Testing EVT-based VaR measures

Mandira Sarma *

November 18, 2001

*Address: Indira Gandhi Institute of Development Research, Goregaon (East), Mumbai 400 065. Phone: +91-22-840-0919. Fax: +91-22-840-2752. Email: mandira@igidr.ac.in

Abstract

Extreme value theory (EVT) provides robust tools for estimating risk of financial portfolios. While the conventional VaR models suffer shortcomings due to distributional assumptions, the extreme value theory based "Peaks-over-threshold" model offers a framework for VaR estimation with miminal distributional assumptions. This paper undertakes a case study of estimating Value-at-Risk (VaR) forecasts using the POT approach based on extreme value theory. The POT framework is applied for estimating the 95% and 99% VaR measures for a short and a long position in the Nifty portfolio. The POT model is then compared with the commonly used historical simulation and the normal distribution models. The results show that the POT model yields statistically valid VaR neasures, and in some cases outperforms the other two competing models in terms of a loss function approach.

KEY WORDS Extreme value theory; Value-at-Risk; pseudo maximum likelihood estimation; correct conditional coverage; loss function approach

JEL Classification: C22, C52, G10, G28

1 Introduction

Extreme value theory (EVT) is a well developed branch of probability theory. It deals with the study of the asymptotic behaviour of extreme (maxima and minima) observations. EVT has been in wide application in the areas of meteorology, hydrology, environment science and other branches of science and engineering, and very recently it has found its place in finance and risk management literature.

Value-at-Risk (VaR), a measure of market risk, is defined as the maximum monetary loss of a portfolio due to market fluctuation, at a pre-specified confidence level over a pre specified risk horizon. Statistically VaR is a quantile on the tail region of a portfolio's return distribution corresponding to the pre-specified confidence level.

The conventional models of VaR estimation, viz., the normal distribution model (also known as the variance-covariance model) and the historical simulation, suffer certain shortcomings. The normal distribution models are based on the assumption of normality of the return distribution which is often found to be incorrect. Moreover the normal distribution assumes the symmetry of the return distribution while in reality financial returns often exhibit asymmetric behaviour. In the historical simulation approach, the observed return distribution is assumed to be drawn from an i.i.d. time series of unknown distribution, which is generally a wrong assumption.

The more recent models, based on extreme value theory, provides a theoretically robust framework for VaR estimation. The extreme value theory exclusively deals with the behaviour of the tails of probability distributions, and hence, can provide a better treatment to the estimation of tail quantiles like VaR. Using EVT one can focus only on the tails of distribution rather than estimating the entire distribution to make inferences about the tails. Another appealing aspect of EVT is that it does not require to make *a priori* assumption about the return distribution. The fundamental result of extreme value theory, known as the "extremal types theorem", identifies the possible classes of distributions for extreme movements, whatever be the actual underlying return distribution. This extremely powerful result of the extreme value theory makes the VaR estimation process free from any *a priori* assumption about the portfolio return distribution. Moreover, EVT based methods inherently incorporates separate estimation of VaR for long and short positions, thus emphasizing the existence of asymmetric tail behaviour of return distributions.

Although EVT primarily deals with i.i.d. random variables, recent works of McNeil and Frey (1999) has developed a procedure for applying EVT to time series processes.

This paper attempts to understand one of the recent state-of-the art techniques of VaR estimation based on EVT. It provides a case study in which VaR forecasts are estimated using the "Peaks-Over-Threshold" (POT) model McNeil and Frey (1999) based on EVT. The EVT based model is then compared with the two most commonly used approaches, viz. the historical simulation (HS) and the normal distribution model. The evaluation of these three approaches are carried out in two stages. The first stage of the evaluation consists in testing these models for statistical precision in terms of "correct conditional coverage"Christoffersen (1998), a necessary statistical property for a well defined VaR model. The second stage of the evaluation process is based on a "loss function approach" Lopez (1998). This approach consists in defining a loss function which reflect the risk management problem. The models which produces smaller losses are assumed to be better.

The case study consist of estimation of 95% and 99% VaR measures for two portfolios–a long and a short position in the Nifty portfolio.

We find that for both these portfolios, all the three approaches seem to be producing the "correct conditional coverage", for both 95% and 99% VaR estimation. For the long Nifty portfolio, for 99% VaR estimation, the EVT-based POT is found to be the best in terms of the loss function approach, followed by historical simulation and then the normal distribution model. However, for 95% VaR estimation for the same portfolio, the loss function approach ranks historical simulation as the best, followed by EVT-based POT model and then the Normal distribution model. For the short Nifty portfolio, the normal distribution model is found to be best in terms of the loss function approach, both for 99% and 95% VaR estimation.

The rest of this paper is organised in this manner. Section 2 presents an overview of extreme value theory and its application in VaR literature. Section 3 describes the POT method in details, along with the two alternative approaches that are carried out in here. Section 5 describes the methodology for comparing the VaR models in terms of statistical precision and the loss function approach. The empirical results of the case studies are presented in Section 4. Section 6 concludes the paper.

2 An overview of Extreme Value Theory

The classical Extreme Value theory (EVT) deals with the study of the asymptotic behaviour of extreme observations (maxima or minima of n random realisations).

Suppose that $X \in (l, u)$ is a random variable with density f and cdf F. Let $X_1, X_2, ...X_n$ be n independent realisations of the random variable X. Define the extreme observations as

$$Y_n = max\{X_1, X_2, ..., X_n\}$$
$$Z_n = min\{X_1, X_2, ..., X_n\}$$

The extreme value theory deals with the distributional properties of Y_n and Z_n as n becomes large.

It can be easily shown that the exact distributions of the extreme observation is degenerate in the limit. In order to find a distribution of interest which is non-degenerate, the extrema Y_n and Z_n are transformed with a scale parameter $a_n (> 0)$ and a location parameter $b_n \in R$, such that the distribution of the standardised extrema

$$\frac{Y_n - a_n}{b_n}$$
$$\text{and} \quad \frac{Z_n - a_n}{b_n}$$

is non-degenerate.

The two extremes, the maximum and the minimum are related by the following relation:

$$min\{X_1, X_2, ..., X_n\} = -max\{-X_1, -X_2, ..., -X_n\}$$

Therefore, all the results for the distribution of maxima leads to an analogous result for the distribution of minima and vice versa. We will discuss the results for maxima only and ignore the same for the minima[1].

2.1 The Fisher-Tippett Theorem

The Fisher-Tippett theorem (1928) is a fundamental result in EVT. The importance of this result is that it exhibits the possible limiting forms for the distribution of Y_n under linear transformations even without the exact knowledge of the underlying distribution F. The "Fisher-Tippett Theorem", also known as the "Extremal type theorem" states thus:

If $\exists$ *constants* $a_n(>0)$ *and* $b_n \in R$ *such that*

$$\frac{Y_n - a_n}{b_n} \to^d \quad H \quad \text{as} \quad n \to \infty$$

for some non-degenerate distribution H*, then* H *must be one of the only three possible 'extreme value distributions'.*

In that case, X (and the underlying distribution F) is said to belong to the (maximum) domain of attraction of the extreme value distribution H. It is denoted by $X \in DA(H)$.

More specifically, this basic result states that if there exist some suitable normalizing constants $a_n(>0)$ and b_n, the transformed maxima $\frac{Y_n - a_n}{b_n}$ has a non-degenerate limiting distribution function $H(x)$, then H must have one of only three possible "forms". The limit laws for maxima were derived by Fisher and Tippett (1928). A first rigorous proof is due to Gnedenko (1943). De Haan (1970) subsequently provided a simpler proof and Weissman (1977) provided a simpler version of de Haan's proof.

The three possible probability laws for suitably normalised extrema are: the Gumbel (or Type I) distribution, the Fréchet (or Type II) distribution and the Weibull (or Type III) distribution[2]. The Gumbel distribution is a limit law for the thin-tailed distributions such as the normal or log-normal distributions. The Fréchet distribution is obtained as a limiting distribution for the fat-tailed distributions such as Student's-t or the Stable Paretian distributions. The marginal distribution of a stationary GARCH process is also in the domain of attraction of the Fréchet family. Finally, the Weibull distribution is obtained when the distribution of returns has no tail.

2.2 The Generalized Extreme Value Distribution

The three families of extreme value distributions, viz. the Gumbel, the Fréchet and the Weibull, can be nested into a single parametric representation, as shown by Jenkinson and Von

[1] A brief description about the minima can be found in Leadbetter et al. (1983)

[2] Details about these distributions can be found in Leadbetter et al. (1983) and Embrechts et al. (1997)

Mises. This representation is known as the "Generalised Extreme Value" (GEV) distribution, and given by

$$H_{\xi}(x) = exp\{-(1+\xi x)^{-\frac{1}{\xi}} \tag{1}$$

where

$$1 + \xi x > 0$$

The support of ξ is

$$\begin{aligned} x &> -\frac{1}{\xi} \text{ if } \xi > 0 \\ x &< \frac{1}{\xi} \text{ if } \xi < 0 \\ x &\in R \text{ if } \xi = 0 \end{aligned}$$

The parameter ξ, called the tail index, models the distribution tails. Each of the three extreme value distributions can be obtained as a special case of the GEV distribution. When $\xi > 0$, we get the Fréchet distribution, when $\xi < 0$ we get the Weibull distribution and $\xi = 0$ is the case of the Gumbel distribution.

These theoretical results show the generality of the extremal type theorem – essentially all the common, continuous distributions of statistics belong to the domain of attraction of a single family H_{ξ}, the extreme value distributions being differentiated only by the value of ξ.

2.3 The Pickands-Balkema-de Haan Theorem

Suppose that $X_1, X_2, ...X_n$ are n independent realisations of a random variable X with a distribution function $F(x)$. Let u be the finite or infinite right endpoint of the distribution F. The distribution function of the excesses over certain (high) threshold k is given by

$$F_k(x) = \Pr\{X - k \leq x | X > k\} = \frac{F(x+k) - F(k)}{1 - F(k)}$$

for $0 \leq x < u - k$.

The Pickands-Balkema-de Haan theorem (Balkema & de Haan 1974; Pickands 1975) states that if the distribution function $F \in DA(H_{\xi})$ then $\exists$ a positive measurable function $\sigma(k)$ such that

$$\lim_{k \to u} \sup_{0 \leq x < u-k} |F_k(x) - G_{\xi,\sigma(k)}(x)| = 0$$

and vice versa, where $G_{\xi,\sigma(k)}(x)$ denote the Generalised Pareto distribution.

The above theorem states that as the threshold k becomes large, the distribution of the excesses over the threshold tends to the Generalised Pareto distribution, provided the underlying distribution F belongs to the domain of attraction of the Generalised Extreme Value distribution.

2.4 The Generalised Pareto Distribution (GPD)

The GPD is given by

$$G_{\xi,\sigma}(x) = \begin{cases} 1-(1+\xi x/\sigma)^{-1/\xi}\ ; & if \xi \neq 0 \\ 1-exp(-x/\sigma); & if \xi = 0 \end{cases} \tag{2}$$

where $\sigma > 0$, and the support of x is $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\sigma/\xi$ when $\xi < 0$.

3 The Peaks-over-Threshold (POT) method of VaR estimation

The POT model provides for a framework of estimating the tails (positive or negative tails) of the return distribution by estimating what is known as the distribution of *excesses* over certain threshold point which identifies the starting of the tail.

The distribution of *excesses* over a high threshold k on the portfolio's loss distribution F is defined by

$$\Phi_k(y) = Pr\{X - k \leq y | X > k\}$$

In terms of the underlying loss distribution F,

$$\Phi_k(y) = \frac{F(y+k) - F(k)}{1 - F(k)} \tag{3}$$

Pickands-Balkema-de Haan theorem says

$$\Phi_k(y) \rightarrow G_{\xi,\beta(k)}(y) \tag{4}$$

for

$$k \rightarrow u$$

Thus, using the Pickands-Balkema-de Haan theorem, one can model the distribution of the excesses over the threshold k as a GPD, provided the threshold is sufficiently high.

Setting $x = k + y$ and using (3) and (4), we can rewrite F as

$$F(x) = (1 - F(k))G_{\xi,\beta}(x-k) + F(k) \tag{5}$$

for $x > k$.

Using HS estimate for $F(k)$ and ML estimates of the GPD parameters gives rise to the following tail estimator formula

$$\hat{F}(x) = 1 - \frac{N_k}{N}\left(1 + \hat{\xi}\frac{x-k}{\hat{\beta}}\right)^{(-\frac{1}{\hat{\xi}})} \tag{6}$$

3.1 Estimating Value-at-Risk

VaR is a measure of extreme risk in terms of the unknown loss distribution $F(x)$ of the portfolio under consideration. VaR is the p^{th} quantile of the distribution F, (where p is very high and pre specified), given by

$$VaR_p = F^{-1}(p)$$

For a given probability $p > F(k)$, the VaR is estimated by inverting the tail estimator formula (6),

$$\hat{V}aR_p = k + \frac{\hat{\beta}}{\hat{\xi}} \left(\frac{N}{N_k}(1-p)^{-\hat{\xi}} - 1 \right) \tag{7}$$

3.2 Handling time dependencies of the return series

EVT is a theory of i.i.d. observations, and all the powerful results of the theory fails when the underlying distribution has a time series structure. The market returns are time series processes whose means and variances are characterized by time dependencies.

In order to apply the results of EVT, we have to remove the time series structure from the return distribution, construct an i.i.d. series and apply the POT method to estimate the tails of the i.i.d. series. These estimates can be translated back into the original return series, given an estimate of the time dependent mean and volatility. This idea has been implemented in McNeil and Frey (1999) in their two stage VaR estimation approach,as described below.

3.2.1 A two stage approach for VaR estimation

Let $\{X_t\}$ is a strictly stationary time series whose dynamics are given by

$$X_t = \mu_t + \sigma_t Z_t \tag{8}$$

where μ_t is the mean process and σ_t the volatility dynamics of X_t, and,

$$Z_t \sim f_Z(z)$$

where $f_Z(z)$ is white noise.

The p^{th} quantile of the distribution of X_t at time t can be obtained by using that of Z_t, as,

$$x_p^t = \mu_t + \sigma_t z_p \tag{9}$$

where z_p is the p^{th} quantile on the distribution of Z_t, which, by assumption, is *iid*.

McNeil and Frey (1999) proposes the following approach to estimate VaR for financial returns

1. Fit a time series model to the return series using a pseudo-maximum likelihood (PML) estimator using normality for $f_Z(z)$. Estimate μ_t and σ_t from the fitted model and extract the residuals Z_t.

The use of the PML approach to estimate the parameters of the time series by using the normal distribution for $f_Z(z)$ does not imply the assumption of normality of $f_Z(z)$. Under standard regularity conditions (Gourieroux, 1997; Gourieroux et al., 1984) the use of normal distribution would yield consistent estimation even if the underlying distribution is not conditionally normal. That is, the consistency of the PML estimator does not depend on the distribution which is used to build the likelihood function (in this case, the normal distribution). Moreover, this estimator is asymptotically normal[3].

2. If the residual series Z_t is found to be strictly white noise, the EVT can be applied to model the tail of the white noise $f_Z(z)$. The EVT based VaR formula (7) can be used to estimate VaR for the Z_t series, say VaR_Z^t.

   Given the estimate of μ_t, σ_t and VaR_Z^t, the VaR for the return series can be estimated as

$$VaR_X^t = \hat{\mu}_t + \hat{\sigma}_t VaR_Z^t \tag{10}$$

3.3 The alternative approaches

The two commonly used alternative approaches to the GPD approximation of the tail are the historical simulation and the normal distribution model. They are described below

- Historical simulation: In HS, the VaR_Z^t of the *iid* white noise is estimated by the historically observed quantiles which are used in the VaR formula (10) instead of the GPD-quantiles.
- Normal distribution model: In the normal distribution model, the standard normal quantiles are used as estimates of VaR_Z^t in formula (10) instead of the GPD-quantiles.

4 Empirical example

In this section, we present an example of the application of the POT method for VaR estimation for the following portfolios:

- A long position of Rs.100 in the Nifty portfolio.
- A short position of Rs.100 in the Nifty portfolio.

For each of these portfolios we estimate 95% and 99% daily VaR.

The subsection 4.1 describes the estimation procedure in details and the subsection 5.1 describes the testing of the VaR measures. In the subsection 5.3 we compare the performance

[3]The PML estimators are obtained by maximising the likelihood function associated with a family of probability distributions which do not necessarily contain the true pdf of the underlying random variable. Gourieroux et al. (1984) obtains families of pseudo maximum likelihood functions providing consistent and asymptotically normal estimators of the parameters involved in the true distribution. In particular, the normal distribution can generate consistent and asymptotically normally distributed estimators of the true distribution, whatever may be the true underlying distribution.

of EVT based VaR measures with historical simulation and the normal distribution models, as described in subsection 3.3.

4.1 Estimation procedure

4.1.1 Data

The data consists of 2562 daily logarithmic returns of the Nifty portfolio (from 3 July 1990 till 13 August 2001). The first 1,250 observations (from 3 July 1990 till 7 May 1996) comprise the estimation window and the rest of 1312 observations (from 8 May 1996 till 13 August 2001) are used for making rolling window "out–of–sample" VaR forecasts (1313 daily VaR forecasts).

4.1.2 Time series model for the Nifty return series

At first we try to ascertain the time series structure of the Nifty return series. A specification search in terms of AIC and SBC criteria leads us to choose the AR(1)-GARCH(1,1) model as the best model[4].

Table 1 presents the estimated parameters of the mean and volatility equations of the Nifty returns. The constant term in the mean equation is found to be insignificant although the AR(1) coefficient is significant.

The parameters in the volatility equations, viz., the constant, the ARCH(1) parameter and the GARCH(1) parameter, are all found to be significant.

Having obtained a proper time series model for the data, we fit this model to the first 1250 observations, which consist the estimation window. We extract the standard residuals from the estimated model and investigate if it can be considered i.i.d. Table 2 presents the values of the test statistics for skewness, kurtosis and autocorrelation (upto lag 35) along with their respective p-values for the original return series and the residual series. The results of Table 2 indicates that the return series has significant skewness, kurtosis and autocorrelation. The residual series, has a significant kurtosis, and insignificant skewness and autocorrelation. Thus, neither the return series nor the residual series can be considered to be normally distributed, since both the series have significant excess kurtosis. However, the residual series is found to be free from autocorrelation and hence we can apply the results of EVT to the residual series.

4.1.3 Modeling peaks-over-thresholds

The Pickands-Balkema-de Haan theorem offers the Generalised Pareto distribution as a natural choice for the distribution of excesses (peaks) over sufficiently high thresholds. However, while choosing an appropriate threshold, one faces a trade off between bias and variance. The

[4]Values of the AIC and SBC criteria for various specifications of the time series can be obtained from the author on request.

theoritical consideration suggests that the threshold should be high enough for the Pickands-Balkema-de Haan theorem to hold good, but in practice, too high a threshold might leave us with very few observations above the threshold for estimating the GPD parameters[5]. The GPD estimators are unbiased if and only if $k \to u$, i.e. if the threshold is sufficiently high. However, if it is chosen very high, there may be very few observations left for estimating the GPD parameters to the tail, leading to statistical imprecision and very high variance of the estimates.

There is no correct choice of the threshold level. While McNeil and Frey (1999), McNeil (1996) and McNeil (1999) use the "mean-excess-plot" as a tool for choosing the optimal threshold level[6], Gavin (2000) uses an arbitrary threshold level of 90% confidence level (i.e. the largest 10% of the positive and negative returns are considered as the extreme observations).

In this paper we follow Neftci (2000) and choose the threshold level as 1.65 times the unconditional variance of the residuals[7]. This represents the 5% of extreme movements if the data were normally distributed. On the both sides of the tails, observations lying beyond 1.65 times the unconditional variance are considered to be extremes[8].

Table 3 presents the estimated threshold point, the number of extreme observations beyond the threshold, and the results of the maximum likelihood estimation of the GPD to the excesses (peaks) over the chosen threshold for the lower tail and the upper tail respectively of the i.i.d. residual series extracted from fitting an AR(1)-GARCH(1,1) model to the Nifty returns. These estimated parameters are used for the VaR estimation. The first column of this table gives the estimated threshold points corresponding to the chosen level of threshold. For the lower tail the threshold point is -1.6493 and for the upper tail it is 1.6494 implying that both the tails seem to be symmetrical. This is also established from the test of skewness in table 2.

The second column of table 3 gives the number of extreme observations beyond the thresholds on both the tails, the third column gives the estimated cdf at the thresholds, the fourth and the fifth columns presents the ML estimation of the GPD parameters (with SEs in parenthesis) fitted to the peaks over the thresholds.

Table 4 presents the estimated 5^{th} and 1^{st} percentiles (for 5% and 1% VaR for the long position) and 95^{th} and 99^{th} percentiles (for 95% and 99% VaR for the short position). These estimates are done by using three alternative models, viz. the GPD approximation where the ML estimations of the GPD parameters fitted to the excesses over the threshold are used (as described in equation (7)), the historical simulation approach where empirically observed quantiles are used as the estimates of the corresponding percentiles, and the normal distribution model where the percentile points on the standard normal curve are used to estimate the quantiles.

These quantiles can be considered as the unconditional VaR measures for the i.i.d. residual series.

[5]For more on this issue, see McNeil and Frey (1999).

[6]Details on mean-excess-plots can be found in McNeil and Frey (1999) and Embrechts et al. (1997).

[7]We tried with the mean excess plots but did not get a well behaved linear mean excess plot.

[8]For analysis of the lower tail, i.e. the minima, we use the negative returns and then apply results for maxima.

4.1.4 Conditional VaR forecasts

In order to make conditional (or dynamic) VaR estimates for the return distribution under consideration, we dynamically estimate $\hat{\mu}_t$ and $\hat{\sigma}_t$ forecasts for the "out of sample" periods, by using a rolling window of size 1250. These daily forecasts and the VaR for the residual series are used in the formula (10) to make one day ahead VaR forecasts for the underlying data.

Figure 1 depicts the forecasted VaR vis-a-vis the actual returns observed over the forecast period for the three different approaches. The graph at the top of this figure depict the VaR forecasts estimated through the POT model, the middle graph is for the historical simulation model and the one at the bottom gives the VaR plot obtained from the normal distribution model. In these figures, the solid line represents the forecasted 95% VaR, the dotted line represents the forecasted 99% VaR, and the dashed line shows the actually observed profit and losses. Each of these graph has two sets of VaR forecasts, one for the long position and the other for the short position. The graphs on the negative side of the returns is for the long position and the one on the positive side is for the short position.

These graphs indicate that all the three models are quite similar in forecasting the VaR measures, and they are able to capture the volatility dynamics of the return process.

5 Evaluation of POT model, Historical simulation and the Normal model

In this section we describe in details the methodology adopted for comparing the EVT based POT model with HS and normal distribution model.

We adopt a two stage framework for the relative evaluation of these three approaches. In the first stage, the alternative models are tested for statistical precision. The tests consists of testing whether the various models display the property of "correct conditional coverage", a necessary condition for statistical precision. The notion of "conditional coverage" and the test for the existence of it are described in details in the subsection 5.1.

The second stage of the performance evaluation is carried out in terms of the ability of the alternative VaR models to minimize a loss function. This approach is described in the subsection 5.2.

The results of application of this two stage performance evaluation to the three approaches are presented in the subsection 5.3.

5.1 Testing for statistical validity

5.1.1 Notion of correct conditional coverage

The notion of "conditional coverage" was formalised by Christoffersen (1998). A correctly specified VaR model should generate the pre specified failure rate *conditionally* at every point in time. This is known as the property of "conditional coverage" of the VaR model. The basic feature of a 99% VaR is that it should be exceeded 1% of the time, and that the probability of the VaR being exceeded at time $t+1$ remains 1% even after conditioning on all information known at time t. This implies that the VaR should be small in times of low volatility and high in times of high volatility, so that the events where the loss exceeds the forecasted VaR measure are spread over the entire sample period, and do not come in clusters. A model which fails to capture the volatility dynamics of the underlying return distribution will exhibit the symptom of clustering of failures, even if (on the average) it may produce the correct unconditional coverage.

Consider a sequence of one period ahead VaR forecasts $\{v_{t|t-1}\}_{t=1}^{T}$, estimated at a significance level $1-p$. These forecasts are intended to be one–sided interval forecasts $(-\infty, v_{t|t-1}]$ with coverage probability p. Given the realisations of the return series r_t and the *ex-ante* VaR forecasts, the following indicator variable may be defined

$$I_t = \begin{cases} 1 & \text{if } \; r_t < v_t \text{ for long position, and } \; r_t > v_t \text{ for a short position} \\ 0 & \text{otherwise} \end{cases}$$

where r_t is observed return and v_t is forecasted VaR measure on day t.

The stochastic process $\{I_t\}$ is called the "failure process". The VaR forecasts are said to be conditionally efficient if they display "correct conditional coverage", i.e., if $E[I_{t|t-1}] = p \; \forall \; t$. This is equivalent to saying that the $\{I_t\}$ series is *iid* with mean p.

Two recent papers, Christoffersen and Diebold (2000) and Clements and Taylor (2000), suggest that a regression of the I_t series on its own lagged values and some other variables of interest, such as day-dummies or the lagged observed returns, can be used to test for the existence of various form of dependence structures that may be present in the $\{I_t\}$ series. Under this framework, conditional efficiency of the I_t process can be tested by testing the joint hypothesis:

$$H : \Phi = 0, \alpha_0 = p \tag{11}$$

where

$$\Phi = [\alpha_1, ...\alpha_S, \mu_1, ..., \mu_S]'$$

in the regression

$$\begin{aligned} I_t &= \alpha_0 + \sum_{s=1}^{S} \alpha_s I_{t-s} + \sum_{s=1}^{S-1} \mu_s D_{s,t} + \epsilon_t & (12) \\ t &= S+1, S+2, ..., T & (13) \\ D_{s,t} & \quad \text{are explanatory variables.} & (14) \end{aligned}$$

The hypothesis (11) can be tested by using an F-statistic in the usual OLS framework.[9]

[9]In view of the fact that the I_t series is binary, a more appropriate way is to do a binary regression rather than an OLS regression. However, there seem to be a technical problem in the implementation of the binary regression as more than 90% of the I_t's are zero and only a few are unity. This asymmetry in the data results in singular Hessian matrices in the estimation process and the maximum likelihood estimation fails as

5.1.2 Testing for correct conditional coverage

To test for the existence of the property of correct conditional coverage, we perform an OLS regression of the I_t series on its five lagged values and five day–dummies representing the trading days in a week. Significance of the F-statistic of this OLS will lead to rejection of a model; otherwise it will lead to its non-rejection. It should be noted that the non-significance of the F-statistic does not necessarily imply non-significance of the t-statistics corresponding to the individual regressors in the OLS. We follow Hayashi (2000) (page 44) and adopt the policy of preferring the F-statistic over the t-statistic(s) in the case of a conflict. Therefore a model will not be rejected if the F-statistic is not significant even though some individual t-statistic(s) may turn out to be significant.

5.2 A Loss function approach

The idea of using loss functions to compare alternative models was due to Lopez (1999, 1998). This approach can be applied to compare different models, which are statistically equivalent but having different ability to minimize losses due to risk management. In this approach, a loss function, appropriate to reflect the risk management problem, is defined. The loss functions are defined with a negative orientation, i.e., low values of the loss functions are preferred because it indicates that the losses due to risk management are low.

The loss function that we consider here is as follows:

$$l_t = \begin{cases} (r_t - v_t)^2 & \text{if } r_t < v_t \text{ for long position, and } r_t > v_t \text{ for a short position} \\ |r_t - v_t| & \text{otherwise} \end{cases}$$

This loss function assigns a quadratic score when there is a failure, and an absolute score when there is no failure. These two components can be interpreted as follows:

- The first component reflects the penalty due to failure of a model. By giving a quadratic penalty, higher magnitude of failures are penalised more than the lower magnitude of failures.
- The second component signifies the "opportunity cost" for the risk manager due to risk management. Risk management always brings about certain opportunity cost to the risk manager. A financial regulator uses VaR measures to set margins and risk capitals for the financial firms. While setting up capital adequacy norms based on VaR measures, the regulator has to be cautious not to set the norms very high in order to retain the attractiveness of the financial market. Therefore, very high VaR measures yields certain cost on the financial regulator.

  From a financial firm's point of view, very high VaR measures has a direct "opportunity cost" of setting aside a high amount of capital to meet the capital adequacy norms.

  Thus, high VaR forecasts on the days of "non-failure" leads to certain disutility for both the regulator and the firm. This disutility has been incorporated in the second term

a result. This problem seems to be more severe in the case of 99% VaR models. Therefore we resort to an OLS regression, which is asymptotically equivalent to a binary regression.

of the loss function defined in (5.2). The absolute difference between the forecasted VaR and the actually observed return can be interpreted as a metric of deviation from reality.

This loss function tries to capture the deviation of the actually observed profit or loss from the forecasted VaR measures. The first component captures the deviation during the failure and the second component incorporates the same during non-failure.

It needs to be mentioned that the definition of a loss function is arbitrary and prone to subjectivity. Use of different loss functions may prove different models to be superior. Hence, the superiority of a model with respect to certain loss function should be viewed with caution.

The superiority of a model with respect to a particular loss function may be tested non-parametrically by using a sign test of the loss functions generated by different models [10]. This is described below in short.

5.2.1 Testing for superiority of a model in terms of the loss function

This test can be applied pairwise. Consider a loss function L and let l_{it} and l_{jt} be the values of the loss function generated by model i and model j respectively on the day t. Given the pairwise values $\{l_{it}, l_{ji}\}_{t=1}^{T}$, the sign test for testing the superiority of model i over model j consists in the following:

- First calculate the loss differentials $z_t = l_{it} - l_{jt}$ between model i and model j for each time period t.
- Define an indicator variable as follows:

$$\psi_t = \begin{cases} 1 & \text{if} \quad z_t \geq 0 \\ 0 & \text{if} \quad z_t < 0 \end{cases}$$

- The sign statistic $S_{i,j}$ is the number of non-negative z's:

$$S_{ij} = \sum_{t=1}^{T} \psi_t$$

- The standardised sign statistic

$$S_{ij}^{a} = \frac{S_2 - 0.5T}{\sqrt{.25T}} \quad \sim \quad N(0,1) \text{ asymptotically.}$$

- The hypothesis of the equivalence of model i and model j against the alternative hypothesis of superiority of model i over model j is rejected if $S_{ij} < -1.66$. This rejection would imply that the model i is significantly better than model j in terms of the particular loss function under consideration; otherwise model i is not significantly better than model j.

[10]For a detailed discussion of the testing of the loss function values by using the sign test see Sarma et al. (2001)

5.3 Results

5.3.1 Test for correct conditional coverage

Table 5 presents the results of the F- test of the hypothesis (11) applied to the VaR forecasts for a long position for the three alternative approaches.

The upper panel of the table pertains to the 95% VaR estimation and the lower panel pertains to 99% VaR estimation. The first column gives the names of the models, the second column the estimated unconditional failure probabilities (with p-values in parenthesis) and the third column presents the estimated value of the F-statistic, along with the respective p-values in parenthesis.

The estimated failure probabilities, both for the 95% and 99% VaR estimation are not significantly different from the respective pre-specified failure rates, viz. 0.05 and 0.01, for each of the three models. Also, the F-statistic of the test of "correct conditional coverage" has been found to be insignificant for all the models. Thus, in terms of statistical precision, all the three approaches are generating valid VaR measures for the long position.

Table 6 presents the results of the testing of the hypothesis (11) to VaR measures estimated from the three approaches for a short position in the Nifty portfolio. For both 95% and 99% VaR estimation, for all the three models considered, the VaR measures indicate existence of "correct conditional coverage".

In summary, all the three models are producing statistically valid VaR measures (both at 95% and 99% levels) for both the long and the short portfolio.

Tables 7 and 8 provides the pairwise test statistics for the test of superiority of a model i over another model j with respect to the loss function defined in (5.2).

For the 95% VaR estimation for the long Nifty, the historical simulation is found to be best, followed by the POT model and then the normal model. However, for the 99% VaR estimation of the long Nifty, the POT model is found to be best in terms of the loss function. (Table 7).

As shown by Table 8, for the short Nifty position, the POT approach occupies the third rank for 95% VaR estimation and second rank for 99% VaR estimation. For both 95% and 99% VaR estimation, the normal distribution model occupies the first rank.

6 Conclusion

In this paper, we carry out a case study of VaR estimation by the recently developed POT approach of VaR estimation. This approach is based on the well established results of the 'Extreme Value Theory'. The approach provides an interesting framework for risk measurement, which does not require to make any a priori assumption about the return distributions. The case study provides strong evidence of the statistical validity of the model. The paper also carries out a comparison of the POT model with the commonly used historical simulation and normal distribution. For this particular case study, POT has found to outperform both the other two approaches once (99% VaR estimation for the long Nifty). For 95% VaR estimation for the long Nifty, the POT outperforms the normal distribution model, but not the historical simulation. Also, the POT outperforms the historical simulation approach for 99% VaR estimation for the short Nifty, but does not outperform the normal model in this case. For the 95% VaR estimation for the short nifty, POT is found to be performing the worst.

References

Christoffersen PF, 1998. Evaluating Interval Forecasts. *International Economic Review* **39**:841–862.

Christoffersen PF, Diebold FX, 2000. How Relevant is Volatility Forecasting for Financial Risk Management. *Review of Economics and Statistics* **82**:1–11.

Clements MP, Taylor N, 2000. Evaluating interval forecasts of high-frequency financial data. Manuscript, University of Warwick.

Embrechts P, Kluppelberg C, Mikosch T, 1997. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag Berlin Heidelberg.

Gavin J, 2000. Extreme value theory- an empirical analysis of equity risk. UBS Warburg working paper.

Gourieroux C, 1997. *ARCH Models and Financial Applications*. Springer.

Gourieroux C, Monfort A, Trognon A, 1984. Pseudo Maximum Likelihood Methods: Theory. *Econometrica* **52**:681–700.

Hayashi F, 2000. *Econometrics*. Princeton Univ Press.

Leadbetter MR, Lindgren G, Rootzen H, 1983. *Extremes and Related Properties of Random Sequences and Process*. Springer-Verlag New York Inc.

Lopez JA, 1998 May June. Testing your Risk Tests. *The Financial Survey* 18–20.

Lopez JA, 1999. Methods for Evaluating Value-at-Risk Estimates. *Federal Reserve Bank of San Francisco Economic Review* (2):3–17.

McNeil AJ, 1996. Estimating the tails of loss severity distributions using extreme value theory. Manuscript, Department Mathematik, ETH Zentrum, Zurich.

McNeil AJ, 1999. Extreme value theory for risk managers. Manuscript, Department Mathematik, ETH Zentrum, Zurich.

McNeil AJ, Frey R, 1999. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Manuscript, Federal Institute of Technology.

Neftci SN, 2000. Value at Risk Calculations, Extreme Events, and Tail Estimation. *The Journal of Derivatives* **Spring**:1–15.

Sarma M, Thomas S, Shah A, 2001. Selection of Value-at-Risk models. Mimeo, Indira Gandhi Institute of Development research.

Table 1 Estimation of AR(1)-GARCH(1,1) model

Parameter	Estimates	SE	Confidence bounds
The mean equation:			
Constant	-0.036	0.040	(-0.114, 0.043)
AR(1)	0.225	0.029	(0.167,0.282)
The variance equation:			
Constant	0.039	0.015	(0.010, 0.067)
ARCH(1)	0.101	0.016	(0.069, 0.132)
GARCH(1,1)	0.893	0.015	(0.863, 0.923)

Table 2 Tests for skewness, kurtosis and auto correlation

This table presents the values of the test statistics and the corresponding p-values of the tests of skewness, kurtosis and autocorrelation for the raw data (Panel A) and the standard residuals (Panel B).

	statistic	p-value
Panel A: The returns series (r_t)		
skewness	-9.5135	0.002
kurtosis	1058.7577	0.000
H.C. Ljung-Box	57.2202	0.01
Panel B: The residual series (z_t)		
skewness	-3.2211	0.073
kurtosis	10.0305	0.002
H.C. Ljung-Box	47.7039	0.074

Figure 1 95% and 99% VaR plots for the POT model, historical simulation and the normal distribution model

This graphs show the dynamic VaR forecasts for a long and a short position of Rs. 100 in the Nifty portfolio. The 95% and 99% VaR forecasts are obtained by using three approaches–the POT approach based on Extreme Value Theory, the historical simulation approach based on empirically observed quantiles and the normal distribution approach based on the normality assumption. The graph at the top is for the POT model, the middle graph is for the historical simulation model and the bottom one is for the normal distribution model. Each graph has two sets of VaR forecasts, one for the long position and the other for the short position. The graphs on the negative side of the observed returns is for the long position and the one on the positive side is for the short position.

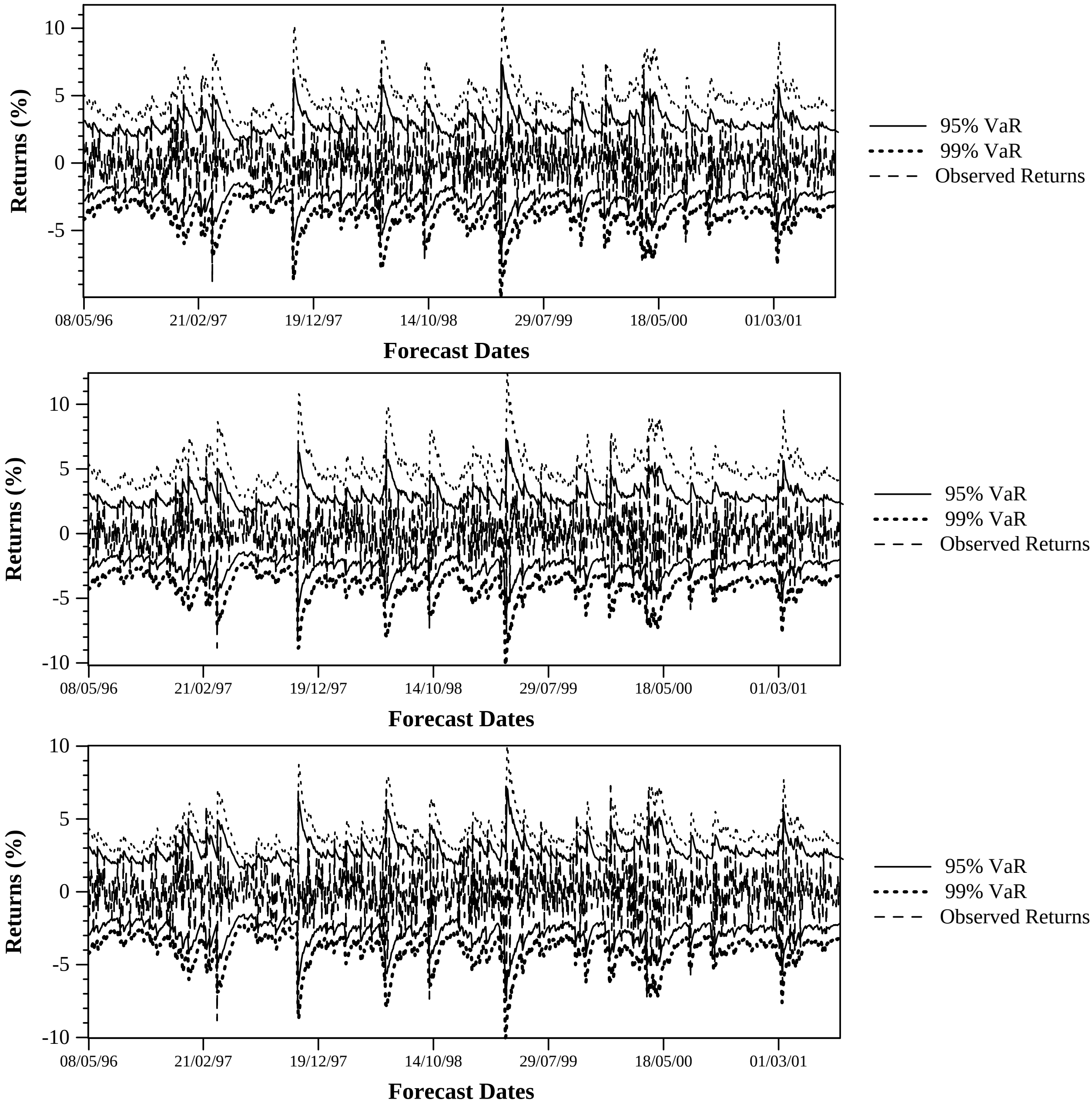

Table 3 Results of the GPD estimation

This table provides the results of the estimated GPD parameters fitted to the excesses over the chosen threshold. The first column gives the threshold points on both the left and the right tails corresponding to the 1.65σ level of threshold. The second column presents the number of observations beyond the threshold level and the third column gives the estimated cdf of the tails at the respective threshold points. The fourth and the fifth columns present the Pseudo-maximum-likelihood estimation of the GPD parameters fitted to the excesses over the thresholds, along with the standard errors of estimation within parenthesis.

	u	N_u	F_u	$\hat{\xi}$	$\hat{\sigma}$
left tail	-1.6493	50	0.0401	0.2027	0.4099
				(0.2095)	(0.1030)
Right tail	1.6494	66	0.9471	-0.0064	0.6460
				(0.2736)	(0.1950)

Figures in parenthesis indicate standard error

Table 4 Estimated quantiles on the i.i.d. residuals

This table gives the estimated percentile points on the tails of the i.i.d. residual series which are used for dynamic VaR forecasts. The quantiles are estimated by three approaches, viz. the GPD approach, the Historical Simulation (HS) approach and the normal distribution approach (normal). The first column gives the name of the models, the second, third, fourth and the fifth columns give, respectively the estimated 1^{st}, 5^{th}, 95^{th} and 99^{th} percentiles. The first two precentiles are used for estimating 1% and 5% VaR for the long Nifty portfolio and the last two are used for 95% and 99% VaR for the short Nifty portfolio.

	$q_{.01}$	$q_{.05}$	$q_{.95}$	$q_{.99}$
GPD	-1.5608	-2.3067	1.6862	2.7202
HS	-1.5092	-2.3615	1.6728	2.8785
Normal	-1.6449	-2.3263	1.6449	2.3263

Table 5 Results of the test of "conditional coverage" for alternative models: The Long Nifty position

This table presents the results of the test of autoregressive and periodic dependence in the failure series generated by the three alternative approaches. The VaR measures are for a long position in the Nifty portfolio. Panel A of the table pertains to 95% VaR estimation and panel B pertains to 99% VaR estimation.
For each of the models, an OLS regression as given in the equation (12) is carried out. The first column gives the name of the model, the second column gives the estimated failure probability and the corresponding p-value in parenthesis, and the third column reports the estimated F-statistics of the hypothesis specified in (11) with the corresponding p-values (within parenthesis).

	$\hat{p}$ (p-value)	F-stat (p-value)
95% VaR estimation		
POT	0.0452 (0.6778)	0.6155 (0.8016)
HS	0.052 (0.4112)	0.6465 (0.7744)
Normal	0.0417 (0.7909)	0.7893 (0.6391)
99% VaR estimation		
POT	0.0087 (0.6397)	0.29651 (0.98208)
HS	0.00865 (0.6397)	0.2965 (0.9821)
Normal	0.016 (0.1222)	0.6098 (0.80651)

Figures in parentheses indicate p-values

Table 6 Results of the test of "conditional coverage" for alternative models: The Short Nifty position

This table presents the results of the test of autoregressive and periodic dependence in the failure series generated by the three alternative approaches. The VaR measures are for a short position in the Nifty portfolio. Panel A of the table pertains to 95% VaR estimation and panel B pertains to 99% VaR estimation. For each of the models, an OLS regression as given in the equation (12) is carried out. The first column gives the name of the model, the second column gives the estimated failure probability and the corresponding p-value in parenthesis, and the third column reports the estimated F-statistics of the hypothesis specified in (11) with the corresponding p-values (within parenthesis).

	$\hat{p}$ (p-value)	F-stat (p-value)
95% VaR estimation		
POT	0.0500 (0.49895)	1.0632 (0.3878)
HS	0.0505 (0.47637)	1.2388 (0.2614)
Normal	0.0564 (0.2647)	1.2882 (0.23193)
99% VaR estimation		
POT	0.0056 (0.83593)	0.7701 (0.6579)
HS	0.0057 (0.88245)	0.73099 (0.6956)
Normal	0.0039 (0.8662)	1.33569 (0.2060)

Figures in parentheses indicate p-values

Table 7 Matrix of $S_{i,j}$ statistics of the test for the superiority of model i over model j: VaR estimation for a long Nifty portfolio

This table provides the pairwise sign statistic for testing the hypothesis of superiority of a particular model over a second model with respect to the loss function defined in (5.2). Panel A table pertains to 95% VaR estimation and Panel B pertains to the 99% VaR estimation for the long Nifty position. If the sign statistic in the $(i, j)^{th}$ cell in the table is significant, it implies that the Model specified in the i^{th} row is significantly superior to the model in the j^{th} column.

95% VaR estimation			
	POT	HS	Normal
POT	-	32.50	-33.11^*
HS	-32.50^*	-	-32.94^*
Normal	33.11	32.94	-
99% VaR estimation			
POT	-	-35.65^*	-35.65^*
HS	35.65	-	35.65
Normal	35.65	-35.65	-

Figures marked by an * are significant at 5% level

Table 8 Matrix of $S_{i,j}$ statistics of the test for the superiority of model i over model j: VaR for a short Nifty portfolio

This table provides the pairwise sign statistic for testing the hypothesis of superiority of a particular model over a second model with respect to the loss function defined in (5.2). Panel A table pertains to 95% VaR estimation and Panel B pertains to the 99% VaR estimation for the short Nifty position. If the sign statistic in the $(i, j)^{th}$ cell in the table is significant, it implies that the Model specified in the i^{th} row is significantly superior to the model in the j^{th} column.

95% VaR estimation			
	POT	HS	Normal
POT	-	32.96	32.91
HS	-32.91^*	-	32.85
Normal	-32.91^*	-32.80^*	-
99% VaR estimation			
POT	-	-35.67^*	35.50
HS	35.67	-	35.50
Normal	-35.50^*	-35.50^*	-

Figures marked by an * are significant at 5% level