18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Extreme Value GARCH modelling with Bayesian inference

Zhao, X. [1] , **L.Oxley** [2] , **C.Scarrott** [1] , **M.Reale** [1]

[1] *Department of Mathematics and Statistics, University of Canterbury, Christchurch*

[2] *Department of Economics, University of Canterbury, Christchurch*

Email: xzh77@student.canterbury.ac.nz

Abstract: Extreme value theory is widely used financial applications such as risk analysis, forecasting and pricing models. One of the major difficulties in the applications to finance and economics is that the assumption of independence of time series observations is generally not satisfied, so that the dependent extremes may not necessarily be in the domain of attraction of the classical generalised extreme value distribution. Even when the dependence satisfies conditions for the sequence to be within the domain of attraction of the generalised extreme value distribution, the traditional modelling approach does not necessarily give full insight into the form of the dependence. The generalized extreme value distribution can be combined with either other time series models or covariates to capture such dependence. This study examines a conditional extreme value distribution with the added specification that the extreme values (maxima or minima) follows a conditional autoregressive heteroscedasticity process. The dependence has been modelled by allowing the location and scale parameters of the extreme distribution to vary with time. The resulting combined model, GEV-GARCH, is developed by implementing the GARCH volatility mechanism in these extreme value model parameters. Bayesian inference is used for the estimation of parameters and posterior inference is available through the Markov Chain Monte Carlo (MCMC) method.

The model is firstly applied to relevant simulated data to verify model stability and reliability of the parameter estimation method. Then real stock returns are used to consider empirical evidence for the appropriate application of the model.

As with most extreme value modelling applications, the shape parameter is the most difficult parameter to estimate. This study also investigates the sensitivity and stability problems in an extension of the GEV-GARCH model to allow a time varying shape parameter. It is demonstrated that a non-constant extreme shape parameter with a GARCH type time varying structure typically leads to over-parameterisation and consequent estimation difficulties.

A comparison is made between the GEV-GARCH and traditional GARCH models. Both the GEV-GARCH and GARCH show similarity in the resulting conditional volatility estimates, however the GEV-GARCH model differs from GARCH in that it can capture and explain extreme quantiles better than the GARCH model because of more reliable extrapolation of the tail behaviour.

Keywords: *Extreme value distribution, dependency, Bayesian, MCMC, Return quantile*

1. INTRODUCTION

Modelling the tails of distributions is important in many fields, such as environmental sciences, hydrology, insurance, finance when there is an interest in the extreme values. Extreme Value Theory (EVT) considers the distributional behaviour of the extremes of random variables. The objective of EVT is to extrapolate 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, relative to that available in the bulk of the distribution. Hence, asymptotically justified parametric models are typically used to represent the data generating process, which can provide reliable extrapolations required in such applications.

The classical Generalised Extreme Value (GEV) distribution is an asymptotically justified model to describe the limiting distribution of the maximum or minimum of a sequence of independent and identically distributed random variables. Under certain conditions (Beirlant, 2004) forms of short range dependence can also lead to the distribution of the maximum or minimum remaining within the same GEV distribution family. However, these results are not necessarily helpful in describing the detailed form of the dependence. . Some recent research has been undertaken on modelling extreme values with covariates in non-stationary conditions (Smith, 1989; Davison and Ramesh, 2000; Pauli and Coles, 2001),. In this paper, we have taken a rather pragmatic approach by amalgamating commonly used GARCH model with the classical GEV model.

Financial data is well known to be heavy tailed and extreme value theory has been shown to be a very useful tool in estimating and predicting the extreme behaviour of actuarial and financial products, such as predicting the largest claim in insurance and the Value at Risk (VaR) in finance. Applying extreme value models however is not straightforward in these applications because of the dependency and seasonality involved. Modifications to classical extreme value models have been implemented to deal with these kind of data. This study examines the dynamics of the maximum and minimum value of intra-day financial returns, using an extreme value model constructed as a combination of classical GEV distribution and well known GARCH model. The dependence is captured by allowing the location and scale parameters of the extreme value distribution to follow a conditional autoregressive heteroscedastic process. The study has been conducted with both simulated and real data. Further results on the model identification of the GEV shape parameter are supplied to show that complications arise if the GARCH structure is also applied to the GEV shape parameter as one might expect due to the inherent difficulty in estimating this parameters due to the typical lack of information in the tails.

2. EXTREMES OF NON-STATIONARY SEQUENCES

2.1. Generalized Extreme Value Distribution

The Generalized Extreme Value distribution (GEV) is an asymptotically motivated approach for describing the distribution of the maxima and minima of a realized sequence of independent random variables having the common distribution function (Beirlant, 2004):

$$G(x \mid \xi, \sigma, \mu) = \exp\left\{ -\left[1 + \xi\left(\frac{x-\mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \qquad -\infty < x < \infty \qquad x:\ 1 + \xi(x-\mu)/\sigma > 0\,. \tag{1}$$

The model has three parameters: a shape parameter ξ, a scale parameter σ and a location parameter μ. The GEV distribution represents three types of tail behaviours according to the value of shape parameter ξ. The Fréchet type (slowly decaying) tail and the Weibull type (upper bounded) tail correspond respectively to $\xi>0$ and $\xi<0$. The Gumbel type (exponentially decaying tail) is considered in the limit as the shape parameter $\xi\rightarrow0$. Classical extreme value theory shows that, if a limiting distribution for the maxima/minima of a sequence of iid random variables exists, then it must fall into these three types.

2.2. Extremes of non-stationary/dependent sequence

The classical extreme value distribution assumes that the time sequence of realizations follow a stationary iid process. Certain forms of (short range) dependence can also be permitted, but still lead to the same type of limiting extreme value distribution. However, this result does not necessarily provide a detailed description of the form of the dependence. We have approached modelling the dependence/non-stationary behaviour by allowing the time varying extreme value parameters. There are a large catalogue of models considered in the literature. The first principle of model selection is the simplicity and attention need to be paid to the structure of shape parameter since it is the most difficult to estimate. It is usually unrealistic to model the shape parameter as the function of either time or other covariate variables in applications. Extra-parameterization in the model has to be supported by the evidence of a need for a more complex model structure to adequately describe variations in the observed data.

The Generalized AutoRegressive Conditional Heteroscedastic (GARCH) process is widely used to model non-stationary in the variance of financial time series. This study considers 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. With the combination of GEV and GARCH process, , we will show that the model is better suited to explain the extreme quantiles than the classic GARCH model alone, which cannot capture the tail behaviours adequately with either Normally distributed or even fatter tailed distributed (e.g. t-distributed) innovations. McNeil and Frey (2000) suggest a two stage model in estimating VaR and related risk measures by applying extreme value theory (EVT) on GARCH residuals, where fit a GARCH type model at the first stage to gain an independent residual sequence of observations and apply Generalized Pareto Distribution on standardized residuals. An important benefit of our approach is that it is a one stage model and has the advantage in accounting for all uncertainty in the estimation which is much more of a challenge in the two stage model.

3. THE GEV-GARCH MODEL

3.1. Model Structure

The proposed model assumes that the observations come from the Generalized Extreme Value (GEV) distribution with a GARCH(1,1) process describing the conditional variance of extremes. Therefore, the distribution function of the observation *x* (maximum values) can be written as:

$$x_t \sim GEV(\xi, \sigma_t, \mu_t)$$

$$G(x_t \mid \xi, \sigma_t, \mu_t) = \exp\left\{-\left[1+\xi\left(\frac{x_t-\mu_t}{\sigma_t}\right)\right]^{-1/\xi}\right\}, \qquad \{-\infty < x_t < \infty, \quad 1+\xi(x_t-\mu_t)/\sigma_t > 0\}. \tag{2}$$

Here:
$$x_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_t$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 \varepsilon_{t-1}^2$$
$$\varepsilon_{t-1} = x_{t-1} - \mu_{t-1}$$

where $|\beta_1| < 1$, $\alpha_0 > 0$, $\alpha_1 > 0$, $\alpha_2 > 0$, and $\alpha_1 + \alpha_2 < 1$. As stated before, it is typically empirically unrealistic to try to model ξ as a function of time unless there is strong evidence that the shape of extreme value distribution does change over time. In this study ξ, the shape parameter, as supported by empirical evidence is constant through time. This simplification does not jeopardize the efficacy of the model and proves very beneficial as the shape parameter is crucial in the GEV distribution and it well known to be challenging to estimate due to the sparsity of tail data. The simulation study in Section 4 shows the sensitivity of the model with a time varying shape parameter and the estimation difficulties with a non-constant shape.

The expected value and variance of x_t are (outside of the shape parameter ranges the moments are infinite):

$$\begin{aligned} E(x_t) &= \mu_t + \sigma_t\left(\Gamma(1-\xi)-1\right)/\xi, \qquad \xi < 1 \\ Var(x_t) &= \sigma_t^2\left(\Gamma(1-2\xi) - \Gamma^2(1-\xi)\right), \qquad \xi < 0.5 \end{aligned} \tag{3}$$

3.2. Bayesian estimation of the model

We use Bayesian inference to estimate the model parameters using Markov Chain Monte Carlo (MCMC) to obtain posterior distributions. The parameter vector $\theta = (\xi, \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1)$ can be decomposed into two component $\theta_1 = (\xi, \beta_0)$ which are defined over the whole real line and $\theta_2 = (\alpha_0, \alpha_1, \alpha_2, \beta_1)$ which must be strictly positive. We use a normal prior on θ_1 and a flat prior on θ_2 for simplicity, to indicate little prior information being available. The posterior distribution of θ is

$$p(\theta \mid \xi, \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1) \propto \prod_{t=1}^{T} f_{GEV}(x_t \mid \xi, \sigma_t, \mu_t) \times \exp\left(-\frac{1}{2}\left(\frac{\theta_1 - \mu_{\theta_1}}{\sigma_{\theta_1}}\right)^2 \right) \times I_{(\theta_G)} \tag{4}$$

where $I(\theta_G)$ is an indicator function reflecting the constraints on the parameters above and $f_{GEV}(x_t \mid \xi, \sigma_t, \mu_t)$ is the GEV density function.

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 the first one to be updated. The rest components are updated as (α_1, α_2, α_0, β_1, β_0). The restrictions on θ ($I(\theta_G)$), are enforced during the sampling procedure by rejecting the draws that violate them. The stationary constrains, $|\beta_1|<1$, $\alpha_1 + \alpha_2 <1$ and $1+\xi(x_t - \mu_t)/\sigma_t > 0$ are also imposed. Without those constraints the chain could simply converge to incorrect values or be computationally inefficient. The convergence of MCMC is checked by monitoring the marginal distributions of the parameters obtained from the parallel chains, using standard diagnostic checks (Gelman, Carlin, Stern and Rubin 2004). The predictive posterior distributions and deviance analysis are imposed for the model fitting.

3.3. Illustration with simulated data

We first check the stability of the model and parameter estimation using simulated data. The simulated data use three different parameter sets with 100 simulates series for each parameter set using the sample size of 2500, which is of similar length to the real application sample in the following section. The first parameter vector set considered is close to the application estimation results in Section 3.4. The second parameter set use 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 feature when the tail getting heavier. A more extensive simulation study is detailed in Zhao (2009)., but have not been included for brevity. We set the priors for parameter θ (ξ, α_0, α_1, α_2, β_0, β_1) as follows:

$$\xi \sim N(0.25, 0.1), \quad \alpha_0 \sim U(0, 0.2), \alpha_1 \sim U(0.35, 0.99), \alpha_2 \sim U(0, 0.5), \beta_0 \sim N(0, 0.2), \quad \beta_1 \sim U(0, 0.5)$$

for simulated samples based on θ_1 and θ_2 .

$$\xi \sim N(0.1, 0.15), \quad \alpha_0 \sim U(0, 0.2), \alpha_1 \sim U(0.30, 0.99), \alpha_2 \sim U(0, 0.5), \beta_0 \sim N(0, 0.3), \quad \beta_1 \sim U(0, 0.5)$$

for simulated samples based on θ_3 .

Table 1 summarizes the results for the simulated data samples. For each individual sample simulated, the estimated parameter value is the mean of 95% highest density interval (HPD) of the posterior samples. The estimated parameter value reported as Table 1 are the mean value over 100 samples. The table also gives the mean square error (MSE) of the estimators compare to the true parameter value. The expected estimators are very close to the true values and the sample variation are small. As the shape increase, the MSE falls, which suggests a better performance of the model on a relative heavier tail distribution.

Table 1: Summary of simulation estimations

θ1	True Value	Estimated Value (mean)	MSE	θ2	True Value	Estimated Value (mean)	MSE	θ3	True Value	Estimated Value (mean)	MSE
ξ	0.08	0.072	0.0006	ξ	0.20	0.201	0.0003	ξ	0.30	0.297	0.00029
$\boldsymbol{\alpha_0}$	0.01	0.009	0.0000	$\boldsymbol{\alpha_0}$	0.01	0.010	0.0000	$\mathbf{A_0}$	0.05	0.051	0.00002
$\boldsymbol{\alpha_1}$	0.45	0.480	0.0066	$\boldsymbol{\alpha_1}$	0.80	0.793	0.0014	$\mathbf{A_1}$	0.50	0.498	0.00036
$\boldsymbol{\alpha_2}$	0.08	0.078	0.0002	$\boldsymbol{\alpha_2}$	0.02	0.021	0.0000	$\mathbf{A_2}$	0.10	0.099	0.00008
$\boldsymbol{\beta_0}$	0.21	0.210	0.0001	$\boldsymbol{\beta_0}$	0.01	0.010	0.0000	$\mathbf{B_0}$	0.05	0.051	0.00012
$\boldsymbol{\beta_1}$	0.32	0.321	0.0003	$\boldsymbol{\beta_1}$	0.10	0.101	0.0001	$\mathbf{B_1}$	0.20	0.201	0.00023

The expected variance of extreme values has been calculated and the *extreme volatility* is defined as the square root of the expected variance. The volatilities from the two models are very close to each other for the simulation data samples as the real application below (refer to Figure 2). This suggests non-homogeneous and GARCH-featured extreme variances.

The similarity of two volatility sequences however induce the question of what is the extra value of GEV-GARCH model respect to GARCH model. The answer is that GEV-GARCH can explain the extreme quantiles better in the way of combining the extreme tail distribution and the conditional variance process. Figure 1 plots the sequence of the quantiles defined by both GEV-GARCH and GARCH model along with the actual observations (dots) for this particular sample. It is clear that the 97% quantile explained by the GEV-GARCH model (solid line) covers more of the extreme values than the corresponding GARCH quantiles (dashed line). Table 2 compares the Mean Square Errors (MSE) of different model quantiles to the true quantiles, including GEV-GARCH, GARCH-GEV (a similar frame as McNeil & Frey's GARCH-GPD) and GARCH. We report only one sample results for each parameter simulation set above for brevity. As the increase of quantile, MSE for all three models increase as expected for less information of further tail. But for all the quantiles, GEV-GARCH has the smallest MSE and GARCH has the largest MSE. The difference among three models gets larger and larger for higher quantiles as well as the tail distribution becomes heavier. All 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, particular if the distribution has relatively heavy tails.

Figure 1. Quantile plot of a sample of θ2

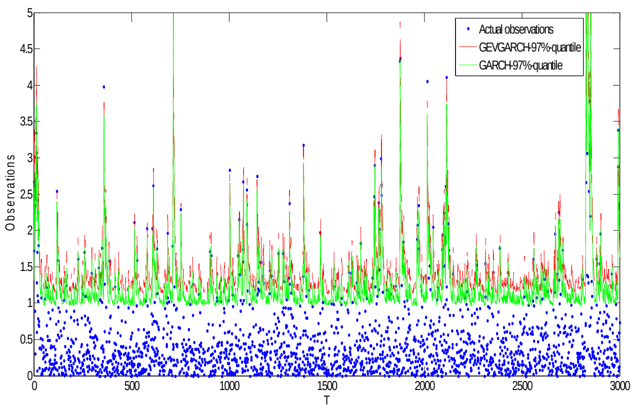


Table 2. Quantile MSE

	Quantiles	0.9	0.95	0.99	0.999
θ1	GEV-GARCH	0.0009	0.0016	0.0047	0.0147
	GARCH-GEV	0.0011	0.0022	0.0073	0.0251
	GARCH	0.0011	0.0073	0.0848	0.5267
θ2	GEV-GARCH	0.0027	0.0058	0.0237	0.1317
	GARCH-GEV	0.0133	0.0267	0.0936	0.3794
	GARCH	0.0182	0.0347	0.6322	6.3343
θ3	GEV-GARCH	0.0014	0.0031	0.0289	0.4748
	GARCH-GEV	0.1005	0.1950	1.4089	18.8870
	GARCH	0.0501	0.5279	11.9866	146.2920

3.4. Illustration with Financial data

Bali and Wenbaum (2007) develop a conditional extreme value volatility estimator (EVT) based on high frequency returns which allows all three parameters of GEV distribution to vary over time. They apply the model to daily 5 minute maximum returns of stock index to get EVT and compare the relative performance of EVT with GARCH and implied volatility (VIX) in forecasts of realized volatility. The paper argues that EVT provides more accurate forecasts. However, there is insufficient support for the model specification and for the assumption that the shape parameter changes according to a GARCH structure as with scale. As stated above, a changing shape parameter can jeopardize the model stability and augment the sensitivity of the estimators. Also the changing shape challenges the estimation method because of the scarcity of information on the extremes. Maximum likelihood estimation can have difficulty in finding the true global mode of the likelihood. In our study, we also apply our GEV-GARCH on daily 5min maximum return of individual stock returns to identify whether the model can supply the additional information in predicting volatility.

Table 3: Model estimation for IBM

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	0.320	0.286	0.356

The estimated results for stock return are very similar to the simulation results of θ1. Table3 reports the parameter estimates for IBM daily 5 minute maximum return and 95% confidence interval of the posterior which was obtained by MCMC. The GARCH coefficient α_1 is less than 0.5.

In Bayesian inference, the posterior predictive graph is used to display the data alongside simulated data from the fitted model (posterior distribution) and to show the discrepancies between real and simulated data to check the model fitting. Since the study is interested the extreme quantiles, the posterior predictive quantile (PPQ) distributions are shown as Figure 4. The figure gives the PPQ distributions and its 95% confidence interval (dot line). All the sample quantiles are located within the confidence interval of PPQ. The sample quantiles (solid line) are all close to the mode of the distributions. As expected, the tail quantile distributions are more skewed and heavy tailed at higer quantiles.

Figure 2. Volatility Plot of stock return

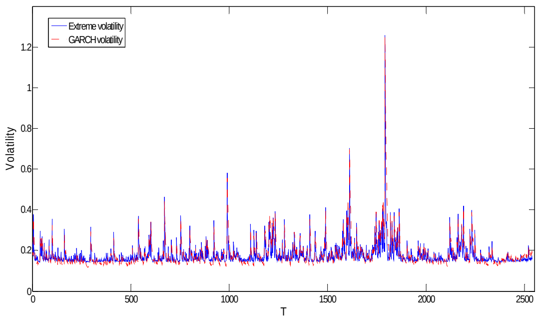
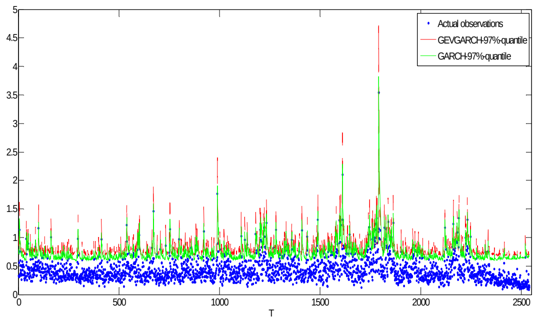
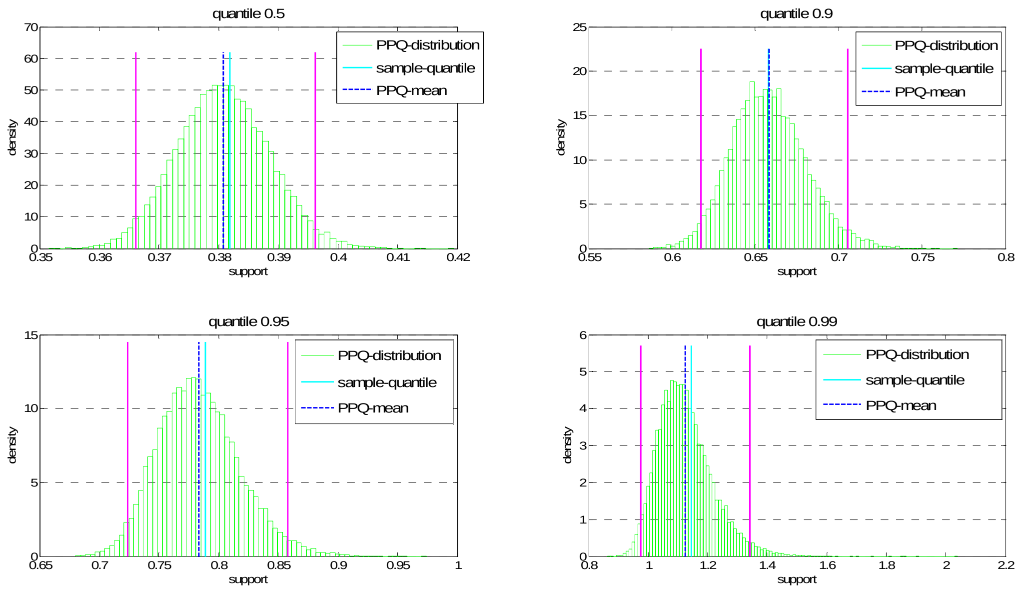


Figure 3 Quantiles plot of stock return



There is no clear difference in volatility estimators (Figure 2). This suggests that a conditional autoregressive heteroscedasticity process of extremes defined through the time varying scale and location parameters can capture the non-homogenous variance of extremes, which the classical GEV cannot do. There is an obvious difference in the extreme quantiles based on two models shown in Figure 3. The GEV-GARCH expected quantile covers more extreme observations since it provides flexibility in modelling the tail behaviour.

Figure 4. Posterior Predictive Quantiles



4. SIMULATED DATA WITH TIME VARYING SHAPE

A simulation study considered a time varying shape parameter of the GEV, for a comparison with Bali and Wenbaum (2007), of the form: $\xi_t = c_0 + c_1\xi_{t-1} + c_2|\varepsilon_{t-1}|$. Considerable estimation problems were encountered, due to a parameter identifiability. As an example, for brevity, a two dimensional profile likelihood plot on c_1, c_2 for the sample simulated, is shown in Figure 5, suggests a flat likelihood surface along parameter c_1. (TRUE VALUE IS WHAT, SAMPLE SIZE?). Hence, even for data generated from the model maximum likelihood and Bayesian inference is problematic. When the maximum daily 5min return series is used, results similar to simulated data have been found, see Zhao (2009) for details. Further, Figure 6 gives the profile likelihoods contour plot on two shape related parameters when applying real stock-IBM returns. The multiple modes of the likelihood surface indicate the sensitivity of the model.

Figure 5. Profile likelihood for shape parameter- simulated data

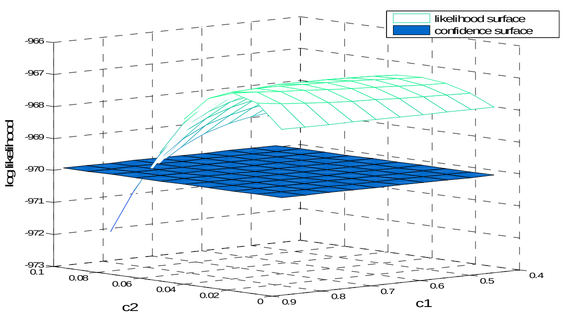
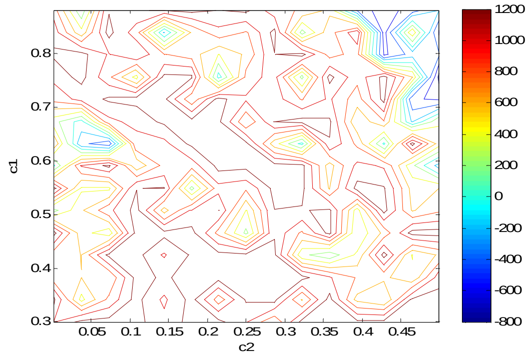


Figure 6. Profile likelihood for shape parameter- IBM



Both simulation and real data results imply that allowing the shape parameter change over time Accoriding to the GARCH structure applied to the scale parameters leads to considerable identifiability problems. Further, it is generally not realistic to assume that shape parameters will change as fast as other two parameters even it is not necessarily a constant over time.

5. DISCUSSION AND CONCLUSIONS

This paper develops a GEV-GARCH model by applying a conditional autoregressive heteroscedastic structure to the classical GEV distribution, through time varying scale and location parameters to capture the effects of temporal dependence present in most financial data. A simulation study and real data application show that the GEV-GARCH can capture the dynamic of conditional variance of extremes and model the tail behaviour of the underlying variables. Further results demonstrate model identification and parameter estimation complications arise when considering a time varying shape parameter with a similar GARCH structure. Although the model considered in the study fixed the shape of GEV as constant in time, it is likely that the real structure of shape will change over time. In the case of market structure change, the expected tail behaviour can change and induce the change on the shape of GEV. However, we have shown that model identification is problematic if the tail behaviour changes as fast as the volatilities. The model can be used in estimation of tail-related risk for heteroscedastic time series and avoids the difficulty in capturing the estimation uncertainty as in the related and commonly used two stage approach by McNeil and Frey (2000).

ACKNOWLEDGMENTS

We are grateful to Marcelo Cunha Medeiros and Marcel Scharth for providing the high frequency data of Dow Jones stocks returns and Dominic Lee for helpful suggestions.

REFERENCES

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

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

Davison, A.C., and N.I. Ramesh, (2000), Impact of urbanization on coastal wetland structure and function. *Journal of the Royal Statistical Society Series B*, 62, 191-208.

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

Green, P.J. (1995), Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82, 711-732.

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

Scharth, M. and M.C. Medeiros, (2006), Asymmetric effects and long memory in the volatility of Dow Jones stocks, *Department of Economics PUC-Rio (Brazil),Textos para discussÃ£o* 532.

Pauli, F. and S.G. Coles, (2001), Penalized likelihood inference in extreme value theory. *Journal of Applied Statistics*, 28, 547-560.

Smith, R.L. (1989), Extreme value analysis of environmental time series: an application to trend detection in ground-level zone. *Statistical Society*, 3, 367-393.

Zhao, X. (2009). Extreme Value Modelling with Application in Finance and Neonatal Research. PhD Thesis (To be submitted). University of Canterbury.