Robust Estimation of the Parameters of the GPD

A Case Study

René Stephan

Klemmstein & Stephan GmbH

ASTIN Colloquium, Lisboa 2016

Agenda

Introduction

- Models are approximations to reality.
- Classical parametric statistics do not provide information on behavior of procedures if model assumptions are just approximately valid.
- Robust statistics describes the behavior of statistical procedures in the neigborhood of strict model assumptions and introduce procedures that are more resilient to deviations from the ideal setup.

Accuracy of Stellar Movements

1914 Eddington proposes mean absolute deviations.

$$\sigma_1 = \frac{1}{n}\sqrt{\frac{\pi}{2}}\sum_{i=1}^{n} | x_i - \bar{x} |, \qquad \text{s.e. } \sigma_1 = \frac{\sigma}{\sqrt{2n}}\sqrt{\pi - 2}$$

1920 Fisher comments that the mean square error is 12% more efficient.

$$\sigma_2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2, \qquad \text{s.e. } \sigma_2 = \frac{\sigma}{\sqrt{2n}}$$

1960 Tukey points out that the mean absolute deviations may be more efficient under slight contamination.

$$F = (1-\varepsilon)\Phi(x/\sigma) + \varepsilon\Phi(x/(3\sigma))$$

Smooth Parametric Model

Family $\mathscr{P}$ of generalized Pareto distributions (GPD) with df

$$F_\theta(x) = 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi}$$

on sample space $\mathscr{X} = [0, \infty)$ and unknown parameter $\theta = (\beta, \xi) \in (0, \infty) \times [0, \infty)$.

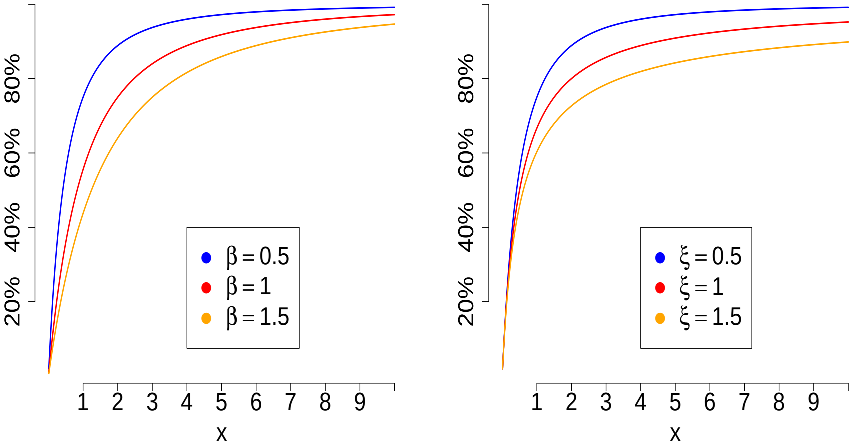
$\mathscr{P}$ is L_2-differentiable (smooth) at $\theta \in \Theta$ with L_2-derivative (scores function)

$$\Lambda_\theta(x) = \frac{d}{d\theta} \ln f_\theta(x),$$

$E_\theta \Lambda_\theta = 0$ and Fisher information of full rank $\mathscr{I}_\theta = E_\theta \Lambda_\theta \Lambda_\theta^t > 0$.

GPD with different Scale and Shape Parameters β and ξ

The shape parameter (tail index) ξ determines essentially the tail of the df.



Influence Curves and Asymptotically Linear Estimators

The set $\Psi_2(\theta)$ of all square integrable influence curves at P_θ is

$$\Psi_2(\theta) = \left\{ \psi_\theta \in L_2^k(P_\theta) \mid E_\theta \psi_\theta = 0, E_\theta \psi_\theta \Lambda_\theta^t = \mathbb{I}_k \right\}.$$

Estimator $\hat{\theta}_n$ is called asymptotically linear at P_θ if there is an influence curve ψ_θ with

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\theta(X_i) + o_{P_\theta^n}\left(n^0\right).$$

Example

Let $\hat{\theta}_n^{ML}$ be the ML estimator for θ. The influence curve is given by

$$\psi_\theta(x) = \mathscr{I}_\theta^{-1} \Lambda_\theta(x).$$

Generalization of ML Calculus

Estimation Problem

Observations $X_1, \ldots, X_n$ i.i.d., $X_i \sim P_\theta$ for some $\theta \in \Theta$.

Estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ for θ.

Maximum Likelihood Estimator $\hat{\theta}_n^{ML}$

$$L_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \ln f_\theta(X_i) = \max_\theta \quad \text{or} \quad l_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \Lambda_\theta(X_i) = 0.$$

Substitute sensitive scores function $\Lambda_\theta = (\partial/\partial\theta) \ln f_\theta$ by more robust function Ψ_θ.

M Estimator $\hat{\theta}_n^{M}$

$$M_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} m_\theta(X_i) = \max_\theta \quad \text{or} \quad \Psi_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \Psi_\theta(X_i) = 0.$$

Examples for Scores Functions

Observations $X_1, \ldots, X_n$ i.i.d., $X_i \sim Exp(\theta)$ with $f_\theta(x) = \theta e^{-\theta x}$, $\theta \in (0, \infty)$.

L_2-derivative $\Lambda_\theta(x) = \frac{\partial}{\partial\theta} \ln f_\theta(x) = \frac{1}{\theta} - x$.

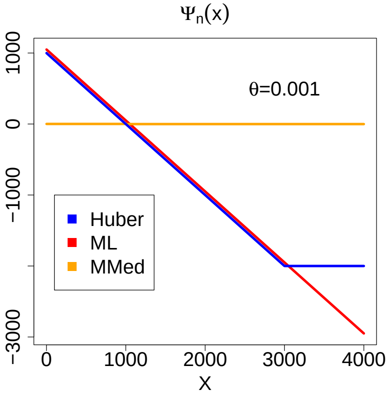
Scores functions to estimate θ

$$\Psi_n^{ML}(x) = \Lambda_\theta(x).$$

$$\Psi_n^{MMed}(x) = \operatorname{sgn}(\Lambda_\theta(x)).$$

$$\Psi_n^{Huber,c}(x) = \begin{cases} c & \text{if } \Lambda_\theta(x) > c \\ \Lambda_\theta(x) & \text{if } |\Lambda_\theta(x)| \le c \\ -c & \text{if } \Lambda_\theta(x) < -c \end{cases}$$

$$= \min\left(1, \frac{c}{|\Lambda_\theta(x)|}\right) \Lambda_\theta(x).$$



Neigborhood and Criteria to assess an Estimator

Observations $X_i \sim Q$ with Q in neighborhood of P_θ.

(Contamination) Neighborhood of P_θ as convex combination

$$B_c(P_\theta, r) = \left\{ (1-r)^+ P_\theta + \min(1,r)\, Q \mid Q \in \mathscr{M}^1(\mathscr{A}) \right\}$$

Criteria for efficiency and (local) robustness of an estimator

i. Trace of asymptotic covariance: trace $Cov_\theta \psi_\theta = E_\theta \parallel \psi_\theta \parallel^2$

ii. Gross error sensitivity (GES): $\sup_x \parallel \psi_\theta(x) \parallel$

Robust optimization problem

(ROP1) $E_\theta \parallel \psi_\theta \parallel^2 = \min!$ subject to $\sup_x \parallel \psi_\theta(x) \parallel \leq b$

(ROP2) $\max \mathrm{MSE}_\theta(\psi_\theta, r) := E_\theta \parallel \psi_\theta \parallel^2 + r^2 \sup_x \parallel \psi_\theta(x) \parallel^2 = \min!$

Estimators

Estimator	Solution to
Maximum likelihood estimator (MLE)	$\frac{1}{n}\sum_{i=1}^{n}\Lambda_\theta(X_i)=0$
Method of medians estimator (MMed)	$\frac{1}{n}\sum_{i=1}^{n} sgn(\Lambda_\theta(X_i)-a_\theta)=0$
under constraint $E_\theta[sgn(\Lambda_\theta(x)-a_\theta)]=0$	
Optimal bias-robust estimator (OBRE)	$\frac{1}{n}\sum_{i=1}^{n} w_b(X_i)(\Lambda_\theta(X_i)-a_\theta)=0,$
with $w_b(x)=\min\left(1,\frac{b}{\lVert A(\Lambda_\theta(x)-a_\theta)\rVert}\right)$, $a_\theta=\frac{\int w_b(x)\Lambda_\theta(x)dF_\theta}{\int w_b(x)dF_\theta}$	
and $A=\left(\int(\Lambda_\theta(x)-a_\theta)(\Lambda_\theta(x)-a_\theta)^t w_b(x)dF_\theta\right)^{-1}$	

Estimators [contd.]

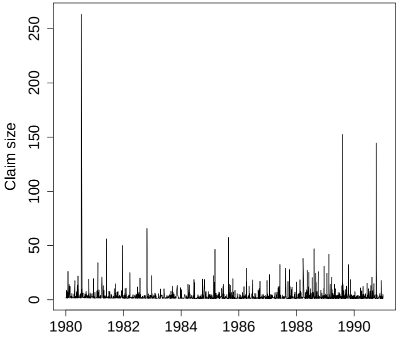
Estimator	Solution to
Method of moments (MOM)	$\overline{X} = E_\theta X$ and $\overline{X^2} = E_\theta X^2$
Minimum distance estimator (CvM)	$d_{CvM}\left(\hat{F}_n, F_\theta\right) = \sqrt{\int \left(\hat{F}_n - F_\theta\right)^2 dF_\theta} = \min_\theta$
Minimum distance estimator (Kol)	$d_{Kol}\left(\hat{F}_n, F_\theta\right) = \sup_{y \in \mathbb{R}} \mid \hat{F}_n(y) - F_\theta(y) \mid = \min_\theta$

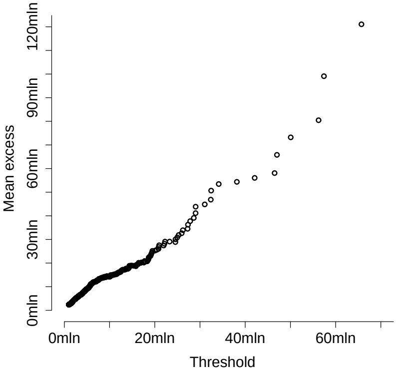
Danish Fire Claims Data revisited

Casestudy

Model 1: Exceedances over threshold of 10mln DKK

Model 2: Introduction of new largest loss of 350mln DKK to dataset





2,156 fire insurance claims over 1mln DKK from 1980 to 1990.

Danish Fire Claims Data revisited

Model 1: Exceedances over threshold of 10mln DKK

Estimator	Scale parameter β	s.e. β	Shape parameter ξ	s.e. ξ	*GES*	Quantiles 99.5%	99.9%
MLE	6.975	1.156	0.497	0.143	∞	181	421
MOM	8.520		0.395		∞	153	309
OBRE ($b = 10$)	6.970	1.161	0.494	0.149	10	179	413
OBRE ($b = 9$)	6.982	1.164	0.488	0.150	9	176	403
OBRE ($b = 8$)	7,032	1.162	0.454	0.150	8	156	341
MDE CvM	7.696	1.382	0.333	0.202	28	112	208

- MLE is efficient but not robust.
- MOM provides a poor fit, the estimator is not asymptotically normal ($\xi > 1/4$).
- OBRE and MDE CvM are relatively efficient and robust. The radius is 5.1% ($b = 10$), 6.4% ($b = 9$), 8.1% ($b = 8$).

Danish Fire Claims Data revisited

Model 1: Exceedances over threshold of 10mln DKK

Model 2: Introduction of new largest loss of 350mln DKK to dataset

Estimator	Model 1 ξ	Model 1 Quantiles 99.5%	Model 1 Quantiles 99.9%	Model 2 ξ	Model 2 Quantiles 99.5%	Model 2 Quantiles 99.9%	Delta 99.9%-Quantile
MLE	0.497	181	421	0.597	257	690	64%
PWM	0.517	191	455	0.613	266	732	61%
OBRE ($b = 10$)	0.494	179	413	0.592	252	672	63%
OBRE ($b = 9$)	0.488	176	403	0.571	234	603	50%
OBRE ($b = 8$)	0.454	156	341	0.551	218	545	60%
MDE CvM	0.333	112	208	0.370	127	248	19%
MDE Kol	0.457	158	347	0.489	179	410	18%

- McNeil (1996) indicates the sensitivity of the shape parameter to contamination applying ML calculus (in context of pricing XL treaties). Robust estimators, e.g. minimum distance estimators, may stabilize the results.

Conclusions

- Robust estimation is not a means to downweight large claims and reduce rates or risk indications.
- Rather, robust statistics may help
  - to fit a distribution to the bulk of the data and to identify outliers,
  - shed light on the sensitivity of estimators and provide information for calibration and validation of models,
  - to establish models that react more resilient to deviations from the ideal setup.