

Hello, everyone.

This is the outline of this paper. First, the introduction section, which introduces the concept of the marginal expected shortfall. Second the main results section, which contains the estimator for the marginal expected shortfall and its asymptotic behavior. And then a simulation study to confirm the finite sample performance of the estimator. And finally, they conclude this paper with a real data application.

Now, let's move to the introduction part.

Before defining the concept of the marginal expected shortfall, we first review the definition of the expected shortfall, which is a very common concept in finance.

Throughout this paper and representation, we use capital Letter to denote the loss of an asset. And without causing any ambiguity, we also use the capital letter to denote this asset.

The expected shortfall of an asset X at level p is defined as the expectation of X conditional on X exceeds a high threshold: the $1-p$ quantile of X . Recall that, we use X to denote the loss. So, we always discuss the right tail.

Now, Suppose a financial institute holds a portfolio. The loss of the whole portfolio, can be seen as weighted average of the loss of the individual assets. The expected shortfall of the whole portfolio is defined as this one, and it can be decomposed as the weighted average of the expectation of R_i conditional on the loss of whole portfolio R exceeds its $1-p$ quantile. The sensitivity to the i -th asset is, this one. We call this, the marginal expected shortfall.

---Page 5

---Page 6

In this paper, we are interested in MES under exceptional stress conditions of the kind that have occurred very rarely or even not at all. In other words, p is at an extreme low level that can be even lower than $1/n$. So, we model p as p_n . Assume that $p \rightarrow 0$, and np_n can tend to infinity or bound. For the case $np_n \rightarrow \infty$, we say that p_n is at an intermediate level. For the case $np_n = O(1)$, we often say that p_n is at an extreme level.

Note that, when p_n is at extreme level, a non-parametric estimator is not reasonable. And we need to rely on extrapolation.

The goal of this paper is to give an estimator for MES and to establish its asymptotic behavior.

Let's move to the second part.

First, some notations used in this paper. The U function is defined as the quantile function at the point $1 - 1/t$. The U function is very useful in the extreme value theory. As long as the distribution function F belongs to the maximum domain of the Frechet distribution, the U function is a regular varying function. If you are not familiar with this notation, you can treat it as a quantile function, but remember with a different point.

Here, use θ_p to denote the marginal expected shortfall.

Next, I will introduce the main idea on how to estimate θ_p . In non parametric theory, when we want to estimate something, we first derive its theoretical expression and replace the items in the expression by its empirical counterpart. In extreme value theory, the idea is similar.

For the time being, we suppose that X is a positive random variable. X is the loss return, so X can be positive and negative. But MES is mainly determined by high, and hence positive values of X . Therefore, we consider the case X greater than 0 first. And we will consider the case X is real later.

Let's see some simple theoretical derivation.

The first equality is the definition of θ_p . The second equality follows from the definition of conditional expectation. We translate the conditional expectation into integration. The third equality follow from the definition of U function. The probability of Y exceeding $U_2(1/p)$ is equal to p . The fourth equality follows from the Integration By Substitution. We change x to xU_1 .

This derivation is simple and we obtain this theoretical expression for θ_p .

First, note that, the item in the integration is a copula if we take $x = 1$.

Since $F_1(X)$ and $F_2(Y)$ both follow the uniform distribution, we have that, this is a copula.

We impose some conditions on this copula as $p \rightarrow 0$.

Suppose that there exists a function $R(x, y)$ such that, $R(x, y)$ is the limit of this expression. This condition is usual in the multivariate extreme value theory. This condition indicates and specifies dependence in the tail. Since, we are always discussing the tail, we do not need to assume the dependence in the bulk.

This condition has some connection, but still different from the integral part in the theoretical expression for θ_p . Let's see this. in the definition of θ_p , we have x in this place. While in the aforementioned condition, we have x here.

In order to connect these two expression, we impose another condition, that is, U_1 is a regular varying function with index $\gamma_1 > 0$. Based on this condition, we can connect the two expressions. The regular varying condition is the basic condition of one-dimensional extreme value theory. Here $\gamma > 0$ means that X is heavy tailed. Since the application is target at finance, so this assumption is quite normal. Financial data are often heavy-tailed.

Under these conditions, we get the first result, the proposition 1 in the paper. From this, we see that, as $p \rightarrow 0$, θ_p will tend to infinity, as the same rate as $U_1(1/p)$, which is the value at risk for X .

This theoretical expression is the basis for extrapolation in extreme value theory, which we will talk later. Till now, we are still discussing the probability. Now, let's go to statistics and look at how to estimate θ_p .

We do that in two stages:

First, we estimate $\theta_{k/n}$. where k tends to infinity and k/n tends to 0 as n tends to infinity, Here, $p = k/n$, in this case, $p \rightarrow 0$ but $np \rightarrow \infty$, so this is at an intermediate level. Thus, $\theta_{k/n}$ can be estimated non-parametrically.

At second stage, we extrapolate from $\theta_{k/n}$ to θ_p .

We now discuss the first stage. The estimation for $\theta_{k/n}$.

Recall the definition of $\theta_{k/n}$. We replace the quantile $U_2(n/k)$ by the corresponding sample quantile $Y_{n-k,n}$ and replace the expectation by the sample mean. We then get this naive non-parametric estimation.

Under some strengthening of the conditions (relating to R and to the sequences $k = k(n)$), we can show the asymptotic normality of the estimator for $\theta_{k/n}$. Here, I will give a non-formal proof for this theorem.

The background of the limit result is this assumption. The $R(x, y)$ is the tail dependence function.

Now, we define that $V = 1 - F_1(X)$, $W = 1 - F_2(Y)$. Then V and W both follow the uniform distribution. Then joint distribution of V and W is a copula.

Now, consider the i.i.d. r.v.'s: $(V_i, W_i) = (1 - F_1(X_i), 1 - F_2(Y_i))$.

Empirical distribution function of this random vector is defined as this. We consider the left tail of (V_i, W_i) , which corresponds to the right tail for (X_i, Y_i) . We define the tail version of the empirical distribution: $T_n(x, y)$.

Now, $T_n(x, y)$ should be close to its mean, this one, and this is close to $R(x, y)$.

Hence, $T_n(x, y)$ converges in probability to $R(x, y)$. Moreover, The normalized version of $T_n(x, y)$ converges in distribution to a mean zero Gaussian Process. And the following relation indeed specifies the structure of this Gaussian process. This "hence" statement indeed requires a lot of work and has been established in a paper by Einmahl, de Haan and Li in 2006.

Next, I will show how this Gaussian process is related to the asymptotic normality of the estimator for $\theta_{k/n}$.

Let's see some derivation.

The first equation follows by the definition of T_n function.

Second, this approximation follows from the assumption U_1 is a regular varying function with index γ_1 .

Third equality follows from the integration by substitution, change $x^{-\gamma_1}$ to x .

Fourth equality is obvious.

Then note that, we only have x in this term and do not have x relating to the indicator function of y , thus, we have this one,

And then, we calculate the integral and then get the second quality.

For the Third equality, Recall that, $Y_{n-k,n}$ is the empirical version of $U_2(n/k)$.

Hence, we have this approximation. Based on this, we have the asymptotic normality of the estimator for $\theta_{k/n}$. It is a function of W_R . And based on this, we can calculate the asymptotic mean and asymptotic variance.

Although there are many details to handle if you want to give a formal proof.

Till now, we have discussed how to estimate $\theta_{k/n}$ and establish the asymptotic behavior for the estimator.

Now, we discuss the extrapolation from $\theta_{k/n}$ to θ_p . Consider the first (non statistical) result again.

In particular, this holds for $p = k/n$. Compare these two equations, we find that, the right hand side of these two equations are the same. Thus, for sufficiently large n , these two terms in the left hand side are close. This provides us a method to estimate θ_p .

Let's see. θ_p is close to $U_1(1/p)/U_1(n/k)\theta_{k/n}$.

Thus, we can estimate θ_p by this. Estimate the the three part independently.

Here, the estimation for $\theta_{k/n}$ has been discussed before. For the estimation of $U_1(n/k)$, we can use its empirical counterpart, the top k order statistics of X . Now, the only issue is how to estimate $U_1(1/p)$. This now is a one dimensional problem.

Recall that, we assume that the function U is a regular varying function with index γ_1 . Thus, for sufficient large t , we have this approximation, $U_1(tx)$ is close to $x^{\gamma_1} U_1(t)$.

Use this relation with $t = n/k$ and $tx = 1/p$, we get this approximation. This also suggests us the estimator for $U_1(1/p)$. In this relation, we still needs to consider the estimation for γ_1 . Since $\gamma_1 > 0$, we use the Hill estimator. For the Hill estimator, we use k_1 top order statistic of X . k_1 may differ from k , but satisfies similar conditions.

The property of the Hill estimator is well know. It is asymptotically normal with speed $1/\sqrt{k_1}$.

The asymptotic behaviour for $\widehat{U_1(1/p)}$ depends on two parts: $X_{n-k,n}$ and $\hat{\gamma}_1$.

The asymptotic behavior for $X_{n-k,n}$ is also easy to obtain.

Then, we combine the two relations. The first equality follows from the definition of the estimator for $U_1(1/p)$. The second approximation follows from the assumption that U_1 is a regular varying function.

The γ term is in the exponential function, so, it is the dominating term. With some additional condition on p . we have that, the asymptotic behavior of estimator for $U_1(1/p)$ is mainly determined by the asymptotic behavior of the estimator for γ .

Let's see this. The last expression is approximate to this one by the asymptotic normality of $X_{n-k,n}$. Now assume *this one*. This means that p cannot be small. Recall that, we are doing extrapolation from k/n to p . If p is very far from k/n , this extrapolation seems not reasonable. So, we require p not to be too small. With this condition, we have the following expansion. $\exp(x)$ is close to $1+x$, when x is close to 0. By comparing the order, we have that, this term is larger than this, thus, the asymptotic behavior of estimator for $U_1(1/p)$ is only related to the asymptotic behavior for γ_1 . Then, we have the asymptotic behavior for the estimator of $U_1(1/p)$.

The limit distribution is the same as the limit distribution for the Hill estimator.

Now, we give the final results, the asymptotic behavior of the estimator for the marginal expected shortfall.

We display these four technical conditions assumed in this paper.

Recall that, we have assumed that, the numerator tends to 0 as $t \rightarrow \infty$. This condition (a) is a second order strengthening. One remark here: if X and Y are asymptotically independent, then $R \equiv 0$. Then, this condition does not hold. So, the method proposed in this paper cannot handle the case when X and Y are asymptotically independent.

Condition (b) is a second order version of the regular varying condition. Recall that, with the condition that U_1 is regular varying with index γ_1 , then for each fixed $x > 1$, this converges to 0. This conditions requires that, after taking supreme over x , this error is still bounded by $A_1(t)$. Condition (C). Recall that the bias for the Hill estimator is $\sqrt{k_1}A(n/k_1)/(1 - \rho)$. By assuming

Condition (C), the Hill estimator is asymptotically unbiased. Condition (d) is similar, the main role is to keep the asymptotic bias equal to 0.

Let's see the statements of Theorem 1. We also require that $\gamma < 1/2$. Recall that, we are calculating the expectation of a heavy-tailed distribution. To make the variance of the estimator exist, we must require $\gamma < 1/2$. But if you only care about the consistency, then we can relax this condition to $\gamma < 1$.

The case for $X \in \mathcal{R}$ will be handled later.

This condition is imposed on p , as we talked before, p cannot be too small.

With these conditions, we have the asymptotic behavior for $\hat{\theta}_p$.

The limit behaviour for $\hat{\theta}_p$ is a combination of the limit behaviour for $\hat{\gamma}_1$ and $\widehat{\theta_{k/n}}$. Two corner cases are $r = 0$ and $r = \infty$. When $r = 0$, the limit behavior is controlled by the limit behavior of $\hat{\theta}_{k/n}$. When $r = \infty$, the limit behavior is controlled by the limit behavior of $\hat{\gamma}_1$.

Next, we will talk about how to handle the case X is real. We need two extra conditions, First, we require thinner left tail. This means that, compared to right tail, left tail can be ignored. Then we can ignore the left tail. And we need a further bound on p .

With these two extra conditions, we can estimate θ_p by this one. The only difference is here.

This estimator has the same asymptotic behavior as in case X is a positive random variable.

Now, Let's move to the simulation study part.

Data are simulated from three distributions. Throughout the simulation section, let (Z_1, Z_2) denotes a standard bivariate Cauchy distribution.

The first distribution is a transformed Cauchy distribution defined as this one. Then, the extreme value index of X is $2/5$ and the tail dependence function $R(x, y) = x + y - \sqrt{x^2 + y^2}$.

The second distribution is student t_3 distribution with density this. Here, the extreme value index of X_1 is $\gamma_1 = 1/3$. and $R(x, y)$ is displayed in this equation, a little bit complex.

The third distribution is also a transformed Cauchy distribution.

Besides the estimator proposed in this paper, the authors construct two other estimators.

The first is the non parametric estimator. It is just an empirical counterpart of θ_p . Note that, this estimator works only for the case $np \geq 1$. Clearly, this estimator is not expected to perform well when np is small.

The second estimator, is based on the proposition 1, we can estimate $R(x, y)$ and $U_1(1/p)$. The estimation for R function is totally a non-parametric estimator, replace the expression for R by its empirical counterpart. The estimator for $U_1(1/p)$ still requires extrapolation, as we talked before, Thus, $\bar{\theta}_p$ is also an EVT estimator.

The simulation is based on 500 samples with sample sizes $n = 500$ and $n = 2000$.

On the basis of each sample, we estimate θ_p for $p = 1/500, 1/5000$ and $1/10000$.

Let see the results.

This figure is the Boxplot of ratios of estimates and true values for the transformed cauchy distribution 1.

They choose k and k_1 based on the so called 'Hill plot'.

First, we observe that the empirical estimator underestimates the MES. It does not perform well.

Second, the two EVT estimators, $\hat{\theta}_p$ and $\bar{\theta}_p$ perform similar and remain stable when p changes from $1/500$ to $1/10000$.

This is the boxplot of ratios of estimates and true values for the second distribution: the student t_3 distribution.

The two evt estimators both overestimate the MES with the $\hat{\theta}_p$ performing slightly better. The patterns for other plots are similar. Thus, we omit these figures.

They also investigate the performance of the estimators when the assumptions are partially violated.

The first distribution is also a transformed cauchy distribution. but with $\gamma = 0.7 > 1/2$.

The second distribution is an asymptotically independent distribution defined as $X = V_1 + W_1, Y = V_2 + W_2$. Here, V_1, V_2 follows the t_3 distribution, we mentioned before. and W_1, W_2 are pareto distributions. Moreover, $(V_1, V_2), W_1, W_2$ are independent.

For the transformed Cauchy distribution 3, Although $\gamma_1 > 0.5$, the two EVT estimators still work, and $\hat{\theta}_p$ performs better, less biased. For the asymptotically independent model, both EVT estimators overestimates the actual MES. They don't perform well in this case. But $\hat{\theta}_p$ still performs slightly better.

Now, let's move to the application part.

They apply the method to estimate the MES for financial institutions. They consider three large investment banks in USA, the Goldman Sachs, Morgan Stanley and T.Rowe Price.

Then, X refers to each of these three banks, and y refers to the market index.

The dataset consists of the loss returns on their equity prices at a daily frequency from 2000 to 2010. The loss return is the minus log return.

They consider to estimate MES for $p = 1/n$, which corresponds to a once-per-decade systemic event. Then $np = 1$, which is at an extreme level.

This is the Hill plot, we see that γ_1 is about 0.4 for Goldman Sachs and T. Rowe Price, and γ_1 for Morgan Stanley is about 0.45.

This is the estimation for MES.

Interpretation:

When the market faces a daily crisis, the expected loss in log return is 0.308 for Goldman Sachs. The interpretation for others banks are similar.

This is the end of this presentation. Thanks.

