Time-Series Analysis and Forecasting of the S&P 500 Index

Please download the file "`sp500_full.csv`". It provides the data on daily prices and returns (among other things) of the S&P 500 index from 1957-01-02 to 2021-12-31. The specific function implementations are hidden as the project mainly focuses on the overall methods.

Extreme Value Modeling and Risk Estimation

The goal is to apply threshold selection, GPD modeling and risk estimation to S&P 500 Index. Use the SP500 time series for the period 1962/07/03 through 2021/12/31.

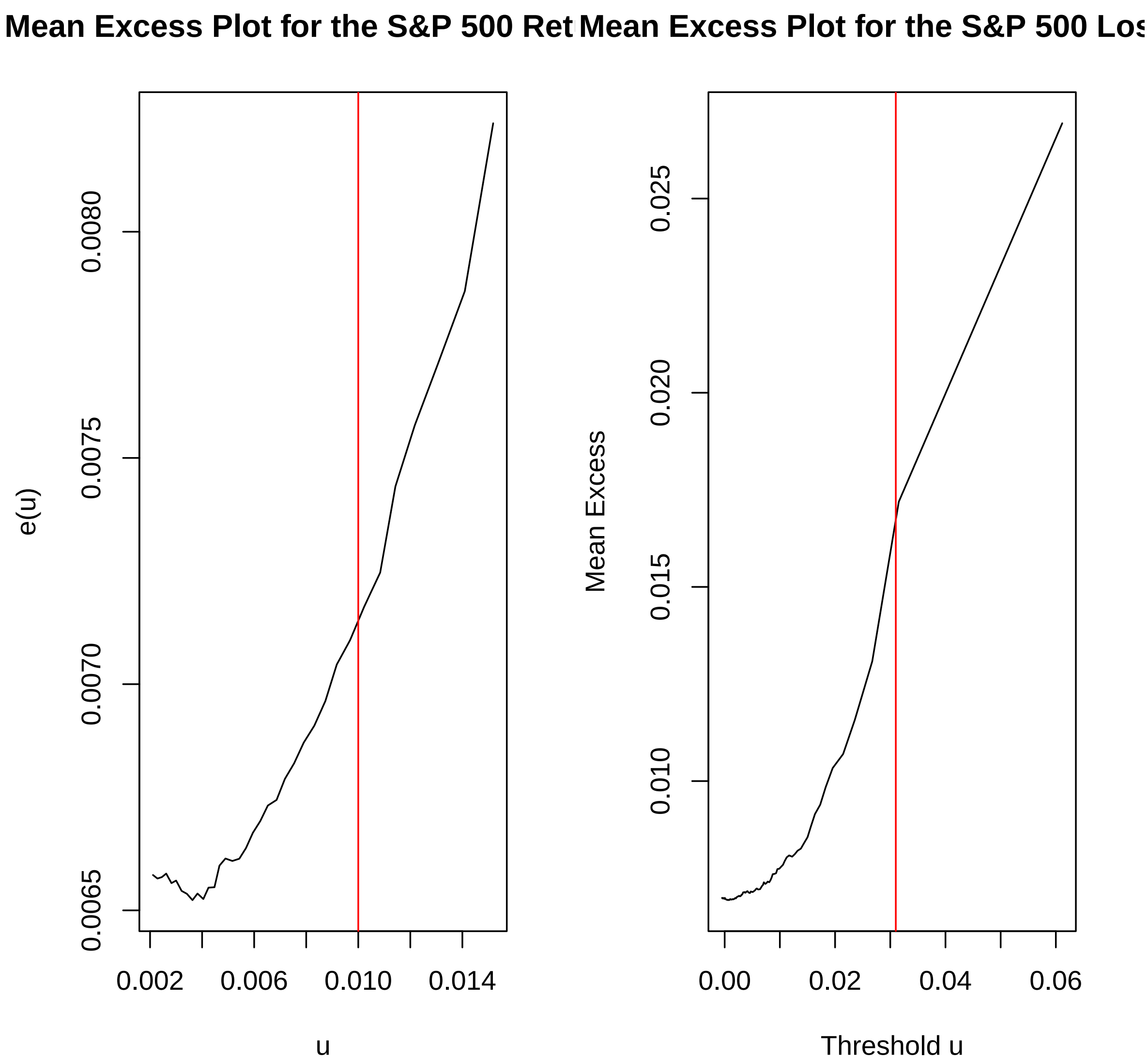
1. Load the sp500 time series and focus on the returns and losses(negative daily returns). Examine the mean-excess plots to determine a suitable threshold u_0, above which the excesses are likely to follow a GPD model.

```
sp = read.csv("sp500_full.csv", header = T)
sp = sp[sp$caldt >= "19620703" & sp$caldt <= "20211231",]

par(mfrow=c(1,2))

# 1. mean excess plot for S&P500 returns
idx = which(!is.na(sp$sprtrn))   # remove null values
return = sp[idx,]
u = quantile(return$sprtrn,p=c(60:95)/100)
m.e = mean.excess(x=sp$sprtrn,u=u)
plot(u,m.e,type="l",xlab = "u", ylab = "e(u)",
main="Mean Excess Plot for the S&P 500 Returns")
abline(v=0.01,col="red")

# 2. mean excess plot for S&P500 losses
pu = seq(from=0.5,to=0.999,length.out=100);
losses = -sp$sprtrn[!is.na(sp$sprtrn)]
u = quantile(losses, pu) # calculate the p-th quantile of the losses
me = mean.excess(losses, u);
plot(u,me,type="l",xlab="Threshold u", ylab="Mean Excess",
     main=paste("Mean Excess Plot for the S&P 500 Losses"));
abline(v=0.031,col="red")
```



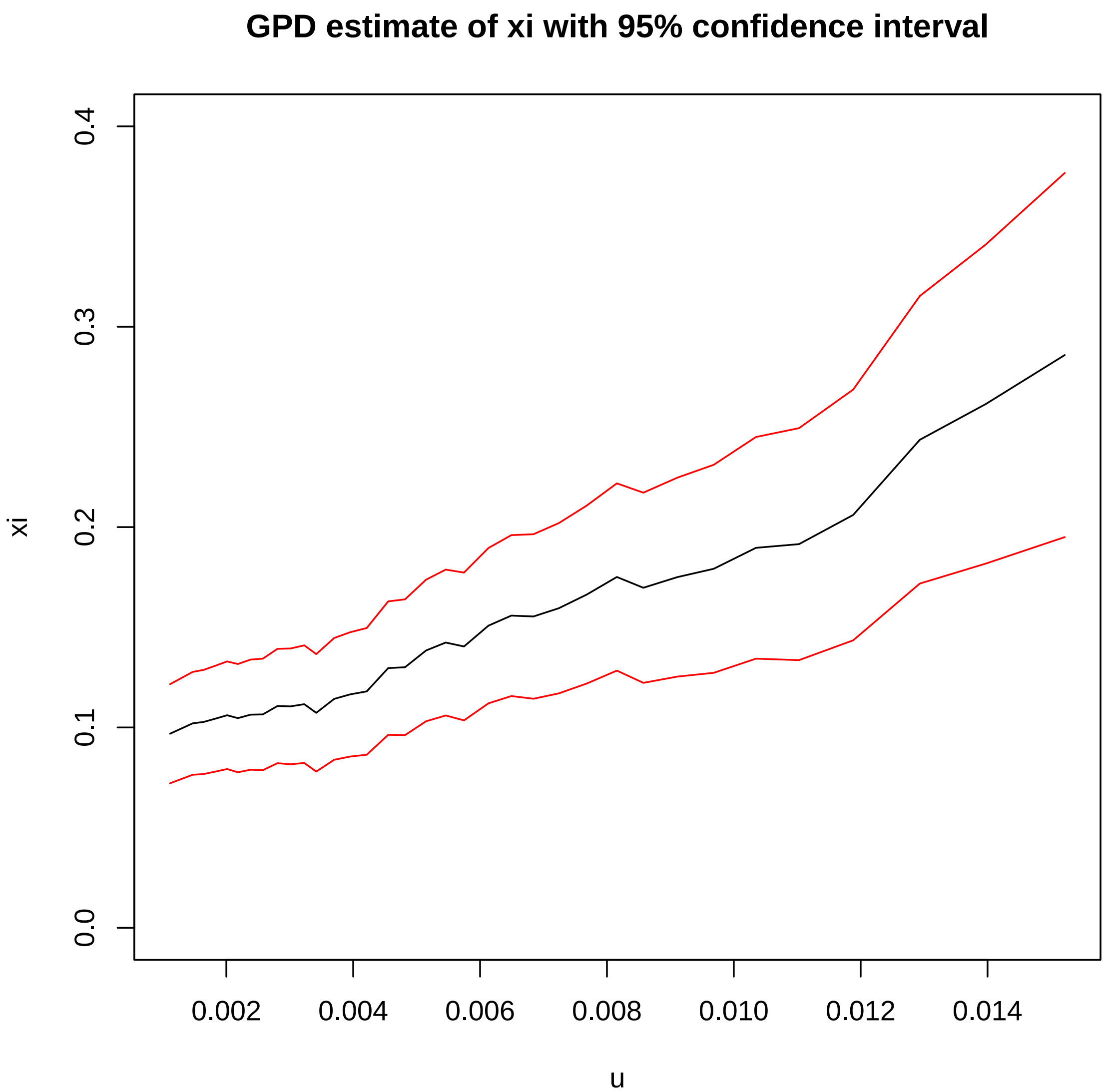
Based on the plot, the best threshold for returns is around 0.1 and that for losses is around 0.031.

2. Fit the GPD model and produce plots of the point estimates and 95% confidence intervals for ξ over a range of thresholds.

```
u = quantile(losses,p=c(60:95)/100)
ci = matrix(0,2,length(u))
xi_line= rep(0,length(u))
for (i in 1:length(u)){
  res=fit.gpd.Newton_Raphson(losses,threshold=u[i]);
  xi_line[i] = res$xi
  ci[1,i]=res$xi-1.96*res$se.xi;
```

```
  ci[2,i]=res$xi+1.96*res$se.xi;
}

plot(u,xi_line,type="l",xlab="u",ylab="xi", ylim =c(0,0.4),
main="GPD estimate of xi with 95% confidence interval")
lines(u,ci[1,],type="l",col="red")
lines(u,ci[2,],type="l",col="red")
```



3. Produce parametric-bootstrap confidence intervals for VaR_α and ES_α for $\alpha = 1/252, 1/(5 * 252), 1/(10 * 252)$, which are levels of risk corresponding to return periods of 1−, 5− and 10-years.

```r
get_VaR_and_ES <- function(data, p=0.95, alpha){
  u = quantile(data,p);
  names(u)="";
  fit <- fit.gpd.Newton_Raphson(data,threshold = u);
  xi.hat <- fit$xi
  sig.hat <- fit$sig
  Cov = fit$Cov[1,,]
  VaR <- ((alpha/mean(data>=u))^(-xi.hat) -1)*sig.hat/xi.hat+u
  ES <- (sig.hat+VaR-xi.hat*u)/(1-xi.hat)
  return(list("VaR"=VaR,"ES"=ES,"alpha"=alpha,
  "Cov"=Cov,"xi"=xi.hat,
  "sig"=sig.hat))
  }

# MC.iter: number of Monte Carlo iterations
# qfactor: the quantile of the normal distribution for Monte Carlo simulation
get_par_boostrap_ci_VaR_and_ES <- function(x,MC.iter=2000,p=0.95,alpha,
p.lower=0.025,p.upper=0.975){
  res <- get_VaR_and_ES(data=x, p=p, alpha)
  u = quantile(x,p); names(u)="";
  pu = mean(x>=u);
  theta.star =
    matrix(c(res$xi,res$sig),2,1)%*%matrix(1,1,MC.iter) +
    mat.power(res$Cov,1/2)%*% matrix(rnorm(2*MC.iter),2,MC.iter);

  xi.star = theta.star[1,]
  sig.star = theta.star[2,]
  #dropping negative sigma.star variates's
  xi.star=xi.star[sig.star>0]
  sig.star=sig.star[sig.star>0]
  ci.VaR = c();
  ci.ES = c();
  for (a in alpha) {
    VaR.a = ((a/pu)^(-xi.star) -1)*sig.star/xi.star+u
    ES.a = (sig.star+VaR.a-xi.star*u)/(1-xi.star)
    ci.VaR = cbind(ci.VaR, quantile(VaR.a, c(p.lower,p.upper)));
    ci.ES = cbind(ci.ES, quantile(ES.a, c(p.lower,p.upper)));
  }

  rownames(ci.VaR)<- c("Lower(VaR)","Upper(VaR)");
  colnames(ci.VaR) <- alpha;
  rownames(ci.ES)<- c("Lower(ES)","Upper(ES)");
  colnames(ci.ES) <- alpha;
  return(rbind(ci.VaR,ci.ES))
}
```

```r
a = c(1/252,1/(5*252),1/(10*252))
cis = get_par_boostrap_ci_VaR_and_ES(losses, alpha=a)
kable(cis, col.names=c('1/250', '1/(5*252)', '1/(10*252)'))
```

	1/250	1/(5*252)	1/(10*252)
Lower(VaR)	0.0357245	0.0562861	0.0673040
Upper(VaR)	0.0405513	0.0735504	0.0953821
Lower(ES)	0.0494042	0.0747396	0.0882853
Upper(ES)	0.0642322	0.1174230	0.1518499

Discuss the effect of the threshold used in the GPD inference on the parametric-bootstrap intervals. Interpret these intervals, i.e., Is it reasonable to expect that over a period of 10 years the SP500 index will see a daily drop of around 7 percent?

- A higher threshold leads to a smaller sample size and wider confidence intervals, while a lower threshold leads to a larger sample size and narrower confidence intervals.
- From the VaR_α for $\alpha = 1/(10 * 252)$, it can be seen that the value 7% falls within the confidence interval. Therefore, it's likely to expect that over a period of 10 years the SP500 index will see a daily drop of around 7 percent.