Chapter 3

Nonparametric Estimation of $100(1-p)$ Percent Expected Shortfall: $p \to 0$ as Sample Size is Increased

From Chapter 1, we recall that under the assumption that $\{X_t\}_{t=1,2,\cdots}$ is a stationary process and $E|X_1| < \infty$, the $100(1-p)$ ES is given by

$$ES_p = -\frac{1}{p}\int_{x>Q_p} x dF(x). \tag{3.1}$$

In this chapter we review several nonparametric ES estimators and compare their known properties. Using Monte Carlo simulations we compare the accuracy of these estimators under the condition that $p \to 0$ as $n \to \infty$ for several asset return time series models, where n is the sample size. Not much seems to be known regarding the properties of the ES estimators under this condition. For p close to zero, the ES measures an extreme loss in the right tail of the loss distribution of the asset or portfolio. Our simulations and real data analysis provide insight into the effect of varying p with n on the performance of nonparametric ES estimators.

3.1 Nonparametric methods for estimating expected shortfall

In this section we review some known nonparametric estimators of ES. Let, F_k^l denote the σ-algebra of events generated by $\{X_t, k \le t \le l\}$ for $l > k$. The α-mixing coefficient introduced by Rosenblatt[88] is defined as

$$\alpha(k) = \sup_i \sup_{A\in F_1^i, B\in F_{i+k}^\infty} |P(AB) - P(A)P(B)|.$$

The series $\{X_t\}_{t\in N}$ is said to be α-mixing if $\lim_{k\to\infty}\alpha(k) = 0$. In the sequel we assume that

1. Assumption 1. $\exists \rho \in (0,1)$ such that $\alpha(k) \leq C\rho^k$ for all $k \geq 1$ and a positive constant C.

2. Assumption 2. The distribution function F of X_t is absolutely continuous with probability density f which has continuous second derivatives in $B(Q_p)$, a neighborhood of Q_p. F_k, the joint distribution function of (X_1, X_{k+1}), have all its second derivatives bounded in $B(Q_p)$ for $k \geq 1$ and $E(|X_t|^{2+\delta}) \leq C$ for some $\delta > 0$ and a positive C.

Let $\gamma(k) = Cov\{(X_1 - Q_p)I(X_1 \geq Q_p), (X_{k+1} - Q_p)I(X_{k+1} \geq Q_p)\}$ for any positive integer k and

$$\sigma_0^2(1-p;n) = \{Var\{(X_1 - Q_p)I(X_1 \geq Q_p)\} + \sum_{k=1}^{n-1} \gamma(k)\}.$$

3.1.1 Empirical estimator

Let $\hat{F}$ denote the empirical distribution of the observed losses $X_1, X_2, \ldots, X_n$ i.e.

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x),$$

where $I(\cdot)$ is the indicator function and X_i is i.i.d. with distribution F. By standard results on empirical distribution (see [103]), the pth quantile can be estimated by:

$$\hat{F}^{-1}(1-p) = X_{(i)},\ 1-p \in \left[\frac{i-1}{n}, \frac{i}{n}\right),$$

where $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ are the order statistics. The empirical estimator of ES is defined as

$$Emp_p = -\frac{\sum_{i=[n(1-p)]+1}^{n} X_{(i)}}{n - [n(1-p)]},$$

where $[x]$ denotes the largest integer not greater than x. The empirical estimator can be re-written as

$$Emp_p = -\frac{\sum_{t=1}^{n} X_t I(X_t \geq \hat{q}_p)}{[np]+1},$$

where $\hat{q}_p = X_{([n(1-p)]+1)}$. Under Assumptions 1-2, Chen [22] obtained the following asymptotic property of the empirical estimator. Chen [22] showed that under the Assumptions 1 and 2 and for some positive κ, as $n \to \infty$

$$Emp_p - ES_p = \frac{1}{p}\left\{\frac{1}{n}\sum_{i=1}^{n}(X_t - Q_p)I(X_t \geq Q_p) - p(ES_p - Q_p)\right\} + o_p(n^{-\frac{3}{4}+\kappa}).$$

The above equation is a Bahadur type expansion which leads to the following theorem regarding the asymptotic normality of the empirical estimator Emp_p.

Theorem 3.1.1. *(Chen [22]) Under Assumptions* 1-2*, as* $n \to \infty$

$$\sqrt{np}\sigma_0^{-1}(1-p;n)(Emp_p - ES_p) \xrightarrow{d} N(0,1).$$

The above theorem indicates that the asymptotic standard deviation of Emp_p equals $\frac{\sigma_0(1-p;n)}{\sqrt{np}}$, which is the standard deviation of $\frac{1}{np}\sum_{i=1}^{n}(X_t - Q_p)I(X_t \geq Q_p)$. Chen [22] argued that the asymptotic variance of Emp_p equals $\frac{2\pi\phi(0)}{np^2}$, where ϕ is the spectral density of $\{(X_t - Q_p)I(X_t \geq Q_p)\}_t$.

3.1.2 Kernel based estimator

Scaillet [91] proposed an estimator which employs kernel smoothing in both the initial VaR estimation and the final averaging of the excessive losses. The kernel estimator proposed by Scaillet [91] is defined as follows. Let K be a kernel function, which is a symmetric probability density function, $G(t) = \int_t^\infty K(u)du$ and $G_h(t) = G(t/h)$, where h is a positive smoothing bandwidth. The kernel estimator of the survival function $S(x) = 1 - F(x)$ is

$$S_h(z) = \frac{1}{n}\sum_{t=1}^{n} G_h(z - X_t).$$

A kernel estimator of Q_p, denoted as $\hat{q}_{p,h}$, is the solution of $S_h(z) = 1 - p$. Then the kernel estimator of ES is given as

$$Chen_{p,h} = -\frac{1}{np}\sum_{t=1}^{n} X_t G_h(\hat{q}_{p,h} - X_t).$$

In the kernel based method the main problem lies with the selection of bandwidth. Azzalini [7], Bowman [16], Scaillet [91] and Chen and Tang [23] provide some choice of the bandwidth parameter. Chen and Tang [23] have obtained the asymptotic bias, variance and the rate of almost sure convergence of their version of $\hat{q}_p$, under the assumption that $\{X_t\}$ is a stationary geometric $\alpha-$mixing process. The authors suggested the following choice for the optimal value of h,

$$h_{opt} = \left\{\frac{2f^3(Q_p)b}{\sigma^4(f^{(1)}(Q_p))^2}\right\}^{1/3} n^{-1/3},$$

where $b = \int uw(u)H(u)du$, and $\sigma^2 = \int u^2 w(u)du$. $H(\cdot)$ is the distribution function of the distribution with density w. h involves unknown constants Q_p, f and its derivative $f^{(1)}$ at Q_p. Chen and Tang [23] suggested to approximate Q_p in h by the corresponding sample quantile. The authors suggested to approximate f and $f^{(1)}$ by the density and the first derivative of the Generalized Pareto distribution.

Theorem 3.1.2. *(Chen [22]) Under the Assumptions* $1-2$ *and assuming that* K *is a symmetric probability density satisfying* $\int_{-1}^{1} uK(u)du = 0,\ \int_{-1}^{1} u^2K(u)du = \sigma_K^2$ *and* K *has bounded and Lipschitz continuous derivative, Chen [22] proved that as* $n \to \infty$

$$\sqrt{np}\sigma_0^{-1}(1-p;n)(Chen_{p,h} - ES_p) \xrightarrow{d} N(0,1)$$

and furthermore, $Bias(Chen_{p,h}) = -\frac{1}{2p}\sigma_K^2 h^2 f(Q_p) + o(h^2)$ *and*

$$Var(Chen_{p,h}) = \frac{1}{np}\sigma_0^2(1-p;n) + o(n^{-1}h),$$

where $h = o(1),\ nh^{3-\beta} \to \infty$ for some $\beta > 0$, and $nh^4\log^2(n) = o(1)$ as $n \to \infty$ (see [22]). The term $\sigma_0^2(1-p;n)$ is the same as in Theorem 1. By comparing Theorems 1 and 2 it is observed that the kernel estimator has the same asymptotic distribution as the unsmoothed sample estimator Emp_p. Both Emp_p and $Chen_{p,h}$ converges to ES_p at the rate of $\sqrt{np}$. The second part of Theorem 2 implies that the kernel estimator does not offer a variance reduction, due to the presence of the second order term $o(n^{-1}h)$ in $Var(Chen_{p,h})$. Also smoothing brings in a bias. Clearly the asymptotic MSE of the $Chen_{p,h}$ is greater than the same for the empirical estimator. Therefore, for the purpose of estimating the ES, the kernel smoothing does not seem to yield an asymptotically efficient estimator.

3.1.3 Estimator of Brazauskas et al.

Let us recall that ES is defined as

$$ES_p = -\frac{1}{p}\int_{1-p}^{1} Q(u)du.$$

Let $\hat{F}$ denote the empirical cumulative distribution function of $X_1, \cdots, X_n$ and $\hat{F}^{-1}$ be its quantile function. Brazauskas et al. [17] defined an empirical estimator of ES_p as follows

$$\widehat{ES}_p = -\frac{1}{p}\int_{1-p}^{1} \hat{F}^{-1}(u)du. \tag{3.2}$$

Under the assumption that $X_1, \cdots, X_n$ are i.i.d. with $E|X_1| < \infty$, $\widehat{ES}_p$ converges to ES_p almost surely as n is increased (see [17]). To construct point-wise and simultaneous confidence intervals for ES_p Brazauskas et al. [17] obtained the following asymptotic result.

Theorem 3.1.3. *(Brazauskas et al. [17]) Let* $X_1, \ldots, X_n$ *be i.i.d. random variables with finite second moment* $E[X_1^2]$. *Let the distribution function* F *be continuous at the point* $F^{-1}(1-p)$. *Then as* $n \to \infty$

$$\sqrt{n}(\widehat{ES}_p - ES_p) \xrightarrow{d} N(0, \sigma^2(1-p)),$$

where $N(0, \sigma^2(1-p))$ denotes a centered normal random variable with the variance

$$\sigma^2(1-p) = \frac{1}{p^2} \int_{Q_p}^{\infty} \int_{Q_p}^{\infty} (F(x \wedge y) - F(x)F(y))dxdy.$$

3.1.4 Tail-trimmed estimator by Hill

Let $X_t^{(-)} = X_t I(X_t < 0)$ and $X_{(1)}^{(-)} \leq X_{(2)}^{(-)} \cdots \leq X_{(n)}^{(-)} \leq 0$ denote the negative numbers in the data ordered in increasing order. Let $\{k_n\}$ be an intermediate sequence, where $k_n \to \infty$ and $\frac{k_n}{n} \to 0$. Hill[53] defined the following tail trimmed estimator of ES.

$$Hill_p = \frac{1}{np} \sum_{t=1}^{n} X_t I(X_{(k_n)}^{(-)} \leq X_t \leq \hat{q}_{n,p}).$$

where $\hat{q}_{n,p} = X_{([pn])}$. k_n is the number of trimmed (omitted) tail extremes, representing an asymptotically vanishing and therefore negligible sample tail proportion $\frac{k_n}{n}$. Under a geometric α-mixing condition on $\{X_t\}$ and few additional regularity conditions (see [53]), $Hill_p$ is a consistent and asymptotically normal estimator, viz. under the assumptions that $\{X_t\}$ is strongly mixing with geometric mixing rate, $E[X_t^2] < \infty$, $k_n \to \infty$ at a slowly varying rate (e.g. $k_n \sim ln(n)$) and some of other assumptions in Hill [53],

$$n^{1/2}(Hill_p - ES_p) \xrightarrow{d} N\left(0, \frac{S^2}{p^2}\right),$$

where $S^2 = \lim_{n\to\infty} S_n^2$, $S_n^2 = \frac{1}{n} E\Big(\sum_{t=1}^{n} \{X_{n,t}^* - E[X_{n,t}^*]\} \Big)^2$ and $X_{n,t}^* = X_t I(-l_n \leq X_t \leq q_{1-p})$. $\{l_n\}$ is a positive sequence such that $P(X_t < -l_n) = \frac{k_n}{n}$.

The advantage of the above tail trimmed ES estimator is that it leads to asymptotically standard inference even for heavy tailed time series with infinite variance (see [53]). Tail-trimming is used to dampen the effect of extremes in a sample, ensuring standard Gaussian inference, and a higher rate of convergence than without trimming when the variance is infinite (see [53]). Hill [53] suggested to use $k_n = max\{1, [0.25n^{2/3}/(ln(n))^{2\iota}]\}$ and $\iota = 10^{-10}$ in his simulation study.

3.1.5 Yamai and Yoshiba's estimator

Yamai and Yoshiba [107] defined the following estimator of ES_p

$$ES_{p,\beta} = -\frac{1}{n(\beta - 1 + p)} \sum_{i=[n(1-p)]}^{n\beta} X_{(i)},$$

where $0 < 1 - p < \beta \leq 1$ is a positive constant. The empirical estimator Emp_p is similar to the above estimator for $\beta = 1$. The estimator is asymptotically normal, with asymptotic variance equal to

$$\frac{1}{n(\beta - 1 + p)^2}\Big[(1-p)Q_p^2 + \beta Q_{1-\beta}^2 + \int_{Q_p}^{Q_{1-\beta}} x^2 f(x)dx$$
$$- \Big\{\beta Q_{1-\beta} + (1-p)Q_p + \int_{Q_p}^{Q_{1-\beta}} x f(x)dx\Big\}^2\Big].$$

The optimal choice β is not specified by Yamai and Yoshiba. The above estimator may also be looked at as a tail trimmed estimator, where we omit the extreme sample quantiles in the right tail beyond $X_{(n\beta)}$. $n - n\beta$ is the number of right tail extremes in the sample that are trimmed to construct $ES_{p,\beta}$.

Remark 4. *1. The asymptotic results obtained in Yamai and Yoshiba [107], Hill [53], Brazauskas et al. [17] and Chen [22] are obtained under the assumption that p is fixed. Not much seems to be known about any of these estimators under the condition that $p \to 0$ as $n \to \infty$.*

2. If $p \to 0$ as $n \to \infty$ tail trimming in Hill's estimator $Hill_p$ seems to be challenging to implement, as for even a large sample size only a few values are below $\hat{q}_{n,p}$ (e.g. if $p = 0.001$ and $n \leq 1000$ there is at most one observation below $\hat{q}_{n,p}$).

3. If $p \to 0$ as $n \to \infty$, we may use $\beta = 1 - r_n$ in the Yamai and Yoshiba's estimator $ES_{p,\beta}$, where r_n converges to zero at a faster rate than p as $n \to \infty$. In this chapter we use $nr_n = max\{1, 0.25(np)^{2/3}/(ln(np+1))^{2\iota}\}$. This choice is motivated by the choice of k_n in Hill's estimator [53]. The difference is that we use np (or $np + 1$) instead of n in the formula for k_n proposed by Hill. The resulting r_n represents the proportion of omitted right tail extremes beyond $X_{[n(1-p)]}$. With this choice of $\beta = 1 - r_n$, $ES_{p,\beta}$ is always defined for any combination of n and p.

3.1.6 Filtered historical method

In this method a suitable time series model, such as an ARMA or a GARCH, is fitted to the asset return data. Let $\hat{e}_i$, $i = 1, 2, \ldots, n$, denote the residuals of the fitted model. Then the filtered historical estimator of ES (Magadia [71]) is given by

$$FH_p = -\frac{\sum_{\eta_t > q} \eta_t}{\sum_{\eta_t > q} I(\eta_t > q)},$$

where $\eta_t = \hat{e}_t - \frac{1}{n}\sum_{t=1}^{n} \hat{e}_t$ and $q = \eta_{([(1-p)n]+1)}$ is the $([(1-p)n]+1)$th order statistic of $\{\eta_1, \ldots, \eta_n\}$. The advantages of the filtered historical method are that it maintains the

correlation structure in the return data without relying on the exact specification of the conditional distribution of asset returns and it takes into account the changing market volatility conditions (see [108]).

3.2 Simulation study

We compare the MSE of six quantile estimators, viz. the empirical quantile estimator Emp_p, the Brazauskas et al.'s estimator $\widehat{ES}_p$, Yamai and Yoshiba's estimator $ES_{p,\beta}$, Filtered historical FH_p, Hill's estimator $Hill_p$ and the kernel estimator $Chen_{p,h}$ with $h = h_{opt}$. It is difficult to compute the exact value of the MSE of these estimators even if the data generating process is completely specified. The Monte Carlo estimate of the MSE of any estimator T_n of a parameter θ is defined as $\frac{1}{B}\sum_{j=1}^{B}(T_{nj} - \theta)^2$, where B is the number of Monte Carlo samples each of size n drawn from a given process and T_{nj} is the estimate based on the jth Monte Carlo sample, $j = 1, \cdots, B$. We consider ten time series models. The first three of these models are as follows

(i) $\{X_i\}_{i=1,2,\cdots}$is an i.i.d. process, marginal distribution GPD with $\xi = 1/3$.

(ii) $\{X_i\}_{i=1,2,\cdots}$is an i.i.d. process, marginal distribution Student's-t with 4 df.

(iii) $\{X_i\}_{i=1,2,\cdots}$is an i.i.d. process, marginal distribution$N(0,1)$.

The first two models are motivated by empirical observations by Cont [26] regarding the extent of tail heaviness of the marginal asset return distributions. Cont [26] mentioned that when sample moments based on asset return data are plotted against sample size, the sample variance seems to stabilize with increase in sample size. But the behavior of the fourth order sample moment seems to be erratic as n is increased. This feature is also exhibited by the sample moments based on i.i.d. draws from the Student's t-distribution with four degrees of freedom, which displays a tail behavior similar to many asset return distributions. Cont also mentioned that the daily return distributions of stocks, market indices and exchange rates seem to exhibit power law tail with exponent α satisfying, $\xi = 1/\alpha$ varying between 0.2 and 0.4. These observations motivate the choice of the marginal distributions in (i) and (ii). The third model (iii) represents the classical Black-Schole's assumption on return time series.

To study the effect of dependence on the above mentioned quantile estimators we consider

the following ARMA(1,1) models in Drees[33]

$$X_i - \phi X_{i-1} = Z_i + \theta Z_{i-1},$$
$$(iv)\ \phi = 0.95,\ \theta = -0.6,$$
$$(v)\ \phi = 0.95,\ \theta = -0.9,$$
$$(vi)\ \phi = 0.3,\ \theta = 0.9.$$

In addition the following GARCH(1,1) models are also considered

$$X_t = \sigma_t Z_t,$$
$$(vii)\ \sigma_t^2 = 0.0001 + 0.9X_{t-1}^2,$$
$$(viii)\ \sigma_t^2 = 0.0001 + 0.4X_{t-1}^2 + 0.5\sigma_{t-1}^2,$$
$$(ix)\ \sigma_t^2 = 0.0386X_{t-1}^2 + 0.9424\sigma_{t-1}^2.$$

The GARCH(1,1) time series is known to model the volatility clustering observed in financial time series data. The first two GARCH models are used in the simulation study in Drees [33]. The model (ix) is the GARCH model fitted to CNX NIFTY daily loss data for the duration 1st April 2012 to 31st March 2015 (details are mentioned in the section on data analysis).

We also consider a small-scale experiment to compare performance of the estimators of ES under netting agreements. The term netting is used to describe the process of offsetting mutual positions or obligations between two parties (see [41]). Suppose a trader borrows money from a broker, takes a long position on a certain equity and also buys a put option (short position) of the market index future to hedge against any random fall in the stock market. The trader can adopt two strategies. In the event of any unforseen downward movement in the market, he may cover the gains in the put option and take delivery of the stocks by paying remaining dues to the broker in cash. Otherwise the trader can exit both the long and short positions at market price, and return the dues to the broker. In this example a sudden downward market movement is the event that causes default. The first strategy is not netted, as only positions with positive gains are used to meet the default obligation. The second strategy involves netting, where overall portfolio gain is used to meet the traders obligation to the broker. Our model (x) represents the loss in the second strategy at time t. To study the effect of netting on ES estimation let us consider a simple portfolio made of a long position in one asset and a short position in another one with the same counter party.

Let $E_{1t},\ E_{2t}$ denote the gains in the long and short positions respectively. The vector (E_{1t}, E_{2t}) is assumed to be Gaussian. m_i and σ_i are mean and standard deviation of E_{it},

$i = 1, 2$ and ρ is the correlation coefficient. Since E_{1t} and E_{2t} are long and short position gains, we assume that ρ is negative. Let D_t be a Bernoulli random variable, independent of (E_{1t}, E_{2t}), such that $D_t = 1$ represents a credit event that causes default at time t (and hence initiation of a netting agreement). In case of default, without any netting arrangement, the loss at time t equals $E_{1t}^+ + E_{2t}^+$. However under netting arrangement, the loss due to default at time t equals $(E_{1t} + E_{2t})^+$ (see [41], page 937). Therefore under this netting arrangement, the loss at time t equals $I(D_t = 1)(E_{1t} + E_{2t})^+ - I(D_t = 0)(E_{1t} + E_{2t})$. This motivates model (x) in our simulation study

$$(x)\ X_t = I(D_t = 1)(E_{1t} + E_{2t})^+ - I(D_t = 0)(E_{1t} + E_{2t}),$$

where $\{(E_{1t},\ E_{2t})\}$ is an i.i.d. Gaussian process, with $m_1 = 10,\ m_2 = -1,\ \rho = 0.89$ and $\sigma_i = 1, i = 1, 2$. And we take $P(D_t = 1) = 0.20$, i.e. the chance of default is assumed to be twenty percent.

From each of the above models (i)-(x) and for each combination of n and p, we draw 1000 Monte Carlo samples of size n. From each of these samples compute the values of the six estimators of ES_p for various values of p. From these values we compute the Monte Carlo estimate of the MSE of that estimator for different choices of n, p and the underlying time series model. In each case, let the Monte Carlo estimates of the MSE of the estimators Emp_p, $\widehat{ES}_p$, $ES_{p,\beta}$, FH_p, $Hill_p$ and kernel based estimator $Chen_{p,h}$ be denoted by $MSE1$-$MSE6$ respectively. In Tables 3.3-3.5, we report the ratios $\frac{MSE2}{MSE1}, \cdots, \frac{MSE6}{MSE1}$ for p varying from 0.05 to 0.001 for the i.i.d, ARMA and the GARCH models respectively. The values of the ratio of the MSEs in Tables 3.3-3.5 are reported for n equal to 100, 250, 500 and 1000. In Table 3.6, we report these ratios for the process generated by model (x).

3.3 Findings

From the Tables 3.1-3.6 we have the following observations.

1. No estimator uniformly out performs the other estimators. However we can identify some conditions under which some of these estimators performs well.

2. *Large n and small p.* For $np > 1$, the Yamai and Yoshiba's $ES_{p,\beta}$ (with our choice of β) clearly outperforms the empirical estimator for the GARCH(1,1) time series model and i.i.d. processes with heavy tailed marginal densities (see the values of the ratio $\frac{MSE3}{MSE1}$ in Tables 3.3 and 3.5).

   If $n \geq 500$ and $p = 0.01, 0.001$ the estimators $\widehat{ES}_p$ and FH_p seem to be more accurate than the empirical estimator Emp_p for data generated by the i.i.d process with Student-t marginal density and ARMA models (iv) and (v) (see Tables 3.3 and 3.4). For ARMA

model (vi) if $n = 1000$ and $p = 0.01, 0.001$ the estimators $\widehat{ES}_p$ and FH_p seem to be more accurate than the empirical estimator Emp_p (see Table 3.4). For the GARCH models (vii) and $(viii)$, the estimators $\widehat{ES}_p$ and FH_p seem to outperform the empirical estimator Emp_p for $p = 0.001$ and $n = 1000$ (see Table 3.5). The FH_p estimator seems to outperform Emp_p for $n \geq 250$ and for all p for ARMA models (iv), (v) and (vi) (see Table 3.4). In general, the estimators $ES_{p,\beta}$, ES_p and FH_p seem to be suitable for estimation of ES_p especially for small p and large sample size, such that $np > 1$.

3. $n \leq 500$. If the marginal density is GPD, the estimators $\widehat{ES}_p$ and $ES_{p,\beta}$ seem to outperform Emp_p for $n \leq 500$ and $p \geq 0.01$ (see Table 3.3). For the ARMA models, the FH_p estimator seems to perform well for $n = 250$ and $p = 0.001$ (See Table 3.4). For the GARCH model (ix) and $n \leq 250$, the FH_p seems to outperform the empirical estimator Emp_p for all values of p (see Table 3.5).

4. *Effect of decreasing p.* The effect of decreasing p keeping n fixed (i.e. as we attempt to estimate more extreme quantiles based on the same sample size), on the relative accuracy of the nonparametric estimators seems to be model specific. For ARMA models, the accuracy of the estimators $\widehat{ES}_p$ and FH_p in comparison to the empirical estimator seems to increase as $p \to 0$, keeping n fixed (see Table 3.4). This implies that for stationary time series data where ARMA models are a good fit, extreme quantiles can be estimated more accurately by $\widehat{ES}_p$ and FH_p than the sample quantile. For the GARCH model (ix) and $n = 1000$, the accuracy of the Yamai and Yoshiba's $ES_{p,\beta}$, compared to Emp_p, seems to improve as $p \to 0$.

5. *Effect of netting.* Under model (x), we see that the estimator $\widehat{ES}_p$ proposed by Brazauskas et al. [17] outperforms the empirical estimator for all choices of n and p. See Table 3.6. For $np < 1$, the other nonparametric estimators perform poorly compared to the empirical estimator. For $np > 1$, Yamai and Yoshiba's estimator $ES_{p,\beta}$ and $\widehat{ES}_p$ outperform the empirical estimator for data generated by model (x). The accuracy of only $\widehat{ES}_p$ seems to increase as $p \to 0$, keeping n fixed, in presence of netting arrangement.

Remark 5. *1. The above observations suggest that estimators $ES_{p,\beta}, \widehat{ES}_p$ and FH_p are preferable choices for estimation of ES for large n and small p, such that $np > 1$. However for $np < 1$, the gain in accuracy using these estimators compared to Emp_p varies widely with the process generating the data. If the data is generated by a $GARCH(1,1)$ model, the filtered historical estimator FH_p seems to perform well.*

2. Performance of $Hill_p$ estimator. *We see that asymptotic variance of $Hill_p$ estimator is $\frac{S^2}{np^2}$ which can be large for fixed n and small p. This explains the poor performance of the $Hill_p$ for small p.*

3. Kernel Smoothing for ES estimation. *The values of the ratio $\frac{MSE6}{MSE1}$ in Tables 3.1-3.3 indicate that the kernel based estimator performs poorly compared to the empirical ES estimator, and that there is no reason to use kernel smoothing for ES estimation. Earlier we observed that Kernel smoothing does not yield an asymptotically efficient estimator. So kernel smoothing is not recommended for ES estimation.*

4. ES estimation in the presence of netting. *In presence of netting, the ES estimator proposed by Brazauskas et al. [17] is recommended. Other nonparametric estimators may perform poorly compared to the empirical estimator in presence of netting, especially for $np < 1$. The FH_p estimator, seems to perform poorly for the data generated by the netting model (x). The FH_p estimator is obtained by fitting an ARMA or a GARCH model to the asset return data. We fitted GARCH model (as this seems to be the appropriate model for Nifty return data) to compute FH_p. The poor performance of the FH_p estimator in the presence of netting is perhaps due to the fact that the GARCH model may not be an appropriate model for the returns generated by the netting model (x).*

3.4 Data analysis

The S & P Nifty is a well diversified 50 stock index accounting for 22 sectors of the Indian economy. For investors in the Indian equity market, it is of natural importance to assess the market risk of the Nifty index for a number of purposes, such as benchmarking performance of mutual funds (See for instance, Biswas and Dutta [13]). We apply the above nonparametric estimators to estimate the expected shortfall of the S & P CNX Nifty index based on the daily closing values of the index from 1st April 2007 to 31st March 2015. These data are collected from national stock exchange(NSE) website (see http://www.nseindia.com/products/content/equities/indices/indices.htm). There are 1983 daily return values in our data. In Appendix A, we have reported the monthly returns of Nifty index from 1st April 2007 to 31st March 2015. The mean and standard deviation of these data are -1.86×10^{-04} and 0.007 respectively (indicating that the Nifty daily returns during 1st April 2007 to 31st March 2015 were closely scattered around zero). However there were 1476 trading days during this period where the Nifty daily return exceeded one percent. In Table 3.2, we report five nonparametric estimates of the 99 and 99.9 percent ES of the Nifty index during the period under study. We do not use the kernel based estimator, as it is found perform poorly in our simulations.

In Table 3.2, we observe discrepancy among the ES estimates for $p = 0.001$. For real data the underlying data generating process is unknown. However, the GARCH model (ix) fits well to this data. Also the asset return data are known to exhibit similar features as an i.i.d. process with Student's t, with 4 degree of freedom, as the marginal density (see [26]). In Table 3.1, we report the ratio of the Monte Carlo estimates of the MSEs of the estimators $\widehat{ES}_p$, $ES_{p,\beta}$, FH_p and $Hill_p$ to the same for the empirical estimator Emp_p based on samples of size $n = 1983$ generated by the GARCH (ix) model (fitted to the Nifty data) and Student's t distribution, with 4 degrees of freedom. From Table 3.1 we see that, the 99.9 percent ES estimates obtained by Yamai and Yoshiba's estimator $ES_{p,\beta}$ and Hill's estimator $Hill_p$ perform poorly for sample of size $n = 1983$ generated by the GARCH (ix) model.

The $\widehat{ES}_p$, FH_p estimates are almost equal for both the 99 and 99.9 percent ES estimates. The 99 and 99.9 percent ES estimates based on the Nifty return data are equal to -1.4 and -1.8 percent respectively. These values serve as important benchmark of (daily) market risk in National Stock exchange, India, during 1st April 2007 to 31st March 2015.

Table 3.1: Ratios estimated for Student's t distribution and GARCH model with varying p.

Model Coeff.	n	1-p	$\frac{MSE2}{MSE1}$	$\frac{MSE3}{MSE1}$	$\frac{MSE4}{MSE1}$	$\frac{MSE5}{MSE1}$
Student's t	1983	0.99	0.890	0.897	0.959	0.931
		0.999	0.874	0.521	0.998	343.110
GARCH($\alpha = 0.039, \beta = 0.942$)	1983	0.99	1.007	3.392	0.964	6.909
		0.999	1.096	5.266	0.873	2631.265

Table 3.2: Estimating ES of nifty returns at 99% and 99.9%.

Index	1-p	Emp_p	$\widetilde{ES}_p$	$ES_{p,\beta}$	FH_p	$Hill_p$
CNX Nifty	0.99	-0.015	-0.014	-0.015	-0.014	-0.015
	0.999	-0.035	-0.018	-0.031	-0.018	-0.028

Table 3.3: Ratios estimated for i.i.d. cases with varying p.

Dist	n	1-p	$\frac{MSE2}{MSE1}$	$\frac{MSE3}{MSE1}$	$\frac{MSE4}{MSE1}$	$\frac{MSE5}{MSE1}$	$\frac{MSE6}{MSE1}$
GPD	100	0.95	0.801	0.547	1.296	0.885	7.808
		0.97	0.785	0.482	1.179	0.855	5.881
		0.99	0.710	0.368	1.094	0.721	3.191
		0.999	1.011	42.282	1.107	3.581	2.051
	250	0.95	0.876	0.969	1.795	0.827	3.498
		0.97	0.897	0.594	1.504	0.780	4.210
		0.99	0.642	0.496	1.097	0.512	5.370
		0.999	2.253	5.914	2.459	15.515	6.001
	500	0.95	0.858	1.572	2.278	1.783	4.900
		0.97	0.868	1.010	1.652	1.037	5.347
		0.99	0.476	0.460	0.822	0.972	6.460
		0.999	1.218	0.824	1.385	57.741	7.205
	1000	0.95	0.899	1.895	3.607	2.409	9.239
		0.97	0.953	1.548	2.344	1.832	10.401
		0.99	0.584	1.115	1.212	1.002	12.865
		0.999	0.395	0.304	0.596	110.798	13.011
Student t	100	0.95	0.833	0.512	0.862	1.023	18.923
		0.97	0.765	0.394	0.909	0.954	12.411
		0.99	0.645	0.226	0.963	0.878	5.119
		0.999	1.010	131.555	0.992	7.425	2.967
	250	0.95	0.781	0.932	0.791	0.775	6.022
		0.97	0.687	0.832	0.770	0.622	7.997
		0.99	0.435	0.790	0.756	0.363	8.094
		0.999	1.690	18.632	1.699	37.739	9.888
	500	0.95	0.814	1.580	0.771	2.008	11.899
		0.97	0.690	0.881	0.744	0.949	12.003
		0.99	0.414	0.403	0.704	0.985	14.058
		0.999	0.728	2.802	0.751	145.212	15
	1000	0.95	0.831	1.744	0.806	2.630	28.870
		0.97	0.741	1.263	0.805	1.658	29.988
		0.99	0.483	0.982	0.897	0.954	31.178
		0.999	0.233	0.171	0.271	386.689	32.001
N(0,1)	100	0.95	1.083	0.413	0.816	1.562	75.720
		0.97	1.103	0.234	0.870	1.682	59.934
		0.99	1.041	0.114	0.937	1.654	35.583
		0.999	1.024	2551.701	0.981	15.991	14.523
	250	0.95	0.727	0.993	0.554	0.841	47.554
		0.97	0.581	3.525	0.462	0.697	49.008
		0.99	1.495	9.432	0.914	0.355	50.451
		0.999	0.081	520.594	0.075	78.565	51.665
	500	0.95	0.741	1.221	0.578	3.059	115.559
		0.97	0.625	0.489	0.497	1.355	120.898
		0.99	1.103	0.508	0.981	1.377	131.968
		0.999	0.028	125.381	0.027	312.661	132.010
	1000	0.95	0.808	1.328	0.654	4.104	104.897
		0.97	0.776	0.871	0.640	3.399	108.433
		0.99	1.493	0.337	1.402	1.180	110.688
		0.999	0.011	0.207	0.014	2251.25	112.001

Table 3.4: Ratios estimated for ARMA model with varying p.

Model Coeff.	n	1-p	$\frac{MSE2}{MSE1}$	$\frac{MSE3}{MSE1}$	$\frac{MSE4}{MSE1}$	$\frac{MSE5}{MSE1}$	$\frac{MSE6}{MSE1}$
(0.95,−0.9)	100	0.95	1.147	0.465	0.675	1.567	134.021
		0.97	1.151	0.280	0.765	1.633	114.420
		0.99	1.072	0.116	0.882	1.699	68.529
		0.999	1.022	1895.102	0.977	16.256	26.177
	250	0.95	0.882	1.0423	0.484	0.980	10.112
		0.97	0.665	2.857	0.413	0.794	12.007
		0.99	1.119	7.273	0.865	0.388	74.978
		0.999	0.117	370.223	0.107	77.841	20.990
	500	0.95	0.878	1.194	0.468	2.626	29.010
		0.97	0.671	0.567	0.425	1.293	30.898
		0.99	0.234	0.424	0.173	1.489	80.121
		0.999	0.043	85.846	0.038	485.339	35.890
	1000	0.95	1.048	1.285	0.5	3.3888	95.119
		0.97	0.829	0.910	0.518	2.7461	99.885
		0.99	0.340	0.370	0.248	1.3596	101.444
		0.999	0.018	0.142	0.016	1921	104.998
(0.95,−0.6)	100	0.95	1.107	0.746	0.559	1.249	11.501
		0.97	1.310	0.543	0.63	1.324	11.277
		0.99	1.182	0.246	0.759	1.457	10.024
		0.999	1.013	684.283	0.922	18.784	5.547
	250	0.95	1.092	1.061	0.554	1.106	9.887
		0.97	1.138	1.109	0.640	1.123	12.001
		0.99	0.833	1.739	0.553	0.956	21.140
		0.999	0.368	139.002	0.315	90.996	17.785
	500	0.95	1.070	1.070	0.539	1.392	20.456
		0.97	1.109	0.829	0.618	1.143	23.665
		0.99	0.790	0.396	0.472	1.298	27.406
		0.999	0.133	32.893	0.110	589.164	30.642
	1000	0.95	1.052	1.081	0.509	1.490	69.897
		0.97	1.084	0.972	0.596	1.445	71.888
		0.99	0.878	0.624	0.597	1.206	74.769
		0.999	0.051	0.108	0.038	8887.656	76.991
(0.3,0.9)	100	0.95	1.090	0.502	0.744	1.352	40.802
		0.97	1.104	0.312	0.82	1.478	34.922
		0.99	1.162	0.133	0.918	1.590	24.574
		0.999	1.02	1602.712	0.979	17.120	9.008
	250	0.95	0.875	0.992	0.624	0.965	8.771
		0.97	0.693	2.407	0.536	0.824	11.877
		0.99	1.068	5.652	0.897	0.456	38.761
		0.999	0.148	296.705	0.137	82.549	18.991
	500	0.95	0.915	1.132	0.610	2.269	19.891
		0.97	0.742	0.589	0.542	1.210	22.654
		0.99	1.021	0.414	0.767	1.381	55.870
		0.999	0.056	68.484	0.050	504.652	29.008
	1000	0.95	0.974	1.204	0.676	2.964	67.909
		0.97	0.873	0.892	0.667	2.393	70.876
		0.99	0.332	0.407	0.326	1.284	73.001
		0.999	0.0234	0.131	0.022	1766.340	77.991

Table 3.5: Ratios estimated for GARCH model with varying p.

Model Coeff.	n	1-p	$\frac{MSE2}{MSE1}$	$\frac{MSE3}{MSE1}$	$\frac{MSE4}{MSE1}$	$\frac{MSE5}{MSE1}$	$\frac{MSE6}{MSE1}$
$\alpha = 0.9$	100	0.95	0.727	0.812	0.909	0.627	3.091
		0.97	0.750	0.878	1	0.651	2.650
		0.99	0.842	0.992	1	0.846	2.368
		0.999	1.004	55.238	0.998	6.057	1.792
	250	0.95	0.714	0.388	0.857	0.519	4.110
		0.97	0.857	0.718	0.929	0.503	3.339
		0.99	0.947	1.066	0.487	0.554	2.166
		0.999	1.237	13.562	1.220	8.926	1.668
	500	0.95	1	0.425	1	0.421	1.988
		0.97	0.857	0.487	1	0.548	2.225
		0.99	0.908	0.892	0.554	0.510	3.400
		0.999	0.835	2.946	0.835	18.612	2.877
	1000	0.95	0.714	0.501	0.857	0.492	1.890
		0.97	0.765	0.477	1	0.475	2.118
		0.99	1.057	0.537	0.914	0.553	2.007
		0.999	0.260	1.073	0.317	34.376	3.007
$\alpha = 0.4$, $\beta = 0.5$	100	0.95	1	0.993	1	0.812	6.250
		0.97	0.923	0.904	1	0.861	5.385
		0.99	1	0.549	0.967	1.072	4.533
		0.999	1.005	118.903	0.995	11.758	2.439
	250	0.95	0.833	0.722	0.833	0.738	4.314
		0.97	0.818	1.007	0.909	0.713	3.388
		0.99	0.881	1.233	0.524	0.668	1.343
		0.999	1.410	18.288	1.371	41.339	1.579
	500	0.95	1	0.657	1	0.763	5.565
		0.97	0.857	0.711	0.857	0.777	4.899
		0.99	0.936	1.040	0.581	0.773	2.454
		0.999	0.721	4.213	1	144.041	2.123
	1000	0.95	1	0.668	1	0.691	3.003
		0.97	0.8	0.669	1	0.663	3.123
		0.99	1	0.732	0.75	0.703	3.243
		0.999	0.256	0.649	0.278	214.734	2.898
$\alpha = 0.039$, $\beta = 0.942$	100	0.95	1.067	0.639	0.878	1.234	28.735
		0.97	1.085	0.412	0.876	1.328	25.773
		0.99	1.132	0.163	0.876	1.532	19.558
		0.999	1.015	1136.792	0.890	16.267	7.963
	250	0.95	1.0190	0.985	0.8932	0.929	52.340
		0.97	1.030	1.784	0.880	0.851	48.019
		0.99	1.054	3.732	0.861	0.536	36.995
		0.999	1.046	198.688	0.867	76.171	15.182
	500	0.95	0.991	1.069	1.101	1.440	94.051
		0.97	0.992	0.806	1.014	1.034	83.831
		0.99	1.008	0.598	0.967	1.130	64.171
		0.999	1.084	44.486	0.861	465.255	27.300
	1000	0.95	0.981	1.063	1.162	1.561	167.376
		0.97	0.973	0.960	1.124	1.346	147.621
		0.99	0.957	0.618	1.056	0.935	105.575
		0.999	1.020	0.147	0.923	8786.198	45.317

Table 3.6: Ratios estimated under netting condition with varying p.

Cond.	n	1-p	$\frac{MSE2}{MSE1}$	$\frac{MSE3}{MSE1}$	$\frac{MSE4}{MSE1}$	$\frac{MSE5}{MSE1}$	$\frac{MSE6}{MSE1}$
Netted	100	0.95	0.002	0.798	11.087	2363.455	92.191
		0.97	0.001	0.528	4.870	6.629	40.244
		0.99	0.001	0.333	1.128	21157.49	9.533
		0.999	0.0003	140900.9	8.424	82.171	903.181
	250	0.95	0.002	37.876	2684.69	248.784	23165.62
		0.97	0.001	292.190	2082.328	132.957	17030.18
		0.99	0.001	684.977	1030.044	76.519	8430.355
		0.999	0.0003	21664.07	194.412	125.840	1927.13
	500	0.95	0.002	0.036	200.635	5670.24	1762.368
		0.97	0.001	0.240	74.402	11885.21	620.636
		0.99	0.001	0.871	10.268	8827.538	84.167
		0.999	0.0004	4412.415	357.420	788.548	3135.653
	1000	0.95	0.002	0.088	755.047	7788.671	6756.13
		0.97	0.001	0.102	279.155	16395.45	2440.59
		0.99	0.001	0.021	34.567	2836.48	290.682
		0.999	0.0005	0.365	1.043	1262574	9.257