0214	-0.0031	0.0190
s	2.00	2.0112	0.0312	2.0022	0.0204	2.0035	0.0132
u_r	2.70	2.4599	0.2791	2.4837	0.2421	2.4483	0.2117
ξ_r	0.00	-0.0181	0.1111	-0.0114	0.0632	0.0005	0.0500
σ_r	1.30	1.2949	0.2716	1.3013	0.0941	1.2777	0.0744
u_l	-2.80	-2.4900	0.2225	-2.5516	0.2125	-2.5718	0.2522
ξ_l	-0.10	-0.1078	0.0796	-0.1132	0.0547	-0.1077	0.0455
σ_l	1.00	1.0514	0.1639	1.0638	0.1079	1.0562	0.1041
6. II-N-I	**True**	**size=1000**		**size=3000**		**size=5000**	
Param	**Value**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
m	0.00	0.0073	0.0468	0.0060	0.0241	0.0026	0.0172
s	2.00	2.0121	0.0342	2.0070	0.0190	2.0086	0.0166
u_r	2.30	2.2478	0.2066	2.2670	0.1923	2.2755	0.1616
ξ_r	0.15	0.1387	0.1029	0.1447	0.0576	0.1500	0.0440
σ_r	1.20	1.1854	0.1781	1.1955	0.1034	1.1971	0.0703
u_l	-2.50	-2.3282	0.2853	-2.3551	0.2367	-2.3419	0.2391
ξ_l	0.00	-0.0219	0.1001	-0.0083	0.0512	-0.0069	0.0403
σ_l	1.20	1.2357	0.1961	1.2154	0.0980	1.1997	0.0727

thresholds. We implement this robust threshold estimation method to check our threshold estimation method and the resultant tail fits from the proposed mixture model.

We applied the robust GPD estimation method on various simulated data with the results consistent from those from the proposed GNG mixture model. We do not report the full results for brevity, however, Table 4 gives an example to show their consistency. The advantage of our method is that it can capture the uncertainty associated with the threshold choice and it is convenient in forecasting and inference since it does not require a manual intervention unlike the robust method.

Table 4: Comparison of the estimated threshold and GPD parameters using the GNG mixture model and Robust estimation procedure of Dupuis (1998).

	Upper Tail			Lower Tail		
	u_r	ξ_r	σ_r	u_l	ξ_l	σ_l
True Parameters	2.00	0.20	1.30	-1.80	0.30	1.40
Mixture Estimator	1.64	0.17	1.39	-1.77	0.29	1.30
Robust Estimator	1.60	0.17	1.40	-1.80	0.26	1.31

4.3. Performance For General Distributions

The previous simulation results show the performance of the estimation method for the GNG mixture model. However, in real applications of this approach the population will be approximated by the GNG model. Therefore in this section, a sensitivity analysis is conducted by applying the GNG model to various population distributions, including both symmetric and asymmetric distributions, to show how the model performs as an approximation in this case.

The distributions chosen for sensitivity analysis are symmetric distributions: Normal (type I tails), t (type II tails) and symmetric Beta (type III tails); and asymmetric distributions: Gumbel (type III and type I tails), inverse gamma (type III and type II tails) and Weibull (type III and type III tails).

For each distribution, we simulated 100 datasets with sample sizes 1000, 3000 and 5000 as before. The full simulation results are reported in Zhao (2009) and available upon request. Table 5 shows the performance of the GNG model to approximate various quantiles of the population distributions. The true and GNG estimated quantiles are shown in the upper part of the table. It is clear that all the GNG quantile estimates for all six distributions are close to the true values.

The RMSE of the posterior quantile estimates for the six distributions are reported as the lower part of Table 5. For comparison purposes, the RMSE of the quantile estimators using maximum likelihood (ML) estimation when using the correct population distributions in the model. The RMSE for the correct model using ML estimation is considered a gold standard, to compare the performance of the approximate GNG model. Although the ML confidence intervals and Bayesian credible intervals are not formally comparable, as very diffuse priors have been used when finding the posterior for the GNG model the two sets of intervals are practically comparable. The ML quantiles are estimated under the correct model and therefore have a smaller RMSE. However, you will notice that ML estimation under the correct model is at most twice as efficient for estimating the various quantiles compared to the GNG model. Overall the GNG approximation is only slightly less efficient for most population quantiles, compared to using the correct population distribution model.

The uncertainty of the quantiles is higher for the heavy tails (t and upper tail of inverse gamma) compared to the short tails (e.g. beta and lower tail of Gumbel). The normal distribution has a slightly higher RMSE than the others due to known slow convergence of the normal tail to the GPD limit (see Beirlant *et al*, 2004). As expected the RMSE increases as the quantile is located further out into the tail of the distribution. The differences between posterior predictive quantiles and the ML quantile estimates are smaller for the heavy tails compare to the short or exponential tails. This result indicates that the GNG model is preferred when describing the tail behavior for the heavy tail applications, which is the typical case found in finance and economic applications.

It is clear from Table 5 that population distributions with highly asymmetric modes result in higher uncertainty, since the normal distribution is used for the bulk of distribution in the GNG model. However, the mixture model can still return reasonable extreme quantile estimates. This results shows that the GNG model is generally applicable as an approximation to a wide range of population distributions.

5. Application to VaR During the Financial Crisis

We use the proposed two stage method to produce the 1-step ahead forecasts of daily return quantiles for the S&P100 index for the period of 07/02/2008 to 19/11/2008, which captures the starting period of the current worldwide financial crisis. Suppose the return sequence is $R_1, ..., R_t, ...R_T$, where $t \leq T$. In providing the forecast we only use historical information from $n = 1000$ daily returns, i.e. approximately 4 years. The forecasted return quantiles are then dependent on all the information up to *t-1*. At time t, we apply

Table 5: Comparison of quantiles estimates for GNG model applied to various general population distributions (fitted using Bayesian inference) and true model distribution (fitted using maximum likelihood estimation). The mean of the posterior predictive quantiles (PPQ) within the 95% highest posterior density have been used as point estimates for each sample. The mean and RMSE of the point estimates are obtained from 100 simulated datasets.

Distributions	**Sample**	**Posterior Predictive Quantiles**											
		1%		**2%**		**5%**		**95%**		**98%**		**99%**	
	Size	**TRUE**	**PPQ**	**TRUE**	**PPQ**	**TRUE**	**PPQ**	**TRUE**	**PPQ**	**TRUE**	**PPQ**	**TRUE**	**PPQ**
	1000	-4.653	-4.709	-4.107	-4.130	-3.290	-3.290	3.290	3.301	4.107	4.132	4.653	4.704
Normal	3000	-4.653	-4.648	-4.107	-4.098	-3.290	-3.278	3.290	3.293	4.107	4.118	4.653	4.672
($\mu = 0, \sigma = 2$)	5000	-4.653	-4.663	-4.107	-4.112	-3.290	-3.287	3.290	3.294	4.107	4.118	4.653	4.667
	1000	-4.541	-4.518	-3.482	-3.495	-2.353	-2.351	2.353	2.370	3.482	3.497	4.541	4.485
t	3000	-4.541	-4.625	-3.482	-3.559	-2.353	-2.375	2.353	2.363	3.482	3.525	4.541	4.563
($v = 3$)	5000	-4.541	-4.576	-3.482	-3.550	-2.353	-2.387	2.353	2.381	3.482	3.543	4.541	4.569
	1000	0.229	0.225	0.256	0.254	0.300	0.299	0.700	0.702	0.744	0.747	0.771	0.776
Beta	3000	0.229	0.227	0.256	0.255	0.300	0.299	0.700	0.700	0.744	0.745	0.771	0.773
($\alpha = 8, \beta = 8$)	5000	0.229	0.227	0.256	0.255	0.300	0.300	0.700	0.700	0.744	0.745	0.771	0.773
	1000	-1.527	-1.562	-1.364	-1.393	-1.097	-1.109	2.970	2.959	3.902	3.891	4.600	4.568
Gumbel	3000	-1.527	-1.568	-1.364	-1.395	-1.097	-1.107	2.970	3.015	3.902	3.949	4.600	4.619
($\sigma = 1$)	5000	-1.527	-1.567	-1.364	-1.394	-1.097	-1.106	2.970	3.018	3.902	3.958	4.600	4.631
Inverse	1000	0.431	0.410	0.473	0.453	0.546	0.536	2.538	2.451	3.269	3.148	3.909	3.740
Gamma	3000	0.431	0.414	0.473	0.458	0.546	0.540	2.538	2.542	3.269	3.281	3.909	3.913
($\alpha = 5, \beta = 5$)	5000	0.431	0.413	0.473	0.458	0.546	0.542	2.538	2.557	3.269	3.287	3.909	3.903
	1000	-8.250	-8.577	-7.307	-7.550	-5.798	-6.007	12.849	12.175	16.154	15.588	18.436	17.953
Weibull	3000	-8.250	-8.473	-7.307	-7.490	-5.798	-5.932	12.849	12.324	16.154	15.738	18.436	18.090
($\lambda = 5, k = 0.1$)	5000	-8.250	-8.483	-7.307	-7.492	-5.798	-5.903	12.849	12.440	16.154	15.846	18.436	18.179

Distributions	**Sample**	**RMSE of Quantile Estimates**											
		1%		**2%**		**5%**		**95%**		**98%**		**99%**	
	Size	**ML**	**PPQ**	**ML**	**PPQ**	**ML**	**PPQ**	**ML**	**PPQ**	**ML**	**PPQ**	**ML**	**PPQ**
	1000	0.130	0.216	0.119	0.168	0.103	0.120	0.107	0.115	0.123	0.156	0.135	0.210
Normal	3000	0.068	0.114	0.063	0.083	0.055	0.061	0.052	0.060	0.059	0.084	0.064	0.104
($\mu = 0, \sigma = 2$)	5000	0.056	0.076	0.051	0.062	0.045	0.053	0.044	0.049	0.051	0.069	0.056	0.083
	1000	0.372	0.390	0.221	0.236	0.098	0.123	0.098	0.132	0.221	0.241	0.372	0.383
t	3000	0.208	0.263	0.123	0.168	0.055	0.082	0.055	0.075	0.123	0.136	0.208	0.209
($v = 3$)	5000	0.162	0.184	0.097	0.131	0.043	0.074	0.043	0.076	0.097	0.138	0.162	0.199
	1000	0.006	0.009	0.006	0.008	0.005	0.006	0.005	0.006	0.006	0.008	0.006	0.010
Beta	3000	0.004	0.006	0.004	0.005	0.003	0.004	0.003	0.003	0.003	0.005	0.003	0.006
($\alpha = 8, \beta = 8$)	5000	0.003	0.004	0.003	0.004	0.002	0.003	0.002	0.003	0.002	0.003	0.003	0.004
	1000	0.053	0.071	0.048	0.060	0.041	0.044	0.105	0.125	0.161	0.167	0.216	0.219
Gumbel	3000	0.028	0.055	0.025	0.044	0.021	0.025	0.049	0.074	0.081	0.098	0.113	0.126
($\sigma = 1$)	5000	0.023	0.050	0.021	0.040	0.018	0.024	0.046	0.072	0.073	0.094	0.099	0.107
Inverse	1000	0.010	0.027	0.010	0.024	0.010	0.016	0.080	0.131	0.128	0.198	0.176	0.275
Gamma	3000	0.005	0.020	0.005	0.017	0.005	0.010	0.044	0.048	0.069	0.088	0.094	0.145
($\alpha = 5, \beta = 5$)	5000	0.004	0.019	0.004	0.016	0.005	0.007	0.031	0.040	0.050	0.070	0.069	0.111
	1000	0.269	0.450	0.241	0.355	0.206	0.304	0.359	0.799	0.491	0.796	0.619	0.805
Weibull	3000	0.149	0.284	0.133	0.231	0.113	0.181	0.236	0.588	0.345	0.551	0.447	0.564
($\lambda = 5, k = 0.1$)	5000	0.132	0.285	0.117	0.227	0.098	0.150	0.164	0.467	0.233	0.426	0.300	0.427

a GARCH(1,1) model on $R_{t-n-1}, ..., R_{t-1}$ to obtain the standardized innovation sequence $x_{t-n-1}, ..., x_{t-1}$. Then, the 1-step ahead forecasting of the expected return $E(\hat{x}_t)$ and the volatility $\hat{v}_t$ are based on the estimates from the first stage. We then fit the GNG model to the standardized innovation sequence $x_{t-n-1}, ..., x_{t-1}$ and forecast the $\hat{x}_t^q$ based on the predictive posterior quantile distributions. The forecasted return quantile of q can then be calculated as $\hat{R}_t^q = E(\hat{x}_t) + \hat{v}_t \hat{x}_t^q$. These forecasted return quantiles are termed "conditional" as they are conditional on the variance being assumed known, using the estimates from the GARCH.

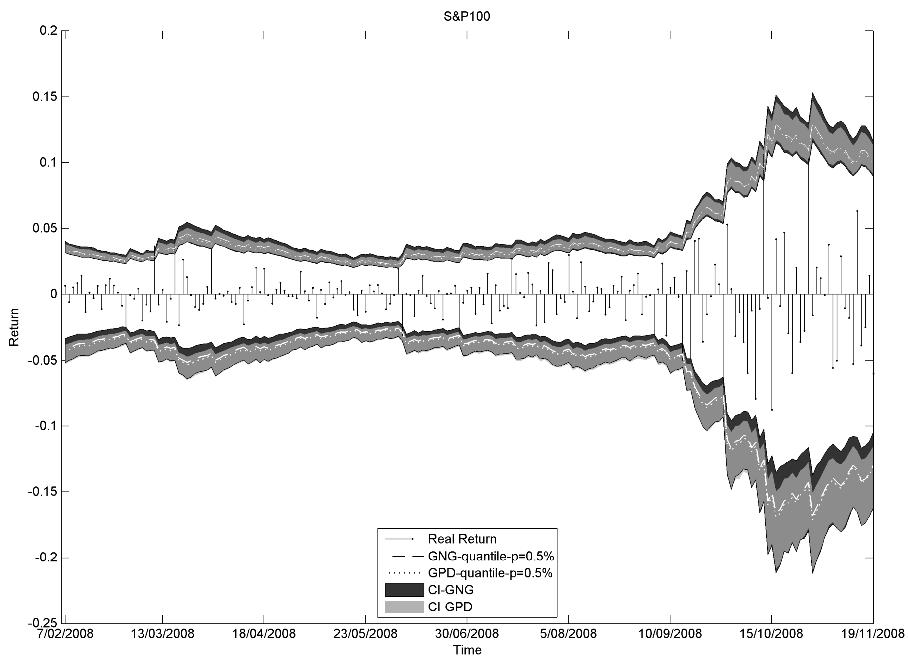

Figure 5: Application of the GNG model and fixed threshold approach to forecasting the 1-step ahead VaR for various quantiles to S&P100 during the 2008 financial crisis.

Figure 5 shows the conditional quantile forecasting results of the S&P100 for the return quantiles at 0.5% and 99.5% using the both the GNG mixture mode and fixed threshold approach with corresponding 95% confidence/credible intervals. The proportion above/below the threshold for the fixed each threshold approach was fixed at 10%. The extreme quantiles for both tails based on the GNG are slightly larger than the fixed GPD based method, with a wider confidence interval as we expect due to accounting for the additional uncertainty about the threshold which is ignored in the fixed threshold approach.

The similarity of the estimates and the only slightly larger credible intervals is extremely pleasing, as the proposed methodology has not required a-priori specification of the threshold which is a major advantage over the traditional fixed threshold approach. An interesting feature of the credible intervals for the mixture model approach is that they tend to be somewhat wider for heavier tails, as shown in Figure 5 by the larger credible intervals for the losses compared to the gains. The shape parameter estimated for the losses (not shown for brevity) is generally larger than for the gains. This result implies that the uncertainty of threshold selection is likely to be more important for the heavy tail distributions relative to the light or short tails, which is commonly the case for financial data (particular for financial crises).

Notice from Figure 5 that the larger credible intervals for the GNG model based quantile estimates, have better coverage of the actual returns than for the fixed threshold approach. The lower bound of the lower tail credible interval is apprximately the same for both models. However, the upper bound of the lower tail credible interval is closer to the mode. For heavier tails (as in the losses in this example), the threshold related uncertainty appears as more uncertainty nearer the mode of the distribution, rather than further out into the lower tail. In contrast the gains credible interval for the GNG model extends further out into the upper tail than the corresponding interval for the fixed threshold model, whereas the lower bounds are very similar. The gains have a short tail (generally a negative shape parameter), hence the uncertainty associated with estimating the threshold leads to more uncertainty about the upper tail.

6. Conclusion

Extreme value theory based models have been widely used in financial applications when assessing financial risk as they supply a statistically justifiable and flexible method for extrapolating tail distributions. In this paper, we propose an approach for forecasting the VaR combining a classical conditional variance model (GARCH) and a new GPD based mixture model, to overcome the difficulty of describing the dependence of the extreme returns (from volatility clustering) and the challenge of threshold selection in traditional GPD applications. The proposed mixture model is able to account for the uncertainty associated with the threshold choice in estimating the VaR, as the threshold is an explicit parameter of the model to be estimated. As the threshold is estimated as part of the inference process, the model fitting (including threshold choice) is easily automated for large scale application to multiple financial time series which is very challenging for the traditional approaches to GPD threshold choice.

The proposed mixture model is very flexible, permitting both symmetric and asymmetric tail behaviours in the gains/losses. A simulation study has shown the performance of a Bayesian inference approach for fitting the new GPD mixture model. The mixture model was also applied to various different population distributions (symmetric and asymmetric) and was shown to provide good approximations to various high quantiles, and in particular being only slightly less efficient compared to using the correct model for estimating these quantiles. This latter simulation study has shown the general applicability of the proposed modelling approach.

The choice of the distribution used to capture the main mode of the distribution in the mixture model (normal considered in this paper) was shown in the simulation study to affect the performance of the model in capturing the quantiles. The normal distribution is suggested for the financial applications in this paper, due to their inherent unimodal, approximately symmetric and quadratic shape around the mode. However, for applications where an asymmetric mode is expected alternative distributions for the bulk should be considered, e.g. a Weibull or gamma distribution. Alternatively, a natural extension for the bulk distribution would be a mixture of uniform distributions extending the approach of Tancredi et al. (2006).

The model was then applied to forecast the VaR for the stock market index S&P100 for the recent financial crisis period (2008) and showed distinct advantages in estimating the extreme quantiles whilst automatically accounting for the uncertainty due to the threshold choice.

References

Bali, T. G., Weinbaum, D., 2007. A conditional extreme value volatility estimator based on high-frequency returns. Journal of Economic Dynamics and Control 31, 361–397.

Behrens, C., Lopes, H., Gamerman, H., 2004. Bayesian analysis of extrme events with threshold estimation. Statistical Modelling 4, 227–244.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. Statistics of Extremes: Theory and Applications. Wiley Series in Probability and Statistics.

Chan, N., Deng, S., Peng, L., Xia, Z., 2007. Interval estimation of Value-at-Risk based on GARCH models with heavy-tailed innovations. Journal of Econometrics 137, 556–576.

Coles, S., 2001. Introduction to Statistical Modelling of Extreme Values. Springer-Verlag.

Davison, A., Ramesh, N., 2000. Impact of urbanization on coastal wetland structure and function. Journal of the Royal Statistical Society Series B 62, 191–208.

Duffie, D., Pan, J., 1997. An overview of Value-at-Risk. Journal of Derivatives Spring, 7–49.

Dupuis, D., 1998. Exceedances over high threshold: A guide to threshold selection. Extremes 1, 251–261.

Embrechts, P., Klüppelberg, C., Mikosch, T., 2003. Modelling Extremal Events for Insurance and Finance. Springer-Verlag.

Ferro, C. A. T., Segers, J., 2003. Inference for clusters of extreme values. Journal of the Royal Statistical Society, Series B 65 (2), 545556.

Frigessi, A., Haug, O., Rue, H., 2002. A dynamic mixture model for unsupervised tail estimation without threshold selection. Extremes 5, 219–235.

Gelman, A., Carlin, J., Stern, H., Rubin, D., 2004. Bayesian Data Analysis. Chapman and Hall.

Gençay, R., Selçuk, F., Ulugülyağci, A., 2003. High volatility, thick tails and extreme value theory in value-at-risk estimation. Insurance: Mathematics and Economics.

Jorion, P., 2000. Value-at-Risk: The new benchmark for managing financial risk. McGraw-Hill: New York.

McNeil, A. J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series an extreme value approach. Journal of Empirical Finance 7, 271–300.

Pauli, F., Coles, S., 2001. Penalized likelihood inference in extreme value theory. Journal of Applied Statistics 28, 547–560.

Smith, R., 1989. Extreme value analysis of environmental time series: an application to trend detection in ground-level zone. Statistical Science 3, 367–393.

Tancredi, A., Anderson, C., O'Hagan, A., 2006. Accounting for threshold uncertainty in extreme value estimation. Extremes 9, 87–106.

Zhao, X., 2009. Extreme value modelling with application in finance and neonatal research. PhD thesis, University of Canterbury.

Zhao, X., Oxley, L., Scarrott, C., Reale, M., 2009. Extreme value GARCH modelling with Bayesian inference. In: Proc. of MODSIM2009 conference. Also Department of Economics, University of Canterbury, Working Paper: 05/2009, http://www.econ.canterbury.ac.nz/RePEc/cbt/econwp/0905.pdf.

Zhao, X., Scarrott, C., Oxley, L., Reale, M., 2010. GARCH dependence in extreme value model with Bayesian inference. To appear in Mathematics and Computers in Simulation.