Lecture 5: Semi-parametric estimation of VaR/ES

Brian Thelen
258 West Hall
bjthelen@umich.edu

Statistics 509 - Winter 2022
Ref: Ruppert/Matteson: Chapter 19.6 – but we are doing a slightly different/simpler framework

Pickands-Balkema-de Haan Theorem I

Adapted from Theorem 3.4.5 in **Embrechts, Kluppelberg and Mikosch**.

Theorem (Pickands-Balkema-de Haan)

Under mild conditions on the distribution of rv X, we have

$$\frac{\mathbb{P}(X > u + x a(u))}{\mathbb{P}(X > u)} \longrightarrow H_\xi(x), \quad \text{as } u \uparrow \infty.$$

for some $a(u) > 0$, where

任何分布的条件概率都服从广义帕利托分布

$$H_\xi(x) = \left(1 + \xi x\right)^{-1/\xi}, \quad \text{for } 1 + \xi x > 0.$$

Interpretation: 没有scale factor因为已经被包含在a(u)里了

Pickands-Balkema-de Haan Theorem II

- Suppose $x > 0$ and $\xi > 0$. Then, as $u \uparrow \infty$,

$$\frac{\mathbb{P}(X > u + x a(u))}{\mathbb{P}(X > u)} \to \frac{1}{(1+\xi x)^{1/\xi}} .$$

- But the left-hand-side equals

$$\mathbb{P}\left(\frac{X-u}{a(u)} > x \mid X > u\right) . \approx \frac{1}{(1+\xi x)^{1/\xi}}$$

- This means that, for large u, the conditional distribution of the excess loss $X - u$, given that the loss is greater than u is

$X \sim GP$ with (u, ξ) scale $= a(u) = \sigma$

附加损耗的xi没有变,但是sigma变了

Estimating the Shape Index

estimate的tail和kernel一致,因此不好用

- Histograms and kernel density estimators can be good estimators in the center of a distribution where most of the data is to be found, but they are rather poor estimators of the tails.
- Peak over threshold (PoT) methods - invoking Theorem, utilize generalized Pareto model. Need to estimate based on "tail" data
  - Pareto: Linear regression
  - GPD: Maximum likelihood estimation
- Issue of selecting the threshold.

通过只考虑大于threshold的分布,来估计tail

tail

Likelihood Function

- Probability models usually depend on unknown parameters θ (here θ can be a vector) – then the joint PDF of iid sample $x_1, \ldots, x_n$ can be written as

$$f_n(x_1, x_2, \ldots, x_n; \theta) = \prod_{1}^{n} f_x(x_i; \theta)$$

- We can view $L(\theta) = f(x_1, \ldots, x_n; \theta)$ as a function of θ with $x_1, \ldots, x_n$ fixed at the observed data, and we call it the likelihood function. It tells us the likelihood of the sample that was actually observed.

Maximum Likelihood Estimation

- Definition: The maximum-likelihood estimates (MLE) are the parameter values that maximize the likelihood function, or equivalently the values that maximize the log-likelihood function.
- The log-likelihood function is the (natural) logarithm of the above, i.e.,

$$\ell(\theta) = \log\left[L(\theta)\right]$$

Steps for Finding MLE

$$l(\mu,\xi,\sigma)=\sum_{i}^{n}\log\left[\frac{1}{\sigma}\left(1+\frac{\xi(x-\mu_i)}{\sigma}\right)^{-1-1/\xi}\right]$$

$$=-n\log\sigma-\left(1+\frac{1}{\sigma}\right)\sum_{i}^{n}\log\left[1+\frac{\xi(x-\mu_i)}{\sigma}\right]$$

$$\hat{\theta}_{MLE} = \arg\max_{\theta} L(\theta) \quad \text{or} \quad \hat{\theta}_{MLE} = \arg\max_{\theta} L(\theta)$$

1. Derive the likelihood or log-likelihood function.
2. Take the derivative with respect to each of the parameters and set the derivatives equal to 0.
3. Solve for unknown parameters – these are the MLEs,

Optimization Procedure GPD没有解析解

Peak over Threshold - Tail Fitting

$P[X \geq x] = \underbrace{P[X \geq x \mid X \geq u]}_{\text{Pareto}} \, P[X \geq u]$

VaR Quantile

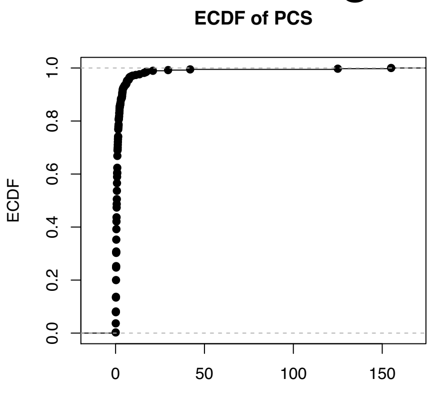
Steps

- Compute empirical cdf and look at the range of quantiles for the tail (.30 to .05 or less) – these are candidate thresholds
- Compute MLE of shape parameter (in the tail) using GPD model over range of threshold values and look at stability
    - Only estimating scale and shape, as the threshold is the location ξ σ threshold就是u,已知了
- Pick threshold based on quantile considerations (not too far out in the tail), but also in the stable region relative to estimation, and compute MLEs of Generalized Pareto parameters
- Look at Goodness of Fit of the tail distribution relative to the estimated model – QQ plots of tails 不断迭代GPD

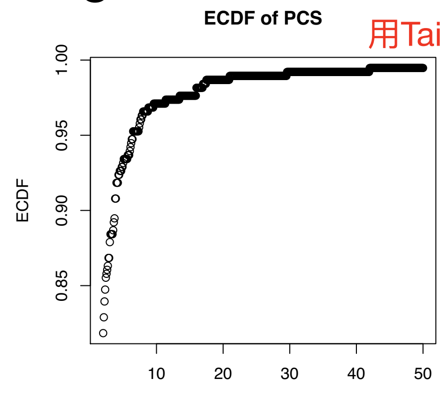
$P[X \leq x \mid X \geq u'] \; (u' > u) \longrightarrow GPD(u', \xi, \xi(u'-u))$

PCS Index: POT Analysis

- Use R evir package for shape-plots/tail plots in POT tail distribution modeling - again using GPD model for tails

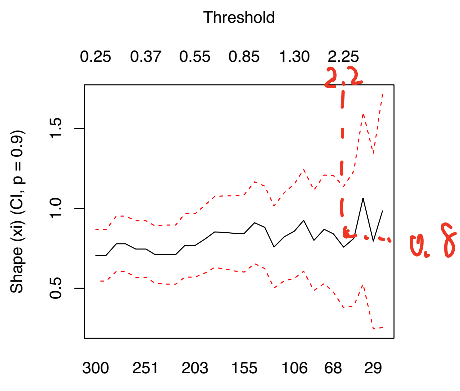


Empirical CDF

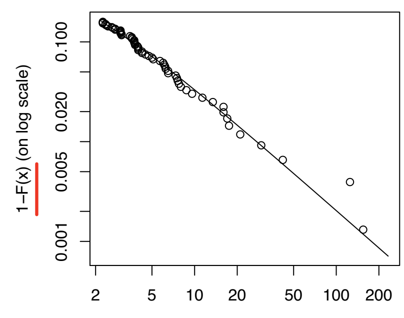


用Tail计算GPD

Empirical CDF - zoom



Shape Plots



Tail plot on log scale

Shape est = .80, thresh = 2.2

R-commands for previous slide

```
> library(evir)
> load('PCS.rda')
> eecdf = ecdf(PCS[,2])
> length(PCS[,2])
[1] 381
```

只取了大于2的部分,因为只考虑tail

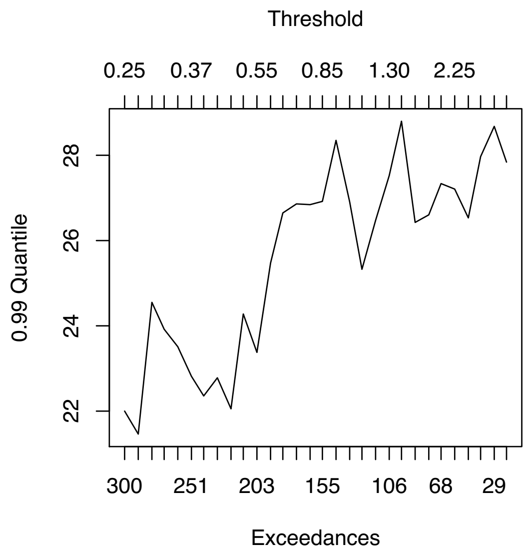
```
> plot(eecdf,main='ECDF of PCS',xlab='Claims',ylab='ECDF')
> uv = seq(from = 2,to = 50, by = .1)
> plot(uv,eecdf(uv),main='ECDF of PCS',xlab='Claims',ylab='ECDF')
>
> shape_out = shape(PCS[,2], models = 30, start = 300,
+ end = 20, ci=.90,reverse = TRUE,auto.scale = TRUE)
> gpd_out = gpd(PCS[,2], threshold = 2.2)
```

设定threshold估计tail Pareto分布的参数

```
> length(PCS[PCS[,2]>2.2,2])
[1] 61
> # Estimates of xi and beta
> gpd_out$par.ests
       xi       beta
0.7989588 2.4645378
> tail_out = tailplot(gpd_out)
```

Estimation of .99 quantile of Claims vs. threshold

- Estimate of .99 quantile for Claims is
- Below is plot of



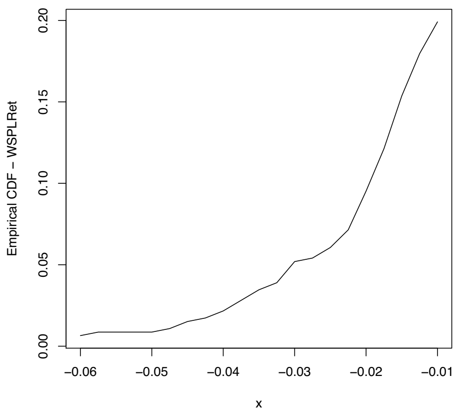
使用ECDF估计的分位数

```
# Rcode
> gpd.q(tail_out,pp=.99)
Lower CI Estimate Upper CI
16.64798 27.39541 66.92750
> quant(PCS[,2], p=.990,models = 30, start = 300,
+ end = 20,reverse = TRUE,ci=FALSE,auto.scale = TRUE)
```

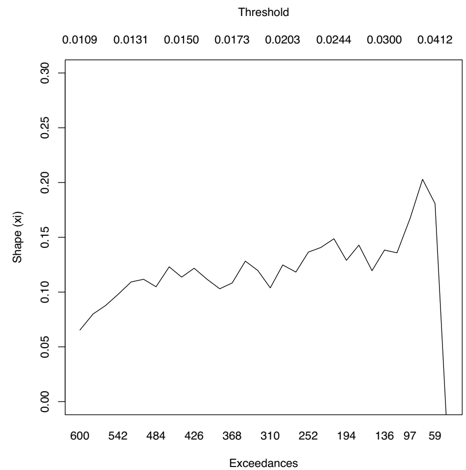
SP500 Weekly Log Returns

To make precise estimate would need to connect result to VaR/returns

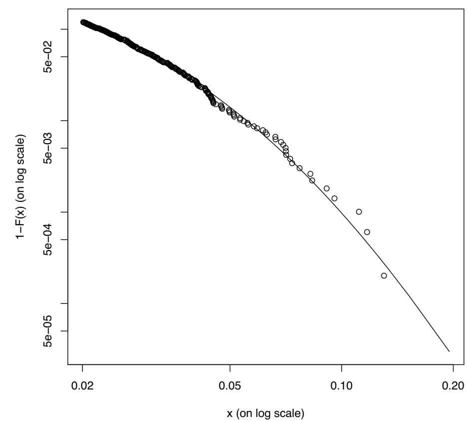
- Results for 1960-2007 data



Emp. CDF of log-returns



Shape Plot



Tail Plot for $\boxed{u = .020}$

R-Commands

```
> WSPRet = SP500wk_return
> library(POT)
> eecdf = ecdf(WSPRet)
> uv = seq(from = -.06,to = -.01, by = .0025)

> plot(uv,eecdf(uv),type='l',xlab='x',ylab='Empirical CDF - WSPLRet')
>
> length(WSPRet)
[1] 2480
> shape(-WSPRet, models = 30, start = 40, end = 600, reverse = TRUE,
+ ci=FALSE,auto.scale='FALSE', ylim=c(0,.3),labels = TRUE)

> gpd_est = gpd(-WSPRet, thresh=.02, method = c("ml"),
+ information = c("observed"))
> gpd_est$par.ests
       xi       beta
0.1210946 0.0122520
> tp = tailplot(gpd_est)
```

Computing VaR: Two Approaches

assume, stationary, independent

Problem: Have historical returns $R_1, \ldots, R_n$ (assuming stationary) with cdf F_R which is unknown and want to estimate VaR for a specified q (e.g., .01, .005) – this corresponds to estimating the q-quantile of the F, and taking the negative. $(1-q)$ Quantile
Alternatively, lets consider the negative returns and denote as $X_1, X_2, \ldots, X_n$ and now want the $(1-q)^{th}$ quantile.

Two approaches: Nonparametric and Semiparametric

- Nonparametric approach – simply use sample quantile of $\hat{\pi}_q = \hat{F}_n^{-1}(1-q)$ from the data: $\tilde{\text{VaR}} = \hat{F}_n^{-1}(1-q)$
  - OK if q is not too small relative to sample size n

如果q太小就是估计tail,
用KDF估计tail效果不好

- Semiparametric approach with threshold u: note that

$$P(X \geq x) = P(X \geq x | X \geq u) P(X \geq u)$$

  - Utilize POT to estimate the parametric tail probability model $F_\theta(x)$ for $P_\theta(X \leq x | X \geq u)$ (e.g., GPD), i.e.,
  - Utilize nonparametric estimate of $P(X \geq u)$ by $1-\hat{F}_n(u)$
  - Estimate quantile from above via

用常规方法估计threshold,用POT估
计threshold后的帕累托条件分布

$$\tilde{\text{VaR}} = F^{-1}_{u,\xi,\sigma}\left[1-\frac{q}{1-\hat{F}_n(u)}\right]$$

Computing VaR for SP500 Weekly Returns

Example. Based on results in previous slides on fitting GPD to the tails and output below, want to derive VaR for $q = .005$ and current investment value of a million dollars.

```
> eecdf = ecdf(-WSPRet)
> eecdf(.02)
0.112
```

Answer. .

R-code for Answer

```
> eecdf = ecdf(-WSPRet)
> qt = 1-.005/(1-eecdf(.02))
> scale = gpd_est$par.ests[2]
> xi = gpd_est$par.ests[1]
> xi
       xi
0.1210946
> m = .02
> # Note that this qgpd is from evir and inputs has different ordering
> # than qgpd in fExtremes
> VaRt = qgpd(qt,xi,m,scale)
> VaRt
0.06745486
> VaR = 1000000*VaRt
> VaR
67454.86
```

Question: What is the the shortfall distribution and how would you compute the expected shortfall?

Answer.

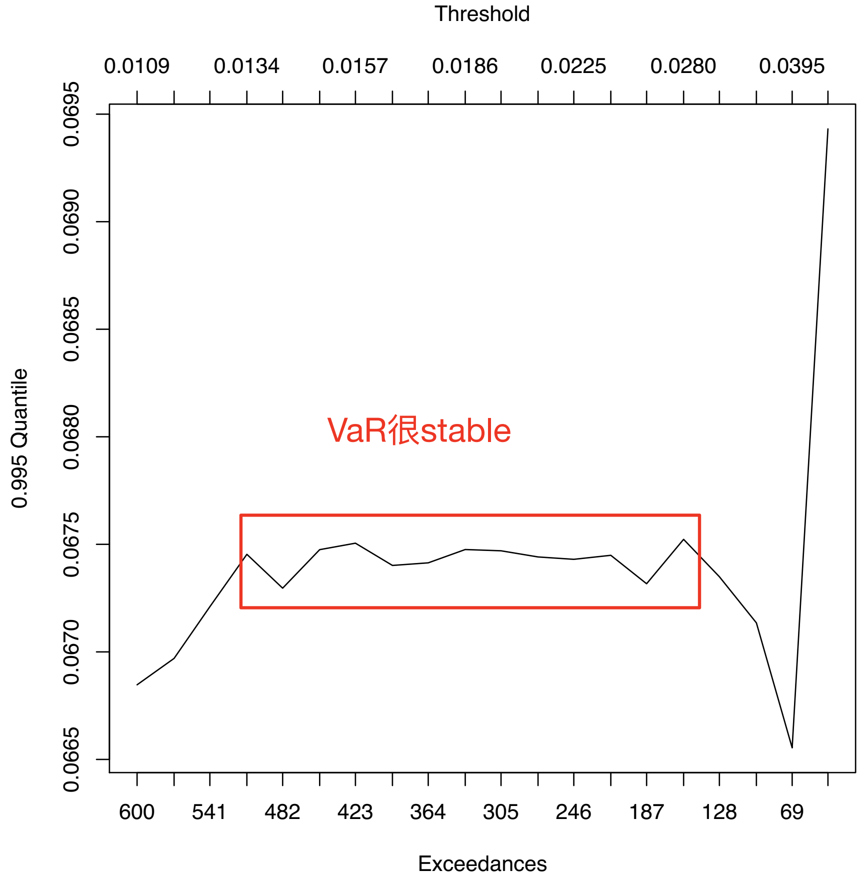
$$P[X \le x \mid X \ge u) \sim GPD(u, \hat{\xi}, \hat{\sigma})$$

$$P[X \le x \mid X \ge \widetilde{VaR}) \sim GPD\left(\widetilde{VaR}, \hat{\xi}, \underbrace{\hat{\sigma} + \hat{\xi}(\widetilde{VaR} - u)}_{\hat{\sigma}'}\right)$$

Expected value of distribution

$$\widetilde{VaR}_q + \frac{\hat{\sigma}'}{1-\hat{\xi}} = \widetilde{VaR}_q + \frac{\hat{\sigma} + \hat{\xi}(\widetilde{VaR} - u)}{1-\hat{\xi}}$$

Relative VaR vs. threshold




```
a=quant(-WSPRet, p = 0.995, models = 20, start = 600, end = 40,
+ reverse =TRUE, ci = FALSE, auto.scale = TRUE, labels = TRUE)
```

Working with Log-Returns

Remark Suppose you have historical log-returns $\tilde{R}_1, \ldots, \tilde{R}_n$ (assuming stationary) with cdf F_R which is unknown and want to estimate VaR for a specified q (e.g., .01, .005).

Important Remark. You cannot just consider the negative log-returns and proceed as usual. Why? What is the approach for doing POT approach for negative log-returns?

Answer.

1. $\tilde{X}_1 \sim \tilde{X}_n$, $\tilde{X}_n = -\tilde{R}_n$ X是negative return (亏损)

   Use POT on $\tilde{X}_{1-q}$, 1-q 是 quantile on negative log ret

2. $\tilde{X}_{1-q} = -\tilde{r}_q$

   $-\tilde{r}_q = -(e^{-\tilde{x}_{1-q}} - 1)$ $= r_q$

Confidence Intervals (CI's)

- **Normal approximation** Normal distribution approximation to MLE-based quantile estimate
- **Nonparametric Bootstrap:** Subsample with replacement from data (past the threshold u), B times, and generate estimates $\{\widetilde{\text{VaR}}_q^{(b)}\}_{b=1}^{B}$ and $\{\widetilde{\text{ES}}_q^{(b)}\}_{b=1}^{B}$. The $100(1-\alpha)\%$ bootstrap CI's are

  $$\left[2\widetilde{\text{VaR}}_q - \pi^V_{1-\alpha/2},\ 2\widetilde{\text{VaR}}_q + \pi^V_{\alpha/2}\right]$$

  $$\left[2\widetilde{\text{ES}}_q - \pi^E_{1-\alpha/2},\ 2\widetilde{\text{ES}}_q + \pi^E_{\alpha/2}\right]$$

  with $\pi^V_{\alpha/2}, \pi^V_{1-\alpha/2}$ being quantiles of $\{\widetilde{\text{VaR}}_q^{(b)}\}_{b=1}^{B}$ and $\pi^E_{\alpha/2}, \pi^E_{1-\alpha/2}$ being quantiles of $\{\widetilde{\text{ES}}_q^{(b)}\}_{b=1}^{B}$.
- **Parametric Bootstrap:** Subsample at random from estimated parametric model, B times, and generate estimates $\{\widetilde{\text{VaR}}_q^{(b)}\}_{b=1}^{B}$ and $\{\widetilde{\text{ES}}_q^{(b)}\}_{b=1}^{B}$, and compute CI's as above

Remark and Assumptions/Issues

Remark. Confidence intervals on $\tilde{\text{VaR}}_q$ and $\tilde{\text{ES}}_q$ are

Disappointing Wide

太wide了可能没有意义

Question. What are key fundamental assumptions (and possible issues) on the returns for the utilizing tail analysis as proposed?

Answer.

Return must be i.i.d for MLE but it's clearly violated to same degree

stationary still work
non-stationary not so much