Applied Econometrics 7, Univariate extreme value theory

Евгений Орлов

21.11.2014

В качестве входных данных взяты цены закрытия по обыкновенным акциям Сбербанка в период с 01.08.2009 по 07.11.2014 включительно.

Источником данных выступила ИТС QUIK.

sber.df <- read.csv("Сбербанк [Price].txt", header=TRUE)  
tail(sber.df)

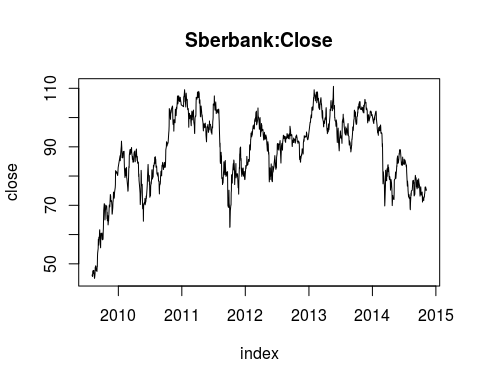
## X.TICKER. X.PER. X.DATE. X.TIME. X.OPEN. X.HIGH. X.LOW. X.CLOSE.  
## 2997 SBER [TQBR] Daily 20141030 0 73.70 75.13 73.43 74.98  
## 2998 SBER [TQBR] Daily 20141031 0 75.45 76.84 75.23 76.23  
## 2999 SBER [TQBR] Daily 20141103 0 76.07 76.66 75.61 76.29  
## 3000 SBER [TQBR] Daily 20141105 0 75.80 75.94 75.02 75.65  
## 3001 SBER [TQBR] Daily 20141106 0 75.76 76.88 74.75 75.05  
## 3002 SBER [TQBR] Daily 20141107 0 75.16 77.65 74.41 75.53  
## X.VOL.  
## 2997 18099518  
## 2998 15467572  
## 2999 4989221  
## 3000 10108565  
## 3001 10972235  
## 3002 17931891

# Цены закрытия начиная с 01.08.2009 по 07.11.2014 включительно  
sber.close <- sber.df[sber.df$X.DATE. >= 20090801, "X.CLOSE."]  
# Длина выборки  
print(paste0("Размер выборки: ", length(sber.close)))

## [1] "Размер выборки: 1319"

График цен закрытия:

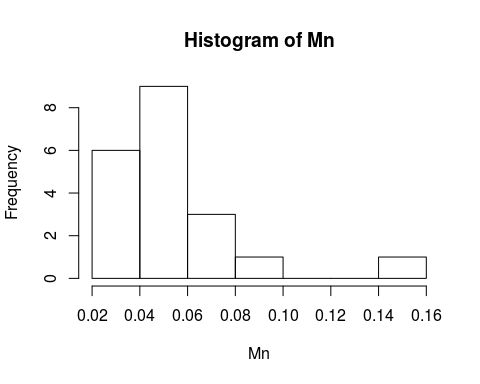
# График цен закрытия  
plot(sber.dates, sber.close,   
 type='l', xlab='index', ylab='close', main='Sberbank:Close')



### Метод блочных максим

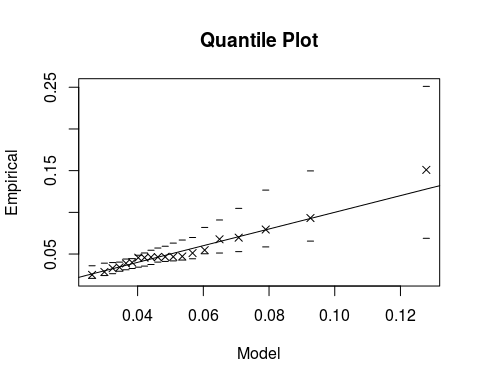
Вместо вектора доходностей нас интересует вектор убытков. Длина вектора убытков выбрана таким образом, чтобы по нему можно было построить выборку из 20 максим, каждая из которых имеет длину 65 (примерное кол-во рабочих дней в квартале).

n <- 65 # кол-во дней в максиме (~ кол-во рабочих дней в квартале)  
m <- 20 # кол-во максим  
size <- n\*m  
sber.loss <- sber.close[2:(size+1)] / sber.close[1:size] - 1  
sber.loss <- -sber.loss # Вектор убытков  
# Расчет максим  
Mn <- rep(0, times=m)  
for (i in 1:m) {  
 Mn[i] <- max(sber.loss[((i-1)\*n + 1):(i\*n)])  
}  
hist(Mn)

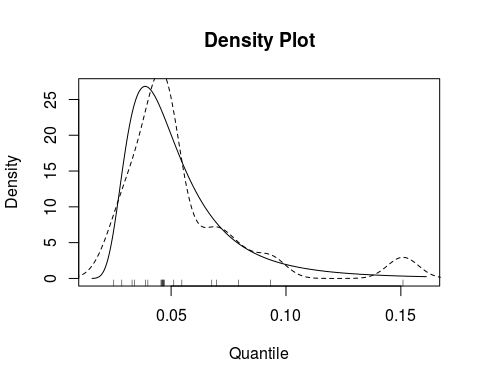


На основе полученной выборки из 20 максим произведен подгон параметров GEV-распределения.

# Распределение максим на основе GEV  
library(evd, quietly=TRUE)  
Mn.fit <- fgev(Mn)  
plot(Mn.fit, which=2) # график "квантиль-квантиль"



plot(Mn.fit, which=3) # график эмпирической плотности



На графиках видно, что в целом полученное распределение хорошо описывает имеющуюся выборку максим (за исключением самого большого выброса, но это связано с недостаточным кол-вом наблюдений в правом хвосте).

На основе полученных параметров распределения, рассчитаем пороговый уровень и средний период наступления события.

mu <- Mn.fit$estimate[1]  
sigma <- Mn.fit$estimate[2]  
xi <- Mn.fit$estimate[3]  
k <- 4  
u <- 0.09  
r.nk <- mu + sigma/xi\*((-log(1 - 1/k))^(-xi) - 1)  
k.nr <- 1 / (1 - pgev(u, loc=mu, scale=sigma, shape=xi))

Согласно вычислениям

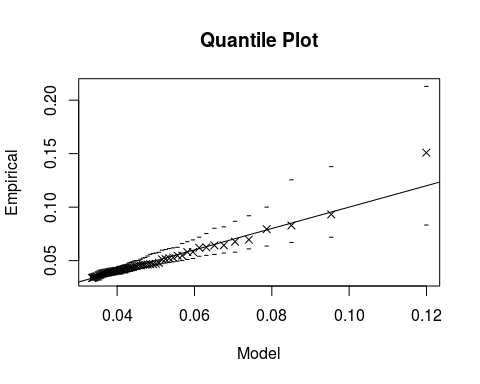
1. Уровень потерь, который будет пройден в среднем 1 раз в год (1 раз в 4 квартала) составляет 6.2544%;
2. Средний период (в кварталах) наступления убытка, правышающего 9% составляет 12.14.

### Расчет VaR, ES с использованием распределения Парето

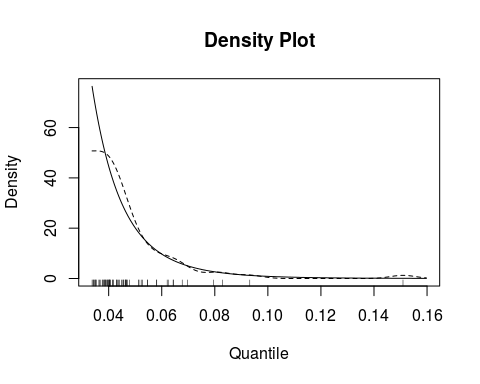
Для расчета VaR и Expected Shortfall по всей выборке в качестве порогового значения возьмём 95-й персентиль.

VaR найдем как соответствующий квантиль распределения превышений порога (в предположении, что оно моделируется распределением Парето - параметр ).

# Пороговое значение  
u <- sort(sber.loss)[0.95\*length(sber.loss)]  
# Подбор параметров  
gpd.fit <- fpot(sber.loss, threshold=u, model='gpd', method='SANN')  
plot(gpd.fit, which=2)



plot(gpd.fit, which=3)



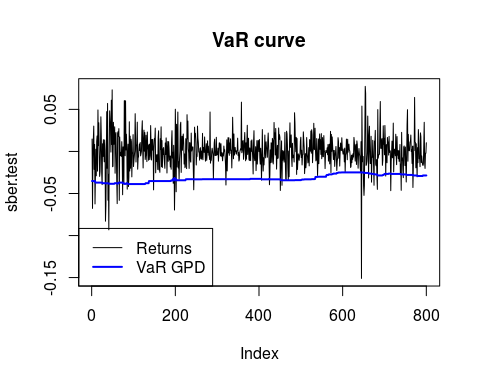
# Оценки параметров  
beta <- gpd.fit$estimate[1]  
xi <- gpd.fit$estimate[2]  
Fu <- gpd.fit$pat  
alpha2 <- 1 - 1/260 # соответствует 1 превышению в год  
VaR <- u + beta/xi\*(((1-alpha2)/Fu)^(-xi) - 1) # раз в год  
ES <- (VaR + beta - xi\*u) / (1 - xi) # раз в год

VaR выбран таким образом, чтобы в течение года ожидается только один день, когда убыток превысит значение VaR. Это позволяет сравнить полученное значение с уже вычисленным уровнем потерь. Эти значения оказались достаточно близки друг к другу (6.25% и 7.22%).

### Кривая VaR с использованием распределения Парето

Кривая VaR строилась для 95%-го VaR с использованием последних 500 известных значений потерь. В качестве порогового значения использовался 95%-й персентиль. VaR расчитывался как соотвествующий квантиль распределения превышений порога. В качестве распределения превышений порога в зависимости от оценки параметров выступало либо распределение Парето (), либо экспоненциальное распределение ()).

T1 <- 500  
T2 <- length(sber.loss)-T1  
alpha0 <- 0.95 # Персентиль порогового значения  
alpha <- 0.95 # VaR  
VaR.gpd <- numeric()  
  
h <- T1 # Длина обучающей выборки  
for (i in (T1+1):(T1+T2)) {  
 sber.train <- sber.loss[(i-h):(i-1)]  
 u <- sort(sber.train)[(alpha0)\*T1]  
 gpd.fit <- fpot(sber.train, threshold=u, model='gpd', method='SANN',   
 std.err=FALSE)   
 beta <- gpd.fit$estimate[1]  
 xi <- gpd.fit$estimate[2]  
 Fu <- gpd.fit$pat  
 if (xi != 0) {  
 VaR.gpd[i-T1] <- -(u + beta/xi\*(((1-alpha)/Fu)^(-xi) - 1))  
 } else {  
 VaR.gpd[i-T1] <- -(u - beta\*log((1-alpha)/Fu))  
 }  
}  
# График кривой VaR  
sber.test <- -sber.loss[(T1+1):(T1+T2)]  
plot(sber.test, type="l", main="VaR curve")  
lines(VaR.gpd, col="blue", lwd=2)  
legend('bottomleft', c("Returns", "VaR GPD"), col=c("black", "blue"),   
 lty=c(1, 1), lwd=c(1, 2))



Для верификации кривой VaR определим функцию для теста Купика.

# Частота пробоев  
kupiec.test <- function(ret, VaR, alpha) {  
 # Тест Купика:  
 # H0: модельная и эмпирическая частоты пробоя VaR совпадают  
 K <- sum(ret < VaR)  
 T2 <- length(ret)  
 alpha0 <- K / T2  
 S <- -2\*log((1-alpha)^(T2-K) \* alpha^K) + 2\*log((1-alpha0)^(T2-K) \* alpha0^K)  
 p.value <- 1-pchisq(S, df=1)  
 return(c(alpha0, p.value))  
}

## [1] "Kupiec test, alpha = 0.05"

## [1] "GPD: alpha0 = 0.05125, p-value = 0.871631"

p-value теста Купика равно 0.87, что значит гипотеза о том, что фактическое кол-во пробоев совпадает с целевым не отвергается, и полученная кривая VaR с поставленной задачей справляется.