FITTING A CONTAMINATED NORMAL DISTRIBUTION TO THE RESIDUALS OF FINANCIAL DATA

S.M. Ellis & H.S. Steyn jr)

Centre for Business Mathematics and Informatics
Potchefstroom University for CHE

Abstract

The family of GARCH models provides a means of estimating and correcting for the heteroscedasticity of the log returns of financial series. The standardised residuals (innovations) of the GARCH fit are, however, heavy-tailed implicating that the assumption of conditional normality does not hold for real data. Fitting a contaminated normal (CN) distribution on the standardised residuals can account for the heavy tails as well as for possible skewness of the innovations. We extended this method further to simultaneous maximum likelihood estimation of all the parameters. Monte Carlo simulations of normal, as well as heavy-tailed error-distributions (skewed and symmetrical), are used to compare Value-at-Risk and Expected Shortfall determined by the CN distribution with other methods. This method is also applied to South African stock exchange data.

Key terms: heteroscedasticity, GARCH models, contaminated normal distribution, Monte Carlo simulations, Value-at-Risk, Expected Shortfall

1 Introduction

During the last ten years the so-called Value-at-Risk (*VaR*) has become a popular measure of risk. We define the $(q \times 100)\%$ *VaR* of a log return *Y* as the lower q^{th} quantile of the probability distribution of *Y*, where q is some small probability (e.g. $q = 0.05$).

Recently, Artzner *et al*. (1999) introduced a number of desirable properties that risk measures should possess in general and showed that *VaR* does not necessarily have these properties. First they showed that *VaR* is not necessarily subadditive and thus not coherent. More so, *VaR* gives only an upper bound on the losses that occur with a given frequency, without referring to the potential size of the loss, given that a loss exceeding this upper bound has occurred. They proposed that the so-called expected shortfall (*ESf*) should be used together with *VaR* since it does have these desirable properties. We denote the *ESf* of *Y* by *ESf*(*Y*) and define it as the conditional expectation of *Y* given that *Y* falls below *VaR*(*Y*), i.e.

$$ESf\,(Y) = E[Y \mid Y < VaR(Y)]\,.$$

Both of these risk measures are determined by the volatility of the return distribution as well a as the value of a specified quantile of the return distribution

A stationary $AR(m)-GARCH(p,q)$ process, as described by Bollerslev (1986), is assumed as a possible model for $\{Y_t,\ t=1,2,\ldots,n\}$, to account for the heteroscedasticity of the log returns on a financial asset price. This model assumes that

$$Y_t = m_t + e_t \ , \text{ with } e_t = \sqrt{h_t} Z_t \tag{1}$$

where the innovations (or noise effects) Z_t are independent and identically distributed over time t with zero expectation, unit variance and marginal distribution function $F(z)$. Further m_t represents the return that would be obtained in the absence of noise effects and h_t represents the heteroscedasticity (volatility) inherent in the noise contributions to the actual returns,

$$m_t = n + \sum_{k=1}^{m} f_k y_{t-k} \text{ and } h_t = w + \sum_{i=1}^{q} d_i \left(y_{t-i} - m_{t-i} \right)^2 + \sum_{j=1}^{p} g_j h_{t-j} \tag{2}$$

with the autoregressive parameters restricted by $\left| f_k \right| < 1, \quad k = 1, 2, \ldots, m$, and the GARCH parameters by $w > 0, \quad d_i > 0, \ i = 1, 2, \ldots, q$, $g_j > 0$, $j = 1, 2, \ldots, p$ and $\sum_i d_i + \sum_j g_j < 1$.

To express the corresponding risk measures *VaR* and *ESf*, we first define these notions for the noise distribution $F(z)$. With q-th quantile $z_q = F^{-1}(1-q)$ and the assumption that $F(z)$ has zero expectation, follows that

$$VaR(Z) = z_q \text{ and } ESf(Z) = E\left[Z \middle| Z < z_q \right] \tag{5}$$

where Z is distributed as the Z_t's. Then it follows from (1) that the conditional *VaR* and *ESf* of Y_{T+1}, given the information up to time T, are given by

$$VaR[Y_{T+1}] = \sqrt{h_{T+1}} VaR(Z) \text{ and } ESf\,[Y_{T+1}] = \sqrt{h_{T+1}} ESf(Z) \tag{6}$$

respectively. Hence we also need estimates of $VaR(Z)$ and $ES(Z)$ in order to be able to estimate conditional risk measures.

Our aim in this paper is to compare risk estimators determined by a contaminated normal (CN) distribution with results from current methods. Monte Carlo simulations of normal, as well as heavy tailed error-distributions, are used to compare this distribution with other methods to calculate *VaR* and *ESf*. It is also applied to South African stock exchange data.

In section 2 an overview of current methods to determine risk is given and in section 3 risk estimation by the CN distribution is described for both PMLE and full MLE. In section 4 different measures of fit, independent of sample size are described. Section 5 gives the results of the simulation study and in section 6 the different methods are applied to South African stock exchange data.

2 Current methods to determine *VaR(Z)* and *ESf(Z)*

2.1 The N-, NE- and T-methods

The N-method assumes that the innovations have a standard normal distribution and thus *VaR(Z)*= $z_q = \Phi^{-1}(q)$ is the q^{th} theoretical quantile of the standard normal distribution, with $\Phi^{-1}(\cdot)$ the inverse standard normal distribution function. The expected shortfall, in this case is:

$$ESf(Z) = E[Z \mid Z < z_q] = \frac{1}{q\sqrt{2p}} \int_{-\infty}^{z_q} y\, e^{-\frac{1}{2}y^2} dy = -\frac{f(z_q)}{q},$$

where $f(\cdot)$ is the density function of a standard normal distribution. PROC AUTOREG of SAS$^{®}$ (SAS Institute Inc., 1999) is implemented to calculate the $AR(m) - GARCH(p,q)$ estimators, by maximising the log-likelihood function $L = \sum_{t=1}^{n} \frac{1}{2}\left[-\ln(2p) - \ln(h_t) - z_t^2\right]$.

The standardised residuals (innovations) of the AR-GARCH fit are, however, heavy-tailed implicating that the assumption of conditional normality does not hold for real data.

For the standardised residuals of the AR-GARCH-fit, the NE-method estimates the $V\hat{a}R(Z) = \hat{z}_{([nq+0.5])}$ as the $[nq+0.5]^{th}$ order statistic of the standardised residuals (or *q%*

empirical quantile as determined by PROC UNIVARIATE of SAS®, SAS Institute Inc., 1999) and $E\hat{S}f(Z) = \sum_{t=1}^{[nq+0.5]} \frac{\hat{z}_{(t)}}{[nq+0.5]}$ is estimated from the mean of the standardised residuals smaller than the *VaR*(*Z*).

By incorporating the *distribution=T* option in PROC AUTOREG of SAS® (SAS Institute Inc., 1999), the $AR(m)-GARCH(p,q)$ estimates of the T-method are calculated, assuming that the standardised residuals have a Student t-distribution to account for the heavy tails of the innovation distribution. The $V\hat{a}R(Z) = \hat{z}_q = \sqrt{(\hat{u}-2)/\hat{u}}\ F_T^{-1}(q),$ the theoretical standardised quantile of the Student t-distribution with $\hat{u}$ the degrees of freedom as determined in the AR-GARCH fit, while the $E\hat{S}f(Z) = \frac{\sqrt{(\hat{u}-2)/\hat{u}}}{q}\int_{-\infty}^{\hat{z}_q} z\ f_T(z)\,dz$ is determined by numerical integration in SAS® (SAS Institute Inc., 1999).

2.2 The E-method

McNeil & Frey (2000) suggested that EVT is used to improve on the estimates for *VaR* and *ESf* as follows. To account for the heteroscedasticity of the data, a PMLE $AR(m)-GARCH(p,q)$ fit is performed in the first stage. The method of PMLE calculates consistent estimates under the assumption of normality for the conditional distribution, Gouriéroux (1997). In the second stage, the standardised residuals from the $AR(m)-GARCH(p,q)$ fit are calculated. These residuals are approximately i.i.d. with heavy tails. We may approximate the tail of $F(z)$ below a suitable threshold, $z_a = F^{-1}(a)$, for $a > q$ by a generalised Pareto distribution (GPD) with parameters b and x, i.e. we have approximately

$$F(z) = a[1-G_{x,b}(z_a - z)] \text{ for } z < z_a,$$

with $G_{x,b}(y)$ the GPD distribution function:

$$G_{x,b}(y) = \begin{cases} 1-(1+xy/b)^{-1/x} & \text{if } x \neq 0, \\ 1\text{-exp}(-y/b) & \text{if } x = 0, \end{cases}$$

Hence also approximately

$$VaR(Z) = F^{-1}(q) = z_a - G_{x,b}^{-1}(1 - q/a)\,.$$

We estimate z_a by $\hat{z}_{([na \quad 0.5])}$ b and x, PROC NLP of SAS® (SAS Institute Inc., 1999) is implemented to maximise the log-likelihood function,

$$L = \sum_{z_i < \hat{z}_{([na+0.5])}} \ln\left[a\, g_{x,b}(\hat{z}_{([na+0.5])} - z_i)\right] ,$$

with $g_{x,b}(\cdot)$ the GPD density function.

From the VaR and ESf of a Pareto distribution, see McNeil & Frey (2000), follows that $VaR(Z)$ is estimated by

$$V\hat{a}R(Z) = \hat{z}_{([na+0.5])} - G_{\hat{x},\hat{b}}^{-1}(1 - q/a) = \hat{z}_{([na+0.5])} - (\hat{b}/\hat{x})\{(q/a)^{-\hat{x}} - 1\}$$

and $$E\hat{S}f(Z) = (Va\hat{R}(Z) - \hat{b} - \hat{x}\hat{z}_{([na+0.5])}) / (1 - \hat{x})\,.$$

McNeil & Frey (2000) took the proportion of data in the tail as $a = 0.1$. This decision was based on a simulation study in which they showed that the mean square error (MSE) of the GPD estimator of k (the number of data smaller than the threshhold, z_a) is robust with respect to the choice of k. The problem of choosing an optimal threshold is, however, a major cause of concern in EVT.

2.3 The PNG-method

The Pareto-Normal-Pareto (PNP) distribution was suggested by Ellis *et al.* (2002) and assumes that the standardised residuals of the AR-GARCH fit, Z_t, are $N(m, s^2)$ distributed between two threshold values with Pareto tails below and above the respective thresholds.

The parameters of this distribution can be estimated either by a two-stage approach, where the AR-GARCH parameters are estimated with PMLE in the first stage and the PNP parameters in the second stage or by simultaneous MLE of all the parameters (see Ellis *et al.*, 2002). We call the last method PNG.

An advantage of this method is that we work with the original returns to estimate the parameters and not with the standardised residuals of a fit that may not be very accurate. The distribution of the standardised residuals of the GARCH fit may differ from that of the innovations because the assumption of normality in the GARCH-fit does not hold for the innovation distribution. {Also, provision is made for thresholds at both tails of the distribution of returns. } A disadvantage of this method is that more parameters have to be estimated as in the current methods to account for the thresholds.

When risk estimation is done by assuming that the standardised residuals have a CN distribution, all the data are used and thus the choice of an optimal threshold is irrelevant and fewer parameters need to be estimated.

4 Methods based on the Contaminated Normal (CN) distribution

If it is assumed that the standardised residuals of the AR-GARCH fit, Z_i, are conaminated normal distributed, then

$$Z = (1-d)e_1 + de_2 \,,$$

with $e_1 \sim N(m_1, s_1^2)$, $e_2 \sim N(m_2, s_2^2)$ and $d = \begin{cases} 1 & \text{with probability} \quad p \\ 0 & \text{with probability} \quad (1-p) \end{cases}$, independent from each other. The density function of Z is then

$$f_Z(z) = (1-p)\,\frac{1}{\sqrt{2p\,s_1^2}}\,e^{-\frac{1}{2}\left(\frac{z-m_1}{s_1}\right)^2} \quad + \quad p\,\frac{1}{\sqrt{2p\,s_2^2}}\,e^{-\frac{1}{2}\left(\frac{z-m_2}{s_2}\right)^2} \;.$$

A constraint of the AR-GARCH process is that innovations must have zero expectation and unit variance. This means that

$$E(Z) = (1-p)\,m_1 + pm_2 = 0 \quad ,$$

so that

$$m_2 = \frac{-(1-p)m_1}{p} \;. \qquad (7)$$

Also

$$
\begin{aligned}
E(Z^2) &= (1-p)\,E_{e_1}(e_1^2) + p\,E_{e_2}(e_2^2) \\
&= (1-p)\left(s_1^2 + m_1^2\right) + p\left(s_2^2 + m_2^2\right) \\
&= (1-p)\left(s_1^2 + m_1^2\right) + ps_2^2 + \frac{\left(1-p\right)^2 m_1^2}{p} = 1.
\end{aligned}
$$

From $E(Z^2) = 1$ follows that

$$
s_2^2 = \frac{1}{p} - \frac{(1-p)}{p}\left(s_1^2 + \frac{m_1^2}{p}\right), \tag{8}
$$

so that we only have to estimate parameters $p,\ s_1^2,\ m_1$.

This parameters can be estimated either by a two-stage approach, where the AR-GARCH parameters are estimated with PMLE in the first stage and the CN parameters in the second stage, or by simultaneous MLE of all the parameters.

3.1 The CN-method

The parameters $p,\ s_1^2,\ m_1$ .of the CN distribution are estimated from the standardised residuals of the AR-GARCH fit by maximising of the log-likelihood function *L* as function of p, m_1 and s_1, using (7) and (8), where

$$
L = \sum_{i=1}^{n} \ln f_Z(z_i) = \sum_{i=1}^{n} \ln\left[\frac{(1-p)}{s_1} f\left(\frac{z_i - m_1}{s_1}\right) + \frac{p}{s_2} f\left(\frac{z_i - m_2}{s_2}\right)\right],
$$

with constraints $0 \le p \le 1,\ s_1 > 0$ and $m_1 \ge 0.$ PROC NLP of SAS® (SAS Institute Inc., 1999) is implemented to calculate the MLE of these parameters in the second stage.

The quantiles are estimated from the contaminated normal distribution function

$$
F(z) = (1-\hat{p})\,\Phi\left(\tfrac{z-\hat{m}_1}{\hat{s}_1}\right) + \hat{p}\ \Phi\left(\tfrac{z-\hat{m}_2}{\hat{s}_2}\right) ,
$$

with $\Phi(\cdot)$ the standard normal distribution function and $\hat{p}, \hat{m}_1, \hat{m}_2, \hat{s}_1, \hat{s}_2$ the MLE of the parameters. The value of *z* are incremented in a suitable interval and $F(z)$ determined until $\hat{z}_q$ is estimated with the desired accuracy for a given $F(z) = q$. The *VaR* estimate for the return on time *T*+1 is then, by definition,

$$V\hat{a}R(Y_{T+1}) = \sqrt{\hat{h}_{T+1}}\ \hat{z}_q \ ,$$

with $\hat{h}_{T+1}$ the variance on time *T*+1, estimated by the MLE of the AR-GARCH fit. The estimated expected shortfall on time *T*+1 is obtained by numerical integration in SAS® (SAS Institute Inc., 1999),

$$E\hat{S}f\,(Y_{T+1}) = \tfrac{\sqrt{\hat{h}_{T+1}}}{q} \int_{-\infty}^{\hat{z}_q} z\ f_Z(z)\,dz \ .$$

This method accounts for the heteroscedasticity of the innovation distribution by the AR-GARCH fit, while the skewness and heavy tails are accounted for by the contaminated normal distribution.

3.2 The CNG-method

This is an extension of the CN-method where the AR-GARCH and CN parameters are estimated simultaneously by MLE. Similarly to the PNG-method, described in section 2.3, the log-likelihood function is maximised under the assumption of CN distributed innovations with zero mean and unit variance.

Parameters $p,\ s_1^2,\ m_1, n, f_1, \ldots, f_m,\ w,\ d_1, \ldots, d_q,\ g_1, \ldots, g_p$ are estimated with constraints as in section 3.1 and for the AR-GARCH models, see section 1. PROC NLP of SAS® (SAS Institute Inc., 1999) is implemented to calculate the MLE of these parameters. The estimates for quantiles, *VaR* and *ESf* are obtained in the same way as in section 3.1.

4 Measures of fit

Ordinary methods of goodness-of-fit of distributions like the Kolmogorov-Smirnov and $c^2 -$tests statistics have the disadvantage that as sample size increases, these tests have a greater tendency to indicate significant deviations from a specified distribution. For a large enough sample a small *p*-value, indicating significance, will be found for any specified distribution. Financial data sets are usually large and therefore we would prefer to use an index, independent of sample size, to determine goodness-of-fit.

4.1 Effect size as a measure of fit

We use effect sizes to define practical significance, where an effect is defined as practically significant if the effect size is large. To determine the goodness-of-fit, a small effect size (defined as of no practical significance) would indicate that the fit is good while a large effect size would indicate a practically significant deviation from the proposed fit. Effect sizes of magnitude 0.5 are said to be large (Cohen, 1988), while effect sizes of magnitude 0.3 are said to be medium and those of magnitude 0.1 small. According to Cohen (1988) the effect size for deviation from a theoretical distribution is $w = \sqrt{\frac{X^2}{n}}$, where X^2 is the usual $c^2 -$statistic for goodness-of-fit. This gives an index, independent of sample size, to determine the goodness-of-fit, (see Ellis *et al.*, 2002 for more details).

4.2 Backtesting of ***VaR***

For a financial return series $y_1, y_2, \ldots, y_n$ where *n*>>*m*, $VaR(y_{t+1})$ is calculated for each day $t \in T = \{m, m+1, \ldots, n-1\}$ using a moving window of *m* days for each calculation, see McNeil & Frey (2000). For each day $t \in T$ a new model is fitted and the estimate of $VaR(y_{t+1})$ is compared with the y_{t+1}-th return for $q \in \{0.005,\ 0.01,\ 0.05\}$. A so-called "violation" occurs if $y_{t+1} < \hat{m}_{t+1} + V\hat{a}R(y_{t+1}) = \hat{m}_{t+1} + \sqrt{\hat{h}_{t+1}}\ \hat{z}_q$.

The indicator for a violation at time $t \in T$ is Bernoulli(*q*) distributed and because z_{t+1} and z_{s+1} are independent for $t, s \in T$ and $t \neq s$, the total number of violations are binomially distributed with parameters *q* and the number of backtests, (*n*-*m*). A two-sided binomial test that the *VaR* gives too few or too many violations is then performed. A *p*-value of less than 0.05 will be interpreted as evidence against the null hypothesis of a proportion of *q* violations.

4.3 Backtesting of *ESf*

McNeil & Frey (2000) also developed a test for *ESf*, for all cases of *VaR* violation. To determine the size of the returns smaller than *VaR*, the shortfall residuals are defined as

$$r_{t+1} = \frac{\hat{m}_{t+1} + E\hat{S}f\,(y_{t+1}) - y_{t+1}}{\sqrt{\hat{h}_{t+1}}} \quad .$$

These residuals are i.i.d. and should have a mean of zero. To test this hypothesis of zero mean, a bootstrap test that makes no assumption about the underlying distribution, is performed (Efron & Tibshirani, 1993).

5 Simulation study

We simulated the return series using the $AR(1)-GARCH(1,1)$ parameters of a financial time series typical of daily returns for stock prices in the South African market. In May 2001, we analysed ten of the important shares in the ALSI40 basket and determined the median values for the estimated AR-GARCH parameters. Thereafter we chose a share with estimated parameters closest to the median - the share, SA-BREWS. We used the estimated values of the AR-GARCH parameters of the SA-BREWS share to simulate the volatility in the simulated return series. The innovations (error term) are simulated to allow for the following different distributions:

i) Standard normal (Normal).

ii) Student's t with 3 degrees of freedom, t(3).

iii) Skewed contaminated normal (CNS) where the standard normal was contaminated with 20% N(-1,16) innovations.

iv) Normal with a Pareto tail (PN) where the lower 10% of the normal innovations was replaced with GPD innovations with x parameter typical of South African daily stock returns.

In all these cases the error terms were obtained after the simulated innovations had empirically been standardised. Note that in cases ii) - iv) we have heavy tailed error distributions in contrast to the normal innovations. The methods, described in sections 2 and 3,, are repeatedly implemented on the same simulated data set for each of the innovation studies.

To obtain the bias of the different methods on the same basis, the mean of the quantiles estimated by the different methods, are compared with the mean of the simulated innovation quantiles (S) and not with the theoretical values.

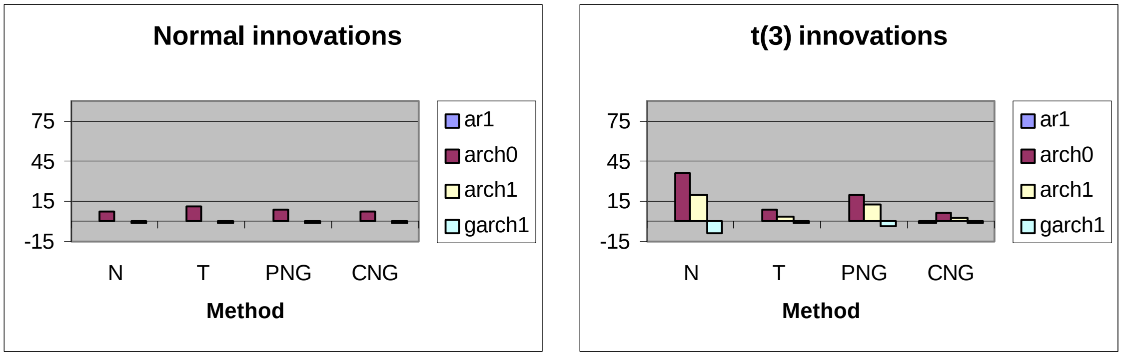
The length of the simulated return series is 1500, corresponding to about 6 years of daily returns for a financial stock price. The number of Monte Carlo simulations, for comparing the different methods, is 2000 for each of the innovation studies.

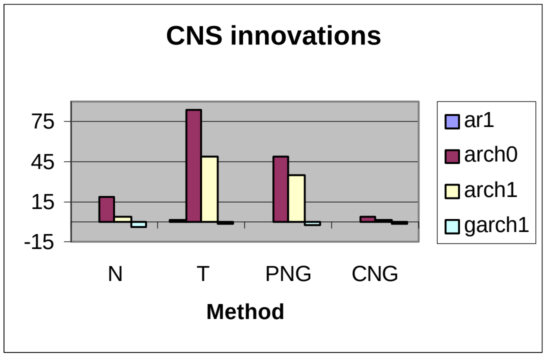
From the simulation results of the T-method, follows that the CNS innovations have the heaviest tails, with MLE of the degrees of freedom (df) assuming a t-distribution, being 2.4. The MLE of the df for the t(3) innovations are 3 (as expected) and for the PN innovations (with shape parameter, $x = 0.35$), df is estimated as 6.7. For the normal innovations the MLE of the df is an eight digit integer, indicating infinity.

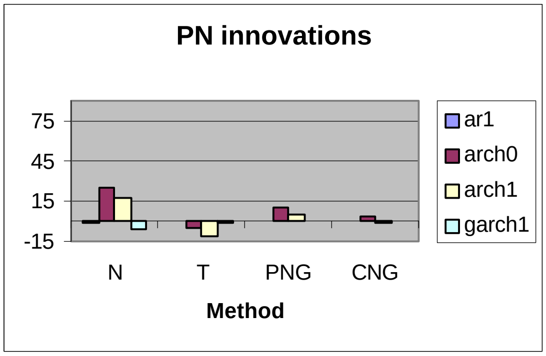
When the CNG-method is used, the MLE of the all parameters, but especially those of the contaminated normal distribution for CNS innovations correspond very well with the standardised values of the simulated innovations.

The following bar charts (in Figure 5.1) compare the percentage bias of the estimators for the AR-GARCH parameters. This comparison is for the different methods for each type of innovation. The percentage bias represents the mean bias as a percentage of the simulated value of the parameter. Note that the biases of the NE-, E- and CN-methods are not reported since these estimates are those of the N-method.

Figure 5.1 % Bias for AR-GARCH parameter estimates







From these bar charts follow that the w (arch0) parameter has the largest bias and the f (ar1) parameter the smallest bias of the AR-GARCH parameters for all the methods for each type of innovation. The T-method has large bias for CNS innovations, while the N-method has a relatively large bias for t(3) and PN innovations. The CNG-method has overall the smallest bias of AR-GARCH parameters for all the methods for each type of innovation.

The bias of contaminated normal parameters for CNS innovations is compared for the CN- and CNG-methods in Tables 5.1.

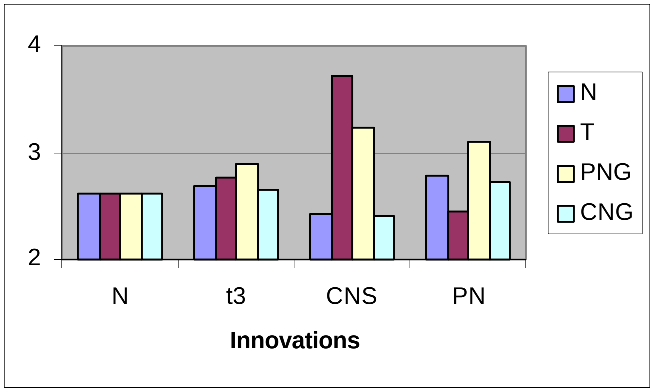
Table 5.1 Bias of contaminated normal parameters, CNS innovations

	Method	
	CN	CNG
p	-0.119	-0.004
m_1 (mu1)	-0.062	-0.010
m_2 (mu2)	-0.035	0.026
s_1 (sigma1)	0.239	-0.004
s_2 (sigma2)	0.496	-0.018

From Tables 5.1 follows that the bias for the CN-method is much larger than for the CNG-method, especially for the parameters s_1 and s_2. In the remainder of this section, only results for the CNG method will further be reported.

The estimated variance, at time *T*+1, of the different methods is computed from the estimated AR-GARCH parameters and is compared in Figure 5.2 for the different types of innovations. Note that the estimated variance of the E- and CN- methods are not reported since the estimates of the N-method are used.

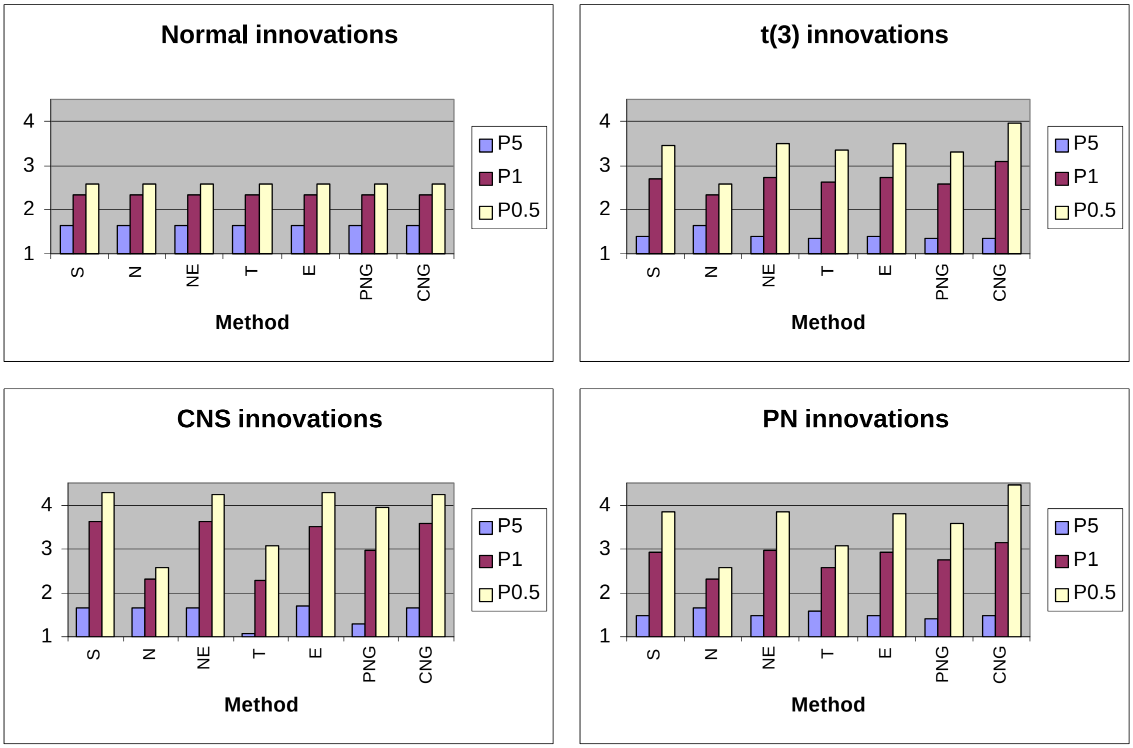
Figure 5.2 Comparison of estimated variance, at time *T*+1, for different methods



From Figure 5.2 follows that all the methods yield approximately the same estimated variance for normal innovations. The estimated variance for the T- and PNG-methods is larger than that of the N- and CNG-method for CNS innovations. The T- method also has large biases for the estimation of AR-GARCH parameters (see Figure 5.1). The CNG-method has the smallest bias of all the methods in estimating the AR-GARCH parameters and we conclude that it will estimate the variance accurately. The variance estimated by the N-method is also approximately the same as that of the CNG-method for all the innovations and can be used alternatively.

The following bar charts (Figure 5.3) compare the mean estimators of the 0.5%, 1% and 5% quantiles (P0.5, P1 and P5) for the different methods for each type of innovation. In these charts S represents the mean quantile of the simulated return series. The NE quantiles are the empirical quantiles of the standardised residuals of the AR-GARCH fit for the N-method. Note that the absolute values of the quantiles are reported.

Figure 5.3 Comparison of quantiles

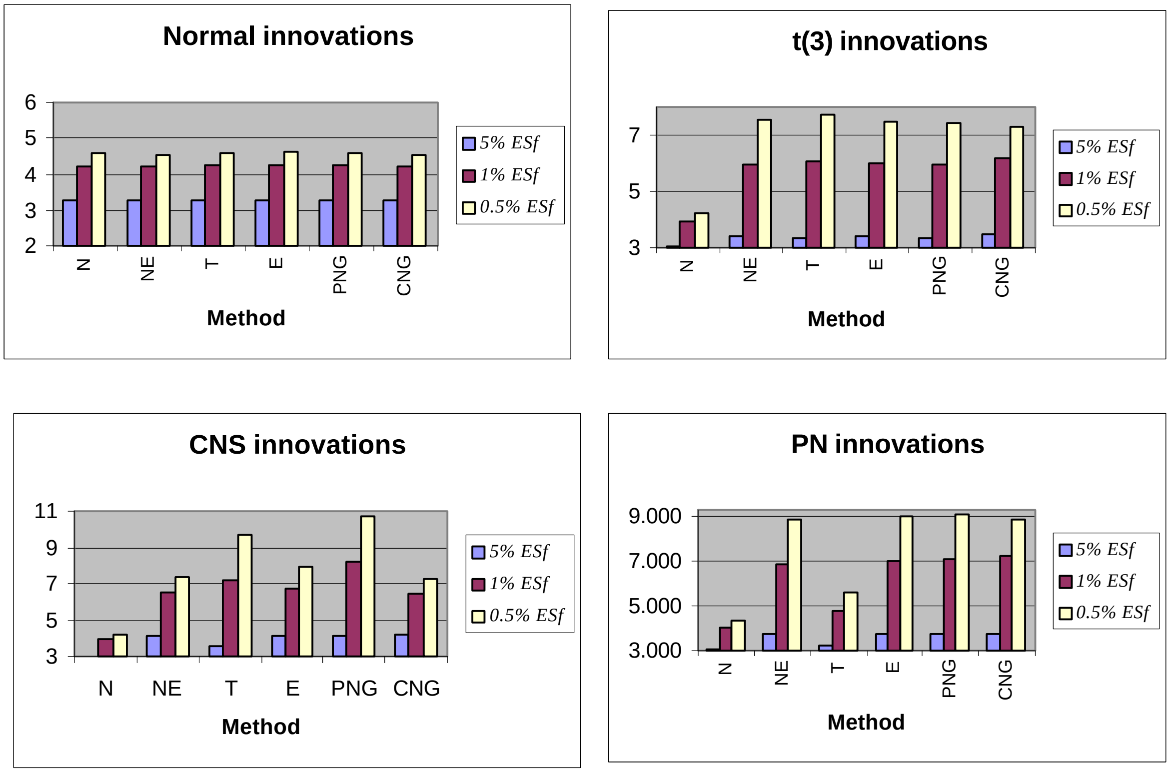


From Figure 5.3 it seems as if the E-method could be successfully applied to most of the innovations, since it compares well to the mean simulated quantiles (S). In most cases the E-method has smallest bias of all the methods for each type of innovation, the highest bias of the E-method are 6.4%.

The results for *VaR* estimates are similar to those of the quantiles due to the fact that the *VaR* is determined from the estimated variance and the estimated quantile. For all the innovations, except normal, the N-method underestimates the 1% and 0.5% *VaR*. It seems therefore as if the E-method are across board the best for estimating the *VaR*.

The following bar charts (Figure 5.4) compare the mean estimators of *ESf* for the different methods for each type of innovation. Note that the absolute values of the *ESf* are given.

Figure 5.4 Comparison of *ESf*



From Figure 5.4 follows that for normal innovations all methods approximately yield the same estimates for *ESf*. For t(3) innovations it seems as if the N-method underestimates the *ESf*, assuming that the T-method yields the most accurate estimates. For CNS innovations the *ESf* is underestimated by the N-method and overestimated by the T-method, assuming that the CNG-method yields the most accurate estimates. For PN innovations the *ESf* is underestimated by the N- and T-methods, assuming that the E-method yields the most accurate estimates. It seems therefore as if the E- and CNG- methods overall yield the most accurate estimates for *ESf*.

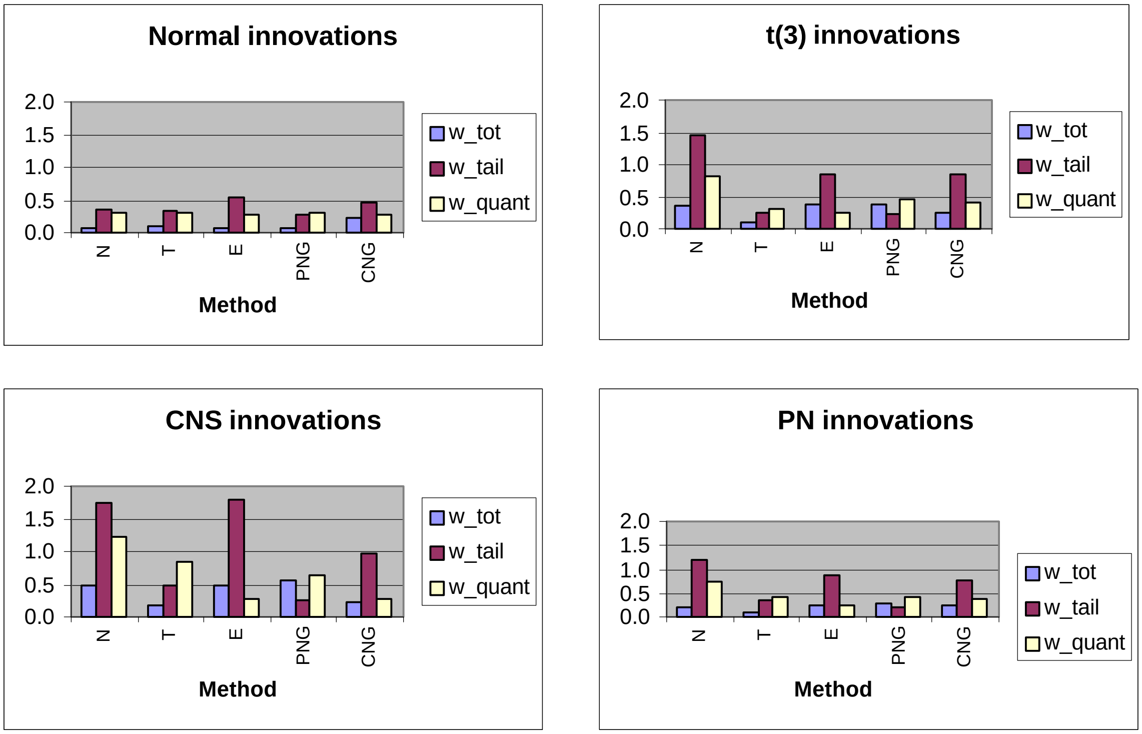
To calculate the effect size, we have to divide the data into a number of intervals of equal theoretical probability. We use Sturge's rule as an indication of the number of intervals. Three types of effect sizes are calculated:

- An effect size (w_tot) for the entire return series to determine the best overall fit on the empirical distribution. In this case we have divided the data into 15 intervals of equal theoretical probability for each of the different methods.

- An effect size for the tail (w_tail) of the return series to determine the best fit for the empirical distribution of the losses. The threshold for all the methods is taken as the MLE of the threshold determined by the PNG-method - this varied between 9% for normal innovations to about 23% for CN innovations. The tail is divided into 10 intervals of equal theoretical probability for each of the different methods.
- An effect size for the quantiles (w_quant) to determine the best fit in the lower 5% of the empirical distribution - this is also the region in which estimates for *VaR* and *ESf* are calculated. In this case the number of intervals of equal theoretical probability is taken as 7 for each of the different methods.

The effect sizes are compared in Figure 5.5 for the different methods for different types of innovations.

Figure 5.5 Comparison of effect sizes



From these bar charts it can be concluded that for the total return series, all the methods seem to have a reasonable fit. We can also conclude that for all the innovations the E- and CNG-methods have a reasonable fit in the lower 5% of the tail.

For the backtesting we simulate 20 returns series of length 1500 for each of the different innovations. We choose a moving window of 1000 observations in which the *VaR* and *ESf* is calculated each day and compared with the following day's return. This means that we have 500 backtestings on each simulated return series and in total 10000 backtestings for each type of innovation. For each simulated return series, the two sided binomial probability that the *VaR* gives too few or too many violations (see section 3.3), is determined. A summary of these results is given in Table 5.2.

Table 5.2 Number of times that the null hypothesis for 5% and 1% *VaR* is rejected

	Method											
	N		NE		T		E		PNG		CNG	
	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%
N	1	1	2	1	1	1	1	2	1	1	0	1
t(3)	4	7	2	1	1	1	3	1	3	2	1	1
CNS	2	20	2	1	7	14	1	4	2	2	1	1
PN	3	5	0	0	0	2	0	0	0	0	1	1
Total	10	33	6	3	9	18	5	7	6	5	3	4

From these results for individual simulations it follows that the CNG-, NE- and E-methods yield the most accurate estimates for *VaR*.

We calculated the shortfall residuals (see section 3.4) from the *ESf* and log return of the following day for those log returns smaller than the *VaR*. From each set of shortfall residuals we simulated 1000 bootstrap samples and determined the achieved significance level as the proportion of samples with test statistic smaller than the observed test statistic. A summary of these results is given in Table 5.3.

Table 5.3 Number of times that the null hypothesis for 5% and 1% *ESf* is rejected

	Method											
	N		NE		T		E		PNG		CNG	
	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%
N	0	1	0	4	0	3	0	1	0	0	0	1
t(3)	16	18	1	0	1	1	2	0	2	3	0	2
CNS	20	20	1	2	2	0	3	0	0	0	0	0
PN	17	16	2	4	13	8	1	2	1	3	2	5
Total	53	55	4	10	16	12	6	3	3	6	2	8

From the results of the individual simulations it follows that the E-, PNG- and CNG-methods yield the most accurate estimates of *ESf*, with the N-method the most inaccurate and the CN-method especially inaccurate for 5% *ESf* and the T-method inaccurate for PN innovations.

6 Results of South African Stock Exchange (JSE) data

We applied the different methods described in sections 2 and 3 to a selection of important daily share prices and the $-RAND exchange rate. Most of the shares were from the ALSI40 basket, but a few other well-known shares were also used. Thus shares from the financial, industrial and mining sectors were included in the study. The data were obtained from ShareNet (2002), updated till 19 February 2002, thus including the effect of the terrorist attack on America on 11 September 2001, and the depreciation of the South African Rand in the last quarter of 2001. We used the ShareNet abbreviations of shares on the Johannesburg Stock Exchange (JSE).

Most of these data sets were large and consisted of about 4000 data points, starting from 1985. Other series started when those shares were listed and were therefore shorter. Return series exceeding 1500 data points (SA-BREWS to ANGLO in the following tables) were considered as large and those below 1500 data points (FIRSTRAND to REMGRO) as short. We deliberately analysed long and short return series, taking the closing price as the price for a specific day.

The absolute values of 1% *VaR* and *ESf* are compared for different methods in Tables 6.1 and 6.2 respectively.

Table 6.1 1% *VaR*

Share	Method					
	N	NE	T	E	PNG	CNG
$-RAND	2.28	2.45	2.66	2.44	2.24	2.51
SA-BREWS	3.34	3.83	3.65	3.98	3.77	4.01
NEDCOR	4.22	4.76	5.15	5.08	5.20	5.98
INVESTEC	4.34	5.06	.	5.19	5.28	3.68
ANGLOPLAT	5.36	5.94	6.50	6.35	6.42	6.88
SASOL	4.54	5.17	5.19	5.26	5.30	5.69
ABSA	6.17	7.72	8.65	7.94	8.64	9.81
RMBH	5.94	6.64	6.68	6.47	6.57	5.36
ALSI40	2.72	3.52	2.92	3.37	3.23	3.44
M-CELL	7.34	8.12	8.74	8.39	8.25	9.06
IMPLATS	7.02	7.92	9.88	8.32	9.40	9.91
SBIC	3.53	4.32	3.83	4.35	4.42	3.10
LIBERTY	5.11	5.60	4.19	6.08	4.82	3.72
ANGLO	5.36	6.14	6.15	6.21	6.22	6.45
FIRSTRAND	5.88	6.11	6.77	6.28	5.90	6.36
ILIAD	8.81	12.03	7.79	11.19	9.85	6.60
IOTA	7.28	9.96	7.61	9.42	8.64	6.68
ALEXFBS	4.55	6.49	5.30	5.85	5.74	5.51
REMGRO	4.12	4.95	4.35	4.87	4.29	4.88

. No meaningful empirical quantiles are found with the T-method. In this cases the T-method fits a distribution with almost 2 degrees of freedom, indicating infinite variance. The garch1 (g) parameter of this fit is zero. An ARCH(q) fit on the data also does not yield meaningful quantiles.

Table 6.2 1% *ESf*

Share	Method					
	N	NE	T	E	PNG	CNG
$-RAND	2.61	3.23	3.77	3.20	3.02	3.18
SA-BREWS	3.83	5.51	5.18	5.44	5.19	5.11
NEDCOR	4.84	7.31	7.93	7.07	7.33	7.33
INVESTEC	4.97	7.79	.	7.60	7.42	4.39
ANGLOPLAT	6.14	8.94	9.15	8.74	9.02	8.73
SASOL	5.20	7.35	7.17	7.17	7.23	7.31

ABSA	7.07	11.67	13.28	11.62	12.93	12.22
RMBH	6.80	8.45	10.72	8.51	8.56	6.28
ALSI40	3.11	4.64	3.85	4.73	4.34	4.55
M-CELL	8.41	11.75	13.28	11.54	11.79	11.49
IMPLATS	8.05	12.30	15.79	12.06	14.03	12.15
SBIC	4.04	6.07	6.48	6.00	5.92	3.66
LIBERTY	5.86	8.47	7.27	8.30	6.49	4.37
ANGLO	6.14	8.29	8.27	8.38	8.40	8.00
FIRSTRAND	6.73	8.22	9.07	7.89	8.38	8.10
ILIAD	10.10	14.45	14.06	14.04	11.63	7.95
IOTA	8.34	12.41	12.18	12.45	11.72	7.80
ALEXFBS	5.22	7.68	8.51	7.53	8.33	6.59
REMGRO	4.72	6.97	6.18	6.83	6.18	6.49

. No meaningful empirical quantiles are found with the T-method. In this cases the T-method fits a distribution with almost 2 degrees of freedom, indicating infinite variance. The garch1 (g) parameter of this fit is zero. An ARCH(*q*) fit on the data also does not yield meaningful quantiles.

From Table 6.1 and 6.2 follows that the estimates for the N-method are in most of the cases lower than that of the other methods.

In Table 6.3 the effect sizes in the lower 5% of the distribution are calculated because *VaR* and *ESf* are calculated in this region. The shaded cells indicate the lowest value per share.

Table 6.3 Effect sizes in lower 5% of the distribution

Share	N	Method				
		N	T	E	PNG	CNG
$-RAND	4089	0.666	0.130	0.187	0.093	0.193
SA-BREWS	4194	0.376	0.248	0.248	1.766	0.255
NEDCOR	4195	0.461	0.455	0.187	1.673	0.615
INVESTEC	3267	0.801	2.449	0.088	1.091	0.787
ANGLOPLAT	4194	0.546	0.165	0.203	1.706	0.199
SASOL	4194	0.441	0.220	0.145	1.747	0.250
ABSA	3770	0.705	0.472	0.246	1.916	0.246
RMBH	2276	0.770	0.522	0.216	1.237	0.564
ALSI40	1638	0.746	0.575	0.239	2.069	0.260

M-CELL	1600	0.572	0.226	0.317	1.801	0.194
IMPLATS	4193	0.680	0.204	0.134	1.689	0.431
SBIC	4084	0.639	0.819	0.056	1.359	0.897
LIBERTY	3441	0.484	0.677	0.206	1.381	0.550
ANGLO	4194	0.488	0.199	0.174	1.812	0.253
FIRSTRAND	911	0.551	0.525	0.544	1.730	0.479
ILIAD	899	0.907	1.246	0.203	1.663	1.457
IOTA	934	0.805	0.955	0.429	2.070	1.317
ALEXFBS	1294	0.577	0.531	0.372	1.723	0.398
REMGRO	1407	0.447	0.273	0.273	1.821	0.389

From Table 6.3 follows that in the lower 5% of the distribution the E-method fits better than other methods, while the T- and CNG-methods also have a good fit for some shares. It also follows that the results for short return series are similar to those for long return series.

For return series exceeding 1500 data points a moving window of 1000 data points and for return series below 1500 a moving window of 500 data points is used to calculate *VaR* and *ESf* estimates for backtesting of the data. Backtesting is performed on a selection of long and short return series of different autoregressive structures. For long return series backtesting is performed on the last 1000 days only. The shaded cells indicate *p*-values smaller than 0.05.

The results for the backtesting of *VaR* for the different methods are given in Table 6.4.

Table 6.4 1% *VaR* violation *p*-values

Share	n	Method					
		N	NE	T	E	PNG	CNG
$-RAND	4089	0.011	0.130	0.075	0.011	0.000	0.011
SA-BREWS	4194	0.000	0.842	0.988	0.720	0.583	0.720
NEDCOR	4195	0.000	0.540	0.362	0.761	0.540	0.001
INVESTEC	3267	0.000	0.000	0.000	0.004	0.001	0.000
ANGLOPLAT	4194	0.006	0.666	0.996	0.826	0.666	0.000
SASOL	4194	0.000	0.041	0.002	0.021	0.041	0.041
ABSA	3770	0.000	0.042	0.132	0.132	0.508	0.076
RMBH	2276	0.003	0.226	0.000	0.079	0.044	0.000
ALSI40	1638	0.117	0.477	0.117	0.777	0.903	0.477

M-CELL	1600	0.038	0.303	0.303	0.303	0.303	0.124
IMPLATS	4193	0.255	0.735	0.507	0.735	0.507	0.035
SBIC	4084	0.000	0.545	0.010	0.545	0.231	0.000
LIBERTY	3441	0.000	0.020	0.205	0.038	0.038	0.205
ANGLO	4194	0.000	0.174	0.032	0.104	0.016	0.032
FIRSTRAND	911	0.778	0.778	0.447	0.830	0.830	0.830
ILIAD	899	0.040	0.877	0.005	0.419	0.877	0.040
IOTA	934	0.000	0.995	0.002	0.656	0.384	0.000
ALEXFBS	1294	0.006	0.209	0.209	0.117	0.349	0.349
REMGRO	1407	0.601	0.894	0.841	0.631	0.601	0.601

From Table 6.4 follows that for 1% *VaR* the NE-, E- and PNG-methods yield better estimates than other methods, while the N- and CNG-methods are inaccurate.

The results for the backtesting of *ESf* for the different methods are given in Table 6.5.

Table 6.5 1% *ESf* violation *p*-values

Share	n	Method					
		N	NE	T	E	PNG	CNG
$-RAND	4089	0.000	0.029	0.495	0.055	0.007	0.209
SA-BREWS	4194	0.000	0.629	0.449	0.142	0.155	0.060
NEDCOR	4195	0.001	0.198	0.999	0.139	0.283	0.008
INVESTEC	3267	0.001	0.961	0.996	0.756	0.798	0.010
ANGLOPLAT	4194	0.000	0.302	0.229	0.163	0.186	0.000
SASOL	4194	0.000	0.452	0.716	0.411	0.469	0.085
ABSA	3770	0.000	0.223	0.515	0.063	0.039	0.011
RMBH	2276	0.004	0.452	0.000	0.440	0.807	0.000
ALSI40	1638	0.002	0.987	0.955	0.989	0.995	0.991
M-CELL	1600	0.596	0.940	0.923	0.925	0.894	0.492
IMPLATS	4193	0.070	0.661	0.616	0.640	0.629	0.137
SBIC	4084	0.000	0.678	0.997	0.556	0.758	0.000
LIBERTY	3441	0.000	0.124	0.998	0.053	0.020	0.053
ANGLO	4194	0.000	0.208	0.596	0.165	0.258	0.037
FIRSTRAND	911	0.102	0.243	0.000	0.092	0.138	0.138
ILIAD	899	0.023	0.877	0.662	0.214	0.222	0.056
IOTA	934	0.012	0.995	0.618	0.289	0.131	0.011

ALEXFBS	1294	0.000	0.209	0.649	0.233	0.601	0.089
REMGRO	1407	0.006	0.894	0.151	0.118	0.184	0.135

From Table 6.5 follows that all the methods except the N- and CNG-methods seem to yield accurate estimates of 1% *ESf*. It also follows that the results for short return series are not different from those for long return series.

7 Conclusions

When recommendations are made about best methods of fit, we can conclude that the E-, PNG- and CNG-methods would all yield accurate estimates for *VaR* and *ESf* in most of the cases (for simulated data as well as JSE data). These methods are, however, difficult to implement in practice because they require extra programming to estimate *VaR* and *ESf* The NE-method also has a good fit for all the *VaR* or *ESf* estimates for all types of data and is easy to implement in practice.

References

ARTZNER, P., DELBAEN, F., EBER, J. & HEATH, D. 1999. Coherent measures of risk. *Mathematical Finance*, 9(3): 203-228.

BOLLERSLEV, T. 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of Exonometrics*, 31: 307-327.

COHEN, J. 1988. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NY : Lawrence Erlbaum Associates.

EFRON, B. & TIBSHIRANI, R. 1993. An introduction to the Bootstrap. New York : Chapman and Hall.

GOURIÉROUX, C. 1997. ARCH-models and financial applications. New York : Springer.

McNEIL, A. & 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.

SAS INSTITUTE INC. 1999. SAS/ETS® User's Guide. Version 8. Cary, NC : SAS Institute Inc.

SHARENET 2002. ShareNet - your key to investing on the Johannesburg stock exchange in South Africa. [Web:] http://www.sharenet.co.za [Date of access: 20 Feb. 2002].