Modelling conditional heteroskedasticity in JSE stock returns using the Generalised Pareto Distribution

Caston Sigauke[a,*], Rhoda M. Makhwiting[b] and Maseka Lesaoana[c]

[a*]School of Statistics and Actuarial Science, University of the Witwatersrand, Private Bag 3, Wits 2050, South Africa. Corresponding author's email: csigauke@gmail.com

[b,c] University of Limpopo, School of Mathematical and Computer Sciences, Department of Statistics and Operations Research, Private Bag X1106, Sovenga, 0727

Abstract

Extreme equity market returns demand the use of specialised techniques for standardised treatment that focuses exclusively on rare tail events. Extreme Value Theory (EVT) is used in this article to model heteroskedastic stock returns of the All Share Index (ALSI) at the Johannesburg Stock exchange (JSE). Daily data of the ALSI at the JSE over the period 2002–2011 is used. A two-stage modelling framework is proposed. In stage one we fit an Autoregressive Moving Average–Generalised Autoregressive Conditional Heteroskedastic (ARMA-GARCH) model to the stock return series. In stage two we filter the residuals from the ARMA-GARCH model. We then fit a Generalised Pareto Distribution (GPD) to the upper tail of the residual series, and refer to this hybrid as the ARMA-GARCH-GPD model. The threshold is estimated using a Pareto quantile plot. Empirical results show that the Weibull class of distributions can be used to model daily returns data. The ARMA-GARCH-GPD model produces more accurate estimates of extreme returns than the ARMA-GARCH model. These results are important to risk managers and investors.

Keywords: *Extreme value theory, GARCH, Generalized Pareto Distribution, risk management*

1 Introduction

It has long been recognised that financial time series data are characterised by a number of stylised facts such as persistence, volatility clustering, time-varying volatility and leptokurtic data behaviour. Accurate modelling of extreme returns is vital to financial risk management. The common assumption in finance theory is that financial returns are normally distributed. Conversely, several studies indicate that most financial time series are fat-tailed (see, e.g., Maghyereh & Al-Zoubi, 2008; Guru, 2012; Song and Song, 2012). Risk managers at a stock exchange are interested in guarding against the risk of high gains/losses due to the rise/fall in the prices of financial assets held by the stock exchange. It turns out that daily returns are approximate quantities which must be investigated. This study focuses on modelling extreme losses. One method of extracting upper extremes from a set of data is to take the exceedances over a predetermined high threshold. This involves the use of Peaks-

Over-Threshold (POT) distributions such as the Generalised Pareto Distribution (GPD) or the Generalised Pareto-type (GP-type) distribution (see Verster & De Waal, 2011) and referred to as the Generalised Single Pareto Distribution (GSPD) in Sigauke *et al.* (2012).

Extreme Value Theory (EVT) is the theory of measuring and modelling extreme events. It is especially well suited to describe the fat-tails of the profit and losses distributions typically found in stock returns. It is important to note that EVT relies on an assumption of independent and identically distributed *(i.i.d)* observations. The *i.i.d* assumption does not hold for financial time series data. This can be corrected by filtering residuals of the return series using time series analysis techniques to get *i.i.d* variables and then apply the EVT method. This study follows a two-stage approach proposed by McNeil and Frey (2000), who estimate an ARMA-GARCH model in stage one with a view to filtering the return series to obtain nearly *i.i.d* residuals. They applied EVT modelling framework to the standardised residuals in stage two. The advantage of the GARCH-EVT combination lies in its ability to capture conditional heteroskedasticity in the data through the GARCH framework, while at the same time modelling the extreme tail behaviour through the EVT method. The GARCH-EVT modelling approach performs better than other models in forecasting VaR for various international stock markets (see, e.g., Gencay & Selcuk, 2004; Fernandez, 2005; Wagner & Marsh, 2005; Chan & Gray, 2006).

The article focuses on modelling the distribution of daily JSE price changes and the estimation of extreme quantiles by fitting traditional time series models and an EVT distribution. The EVT approach captures the features of the innovation distribution well and provides more accurate estimates of risk measures, compared to other approaches (Fernandez, 2005). The GPD was first introduced by Pickands (1975) in the extreme value framework as a distribution of the sample excesses (or exceedances) above a sufficiently high threshold. EVT is discussed and used in literature to estimate high quantiles. The POT method is one of the most widely used modelling approaches for fitting distributions above a sufficiently high threshold (see, e.g., Chan & Gray, 2006; Gilli and Kellezi, 2006; Magheyereh & Al-Zoubi, 2008; Castillo & Daoudib, 2009; Song & Song, 2012). The GPD is the distribution which is normally used. Recent work includes the use of the GSPD. Sigauke *et al.* (2012) use the Autoregressive Moving Average-Exponential Generalised Autoregressive Conditional Heteroskedastic-Generalised Single Pareto Distribution (ARMA-EGARCH-GSPD) modelling framework to model under demand estimation in daily peak electricity demand forecasting, using South African data. Empirical results from this study show that the ARMA-EGARCH-GSPD model produces more accurate results than an ARMA-EGARCH model.

The article explores the usefulness of EVT in modelling extreme events in stock markets. The remainder of the article is organised as follows: the section hereunder describes the data while the subsequent section discusses the modelling framework.

Next follow the results and discussion, while the final section gives concluding remarks with some ideas for future work.

2 Data

The time series data used for modelling conditional heteroskedasticity in this article is the All Share Index (ALSI) at the Johannesburg Stock Exchange (JSE) over the period 7 January 2002 to 30 December 2011, resulting in a total number of 2 495 observations. A visual inspection of Figure 1 shows that daily stock prices are not stationary. Formal unit root tests are carried out using the Augmented Dickey–Fuller (ADF) test, as shown in Table 1. Since the computed ADF test-statistics (-0.6326) is greater than the critical values at one per cent, five per cent and ten per cent significant levels respectively, we fail to reject the null hypothesis that there is a unit root and that the series needs to be differenced in order to make it stationary.

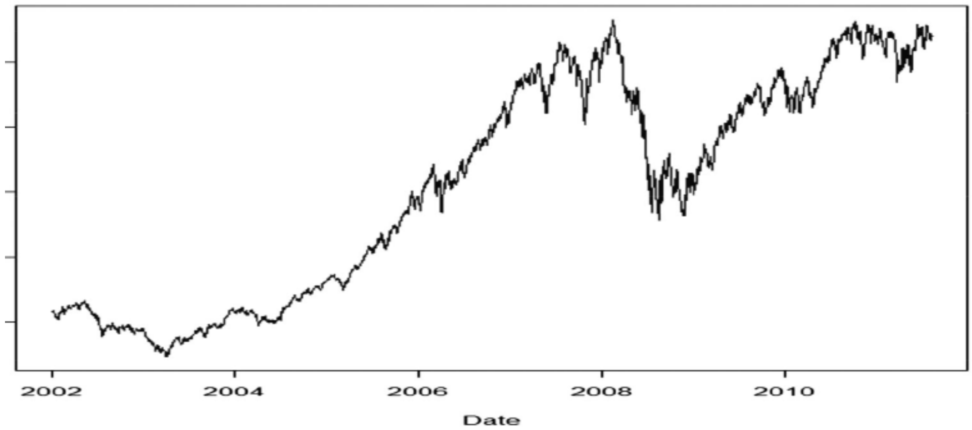

Figure 1: Plot of the all-share JSE stock index (2002-2011)

Table 1: Augmented Dickey-Fuller test of the JSE stock index

ADF Test Statistic -0.6326	1% Critical value	-3.4360
	5% Critical value	-2.8632
	10% Critical value	-2.5677

A plot of the return series given in Figure 2 shows periods of high volatility, occasional extreme movements and volatility clustering. The plot indicates that the logarithm of stock prices is stationary after taking the first-difference, and the ADF test results in Table 2 confirm the stationarity of the return series data. The daily returns (r_t) are calculated as

$$r_t = 100 \text{ In} \left(\frac{P_t}{P_{t-1}} \right) \tag{1}$$

where P_t denotes the current stock price on day t and P_{t-1} denotes one lagged stock price on day $t-1$.

Table 2: Augmented Dickey-Fuller test of the daily returns

ADF Test Statistic -24.57	1% Critical value	-3.4360
	5% Critical value	-2.8632
	10% Critical value	-2.5677

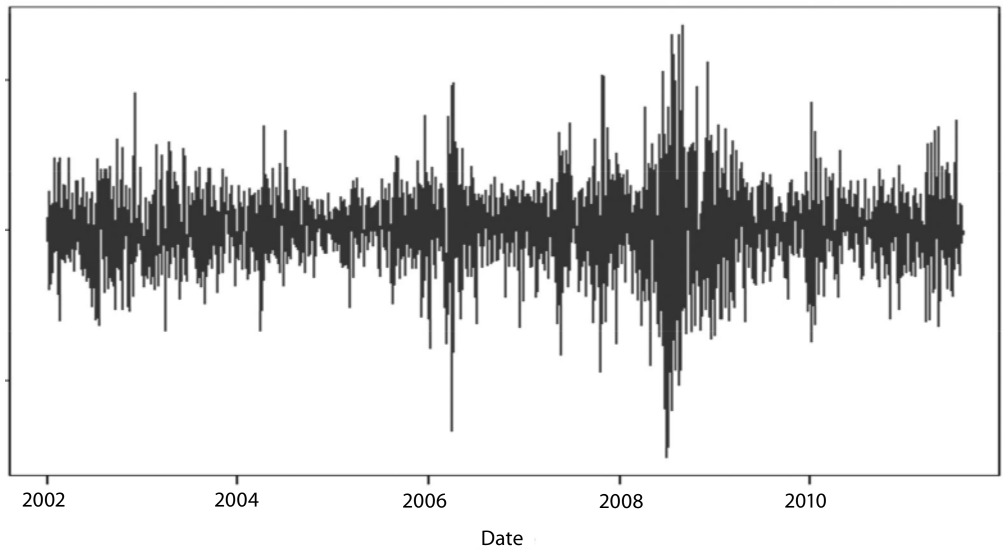

Figure 2: Plot of daily returns for JSE stock index (2002–2011)

A summary of the statistics of the return series data is given in Table 3, with p-values in parentheses. The mean is positive, suggesting that stock returns increase slightly over time. The coefficient of skewness indicates that returns have asymmetric distribution, i.e., they are skewed to the left. The kurtosis of returns is 5.9736 which is greater than three, indicating that the distribution of returns follows a fat-tailed distribution, thereby exhibiting one of the important characteristics of financial time series data, namely that of leptokurtosis. The non-normality condition is supported by a Jarque-Bera test which shows that the null hypothesis of normality is rejected at the five per cent level of significance.

Table 3: Summary statistics of the returns series

Mean	0.0435
Minimum	-7.581
Maximum	6.834
Standard deviation	1.332
Skewness	-0.132
Kurtosis	5.974
Jarque-Bera	927.2 (0.000)

3 Modelling framework

3.1 Mean equation

Initially we specify an ARMA (p, q) model for the mean returns, as it provides a flexible and parsimonious approximation to conditional mean dynamics. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are used to determine the order of ARMA (p, q) models. The ACF and PACF plots given in Figure 3 suggest that the returns may be modelled by an ARMA (1, 0) process, that is:

$$r_t = \varepsilon_t - \phi r_{t-1} + \mu \quad (2)$$

where r_t is the return series, μ is the mean value of the returns and ε_t is the error term with zero mean and variance σ_t^2.

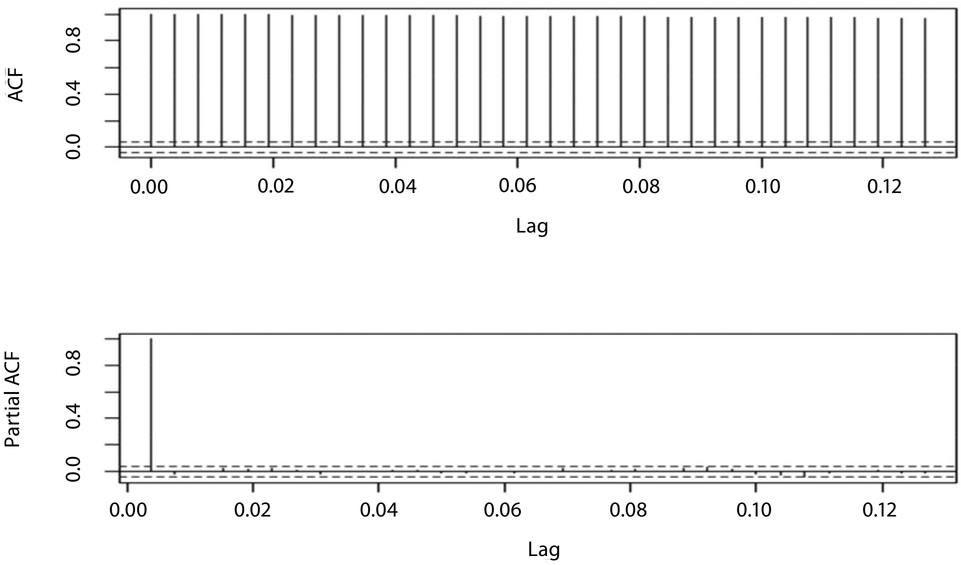

Figure 3: Plot of ACF and PACF of the logged ALSI

3.2 GARCH (p, q) model

Following the natural extension of the ARMA process as a parsimonious representation of a higher order AR process, Bollerslev (1986) extended the work of Engle (1982) to the GARCH process. The GARCH (p, q) process is defined as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{3}$$

where σ_t^2 is the conditional variance, which is a linear function of q lags of the squares of the error terms ε_t^2 or the ARCH terms and p lags of the past value of the conditional variances σ_t^2 or the GARCH terms, and , α_i β_j and ω are parameters. Error term ε_t is assumed to be conditionally normally distributed with zero mean and conditional variance σ_t^2. The GARCH model is estimated with a view of filtering the residuals of the return series to obtain nearly *i.i.d* series. Figure 4 shows the plot of the residual demand, the Q-Q plot and the probability density of the residual demand, together with the enlarged right tail. The density is estimated using kernel density estimation (Silverman, 1986).

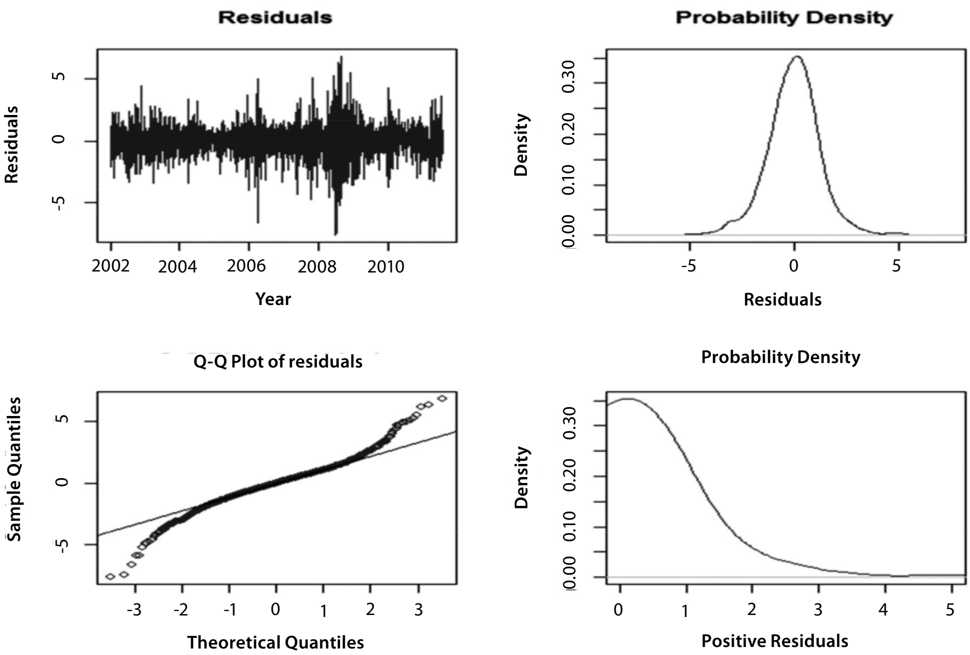

Figure 4: Plot of residuals, probability density and the Q-Q plot

3.3 GPD

We consider a peaks over threshold (POT) distribution to model the positive residuals above a high threshold. The POT method involves the selection of a sufficiently high threshold, denoted by τ .This method consists of fitting the GPD to the distribution of excesses over the threshold. Several methods are proposed in literature for selecting the threshold. Here, we use the Pareto quantile plot discussed in Beirlant *et al.* (2004). We estimate the parameters of (σ, ξ) using the Maximum Likelihood (ML) method. The R statistical package is used. The distribution and survival functions of the GPD are given in equations (4) and (5) respectively.

$$W_\xi(\varepsilon_t) = \begin{cases} 1 - \left(1 + \dfrac{\xi(\varepsilon_t - \tau)}{\sigma}\right)^{-\frac{1}{\xi}} & , \text{if } \xi \neq 0 \\ 1 - \exp\left(-\dfrac{\varepsilon_t - \tau}{\sigma}\right), \text{if } \xi \approx 0 \end{cases} \quad (4)$$

$$P(\ _t > \varepsilon_t |\ _t > \tau) = 1 - W_\xi(\varepsilon_t) = \begin{cases} \left(1 + \frac{\xi(\varepsilon_t - \tau)}{\sigma}\right)^{-\frac{1}{\xi}}, & \text{if } \xi \neq 0 \\ \exp\left(-\frac{\varepsilon_t - \tau}{\sigma}\right), \text{if } \xi \approx 0 \end{cases} \quad (5)$$

where ξ is the shape parameter or Extreme Value Index (EVI) and σ is the scale parameter. The EVI, ξ gives an indication of the heaviness of the tail. If $\xi > 0$ then belongs to the Pareto class of distribution, for $\xi = 0$ W_ξ (ε_t) belongs to the exponential class and if $\xi < 0$ then W_ξ (ε_t) belongs to the Weibull class of distributions. The quantile function of the GPD is given by:

$$\varepsilon_{t,p} = \begin{cases} \dfrac{\sigma}{\xi}\left[(p)^{-\xi} - 1\right], \xi \neq 0 \\ -\sigma \ln(p), \xi = 0 \end{cases} \quad (6)$$

where p is the tail probability. The derivation of the quantile function of GPD is given in the appendix.

4 Empirical results and discussion

4.1 Fitting ARMA (1, 0)-GARCH (1, 1) model to the returns

The results of fitting an ARMA (1, 0)-GARCH (1, 1) model to the JSE return series are presented in Table 4. The econometric package EVIEWS is used for estimating the parameters. The estimates are obtained by the Berndt *et al.* (1974) algorithm using numerical derivatives.

Table 4: ARMA (1, 0)-GARCH (1, 1) model for returns

Mean equation	Variance equation	Model diagnostics
μ = 0.0896 (0.0000)	α_0 = 0.0252(0.0002)	$\alpha + \beta = 0.9854$
ϕ = 0.0487 (0.0172)	α = 0.0974 (0.0000)	*$Q(20)$ = 17.030(0.588)
	β = 0.8880 (0.0000)	Q^2 (20) = 16.469 (0.626)

*Q (20) and Q^2 (20) are the Ljung-Box statistics for testing autocorrelation in return and squared return series data respectively for the 20 lags. In all cases five per cent level of significance is used. P-values are shown in parentheses.

The estimate of ϕ is significant, supporting the use of the ARMA (1, 0) model for the returns. Volatility shocks are persistent since the sum of the ARCH and GARCH coefficients are very close to one. The estimates for α and β are highly significant. The Box-Pierce Q statistics is insignificant up to lag 20, indicating that there is no excessive autocorrelation left in the residuals.

4.2 Threshold and GPD parameter estimation

The Pareto quantile plot is a graphical method for inspecting the parameters of Pareto distribution. The logarithm of the observed positive residuals is plotted against the theoretical quantiles. Figure 5 displays the Pareto quantile plot for the positive residual data. The threshold is $\tau = \exp(0.9634) = 2.6206$. There are 58 observations above the threshold (see Figure 6). We now assume observations above the threshold to be generalised Pareto distributed. The maximum likelihood estimation is used for the determination of the GPD parameters from 58 exceedances. The estimated GPD parameters are $\hat{\sigma}$ = 1.0896 (0.2442) and $\hat{\xi}$ = –0.0321 (0.1824) respectively, with standard errors in the parentheses. The results show that residuals can be modelled using the Weibull class of distributions, since ξ , 0.

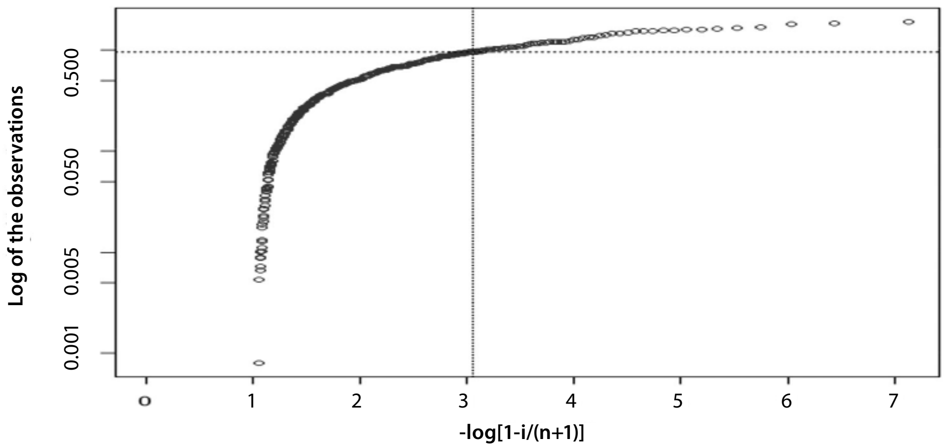

Figure 5: Pareto quantile plot

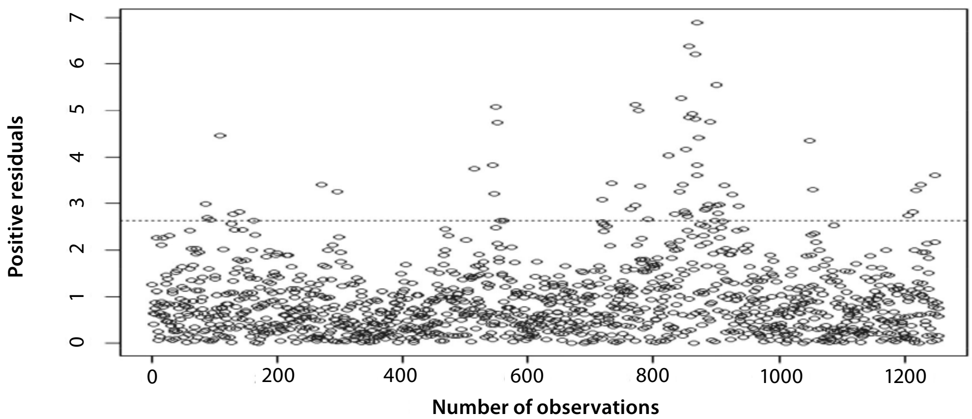

Figure 6: Plot of positive residuals (the threshold is 2.6206 – horizontal line)

Diagnostic plots for the fitted GPD are shown in Figure 7. If the QQ-plot follows a 45^0 line it indicates a good fit. Figure 7 indicates a fairly good fit of the GPD to the exceedances.

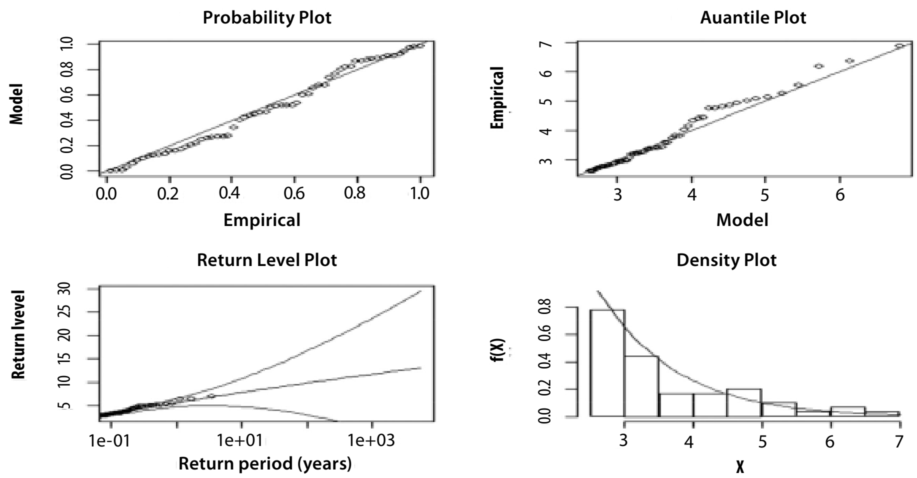

Figure 7: Diagnostic plots illustrating the fit of the exceedances to the GPD

4.3 Estimation of extreme quantiles

Table 5 shows the number of the exceedances related to the corresponding tail probabilities.

Table 5: Estimated tail quantiles at different probabilities (Number of exceedances)

Tail probability (*p*)	Expected Observation	ARMA-GARCH	Conditional GPD
0.05	63	107	34
0.01	13	53	12
0.005	6	40	3
0.001	1	17	0

The number of expected observations can be calculated by multiplying the number of residuals by tail quantiles (Bystrom, 2005). The theoretical number of exceedances of a 95 per cent tail quantile over positive residuals of 1 255 is (0.05*1255) = 63. The ARMA-GARCH model is presented with conditionally normally distributed standardised residuals. Using the quantile function given in equation (6) for the GPD we get

$$\varepsilon_{t,0.05} = \frac{1.0896}{-0.0321}[0.05^{0.0321} - 1] = 3.112$$

We then count the number of observations above the estimated tail quantile ($\varepsilon_{t,0.05} = 3.112$) and get 34. The ARMA-GARCH model significantly underestimates all tail quantiles, resulting in an excessive number of exceedances. The conditional GPD approach produces better forecasting results, thus the ARMA-GARCH-GPD model yields more accurate estimates of extreme tail quantiles.

The findings of this study also show that the ARMA-GARCH-GPD model performs well, especially in financial markets where the distribution of returns exhibits large movements. A plot of exceedances against tail probabilities for expected observations is given in Figure 8.

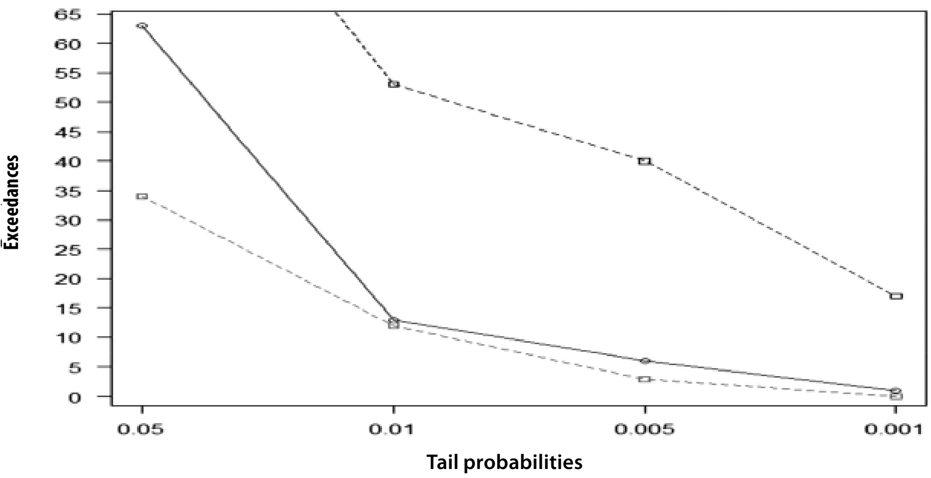

Figure 8: Exceedances against tail probabilities for expected observations

In Figure 8 the exceedances against the tail probabilities is given by the middle solid line, the conditional GPD is shown by the dashed line below and the ARMA-GARCH is indicated by the dashed line above**.** Figure 9 reflects the monthly frequency of occurrence of exceedances. Note that October has the highest number of exceedances above the threshold. This investigation provides an important implication to investors and risk managers when modelling extreme events at the JSE.

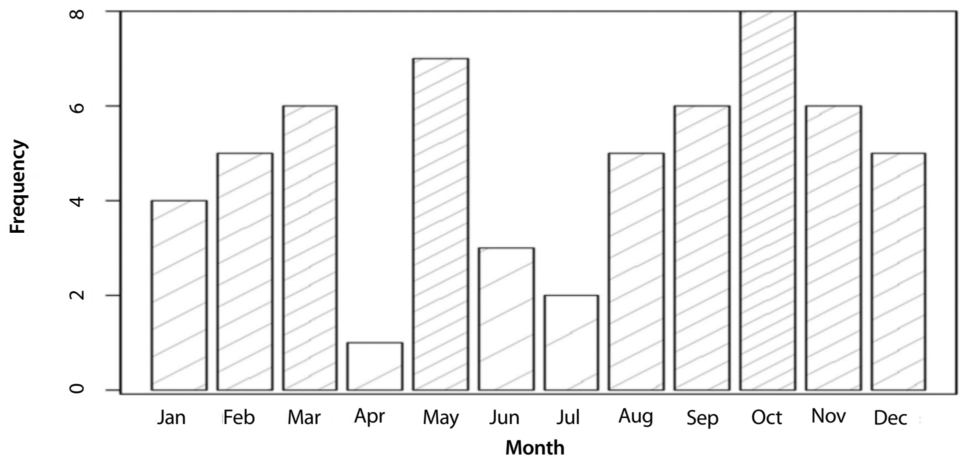

Figure 9: Monthly frequency of occurrence of exceedances

The bar chart of the yearly frequency of occurrence of exceedances is given in Figure 10, where 2008 has the highest number of exceedances, followed by 2009. This is probably due to the global recession of 2008/2009.

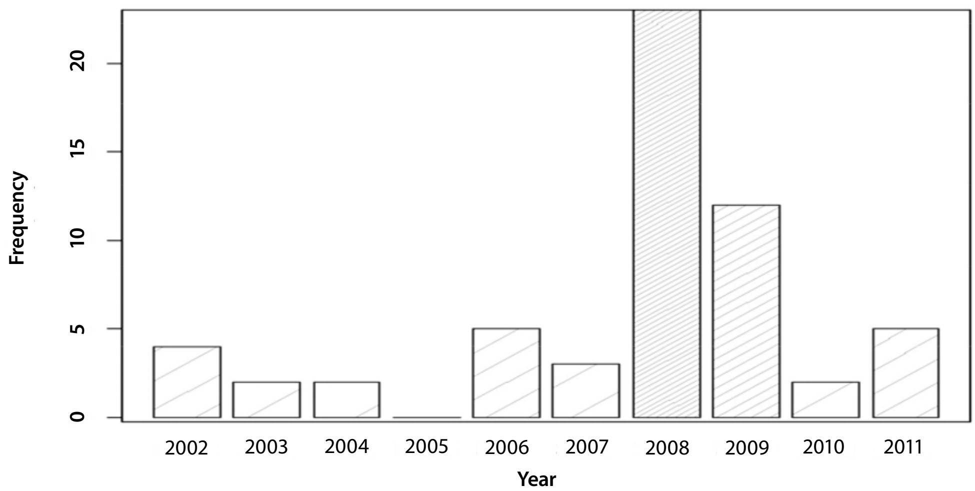

Figure 10: Annual frequency of occurrence of exceedances

5 Conclusion

The article has modelled conditional heteroskedastic stock returns at the JSE for the period 7 January 2002 to 30 December 2011. An ARMA-GARCH model is applied in stage one, with a view to filtering the return series to obtain nearly *i.i.d* residuals. In stage two, the EVT framework is applied to the standardised residuals from the ARMA-GARCH model. The results show that residuals can be modelled using Weibull class distributions. The ARMA-GARCH model overestimates all tail quantiles, thus the distribution fails to model the positive tail accurately. The ARMA-GARCH-GPD model produces more accurate estimates of extreme returns than the ARMA-GARCH model.

In summary, the results of this article support the combination of the ARMA-GARCH model with EVT for estimating upper extreme quantiles. In particular, the results show that the participants in the JSE market can rely on EVT-based models such as GPD when modelling the conditional heteroskedasticity of extreme events. The article also has important implications for investors and risk managers. Bystrom (2005) indicates that value-at-risk (VaR) performance under a GARCH-EVT framework is superior to a number of parametric approaches. The findings of this study show that the ARMA-GARCH-GPD model performs well, especially in financial markets where the distribution of returns exhibits large movements. It would be interesting to see what sort of results we get if we use other threshold selection methods, including comparative analysis with ARMA-GARCH-GSPD and ARMA-GARCH-GEVD models, where GEVD denotes the Generalised Extreme Value Distribution and GSPD the Generalised Single Pareto Distribution. Another interesting areas for further research will include a Bayesian estimation of the GARCH (1, 1) model with Student-*t* innovations. These will be studied elsewhere.

Biographical notes

Dr **Caston Sigauke** is a lecturer in the School of Statistics and Actuarial Science. He has an MSc in Operations Research and a PhD in Statistics. His research interests are in forecasting, applied statistical modelling, Bayesian statistics, extreme value theory, econometric modelling and time series analysis.

Rhoda Makhwiting has an MSc in Statistics and is a statistician with Statistics South Africa. Her research interests are extreme value theory and volatility modelling using GARCH type models.

Professor **Maseka Lesaoana** is Associate Professor and Director in the School of Mathematical and Computer Sciences at the University of Limpopo. She has an MMath (Combinatorics and Optimisation) and a PhD (Operational Research). Her areas of interest and expertise are statistical analysis and operations research.

Acknowledgements

The authors thank the JSE officials for their respective company information and data, and the numerous people who made valuable comments pertaining to this article. Thanks also go to the National Research Foundation for their financial support.

References

Beirlant, J., Goedgebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of extremes: Theory and applications*, Wiley Series in Probability and Statistics, England.

Berndt, E.K., Hall, B.H., Hall, R.E., and Hausman, J.A. (1974). 'Estimation and inference in nonlinear structural models', *Annals of Economic and Social Measurement*, vol. 4, no. 4, pp. 653–665.

Bollerslev, T. (1986). 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327.

Bystrom, H.N.E. (2005). 'Extreme value theory and extremely large electricity price change', *International Review of Economics and Finance*, vol. 14, no. 1, pp. 41–55.

Castillo, E., and Daoudib, J. (2009). 'Estimation of the generalized Pareto distribution', *Statistics and Probability Letters*, vol. 79, no. 5, pp. 684–688.

Chan, K.F., and Gray, P. (2006). 'Using extreme value theory to measure value-at-risk for daily electricity spot prices', *International Journal of Forecasting*, vol. 22, no. 2, pp. 283–300.

Engle, R.F. (1982). 'Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation', *Econometrica*, vol. 50, no. 4, pp. 987–1008.

Fernandez, V. (2005). 'Risk management under extreme events', *International Review of Financial Analysis*, vol. 14, no. 4, p. 113–148.

Gencay, R., and Selcuk F. (2004). 'Extreme value theory and value-at-risk: Relative performance in emerging markets', *International Journal of Forecasting*, vol. 20, no. 2, pp. 287–303.

Gilli, M., and Kellezi, E. (2006). 'An application of extreme value theory for measuring financial risk', *Computational Economics*, vol. 27, no. 2, pp. 207–228.

Guru, A. (2012). 'Extreme movements in the Indian equity markets: An extreme value theory approach', *International Journal of Social Sciences & Interdisciplinary Research*, vol. 1, no. 2, pp. 6–18.

Maghyereh, A.I., and Al-Zoubi, H.A. (2008). 'The tail behaviour of extreme stock returns in the Gulf emerging markets', *Studies in Economics and Finance*, vol. 25, no. 1, pp. 21–37.

Mcneil, A., and Frey, R. (2000). 'Estimation of tail-related risk measures for heteroskedastic financial time series: An extreme value approach', *Journal of Empirical Finance*, vol. 7, no. 3–4, pp. 271–300.

Pickands, J. (1975). 'Statistical inference using extreme order statistics', *Annals of Statistics*, vol. 3, no. 1, pp. 119–131.

Sigauke, C., Verster, A., and Chikobvu, D. (2012). 'Tail quantile estimation of heteroskedastic intraday increases in peak electricity demand', *Open Journal of Statistics*, vol. 2, pp. 435–442.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis*, Chapman and Hall, London.

Song, J., and Song, S. (2012). 'A quantile estimation for massive data with generalized Pareto distribution', *Computational Statistics and Data Analysis*, vol. 56, no. 1, pp. 143–150.

Verster, A., and De Waal, D.J. (2011). 'A method for choosing an optimum threshold if the underlying distribution is Generalized Burr-Gamma', *South African Statistical Journal*, vol. 45, no. 2, pp. 273–292.

Wagner, N., and Marsh, T.A. (2005). 'Measuring tail thickness under GARCH and an application to extreme exchange rate changes', *Journal of Empirical Finance*, vol. 12, no. 1, pp. 165–185.

Appendix

Derivation of the quantile function of the GPD

The cumulative distribution function of GPD is given by:

$$W_{\xi}(\varepsilon_t) = \begin{cases} 1 - \left[1 + \dfrac{\xi(\varepsilon_t - \tau)}{\sigma}\right]^{\frac{-1}{\xi}}, & \xi \neq 0 \\ 1 - e^{-\frac{\varepsilon_t - \tau}{\sigma}}, & \xi = 0 \end{cases}$$

The survival function of the GPD is $P(\mathcal{E}_t > \varepsilon_t) = 1 - W_{\xi(\varepsilon_t)}$

$$P(\mathcal{E}_t > \varepsilon_t) = \begin{cases} 1 - \left[1 + \dfrac{\xi(\varepsilon_t - \tau)}{\sigma}\right]^{\frac{-1}{\xi}}, & \xi \neq 0 \\ 1 - e^{-\frac{\varepsilon_t - \tau}{\sigma}}, & \xi = 0 \end{cases}$$

Let $p = P(\mathcal{E}_t > \varepsilon_t)$

$$p = 1 - \left[1 + \frac{\xi(\varepsilon_t - \tau)}{\sigma}\right]^{\frac{-1}{\xi}}, for\ \xi \neq 0$$

$$(1-p)^{-\xi} = \left[1 + \frac{\xi(\varepsilon_t - \tau)}{\sigma}\right]$$

$$\varepsilon_t = \tau + \frac{\sigma}{\xi}\left[(p)^{-\xi} - 1\right]$$

Then

$$\varepsilon_{t,p} = \frac{\sigma}{\xi}\left[(p)^{-\xi} - 1\right]$$

After dropping and is the quantile function.

Similarly when $\xi = 0$ we have $\varepsilon_{t,p} = -\sigma 1n(p)$ (For more details see Beirlant *et al.* 2004.)