Another Look at Empirical Estimation of Actuarial Risk Measures

Vytautas Brazauskas

University of Wisconsin-Milwaukee

39th Actuarial Research Conference
Iowa City, Iowa, August 5–7, 2004
Outline

Part I  Introduction and Preliminaries

- Introduction
- Risk Measures
- Interval Estimation
- Simulation Study

Part II  Estimation Techniques

- Nonparametric Approach
- Parametric Methods
- Robust Procedures

Part III  Comparisons and Conclusions

- Comparisons
- Conclusions
1. Introduction and Preliminaries

Introduction

- Tools for measuring “riskiness”

  *Problem of interest*: price determination of an insurance risk; “riskiness” of a risk: certain properties of claims distribution; consider risk measures defined in terms of expectations with respect to distorted probabilities. (See Wang, Young, Panjer (1997), Wang (1998), Artzner (1999).)

- Unifying representation of risk measures

  \[ R(F, \psi) = \int_0^1 F^{-1}(s) \psi(s) \, ds \]

  (See Jones and Zitikis (2003).)

- Estimators of risk measures
  - Nonparametric (based on L-statistics)
  - Parametric (based on MLEs)
  - Robust (based on trimmed means)
Risk Measures

• Proportional Hazard Transform (PHT)

\[ \psi(s) = r(1 - s)^{r-1} \]

Here constant \( r \) (0 ≤ \( r \) ≤ 1) can be interpreted as the degree of distortion.

• Right-Tail Deviation (RTD)

\[ \psi(s) = r(1 - s)^{r-1} - 1 \]

For \( r = 1/2 \) this measure corresponds to Wang’s Right-Tail Deviation (Wang (1998)).

• Wang Transform (WT)

\[ \psi(s) = \exp\{-\lambda \Phi^{-1}(s) - \lambda^2/2\} \]

Here parameter \( \lambda \) reflects the systematic risk and \( \Phi(\cdot) \) is the cdf of \( N(0, 1) \) distribution.
Interval Estimation

• **General objective**

For a fixed sample size, favorable statistical procedures are those that yield the *shortest* interval while maintaining the desired *(high)* confidence level.

• **Specific aims**

  – *Convergence rates*

    How fast do the proposed (asymptotic) intervals attain the intended confidence level?

  – *Comparison of procedures at the model*

    Under strict distributional assumptions, how much do we gain if, instead of empirical intervals, parametric or robust confidence intervals are used?

  – *Sensitivity to assumptions*

    How bad are the consequences if the underlying assumptions necessary for the theoretical statements to hold are ignored or cannot be verified?
Simulation Study

• Contamination model

\[ H_\varepsilon = (1 - \varepsilon) F + \varepsilon G \]

*F* is central distribution, *G* is contaminating distribution, *\varepsilon* is level of contamination.

• Choices for the central distribution *F* 

  – Exponential distribution \((x > x_0, \theta > 0)\)

\[ F_1(x) = 1 - e^{-(x-x_0)/\theta} \]

  – Pareto distribution \((x > x_0, \gamma > 0)\)

\[ F_2(x) = 1 - \left(\frac{x_0}{x}\right)^\gamma \]

  – Lognormal distribution \((x > x_0, \mu \in \mathcal{R})\)

\[ F_3(x) = \Phi \left( \log(x - x_0) - \mu \right) \]

**NOTE:** Due to \(x_0\), distributions *F*₁, *F*₂, and *F*₃ have the same support.
• **Choices for the contaminating distribution** $G$
  
  – *Uniform distribution* ($10x_0 < x < 50x_0$)
  
  – *Pareto distribution* ($x > x_0, 0 < \gamma_1 < \gamma$)

  NOTE: Since $0 < \gamma_1 < \gamma$, this distribution has heavier upper-tail than Pareto with parameter $\gamma$.

• **Choice of parameters**

  Parameters $\theta$, $\gamma$, and $\mu$ are chosen so that all three distributions have the same level of “riskiness”. That is, for selected function $\psi$,

  $$R(F_1, \psi) = R(F_2, \psi) = R(F_3, \psi)$$

• **Study design**

  – Sample size: $n = 25, 50, 100, 250$.
  
  – Confidence level: $1 - \alpha = 0.90, 0.95, 0.99$.
  
  – Distortion level: $r = 0.50, 0.70, 0.85, 0.95$.
  
  – Systematic risk: $\lambda = 0.25, 0.50$. 
2. Estimation Techniques

Nonparametric Approach

100(1 − α)% confidence interval (based on the empirical estimator of a risk measure) is

\[
L_n[X] \pm z_{\alpha/2} \sqrt{\frac{Q_n(\psi, \psi)}{n}},
\]

where

\[
Q_n(\psi, \psi) = \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \left[ c_n(i, j) \psi(i/n) \psi(j/n) \times \right.
\]

\[
\times (X_{i+1:n} - X_{i:n})(X_{j+1:n} - X_{j:n}) \left. \right]
\]

with \( c_n(i, j) = \min\{i/n, j/n\} - (i/n)(j/n) \) and

\[
L_n[X] = \sum_{i=1}^{n} c_{in} X_{i:n}
\]

with \( c_{in} = \int_{i/n}^{(i-1)/n} \psi(s) \, ds \), and \( z_{\alpha/2} \) is the \( \alpha/2 \)-critical value of \( N(0, 1) \), and \( X_{1:n} \leq \cdots \leq X_{n:n} \) denote the ordered values of data \( X_1, \ldots, X_n \).
Part II

Estimation Techniques

Parametric Methods

• Exponential distribution \((x > x_0, \theta > 0)\)
  
  - \((1 - \alpha)100\%\) confidence interval for \(\theta:\)
    \[
    \hat{\theta}_{\text{ML}} \left( 1 \pm z_\alpha / 2 \sqrt{1/n} \right)
    \]
  
  - MLE of \(\theta:\)
    \[
    \hat{\theta}_{\text{ML}} = \frac{1}{n} \sum_{i=1}^{n} (X_i - x_0)
    \]

• Pareto distribution \((x > x_0, \gamma > 0)\)
  
  - \((1 - \alpha)100\%\) confidence interval for \(\gamma:\)
    \[
    \hat{\gamma}_{\text{ML}} \left( 1 \pm z_\alpha / 2 \sqrt{1/n} \right)
    \]
  
  - MLE of \(\gamma:\)
    \[
    \hat{\gamma}_{\text{ML}} = \left[ \frac{1}{n} \sum_{i=1}^{n} \log(X_i/x_0) \right]^{-1}
    \]

• Lognormal distribution \((x > x_0, \mu \in \mathcal{R})\)
  
  - \((1 - \alpha)100\%\) confidence interval for \(\mu:\)
    \[
    \hat{\mu}_{\text{ML}} \pm z_\alpha / 2 \sqrt{1/n}
    \]
  
  - MLE of \(\mu:\)
    \[
    \hat{\mu}_{\text{ML}} = \frac{1}{n} \sum_{i=1}^{n} \log(X_i - x_0)
    \]

NOTE: Corresponding intervals for the risk measures are found by appropriately transforming the above intervals.
Robust Procedures

- **Exponential distribution** \((x > x_0, \theta > 0)\)
  
  - \((1 - \alpha)100\%\) confidence interval for \(\theta\):
    \[
    \hat{\theta}_{TM} \left( 1 \pm z_{\alpha/2} \sqrt{k/n} \right)
    \]
  
  - Trimmed Mean (TM) estimator of \(\theta\):
    \[
    \hat{\theta}_{TM} = \frac{1}{d} \sum_{i = [n\beta_1] + 1}^{n - [n\beta_2]} (X_{i:n} - x_0),
    \]
    
    where
    \[
    d = d(\beta_1, \beta_2, n) = \sum_{j = [n\beta_1] + 1}^{n - [n\beta_2]} \sum_{i = 0}^{j-1} (n - i)^{-1}
    \]
    
    and \(\beta_1\) and \(\beta_2\) are trimming proportions.

- Efficiency constants \(k\):
  
  \[
  \begin{array}{c|cccc}
  \beta_1 = \beta_2 & 0.00 & 0.05 & 0.15 & 0.45 \\
  k & 1.00 & 1.090 & 1.271 & 1.946 \\
  \end{array}
  \]
Part II  Estimation Techniques

- **Pareto distribution** \((x > x_0, \gamma > 0)\)
  - \((1 - \alpha)100\%\) confidence interval for \(\gamma:\)
    \[
    \hat{\gamma}_{TM} \left( 1 \pm z_{\alpha/2} \sqrt{\frac{k}{n}} \right)
    \]
  - TM estimator of \(\gamma:\)
    \[
    \hat{\gamma}_{TM} = \left[ \frac{1}{d} \sum_{i = [n\beta_1] + 1}^{n - [n\beta_2]} \log(X_{i:n}/x_0) \right]^{-1}
    \]
    
    **NOTE:** Constants \(d\) and \(k\) are the same as for the exponential distribution.

- **Lognormal distribution** \((x > x_0, \mu \in \mathbb{R})\)
  - \((1 - \alpha)100\%\) confidence interval for \(\mu:\)
    \[
    \hat{\mu}_{TM} \pm z_{\alpha/2} \sqrt{\frac{K_\beta}{n}}
    \]
  - TM estimator of \(\mu\) \((\beta_1 = \beta_2 = \beta):\)
    \[
    \hat{\mu}_{TM} = \frac{1}{n - 2 \lfloor n\beta \rfloor} \sum_{i = \lfloor n\beta \rfloor + 1}^{n - \lfloor n\beta \rfloor} \log(X_i - x_0)
    \]
    
    **NOTE:** Efficiency constants \(K_\beta\) are: \(K_0 = 1, K_{0.05} = 1.026, K_{0.15} = 1.100, K_{0.45} = 1.474.\)
3. Comparisons and Conclusions

Comparisons

PHT measure ("clean" data scenario)

TABLE 1. Length ($L$) and coverage ($C$) of 95\% CI’s, for selected $F$ and $\varepsilon = 0$, $r = 0.85$, $n = 100$.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$F_1$ (exponential)</th>
<th>$F_2$ (Pareto)</th>
<th>$F_3$ (lognormal)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$L$</td>
<td>$C$</td>
<td>$L$</td>
</tr>
<tr>
<td>EMPIRICAL</td>
<td>.10</td>
<td>.92</td>
<td>.13</td>
</tr>
<tr>
<td>MLE($F_1$)</td>
<td>.11</td>
<td>.94</td>
<td>.10</td>
</tr>
<tr>
<td>MLE($F_2$)</td>
<td>.16</td>
<td>.92</td>
<td>.15</td>
</tr>
<tr>
<td>MLE($F_3$)</td>
<td>.10</td>
<td>.86</td>
<td>.09</td>
</tr>
<tr>
<td>TM(5%,$F_1$)</td>
<td>.11</td>
<td>.95</td>
<td>.10</td>
</tr>
<tr>
<td>TM(5%,$F_2$)</td>
<td>.18</td>
<td>.84</td>
<td>.15</td>
</tr>
<tr>
<td>TM(5%,$F_3$)</td>
<td>.11</td>
<td>.87</td>
<td>.10</td>
</tr>
<tr>
<td>TM(15%,$F_1$)</td>
<td>.12</td>
<td>.95</td>
<td>.10</td>
</tr>
<tr>
<td>TM(15%,$F_2$)</td>
<td>.21</td>
<td>.77</td>
<td>.17</td>
</tr>
<tr>
<td>TM(15%,$F_3$)</td>
<td>.13</td>
<td>.76</td>
<td>.11</td>
</tr>
<tr>
<td>TM(45%,$F_1$)</td>
<td>.15</td>
<td>.94</td>
<td>.13</td>
</tr>
<tr>
<td>TM(45%,$F_2$)</td>
<td>.29</td>
<td>.75</td>
<td>.22</td>
</tr>
<tr>
<td>TM(45%,$F_3$)</td>
<td>.16</td>
<td>.67</td>
<td>.13</td>
</tr>
</tbody>
</table>

NOTE: Standard errors for all entries are between .0001 and .0009 (for $L$) and between .001 and .009 (for $C$).
**PHT measure ("contaminated" data scenario)**

### TABLE 2. Performance of 95% CI's for selected $F$, $G = U(10x_0, 50x_0)$, and $\epsilon = 0.05, r = 0.85, n = 100.$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$F_1$ (exponential)</th>
<th>$F_2$ (Pareto)</th>
<th>$F_3$ (lognormal)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$L$</td>
<td>$C$</td>
<td>$L$</td>
</tr>
<tr>
<td><strong>EMPIRICAL</strong></td>
<td>3.66</td>
<td>.19</td>
<td>3.55</td>
</tr>
<tr>
<td>MLE($F_1$)</td>
<td>.79</td>
<td>.01</td>
<td>.76</td>
</tr>
<tr>
<td>MLE($F_2$)</td>
<td>.65</td>
<td>.02</td>
<td>.55</td>
</tr>
<tr>
<td>MLE($F_3$)</td>
<td>.14</td>
<td>.40</td>
<td>.12</td>
</tr>
<tr>
<td>TM(5%,$F_1$)</td>
<td>.22</td>
<td>.53</td>
<td>.21</td>
</tr>
<tr>
<td>TM(5%,$F_2$)</td>
<td>.30</td>
<td>.32</td>
<td>.26</td>
</tr>
<tr>
<td>TM(5%,$F_3$)</td>
<td>.13</td>
<td>.57</td>
<td>.11</td>
</tr>
<tr>
<td>TM(15%,$F_1$)</td>
<td>.13</td>
<td>.89</td>
<td>.12</td>
</tr>
<tr>
<td>TM(15%,$F_2$)</td>
<td>.24</td>
<td>.44</td>
<td>.20</td>
</tr>
<tr>
<td>TM(15%,$F_3$)</td>
<td>.14</td>
<td>.52</td>
<td>.12</td>
</tr>
<tr>
<td>TM(45%,$F_1$)</td>
<td>.16</td>
<td>.95</td>
<td>.14</td>
</tr>
<tr>
<td>TM(45%,$F_2$)</td>
<td>.33</td>
<td>.54</td>
<td>.25</td>
</tr>
<tr>
<td>TM(45%,$F_3$)</td>
<td>.17</td>
<td>.46</td>
<td>.14</td>
</tr>
</tbody>
</table>

**NOTE:** Standard errors for all entries are between .0002 and .0151 (for $L$) and between .001 and .013 (for $C$).
Part III
Comparisons and Conclusions

PHT measure (overall performance)

FIGURE 1. Proportions of coverage of 95% CI’s for selected $F, G = U(10x_0, 50x_0)$, $r = 0.85$, $n = 100$, $\varepsilon = 0$ ("clean" model) and $\varepsilon = 0.05$ ("contaminated" model).
WT measure (overall performance)

![Graphs showing coverage of 95% CI's for selected distributions](image)

**FIGURE 2.** Proportions of coverage of 95% CI’s for selected $F$, $G = U(10x_0, 50x_0)$, $\lambda = 0.25$, $n = 100$, $\varepsilon = 0$ (“clean” model) and $\varepsilon = 0.05$ (“contaminated” model).
Conclusions

- Convergence of the proportion of coverage of the empirical intervals is slow and depends on the function $\psi$. For “light” $\psi$, the coverage levels of these intervals get reasonably close to the nominal level for $n \geq 100$ and for all distributions $F$ that we considered. For “severe” $\psi$, however, their performances are unacceptable even for $n = 1500$. Parametric and robust intervals attain the intended confidence levels for all $\psi$ and $F$, and for sample sizes as small as $n = 50$.

- At the assumed model $F$, robust and parametric intervals perform better than empirical intervals with respect to the coverage criterion. Also, for $n \geq 250$, parametric intervals dominate robust and empirical counterparts with respect to the length criterion.

- When the assumed model $F$ is contaminated or misspecified, both parametric and empirical procedures perform poor. In such situations, only sufficiently robust estimators, designed for model $F$, yield intervals with consistently satisfactory performance.