GARCH dependence in extreme value models with Bayesian inference

Xin Zhao [a], Carl John Scarrott [a,*], Les Oxley [b], 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*

Abstract

Extreme value methods are widely used in financial applications such as risk analysis, forecasting and pricing models. One of the challenges with their application in finance is accounting for the temporal dependence between the observations, for example the stylised fact that financial time series exhibit volatility clustering. Various approaches have been proposed to capture the dependence. Commonly a two-stage approach is taken, where the volatility dependence is removed using a volatility model like a GARCH (or one of its many incarnations) followed by application of standard extreme value models to the assumed independent residual innovations.

This study examines an alternative one stage approach, which makes parameter estimation and accounting for the associated uncertainties more straightforward than the two-stage approach. The location and scale parameters of the extreme value distribution are defined to follow a conditional autoregressive heteroscedasticity process. Essentially, the model implements GARCH volatility via the extreme value model parameters. Bayesian inference is used and implemented via Markov chain Monte Carlo, to permit all sources of uncertainty to be accounted for. The model is applied to both simulated and empirical data to demonstrate performance in extrapolating the extreme quantiles and quantifying the associated uncertainty.
© 2010 IMACS. Published by Elsevier B.V. All rights reserved.

Keywords: Extreme values; Dependence; Bayesian inference; GARCH

1. Introduction

Modelling the tails of distributions is important in many fields, such as environmental sciences, hydrology, insurance, finance where the risk of unusually large or small events are of interest. Extreme value theory considers the distributional behaviour of the extremes of random variables, towards extrapolating the stochastic dynamics of a process to states with small chances of realization and typically beyond the range of observed data. It is always challenging to justify the form of extreme models and to estimate parameters due to the inherent sparsity of observations in the tails. Hence, asymptotically justified parametric models are typically used to represent the data generating process, which can provide the reliable extrapolations required in such applications.

* Corresponding author at: Department of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch, Canterbury, New Zealand. Tel.: +64 03 3642600; fax: +64 03 3642587.
  E-mail address: Carl.scarrott@canterbury.ac.nz (C.J. Scarrott).

The generalised extreme value (GEV) distribution is one such asymptotically justified model used to describe the limiting distribution of the maximum or minimum of a sequence of independent and identically distributed random variables. Under certain conditions [4], which include short range dependence between the observations, the distribution of the maximum or minimum remains within the same GEV distribution family. However, these results are not necessarily helpful in describing the detailed form of the dependence and cannot always describe all the impacts of the dependence on inferences for tail related measures. Financial data is well known to exhibit heavy tails and extreme value theory has been shown to be a very useful tool in estimating and predicting the extremal behaviour of actuarial and financial products, such as predicting the largest claim in insurance and the Value at Risk (VaR), see [11]. However, applying extreme value models is not always straightforward because of dependence and/or nonstationary behaviour. For example, financial returns typically show clusters of observations in the tails, often termed volatility clustering.

There are three common approaches to account for the impact of dependence on inferences. One of these approaches is to decluster the dependent extremes and then apply standard extreme models to the non-dependent sequence using statistical declustering algorithms, see [12] for recent developments. Peaks-over-threshold (POT) is one such declustering approach, see [10]. A second common approach is to apply standard extreme value models, but adjust the uncertainty estimates for the dependence using techniques like block bootstrapping, see [7]. However, the substantial drawback with both these approaches is although they do not usually provide additional understanding of the dynamics of the dependence structure.

A common (two-stage) approach in the finance literature (see [17,8,23]) is to capture the dependence in the returns induced by the volatility clustering using a generalised autoregressive conditional heteroscedastic (GARCH) type model (see [5] and excellent glossary of ARCH/GARCH variants in [6] and references therein), followed by extreme value modelling of the approximately independent residual innovations. This approach overcomes the dependence in the extremes and explicitly captures the dependence structure itself, allowing further insight to be gained. The principal drawback with this approach is that accounting for the uncertainties associated with both stages is not straightforward.

In this paper, we explicitly model the dependence behaviour of extremes via time-varying extreme value parameters. We have amalgamated the GARCH model, which is widely used to model dependence in the volatility of financial time series (see [5]), with the classical GEV model. Specifically, we consider a conditional GEV distribution with the specification that the extreme value sequence (maxima or minima) following an autoregressive process with a GARCH-type conditional variance structure. The dependence is captured by appropriate temporal dependence structure in the location and scale parameters of the GEV. The combination of GEV and GARCH models is shown to be better suited to explain the extreme quantiles than the classic GARCH model alone, which cannot capture the tail behaviour adequately with either normally distributed or even fatter tailed distributed (e.g. *t*) innovations as suggested by [17]. Also, an important benefit of our approach is that it is a one stage model and thus making inferences and accounting for all uncertainties much easier.

Section 2 gives the review of classical GEV model for stationary sequences. The proposed GEV-GARCH model is outlined in Section 3. The estimation method of proposed model is discussed in Section 4. This study examines the dynamics of the maximum and minimum value of intra-day financial returns with focus on explaining the extremal quantiles, which can be useful in the control of risk. The study has been conducted with both simulated and real data, the results of which are presented in Sections 5 and 6. Further results on an extension to our approach considered by [3] are provided in Section 7, which considers what happens when the shape parameter of the GEV is also permitted to vary according to a GARCH-type structure. It is demonstrated that model identification is very challenging if the GEV shape parameter is allowed to vary over time in such a manner, as one might expect, due to the inherent difficulty in estimating this parameter because of the lack of information in the tails. Conclusions follow in Section 8.

2. Generalised extreme value distribution

The generalised extreme value distribution (GEV) is an asymptotically motivated model for describing the distribution of the maxima and minima of a realized sequence of independent random variables having the distribution function:

$$G(x|\xi, \sigma, \mu) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \quad \text{where } 1 + \xi(x-\mu)/\sigma > 0 \tag{1}$$

The GEV model has three parameters: the shape ξ, scale $\sigma > 0$ and location μ. The density of GEV is defined as

$$g(x|\xi, \sigma, \mu) = \frac{1}{\sigma} G(x|\xi, \sigma, \mu) \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-(1+1/\xi)} \text{ where } 1 + \xi(x - \mu)/\sigma > 0 \tag{2}$$

The GEV model can represent three types of tail behaviour determined by the value of shape parameter ξ: Gumbel type (exponentially decay, like upper tail of normal distribution) which is defined in the limit as $\xi \to 0$, Fréchet type (heavier than an exponential) tail ($\xi > 0$) and the Weibull type (bounded) tail ($\xi < 0$). Classical extreme value theory shows that, if a limiting distribution for the maxima/minima of a sequence of independent and identically distributed random variables exists, then it must fall into one of these three types.

3. GEV-GARCH model

The proposed model assumes that the observations come from the GEV distribution following an AR(1)-GARCH(1,1) process which describes the conditional variance of extremes. Therefore, the distribution function of the observation x_t can be written as:

$$x_t \sim GEV(\xi, \sigma_t, \mu_t) \tag{3}$$

$$G(x_t|\xi, \sigma_t, \mu_t) = \exp \left\{ - \left[1 + \xi \left(\frac{x_t - \mu_t}{\sigma_t} \right) \right]^{-1/\xi} \right\},$$

where $\sigma_t > 0$ and $1 + \xi(x_t - \mu_t)/\sigma_t > 0$, with conditional dependence:

$$x_t = \beta_0 + \beta_1 x_{t-1} + \epsilon_t$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 \epsilon_{t-1}^2$$
$$\epsilon_{t-1} = x_{t-1} - \mu_{t-1}$$

where $|\beta_1| < 1, \alpha_0 > 0, \alpha_1 > 0, \alpha_2 > 0$, and $\alpha_1 + \alpha_2 < 1$ to ensure stationarity. As previously stated, it is typically unrealistic to model ξ as a function of time, due to this parameter requiring very large amounts of the data to get reliable estimates. Therefore the shape parameter ξ, as supported by empirical evidence, is constant. A simulation study in Section 7 will show model identification and estimation problems associated with a time-varying shape parameter, as was proposed in related work by [3].

The assumed GARCH process in the above model exhibits symmetry in how the scale of the GEV distribution (and therefore variance of the returns) responds to both negative and positive shocks, due to the use of ϵ_{t-1}^2. In the IBM example considered in the application below there was only marginal evidence for asymmetric effects and hence a symmetric model (which is commonplace in the literature) was adopted. In general, however we would expect that if the intra-day returns for a particular stock or index behave asymmetrically in terms of their volatility response to the negative and positive shocks that this would be shown strongly in the tails. It is straightforward to extend the above specification to other GARCH-type formulations with asymmetry in the volatility response to negative or positive shocks, e.g. GJR-GARCH, EGARCH or QGARCH [14,19,20]. General discussion of asymmetric and symmetric effects are provided by [15,2,22,16] and catalogue of ARCH/GARCH-type models by [6]. Comparison of the model fit between the symmetric and asymmetric volatility models would also allow evaluation (and potentially formal testing) of the evidence for asymmetry in the shock effects.

The expected value and variance of x_t are:

$$E(x_t) = \mu_t + \frac{\sigma_t(\tau(1 - \xi) - 1)}{\xi}, \quad \xi < 1 \tag{4}$$

$$Var(x_t) = \sigma_t^2(\tau(1 - 2\xi) - \tau^2(1 - \xi)), \quad \xi < 0.5, \tag{5}$$

where outside of the ranges for the shape ξ the moments are infinite.

4. Bayesian inference

Bayesian inference is used for fitting the GEV-GARCH model as we can potentially take advantage of any expert prior information, which can be important in tail estimation due to the inherent sparsity of extremal data, and to account for all uncertainties in the estimation. Although, in this study we have deliberately used diffuse priors, to show our lack or prior information allowing the data to speak for themselves which demonstrates the worst case for estimation performance. Markov chain Monte Carlo (MCMC) has been used to obtain posterior distributions.

4.1. Prior distributions

The parameter vector $\theta = (\xi, \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1)$ can be decomposed into two components $\theta_a = (\xi, \beta_0)$ which are defined over the whole real line and $\theta_b = (\alpha_0, \alpha_1, \alpha_2, \beta_1)$ which are bounded. We use a normal prior on θ_a and a flat prior on θ_b, to indicate little prior information, giving:

$$\pi(\theta_a) \propto \exp\left\{-\frac{1}{2}\left(\frac{\theta_a - \mu_{\theta_a}}{\sigma_{\theta_a}}\right)^2\right\}$$

$$\pi(\theta_b) \propto I_{(\theta_b)}$$

where μ_{θ_a} and σ_{θ_a} are the location and standard deviation. $I_{(\theta_b)}$ is an indicator function which will reflect the constraints on the parameters defined in Section 3.

4.2. Posterior distribution

Given $x_t \sim GEV(\xi, \sigma_t, \mu_t)$ and the density of GEV given in Eq. (2), the posterior is given by:

$$p(\theta|x) \propto \prod_{t=1}^{T} g(x_t|\xi, \sigma_t, \mu_t) \times \exp\left\{-\frac{1}{2}\left(\frac{\theta_1 - \mu_{\theta_a}}{\sigma_{\theta_a}}\right)^2\right\} \times I_{(\theta_b)} \tag{6}$$

A random walk Metropolis-Hastings algorithm has been used to sample the posteriors, as it has the advantage of being free of functional form since the posterior distribution function is not a proper probability function. MCMC has been used to update component by component, in order of importance of the parameters. Since the shape parameter is the most important it is updated first. The rest of the components are updated as follows: ξ, α_0, α_1, α_2, β_0 and β_1. The parameter restrictions within θ_b are enforced during the sampling procedure by rejecting the draws that violate them. The stationarity constraints, $|\beta_1| < 1$, $\alpha_1 + \alpha2 < 1$ and constraint on range of support of the GEV $1 + \xi(x_t - \mu)/\sigma_t > 0$ is also imposed. The convergence of the final chain is checked by monitoring the marginal distributions of the parameters obtained from parallel chains, see [13].

5. Model performance assessment by simulation

We first check the stability of the parameter estimation method using simulated data. Three different parameter sets were considered, with 100 simulated series for each parameter set using the sample size of 2500, which is of similar length to the application in the following section. The first parameter vector set considered is close to those found for the application in Section 6. The second parameter set uses a high coefficient of the lag variance term, α_1, and the third parameter use a relative lower value compared to the first set. The different shape parameter values used for these three sets are to identify the model features for heavier tails. Only positive shape parameters are considered as negative shape parameters are typically not observed in financial data. The priors were defined as to be very diffuse: $\xi \sim N(0.1, 0.15)$, $\alpha_0 \sim U(0, 0.2)$, $\alpha_1 \sim U(0.30, 0.99)$, $\alpha_2 \sim U(0, 0.5)$, $\beta_0 \sim N(0, 0.3)$, $\beta_1 \sim U(0, 0.5)$, indicating out lack of prior knowledge of these parameters. Fig. 1 describes the posterior distribution of a simulated sample with parameter θ_2 given in Table 1 as an example. The estimated parameter value (posterior mean) and true parameter are consistent for all parameters and the true parameter value are all contained by the credible interval of posteriors (95% highest density interval of posteriors). The difference between true parameter value and estimated value is due to sample variability.
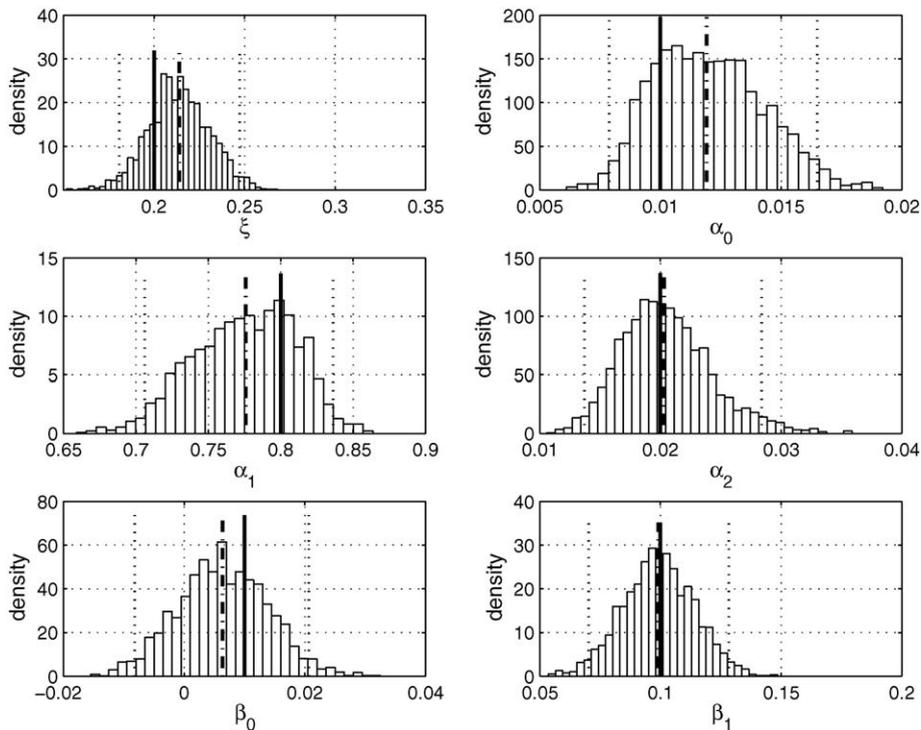
Fig. 1. Posterior distribution of parameters for simulated sample with θ_2 in Table 1. The solid line indicate the true parameter values and the dashed line represent the mean of posteriors over the credible interval, which is denoted by dotted lines.

Table 1 summarizes the results for all the simulated data. The expected posterior predictive estimates are very close to the true values (indicating low bias) and the sample variation are small. Generally, the root mean square error (RMSE) for ξ and α_1 (the coefficient of lag 1 term on the scale parameter) are higher than others, as expected, since they are the most difficult parameters to estimate. It is interesting to note that as the shape parameter value increases, the RMSE falls.

The expected variance of extreme values has been calculated according to Eq. (4) and the extreme volatility is defined as the square root of the expected variance. The volatilities from the GEV-GARCH are very close to that from a pure GARCH (AR(1)-GARCH(1,1)) for the simulated data samples, as shown in the simulated example in Fig. 2. This graph demonstrates the empirical consistency of the GARCH structure on the GEV parameters with the traditional GARCH variance estimates.

The similarity of the two volatility sequences in Fig. 2 induces the question of what is the added value of the proposed GEV-GARCH model with respect to GARCH model alone? The GARCH is traditionally combined with a normal distribution for the innovations. The GEV-GARCH can better capture the extreme quantiles by combining

Table 1

Results of fitting GEV-GARCH to 100 simulated datasets (with sample size of 2500). For each individual sample, the estimated parameter value is the mean of 95% highest density interval (HPD) of the posterior samples. The estimated parameter value reported as the mean value over 100 samples and the RMSE is also provided.

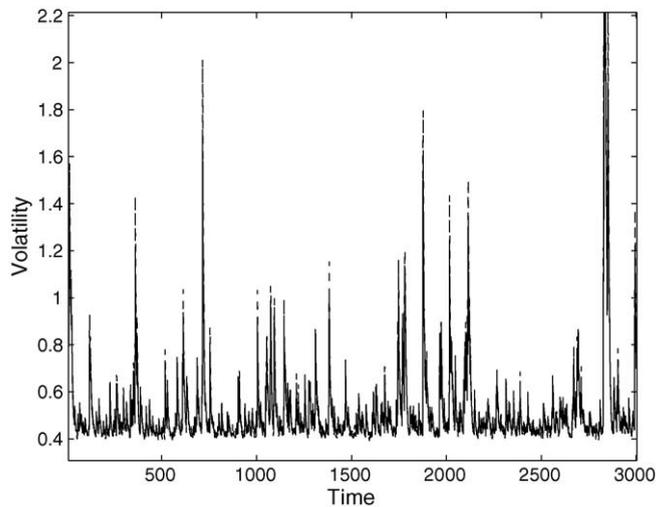| θ_1	TRUE	Estim.	RMSE	θ_2	TRUE	Estim.	RMSE	θ_3	TRUE	Estim.	RMSE
ξ	0.08	0.072	0.024	ξ	0.2	0.201	0.016	ξ	0.3	0.2970	0.017
α_0	0.01	0.009	0.002	α_0	0.01	0.010	0.002	α_0	0.05	0.051	0.004
α_1	0.45	0.480	0.082	α_1	0.80	0.793	0.038	α_1	0.50	0.498	0.019
α_2	0.08	0.078	0.013	α_2	0.02	0.021	0.004	α_2	0.10	0.099	0.009
β_0	0.21	0.210	0.007	β_0	0.01	0.010	0.006	β_0	0.05	0.051	0.011
β_1	0.32	0.321	0.016	β_1	0.10	0.101	0.012	β_1	0.20	0.201	0.015

Fig. 2. Volatilities of a single simulated dataset with θ_2 in Table 1. The solid line represents the volatilities from GEV-GARCH model and the dashed line represents the volatilities from the GARCH (with normal innovations) model.

the flexible extreme value tail behaviour and the conditional variance process. Further, the one stage approach ensures it much easier to account for uncertainties due to the volatility estimation which is more complex for the traditional two-stage approach. Fig. 3 plots the sequence of estimated quantiles using both the GEV-GARCH and GARCH (with assumed normally distributed innovations) models, along with the actual observations for this particular sample. It is clear that the 95% quantile estimates from the GEV-GARCH model has a better coverage rate of the extreme values than the corresponding GARCH quantiles, since the positive shape indicates a heavy tail from normal distribution typically assumed when using the GARCH.

The second part of our simulation study compares the GEV-GARCH to other commonly used estimators. We compare the proposed GEV-GARCH model, to the two-stage model approach of [17] and a pure GARCH model (with normal innovations) to assess the general performance in terms of explaining the extreme quantiles. In the second stage of the [17] approach, they consider only the upper tail of the innovation distribution which can be effectively modelled using a generalised Pareto distribution (GPD), which can be derived as a tail expansion of the GEV distribution outlined
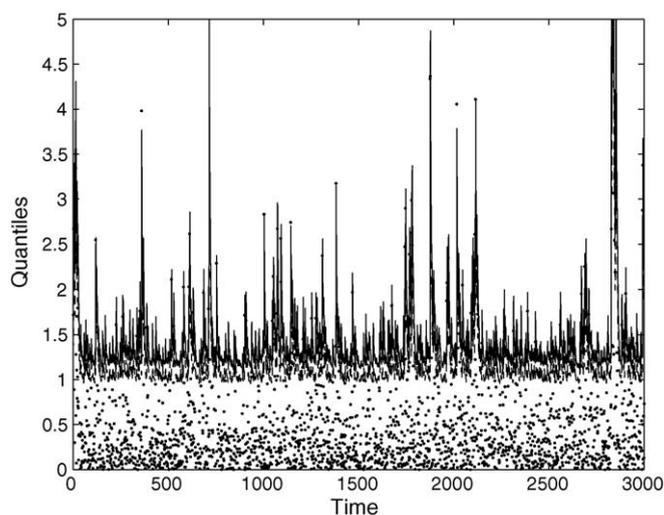

Fig. 3. Simulated sample with θ_2 in Table 1 with 95% quantiles estimates from GEV-GARCH and GARCH (with normal innovations). The dots indicate the simulated data points. The solid line represents the quantiles from GEV-GARCH model and the dashed line represents the quantiles from GARCH model (with normal innovations).

Table 2
RMSE of quantile estimates from 100 simulated datasets of length 2500, using the parameter sets defined in Table 1.

	Quantiles	0.9	0.95	0.99	0.999
θ_1	GEV-GARCH	0.0298	0.0406	0.0688	0.1214
	GARCH-GPD	0.0340	0.0471	0.0824	0.1430
	GARCH	0.0339	0.0855	0.2913	0.7257
θ_2	GEV-GARCH	0.0520	0.0760	0.1540	0.3630
	GARCH-GPD	0.1154	0.1639	0.3051	0.6110
	GARCH	0.1347	0.1862	0.7951	2.5168
θ_3	GEV-GARCH	0.0372	0.0560	0.1701	0.6891
	GARCH-GPD	0.2585	0.3338	1.1061	5.1489
	GARCH	0.2238	0.7266	3.4622	12.0951

above. The excesses above some suitably high threshold u can be shown to be well approximated by the GPD with distribution function given by:

$$
F(x|\xi, \sigma, u) = \begin{cases} 1 - \eta_u \left[1 + \xi \left(\dfrac{x - u}{\sigma} \right) \right]^{-1/\xi} & \xi \neq 0 \\[2em] 1 - \eta_u \exp \left(-\dfrac{x - u}{\sigma} \right) & \xi = 0 \end{cases} \tag{7}
$$

where $\eta_u = P(x > u)$ is the probability of exceeding the threshold u and $1 + \xi \left(\frac{x-u}{\sigma} \right) > 0$. The maximum likelihood estimate of η_u is the sample proportion. We will refer to the two-stage estimates as the GARCH-GPD.

[17] use maximum likelihood estimation when fitting the GPD to the upper tail, whereas we use Bayesian inference (except for η_u which is well estimated using the sample proportion) for direct comparison to the GEV-GARCH. The threshold choice can be problematic for GPD (see [18,21,24]), however discussion of these issues is beyond the scope of the paper. We simply chose to use the upper 25% of the data. A sensitivity analysis was carried out to assess this prechosen threshold, which showed no spurious artefacts were introduced. We carried out a complete simulation study with 100 simulated datasets for each parameter set mentioned in Table 1, the full results of which are available upon request or [24]. The key results are shown in Table 2, which compares the RMSE of the quantile estimates based on the three different models.

In Table 2 the RMSE for all three models increases with the quantile as expected, due to the reduced information in the tails. The data is simulated from the GEV-GARCH model, so it is expected that this model will have the lowest RMSE, as observed in Table 2. Notice that the GEV-GARCH and GARCH-GPD have similar RMSE and both outperform the standard GARCH. These results show that the commonly used two-stage approach of [17] and our proposed approach have similar performance for estimating high quantiles. However, our approach has an important benefit that it is a one stage model which simplifies the parameter estimation and makes it easier to account for all sources of uncertainty, which is much more challenging in the two stage approach. All the other simulated samples have similar results. These results suggest that the GEV-GARCH should be preferred if the interest is in modelling or predicting extremes with conditional variance.

6. Empirical results with financial returns

A similar conditional extreme value model based on high frequency returns which allows all three parameters of GEV distribution to vary over time is developed by [3]. Their study applies the model to daily 5 min maximum returns of stock index to compare the relative performance of with GARCH and implied volatility (VIX) in forecasts of realized volatility. We will return to assess the performance of their approach in Section 7. In our study, the GEV-GARCH model is also applied to daily 5 min maximum of individual stock returns. The dataset used in this section includes all intra-day quotes of IBM and the period is from January 3, 1994 to December 31, 2003, leaving a total of 2540 days.

Table 3
Mean of posterior parameter estimates for 5 min returns of IBM. *Note:* the CI is the 95% highest posterior credible interval of the posterior.

θ	Estimated value	CI	
		Low	High
ξ	0.076	0.048	0.104
α_0	0.006	0.004	0.007
α_1	0.453	0.309	0.598
α_2	0.083	0.055	0.113
β_0	0.207	0.193	0.221
β_1	0.32	0.286	0.356

Followed by [1], we use two time scales to construct 5 min log return to minimize the microstructure noise and the daily maxima of them is used in the following analysis.

The estimated parameter values for the stock returns are very similar to the simulation results of θ_1 reported in the previous section. Table 3 reports the parameter estimates for the IBM daily 5 min maximum returns and 95% credible interval of the posterior obtained using MCMC. The GARCH coefficient α_1 is less than 0.5. The similarity of GEV-GARCH volatilities and GARCH volatilities is shown in Fig. 4, indicating that the GEV-GARCH model has captured the volatility clustering via imposing a time-varying scale parameter. When the quantile estimates are of concern as Fig. 5, the GEV-GARCH expected quantile (of 95%) has a better coverage rate than the GARCH (with normal innovations) due to its flexibility in capturing the heavy tail behaviour. Also we can expect a even larger difference between two quantile estimators when compare higher quantiles since the tail quantile distributions are more skewed with higher variation at higher quantiles which is showing the asymmetry in the information (and lack of information) in estimating higher quantiles.

7. Exploration of time-varying shape parameter

In the proposed GEV-GARCH model, the shape parameter of the GEV distribution is constant in time. The shape parameter is well known to be difficult to estimate requiring very large datasets to get reliable estimates, see [9]. Further, it is unclear whether it is physically meaningful to allow the shape to rapidly vary in time, as it is driven by market trading behaviour. [3] consider an extension of the GEV, where they also the shape parameter to vary in time according
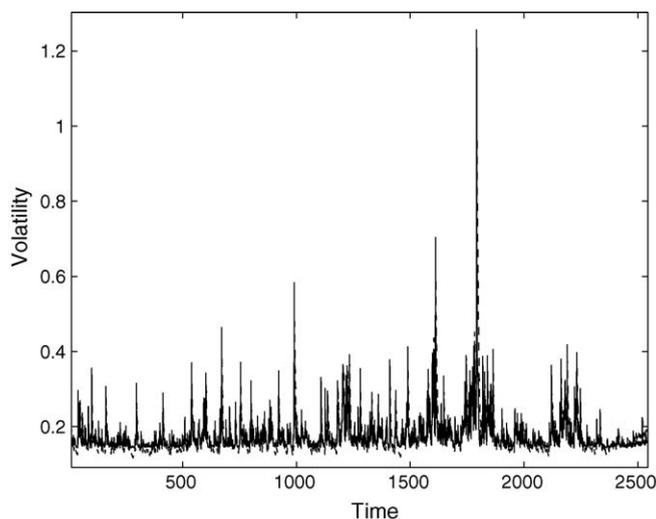

Fig. 4. Volatilities of IBM return (5 min maximum). The solid line represents the volatilities from GEV-GARCH model and the dashed line represents the volatilities from the GARCH model (with normal innovations).
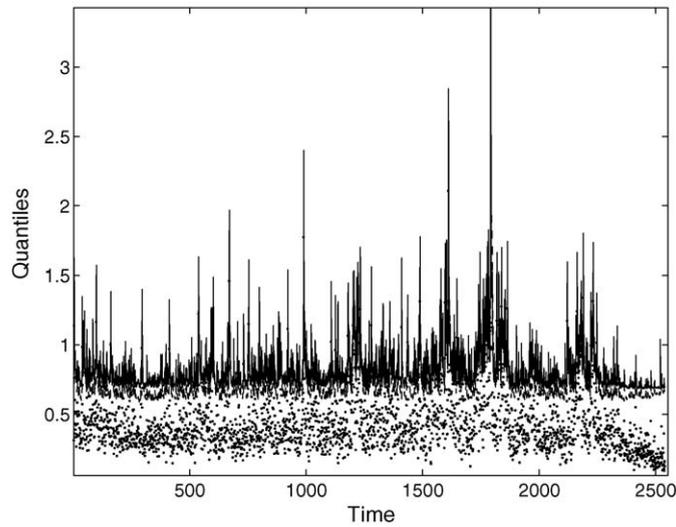
Fig. 5. Estimated 95% quantiles of IBM returns (5 min maximum). The solid line represents the quantiles from GEV-GARCH model and the dashed line represents the quantiles from the GARCH model (with normal innovations).

to a GARCH-type structure namely:

$$M_t \sim G(\xi_t, \sigma_t, \mu_t)$$
$$\mu_t = \mu_0 + \phi M_{t-1}$$
$$\sigma_t = \sigma_0 + \lambda_1 \sigma_{t-1} + \lambda_2 |\epsilon_{t-1}|$$
$$\xi_t = \xi_0 + \gamma_1 \xi_{t-1} + \gamma_2 |\epsilon_{t-1}|$$

where M_t denote the block maxima/minima and $\epsilon_{t-1} = M_{t-1} - \mu_{t-1}$.

In attempting to compare our approach to that of [3] we attempted a simulation study, which frequently failed due to problems of identifiability of the shape parameter. Fig. 6 gives the profile likelihoods plot for the two shape related parameters when applying the model of [3] to the real IBM returns above as an example. You will notice that the profile likelihood has multiple modes, which would clearly cause problems for standard numerical optimization procedures used for likelihood maximisation, and is also generally very flat which is indicative model identifiability problems when the shape parameter is permitted vary in time in such dynamic manner.
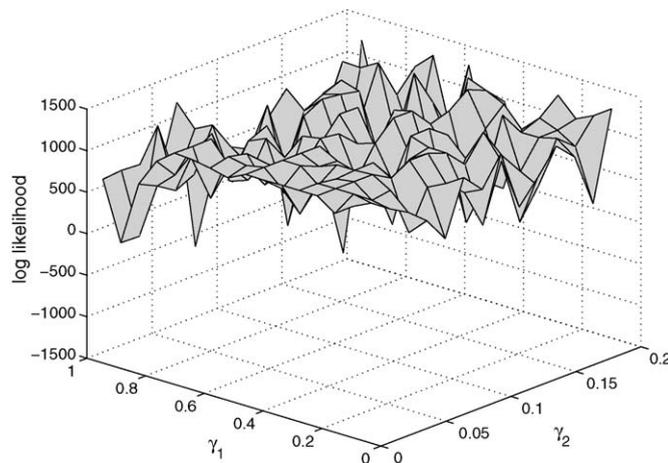

Fig. 6. Profile likelihood for shape related parameters of Bali and Weinbaum [3] model for IBM stock returns.

8. Discussions and conclusions

We have proposed a new extreme value model for dynamic estimation of extremal quantiles. To capture the temporal dependence typically observed in financial applications, that of volatility clustering, the scale and location parameters of the generalised extreme value distribution (GEV) have been defined to exhibit a conditional autoregressive heteroscedastic type structure. Arguably the most commonly used approach in the finance literature to overcome this temporal dependence is the two-stage approach of [17], where the volatility dependence is captured by a GARCH model [5] followed by applying standard extreme value modelling approach to the approximately independent residual innovations. The proposed approach overcomes the challenges associated with uncertainty estimation in this two-stage approach, and also avoids the problems associated with threshold determination.

A simulation study and real data application have been used to demonstrate the performance of the GEV-GARCH in capturing the dynamic of conditional variance of extremes and to model the tail behaviour of the underlying variables. The GEV-GARCH performance in estimating the conditional variance is compared to the standard GARCH model with normal residual innovations, and is shown to provide estimates with much smaller RMSE. The performance in estimating the extremal quantiles is compared the two-stage approach of [17] and standard GARCH approach with normal residual innovations. The proposed model is shown to outperform the standard GARCH for estimating the tail quantiles, due to flexibility of the extreme value distribution for capturing the heavy tail behaviour typically observed in financial times series. The GEV-GARCH provides similar performance to the two-stage approach where the generalised Pareto distribution is used to describe the tail of the innovation distribution, but has an important benefit in that it is a one stage procedure, thus making parameter estimation and accounting for the uncertainties more straightforward.

Further results demonstrate model identification and parameter estimation complications when permitting a time-varying shape parameter with a similar GARCH structure, as proposed by [3]. The shape parameter is known to be challenging to estimate requiring very large sample sizes, see [9]. It is possible that the real structure of shape could change over time, for example if the market structure changes. However, it is clear in the results presented that model identification is problematic if the tail behaviour changes according to a GARCH-type structure. Further work could investigate a smoothly varying or structural break type model for the shape parameter if there is believed to be some form of change in market trading behaviour.

Acknowledgements

References

[1] Y. Aït-Sahalia, P. Mykland, L. Zhang, How often to sample a continues-time in the presence of the market microstructure noise, The Review of Financial Studies 18 (2005) 351–416.
[2] M. Asai, M.J. McAleer, M.C. Medeiros, Asymmetry and leverage in realized volatility. Econometric Institute Report EI 2008–31, Erasmus University Rotterdam, Econometric Institute, 2008.
[3] T.G. Bali, D. Weinbaum, A conditional extreme value volatility estimator based on high-frequency returns, Journal of Economic Dynamics and Control 31 (2007) 361–397.
[4] J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels, Statistics of Extremes: Theory and Applications, Wiley Series in Probability and Statistics, 2004.
[5] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31 (1986) 307–327.
[6] T. Bollerslev, Volatility and time series econometrics: essays in honor of Robert F. Engle, chapter 3. Glossary to ARCH (GARCH), in: T. Bollersley, J.R. Russell, M.W. Watson (Eds.), Advanced Texts in Econometrics, OUP, Oxford, 2010.
[7] T.A. Buishand, Rainfall depth-duration-frequency curves: a problem of dependent extremes, in: V. Barnett, K.F. Turkman (Eds.), Statistics for the Environment, Wiley, 1993, pp. 183–197.
[8] N.H. Chan, S. Deng, L. Peng, Z. Xia, Interval estimation of Value-at-Risk based on GARCH models with heavy-tailed innovations, Journal of Econometrics 137 (2007) 556–576.
[9] S.G. Coles, Introduction to Statistical Modelling of Extreme Values, Springer-Verlag, 2001.

[10] A. Davison, R.L. Smith, Models for exceedances over high thresholds (with discussion), Journal of the Royal Statistical Society Series B 52 (1990) 393–442.

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

[12] C.A.T. Ferro, J. Segers, Inference for clusters of extreme values, Journal of the Royal Statistical Society Series B 65 (2) (2003) 545–556.

[13] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, Chapman and Hall, 2004.

[14] L.R. Glosten, R. Jagannathan, D.E. Runkle, On the relationship between the expected value and volatility of nominal excess return on stocks, Journal of Finance 48 (1993) 1779–1801.

[15] L.E. Hentschel, All in the family: nesting symmetric and asymmetric GARCH models, Journal of Financial Economics 39 (1995) 71–104.

[16] M. McAleer, F. Chan, D. Marinova, An econometric analysis of asymmetric volatility: theory and application to patents, Journal of Econometrics 139 (2) (2007) 259–284.

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

[18] B.B.M. Mendes, H.F. Lopes, Data driven estimates for mixtures, Applied Statistics 45 (4) (1996) 463–478.

[19] D.B. Nelson, Conditional heteroskedasticity in asset returns: a new approach, Econometrica 59 (1991) 347–370.

[20] E. Sentana, Quadratic ARCH models, Review of Economic Studies 62 (1995) 639–661.

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

[22] P. Verhoeven, M. McAleer, Fat tails and asymmetry in financial volatility models, Mathematics and Computers in Simulation 64 (3–4) (2004) 351–361.

[23] X. Zhao, C.J. Scarrott, L. Oxley, M. Reale, Extreme value modelling for forecasting market crisis impacts, Applied Financial Economics (2009).

[24] X. Zhao, Extreme Value Modelling with Application in Finance and Neonatal Research, PhD thesis, University of Canterbury, 2010.