Applied Financial Economics, 2010, **20**, 63–72

Extreme value modelling for forecasting market crisis impacts

Xin Zhao[a], Carl Scarrott[a,*], Les Oxley[b] and Marco Reale[a]

[a]*Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand*
[b]*Department of Economics, University of Canterbury, Christchurch, New Zealand*

This article introduces a new approach for estimating Value at Risk (VaR), which is then used to show the likelihood of the impacts of the current financial crisis. A commonly used two-stage approach is taken, by combining a Generalized Autoregressive Conditional Heteroscedasticity (GARCH) volatility model with a novel extreme value mixture model for the innovations. The proposed mixture model permits any distribution function for the main mode of the innovations, with the very flexible Generalized Pareto Distribution (GPD) for the upper and lower tails. A major advance with the mixture model is that it overcomes the problems with threshold choice in traditional methods as it is treated as a parameter in the model to be estimated. The model describes the tail distribution of both the losses and gains simultaneously, which is natural for financial applications. As the threshold is treated as a parameter, the uncertainty from its estimation is accounted for, which is a challenging and often overlooked problem in traditional approaches. The model is shown to be sufficiently flexible that it can be directly applied to reliably estimate the likelihood of impact of the financial crisis on stock and index returns.

I. Introduction

Investment banks and financial corporations often want to reliably estimate the Value at Risk (VaR) for their portfolios. The developing 2008–2009 financial crisis is not only the worst and largest crisis in financial history, but also one of the most underestimated events in modern financial history. The measurement of risk is therefore attracting more attention from financial investors than ever before. VaR is a measure of the market risk as the expected financial reverse movements over a certain period, or the expected tail distributions for both losses and gains. Statistically, the key issue of VaR is to reliably describe the distribution of the tail behaviour for the holding portfolios. There are three common statistical approaches to estimate VaR: nonparametric methods which frequently use sample quantiles; parametric methods such as myriad of volatility models; and various Extreme Value Theory (EVT) based methods, where commonly the Generalized Pareto Distribution (GPD) is assumed for the tail distribution.

Nonparametric methods are rather uncertain due to the inherent lack of sample information in the tails of the distribution. Parametric methods such

*Corresponding author. E-mail: carl.scarrott@canterbury.ac.nz

Applied Financial Economics ISSN 0960–3107 print/ISSN 1466–4305 online © 2010 Taylor & Francis
http://www.informaworld.com
DOI: 10.1080/09603100903262947

as the Autoregressive Conditional Heteroscedasticity (ARCH) and Generalized ARCH (GARCH) model are known to frequently underestimate or overestimate tail risk (McNeil and Frey, 2000) due to the residuals often exhibiting heavier tailed rather than the normal or *t*-distribution which are usually assumed. Traditional EVT approaches are challenging to adapt to cope with serial dependence seen in financial return series.

There are several solutions to these issues, some of which are now discussed. Parametric financial models and the EVT theory have been combined in various forms to capture the impact of the heteroscedastic/dependence on the tail behaviour of the return series (see for example Pauli and Coles, 2001). One approach is to introduce a time structure on the parameter of the GPD to capture the heteroscedastic process, see Bali and Weinbaum (2007) or Zhao *et al.* (2009), for example. However, among other difficulties with this approach is defining a stationary time series structure based on irregularly spaced time series. McNeil and Frey (2000) put forward a two-stage approach by combining the GARCH model to estimate volatilities and then use the EVT based GPD to model the tail of the standard innovations of the GARCH above some chosen thresholds. After removal of the stochastic volatility, the standard innovations are approximately independent and identically distributed (i.i.d.), permitting the use of the GPD as an asymptotically justified model for the tail.

A key problem with the above two-stage method is the threshold selection for the GPD (see Beirlant *et al.*, 2004) and accounting for the uncertainty due to the selection. The threshold must be sufficiently high to ensure the threshold excesses asymptotic arguments used to justify the GPD provide a reliable approximation (see Smith, 1989). However, the variance of the parameter increases if the threshold is too high, due to the reduced sample information. There is always trade off between these two conditions. Various graphical techniques (see Beirlant *et al.*, 2004) based on properties of the GPD are typically used for threshold selection. However, these techniques are often criticized for being rather subjective in nature. In many applications, automated threshold estimation would be beneficial, e.g. in forecasting or applications with many series for the model to be applied to. Further, once the threshold is selected, it is typically treated as a known fixed quantity and so the uncertainty surrounding its estimation is not accounted for.

Several approaches have recently been taken to estimate the threshold. Dupuis (1998) put forward a semi-automated method of robust threshold selection. However, some subjective assessment is still required. Frigessi *et al.* (2002) suggest a dynamically weighted mixture model of a GPD and a light-tailed distribution for the bulk of the distribution, with a smooth weight function to transition between the two distributions, to replace the usual threshold choice. Another similar mixture model is developed by Behrens *et al.* (2004) by using a mixture of truncated Gamma for nonextremes and the GPD for one side extremes. The threshold is a model parameter that is estimated directly with the others. Both of these mixture models assume a certain form for the distribution below the threshold. Tancredi *et al.* (2006) propose a less-restrictive approach to overcome the lack of a natural model below the threshold distribution as adapting an unknown number of uniform distributions. The threshold itself is again a parameter of the model to be estimated.

These mixture models essentially combine the GPD for one of the tails (typically the upper tail) with another distribution for the bulk of the distribution. In some applications, interest lies in both tails (e.g. loss and gain for finance applications). A mixture model, where both tails are GPD is developed in this article to capture both tails simultaneously and account for the uncertainty about the thresholds in any inferences made. Clearly this extension provides a very flexible model for capturing all forms of tail behaviour, potentially allowing for asymmetry in the distribution of two tails.

In this study, we are interested in quantifying the tail related VaR risk measure using the two-stage McNeil and Frey (2000) approach, but with our two GPD tail mixture model. More specifically, we define a mixture distribution with three components: a GPD for losses tail, a GPD for gains tail and the normal for the innovations between main mode between the two GPD thresholds. The distribution selected for the nonextreme data can affect estimation of the tail distribution and therefore it is necessary to choose it according to the application. The normal distribution is suggested for the financial applications in this article, due to their inherent unimodal, approximately symmetric and quadratic shape around the mode. A Bayesian method of estimation is used for the mixture model as it can take the advantage of any expert prior information (see Gelman *et al.*, 2004), which is important in tail estimation due to the inherent sparsity of data.

The proposed method is applied to the forecasting of VaR on daily return series of both individual stock and market index including the period of the most recent world financial crisis (year 2008). The two-stage mixture modelling approach is described in Section II. Section III gives the empirical results from application to returns from a stock and index. Conclusions and discussion are supplied in Section IV.

II. Model and Estimation

The GARCH–GPD mixture model

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

(1) Fit the $\{R_t\}$ with a GARCH volatility model to obtain the standardized innovation term x_t as

$$R_t = E(R_t) + v_t x_t \tag{1}$$

here, the $E(R_t)$ is the expected return at time t and v_t is the volatility estimators from GARCH model. The form of GARCH can be changed according to applications.

(2) Fit the GPD–Normal–GPD mixture model (henceforth denoted GNG) to $\{x_t\}$ (the standardized innovation sequence). The distribution function of the GNG model is defined as

$$F(x|m,s,\xi_r,\sigma_r,u_r,\xi_l,\sigma_l,u_l) = \begin{cases} \Phi(u_l|m,s)[1-G(-x|\xi_l,\sigma_l,-u_l)] & x \le u_l \\ \Phi(x|m,s) \quad u_l < x < u_r \\ \Phi(u_r|m,s) + [1-\Phi(u_r|m,s)]G(x|\xi_r,\sigma_r,u_r) \\ \quad x \ge u_r \end{cases} \tag{2}$$

where $\Phi(x|m,s)$ 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 with shape ξ, scale σ and threshold u.

The subscripts on the GPD parameters are l for the left 'losses' tail and r the right 'gains' tail, e.g. u_r is the threshold for the gains tail. For brevity we have dropped the subscripts in the GPD cumulative distribution function:

$$G(x|\xi,\sigma,u) = 1 - \left[1+\xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi}, \quad \left\{x\in\Re,\ x>u,\ 1+\xi(x-u)/\sigma>0\right\} \tag{3}$$

where the location $u \in \Re$, scale $\sigma > 0$ and $\xi \in \Re$. Therefore, the parameter vector of the GNG model is $\theta(m,s,\xi_r,\sigma_r,u_r,\xi_l,\sigma_l,u_l)$. The special case of $\xi = 0$ is considered in the limit as $\xi \to 0$ as:

$$G(x;\sigma,u) = 1 - \exp\left(-\frac{x-u}{\sigma}\right), \quad \{x \ge u\} \tag{4}$$

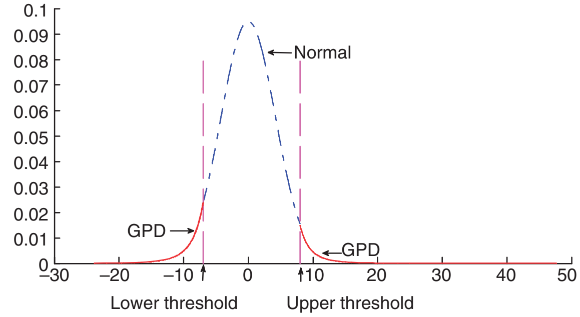

Fig. 1. Example of GNG mixture density

The components of the GNG in Equation 2 given by $\Phi(u_l|m,s)$ is the proportion of the population below the lower threshold and $1-\Phi(u_r|m,s)$ the proportion above the upper threshold. The proportion below the lower threshold supplies the information of the unexpected shortfalls and the proportion above the upper threshold supplies the information of the unexpected rise, respectively, in the VaR.

Figure 1 gives an example of a GNG density. The two vertical dashed lines represent the thresholds for the GPD tails. The density is not necessarily continuous at the thresholds. However, frequently in applications the density is close to continuous and any lack of continuity is typically of no concern if interest is only in the risk of extremes.

Although the extra parameters of the GNG model increase the complexity for statistical inference, the model has numerous advantages:

(1) The application of traditional GPD requires what is often a rather subjective choice of threshold. The proposed mixtures explicitly define the threshold as a parameter to be estimated using the usual inference techniques.
(2) Relatedly, the uncertainty of estimating the threshold is directly accounted for in estimation of the mixture model parameters, which is more complex for the traditional fixed threshold approach.
(3) The model allows both tails to follow the GPD distribution, explicitly allowing for asymmetry in the tails. It therefore has the flexibility in dealing with a variety of distributions with or without the symmetric features.
(4) The proposed mixture GPD model is able to capture two-sided tails distribution simultaneously, which is more applicable than a one-sided tail mixture GPD.

Estimating the extreme return quantile

If we consider only the observation in the upper tail modelled using the GPD, then its distribution function in Equation 4 can be used to show that the $1-p$ quantile can be expressed as

$$q_p(x) = \begin{cases} u - \frac{\sigma}{\xi}(1 - p^{-\xi}) & \xi \neq 0 \\ u - \sigma \log(p) & \xi = 0 \end{cases} \quad (5)$$

where q_p is commonly referred to as the return level with a return period of $1/p$. An equivalent formula can also be determined for the lower GPD tail.

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.

$$E(R_{p^*,t}) = E(R_t) + v_t q_p(x) \quad (6)$$

where $p^* = p_u \cdot p$ and p_u is the proportion below the lower threshold (for the fall of returns) or the proportion above the upper threshold (for the rise in returns). Typically, p_u is estimated using the sample proportion above/below the threshold.

Since the proposed GNG model describes the entire sample distribution we simply use the inverse of the distribution in Equation 2 to get the expected return quantile $q_p(x)$ with a return period of $1/p$ directly:

$$E(R_{p,t}) = E(R_t) + v_t q_p(x) \quad (7)$$

As the normal distribution is used for the main mode, standard numerical approximations are used in calculating the inverse.

When forecasting, 1-step prediction of the conditional quantile at q, which is the lower quantile p and upper quantile $1-p$ of return, is

$$R_{q,t}(1) = \inf\{F(R_t) \geq q | \varphi_{t-1}\} \quad (8)$$

where φ_{t-1} is the information up to day $t-1$.

Method of estimation

Stage 1: GARCH model. After choosing the form of GARCH type volatility model, we fitted the model using Maximum Likelihood (ML), as this is the most common inference approach used in the literature. The expected return $E(R_t)$ and the volatility v_t are the 1-step forecasts. The observed standard innovation series are then used for the second stage.

Stage 2: GNG model. The inherent sparsity of extreme data means that it is often sensible to try and utilize information from all sources for inference purposes, e.g. expert knowledge. We have therefore used Bayesian inference for the GNG model parameters so that prior information, along with the sample data can be utilized. Markov Chain Monte Carlo (MCMC) has been used to obtain posterior distributions (see Green, 1995). A brief outline of the priors is given below for brevity, as those chosen are fairly standard in the extremes literature with full details in Zhao (2009) and available upon request.

Specify the prior distributions

The parameter vector $\theta(m, s, \xi_r, \sigma_r, u_r, \xi_l, \sigma_l, u_l)$ can be decomposed into three components, those associated with the normal, GPD parameters and the thresholds. A normal prior is for the location m and a gamma prior for scale parameter s, assuming m and s are independent. Following Coles and Tawn (1996) the GPD priors are specified on the quantile differences, which is becoming standard in the extremes in the literature. A truncated normal distribution is used as the prior distribution for thresholds of both tails, which are truncated at the minimum and maximum of the sample data respectively, due to Behrens *et al.* (2004). Hyperparameters are chosen to ensure relative diffuse priors to indicate our lack of prior knowledge of the parameters for these applications.

III. Application in Financial Crises

Comparison of models

We compare the fixed GPD and the proposed GNG based methodologies using results from application to the daily returns of the Citigroup and S&P100 index over the period 17 January 2001 to 30 December 2008. These include 2000 daily prices and two major financial crises, namely 9/11 and the developing 2008–2009 financial crisis. We use a GARCH(1,1) for the volatility model. Although not shown for brevity, after removing the conditional variance (volatility) the standardized innovations $\{x_t\}$ are essentially an independent sequence, which is beneficial for inference purposes for the second stage which assumes independence.

It is clear from Fig. 2 that both the Gaussian and t-distributions do not adequately represent the tail innovation distributions of the Citigroup, as they are sufficiently inflexible at representing the tail shape. The left plot in Fig. 2 is the empirical quantiles versus the normal quantiles. It is clear that both tails are heavier than the normal tail. The right plot is the empirical quantiles versus the t-quantiles and it shows a lighter tail than the t-distribution. The Q–Q plots for the S&P100 index gives similar results as the Citigroup, so are not shown for brevity. It is clear that a model which provides more flexibility in

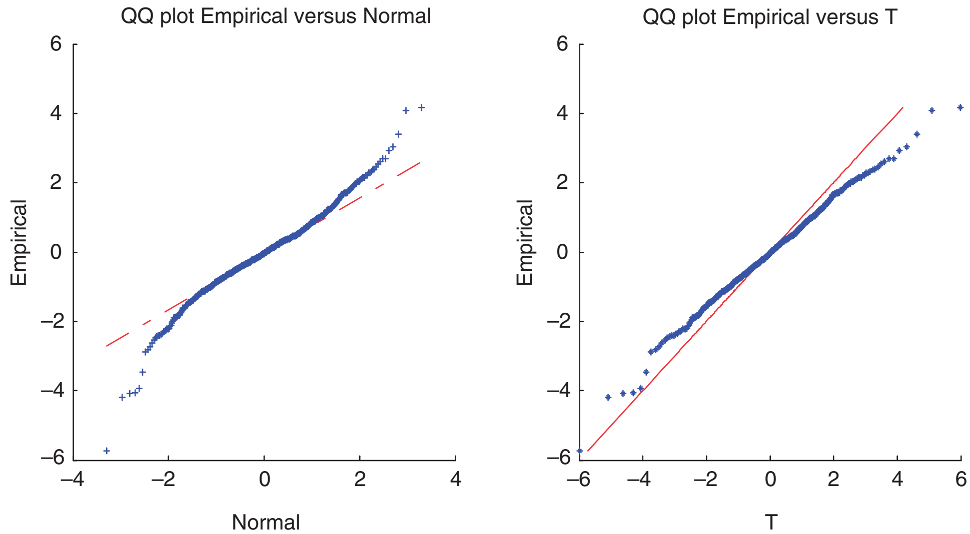

Fig. 2. QQ plot of the standardized innovation from the Citigroup 2001–2008

the tails of the distribution of the standardized innovations is required, particularly as it is the tails that are key for reliably estimating the expected fall or rise from the expected return.

The mixture GPD model GNG proposed in these articles, along with the traditional fixed GPD approach, provide flexibility in capturing different tail behaviours. The model is suitable for any combinations of three types of tail behaviours with the typed determined by the shape parameter of the GPD, namely exponential type ($\xi = 0$), heavy tailed ($\xi > 0$) and short tailed ($\xi < 0$). The models also permits asymmetric tail behaviours (e.g. lower tail could be of exponential form and upper tail a heavy tail). The tails can be constrained to be symmetric if required, and various tests for evidence of asymmetry can be derived, see Zhao (2009).

In order to show the problem of the threshold selection of the fixed threshold method with the model proposed, we compare the estimation results of the two approaches in fitting the innovations of these financial time series. We use the ML Estimation (MLE) for the fixed threshold GPD, since this is the common approach in the literature. The use of relatively diffuse priors in the Bayesian inference for the GNG model means the results are directly comparable. Table 1 illustrates the effect of threshold selection in the traditional fixed threshold GPD approach. The GPD is fitted to the standardized innovation term of the Citigroup data with different thresholds and compared to the GNG mixture model. The choice of the threshold value results in variation of the GPD shape and scale estimators, and consequently for estimates of the VaR. Even though these variations appear rather slight for this data, the differences can have a substantial impact on the quantile and VaR estimates.

Although, the Bayesian Credible Interval (CI) and MLE based confidence intervals have philosophically different interpretations, they are to some degree heuristically comparable. The CIs for the shape and scale parameters of the GNG model, are wider than the confidence interval of the fixed threshold GPD as expected, as this in part comes from the extra source of threshold uncertainty. For heavy tailed distributions, a wider confidence interval is expected since the uncertainty about the threshold is much higher compared to the short tailed distributions, where there is a natural end-point for the threshold value.

Table 2 reports the estimated expected quantiles of the standardized innovations for the threshold choices in Table 1, to show the impact on threshold choice. The expected quantiles of the fixed threshold GPD are calculated as the return level of the GPD explained in Section II. The expected quantiles for the GNG are calculated as the mean of the posterior predictive quantiles. It is clear that the expected quantile values vary when the threshold changes and the differences among them gets larger for quantiles further out into the (upper and lower) tails of the distribution. As mentioned previously, these differences tend to be magnified for heavier tailed distributions, which are commonly observed in financial data. Different choices of threshold can clearly lead to very different estimates of the extreme quantiles used in risk measures like VaR. The GNG explicitly estimates both thresholds at the same time, and accounts for the uncertainty in the estimates.

Table 1. Parameter estimates for the fixed threshold GPD and the GNG (two-tail mixture GPD) models

	p_u	u_r	ε_r	CI		σ_r	CI	
Fixed threshold GPD – right	0.20	0.6716	−0.0607	−0.1498	0.0284	0.6691	0.5862	0.7638
	0.15	0.8827	−0.0194	−0.1336	0.0949	0.6076	0.5173	0.7135
	0.10	1.1178	−0.0429	−0.1769	0.0912	0.6296	0.5192	0.7634
	0.05	1.5721	0.0790	−0.1721	0.3300	0.5054	0.3677	0.6947
GNG model – right		1.4604	0.0304	−0.1897	0.2633	0.5991	0.4163	0.8012
	CI of u	0.9612	1.9182					
	p_u	u_l	ε_l	CI		σ_l	CI	
Fixed threshold GPD – left	0.20	−0.7788	0.0438	−0.0566	0.1443	0.6255	0.5437	0.7197
	0.15	−0.9649	0.0505	−0.0661	0.1671	0.6246	0.5309	0.7347
	0.10	−1.2481	0.1177	−0.0435	0.2790	0.5655	0.4575	0.6991
	0.05	−1.6410	0.1165	−0.1376	0.3705	0.6240	0.4531	0.8594
GNG model – left		−1.6149	0.1644	−0.1172	0.4323	0.6146	0.3909	0.9337
	CI of u	−1.9032	−1.1093					

Notes: The fixed GPD is fitting using MLE so confidence intervals are shown. The GNG is fitted using Bayesian inference with diffuse priors, so CIs are shown. In the table, p_u is the proportion of innovations above/below the upper/lower thresholds respectively. u denotes the thresholds; ξ is the shape; σ is the scale of GPD. 95% confidence/CIs are shown.

Table 2. Expected quantile estimators for the fixed-threshold-GPD and the GNG (two-tail mixture GPD) models, along with CIs for the GNG model

	Quantiles				
Fixed GPD – right	u	0.9	0.99	0.999	0.9999
	0.6716	1.1257	2.5043	3.7029	4.7451
	0.8827	1.1281	2.4857	3.7841	5.0259
	1.1178	1.1178	2.4983	3.7490	4.8823
	1.5721	1.2312	2.4395	3.8887	5.6269
GNG model – right	u	0.9	0.99	0.999	0.9999
	1.4604	1.1359	2.4460	3.9691	5.6174
	CI	1.0608	2.1760	3.2827	3.8861
		1.1817	3.1620	6.7840	13.5697
Fixed GPD – left	u	0.1	0.01	0.001	0.0001
	−0.7788	−1.2190	−2.7812	−4.5093	−6.4209
	−0.9649	−1.2208	−2.7774	−4.5260	−6.4903
	−1.2481	−1.2481	−2.7438	−4.7053	−7.2776
	−1.6410	−1.2254	−2.7456	−4.7333	−7.3323
GNG model – left	u	0.1	0.01	0.001	0.0001
	−1.6149	−1.2376	−2.6779	−5.0217	−8.8012
	CI	−1.1647	−2.3432	−3.8287	−4.7203
		−1.3070	−3.0970	−7.4052	−21.9675

The variation considered by the GNG model represents the full uncertainty for all the model parameters. Figure 3 gives the sample quantiles from the predictive posterior quantile distributions based on the GNG estimates for the Citigroup data. As expected, the posterior quantile is more skewed for the higher quantiles, due to the reduced information in the tails. The predictive posterior for the upper tail is right skewed, summarizing the lack of information for higher levels, and *vice versa* for the lower tail quantiles. Notice that the sample quantiles are included for comparison and are positioned at the mode and well within the CI, which demonstrates that the predictive quantile distributions are consistent with the sample information.

Forecasting the extreme quantiles

We use the proposed two-stage method to conduct the 1-step forecasts of daily return quantiles R_t^q for

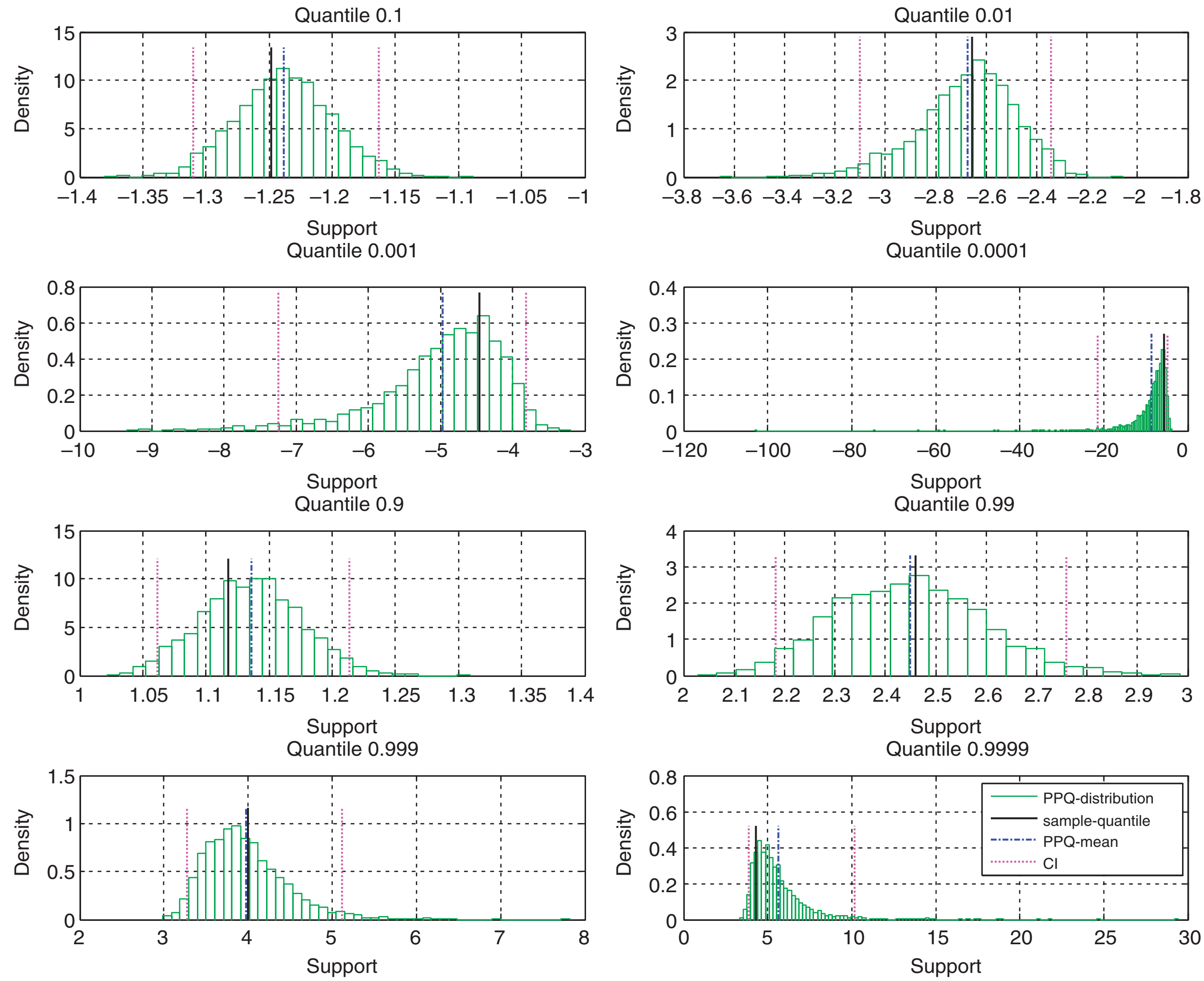

Fig. 3. Predictive posterior quantile distributions for the Citigroup 2001–2008 innovations

S&P100 index and Citigroup. Suppose the return sequence is $R_1, \ldots, R_t, \ldots, R_T$, 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 quantile $\hat{R}_t^q$ are then dependent on all the information up to $t - 1$.

The scheme is as follows at time t:

(1) Apply GARCH(1, 1) model on $R_{t-n-1}, \ldots, R_{t-1}$ to obtain the standardized innovation sequence $x_{t-n-1}, \ldots, x_{t-1}$.
(2) Obtain the 1-step forecasting of the expected return $E(\hat{R}_t)$ and the volatility $\hat{v}_t$ based on the estimators of step 1.
(3) Apply the GNG model on the standardized innovation sequence $x_{t-n-1}, \ldots, x_{t-1}$, and forecast the $\hat{x}_t^q$ based on the predictive posterior quantile distributions.
(4) The forecasted return quantile can be calculated as

$$\hat{R}_t^q = E(\hat{R}_t) + \hat{v}_t \hat{x}_t^q \tag{9}$$

using the obvious notation developments from Equation 7.

Repeat all the steps for each time point to obtain a sequence of conditional forecasted return quantiles $\{\hat{R}_t^q\}$ for the forecasting period. These forecasted return quantiles are termed 'conditional' as they are conditional on the variance being assumed known, using the estimates from the GARCH. We also estimate the unconditional quantiles for comparison, which are the quantile estimators by applying the GNG model directly on the return series. The unconditional quantiles do not account for the dependence due to the volatility.

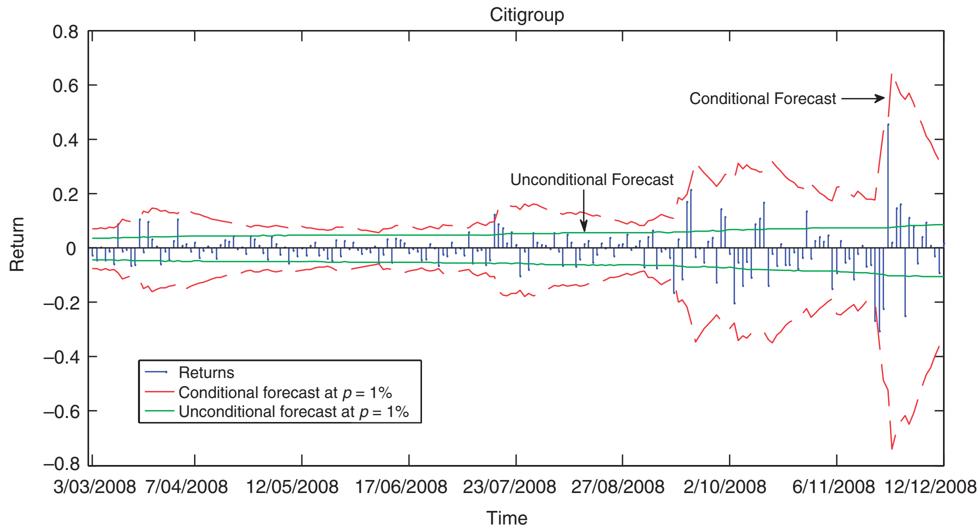

Fig. 4. VaR of the Citigroup standardized innovations in 2008, comparing the conditional GNG and an unconditional approach (where the stochastic volatility is ignored)

Figure 4 gives the plot of real return observations of the Citigroup and forecasted conditional and unconditional quantiles for the 1% upper and lower quantiles. We have focused on the period from March 2008 to December 2008, as this covers the recent financial crisis. As expected, the unconditional quantile estimators fail to capture the changing volatility and dependence of the returns. The conditional forecasts, on the other hand, can describe both the dynamic of the conditional variance and the tails behaviour of the expected falls and rises. The conditional approach based on the GARCH(1, 1) is therefore clearly warranted due to it providing more realistic description of the return quantiles. Similar results were obtained for the S&P100 index and are not shown for brevity.

A comparison between our method and the fixed threshold approach is now conducted. To implement this method in a forecasting framework it is necessary to choose the thresholds at each timepoint, which is clearly impractical. It is common in the literature to determine a threshold in advance, by specifying what proportion of the observations at each timepoint should above/below the threshold, using graphical diagnostics the details of which are not shown here for brevity. Figures 5 and 6 show the comparison between the GNG and the fixed threshold GPD for the upper and lower quantiles at the 0.5% level with associated confidence intervals. The proportion above/below the threshold for the fixed each threshold approach was fixed at 10%. Extensive comparisons with other sensible proportions above/below the thresholds provided no change in the conclusions drawn. Notice that 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 uncertainty about the threshold.

The similarity of the estimates and the only slightly larger CIs 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 GPD based method.

An interesting feature of the CIs for the GNG approach is that they tend to be somewhat wider for heavier tails, as shown in Figs 5 and 6 by the larger confidence intervals for the losses compared to the gains. Table 1 confirms that the losses have a larger shape parameter than the gains, thus indicating a heavier tail. 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.

Summary

To summarize the above empirical results the proposed GNG model based method of forecasting has the following advantages over the traditional fixed threshold GP approach:

- The model resolves the difficulty in threshold selection by treating the threshold as a further parameter to be estimated. This advantage becomes extremely beneficial for problems where automated application is required, as in the above forecasting example and applications containing a large number of series to apply the approach to. The automated threshold

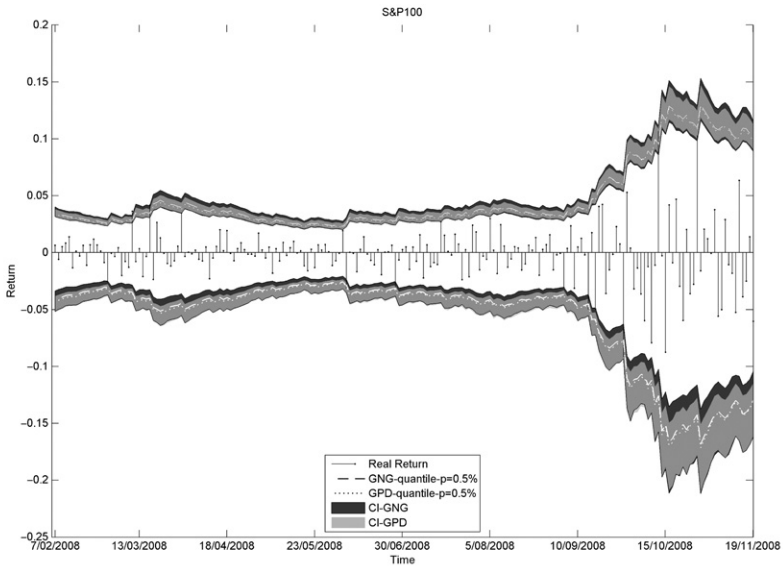

Fig. 5. Conditional quantile forecast from the GNG and the fixed-threshold GPD model for the S&P100 index standardized innovations in 2008

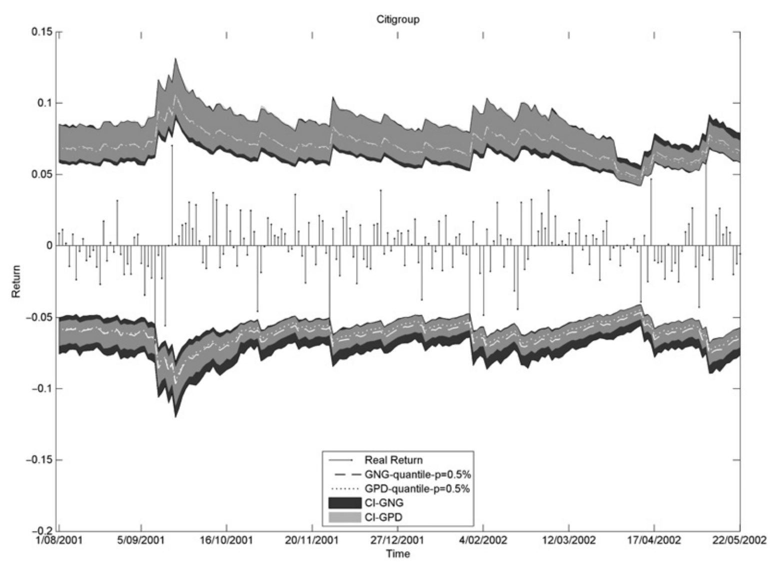

Fig. 6. Conditional quantile forecast for the GNG and the fixed-threshold GPD model for the Citigroup standardized innovations in 2001

estimation also removes the subjectivity employed in the common methods used for threshold choice (e.g. using the mean residual life plot).

- The model explicitly accounts for the uncertainty due to the threshold choice.
- The model fits to both the gain and loss tails simultaneously which is more natural and convenient when the interest lies in both tails.
- The model supplies a more objective approach in predicting the extreme shortfall and rise in the VaR along with a more precise uncertainty interval associated with the unexpected returns.

The obvious drawback with this type of approach is the *a priori* use of the normal distribution to capture the main peak. In many applications this will be appropriate, as the density around the mode is often approximately symmetric and quadratic in shape. However, it is possible that in some applications that other distributions, e.g. an unknown number of uniform distributions as Tancredi *et al.*

(2006), may be more appropriate for which it would be trivial to extend the outlined methodology. Further research (Zhao, 2009) is underway to show the general performance of the GNG model at approximating a wide variety of distributions, with different tail behaviours and modal shapes, which is available upon request. Another drawback of the model is the discontinuity of the density at thresholds, which is a fairly minor problem particularly as in most application interest in extreme quantiles well away from the threshold.

IV. Discussion and Conclusions

In this article, we proposed an approach for forecasting the VaR in combining the classical GARCH model and a new GPD mixture model. The proposed model overcomes the difficulty in threshold selection in traditional approaches. Further, the threshold is an explicit parameter of the model to be estimated, and therefore the uncertainty associated with the threshold selection is accounted for in forecasting the VaR. The mixture model for the GARCH innovations gives a flexible asymmetric distribution, with the GPD for capturing the gains and losses, tail behaviours and a normal distribution for the main mode. We applied the model in forecasting the VaR for the Citigroup and S&P100 for two financial crisis periods, 2001 and 2008, to demonstrate the advantages of the model for estimating the VaR.

A natural extension/adaption of the model is to consider an alternative distribution for the main mode of density, as discussed above. In this application a normal distribution was shown to be appropriate, however, in some applications it is likely that an asymmetric distribution may be appropriate, which is trivially captured in the above framework. A more flexible approach may be to adopt or extend the approach of Tancredi *et al.* (2006) to allow an unknown number of uniform distributions for the main mode of the density between the thresholds.

References

Bali, T. G. and Weinbaum, D. (2007) A conditional extreme value volatility estimator based on high-frequency returns, *Journal of Economic Dynamics and Control*, **31**, 361–97.

Behrens, C. N., Lopes, H. F. and Gamerman, D. (2004) Bayesian analysis of extreme events with threshold estimation, *Statistical Modeling: An International Journal*, **4**, 227–44.

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

Coles, S. and Tawn, J. A. (1996) A Bayesian analysis of extreme rainfall data, *Applied Statistics*, **45**, 463–78.

Dupuis, D. J. (1998) Exceedances over high thresholds: a guide to threshold selection, *Extremes*, **1**, 251–61.

Frigessi, A., Haug, O. and Rue, H. (2002) A dynamic mixture model for unsupervised tail estimation without threshold selection, *Extremes*, **5**, 219–35.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004) *Bayesian Data Analysis*, Chapman & Hall/CRC, Boca Raton.

Green, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, **82**, 711–32.

McNeil, A. J. and 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. and Coles, S. G. (2001) Penalized likelihood inference in extreme value theory, *Journal of Applied Statistics*, **28**, 547–60.

Smith, R. L. (1989) Extreme value analysis of environmental time series: an application to trend detection in ground-level zone, *Statistical Society*, **3**, 367–93.

Tancredi, A., Anderson, C. and 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 (to be submitted), University of Canterbury, New Zealand.

Zhao, X., Scarrott, C. J., Reale, M. and Oxley, L. (2009) Extreme value GARCH modeling with Bayesian inference, in *Proceedings of the 18th World IMACS/MODSIM Congress*. Modelling and Simulation Society of Australia and New Zealand, pp. 1391–7.