Robust Statistics vs. MLE for OpRisk Severity Distribution Parameter Estimation (with and without truncation)

John Douglas (J.D.) Opdyke*, President
DataMineIt, JDOpdyke@DataMineIt.com

Presented at ORX Analytics Forum, San Francisco, CA, September 27-29, 2011

*The views presented herein are the views of the sole author, J.D. Opdyke, and do not necessarily reflect the views of other conference participants or discussants. All derivations, and all calculations and computations, were performed by J.D. Opdyke using SAS®.

Robust Statistics vs. MLE for OpRisk Severity Distribution Parameter Estimation (with and without truncation)

Presented at the ORX Analytics Forum 2011
September 27-29, 2011 • San Francisco, CA

J.D. Opdyke
President, DataMineIt
JDOpdyke@DataMineIt.com

ABSTRACT:

In operational risk measurement, severity distribution parameter estimation is the main driver of the aggregate loss distribution, yet it remains a non-trivial challenge for many reasons. Maximum likelihood estimation (MLE) does not adequately meet this challenge because of its well-documented non-robustness to modest violations of idealized textbook model assumptions (e.g. independent and identically distributed (i.i.d.) data, which OpRisk loss event data clearly violate). Even under i.i.d. data, the expected value of capital estimates based on MLE is biased upwards due to Jensen's inequality. This overstatement of the true risk profile increases as the heaviness of the severity distribution tail increases, so dealing with data collection thresholds by using truncated distributions, which have thicker tails, increases MLE's bias considerably. In addition, truncation typically induces dependence between a distribution's parameters (if not there already), and this exacerbates the non-robustness of MLE. This paper derives influence functions for MLE under a number of severity distributions, truncated and not, to analytically demonstrate its non-robustness. Simulations and empirical influence functions are then used to empirically compare its statistical properties (robustness, efficiency, and unbiasedness) to those of robust alternatives such as OBRE and a common minimum distance estimator (CvM). SLA (single-loss approximation) translates these parameter estimates into (VaR) estimates of regulatory capital requirements. These results show that OBRE estimators are very promising alternatives to MLE for use with actual OpRisk loss event data, whether truncated or not, when the ultimate goal is to obtain accurate (non-biased) capital estimates.

Keywords:

- OBRE
- Basel II
- MLE
- Severity Distribution
- Robust Statistics
- Capital Quantification
- Economic Capital

Contents

1. The OpRisk Setting and the Specific Objective

Operational Risk
- **Basel II/III**
  - **Advanced Measurement Approach**
    - **Risk Measurement & Capital Quantification**
      - **Loss Distribution Approach**
        - **Frequency Distribution**
        - **Severity Distribution*** (arguably the main driver of the aggregate loss distribution)

Specific Objective:

Develop a method to estimate the parameters of the severity distribution based on the following criteria – unbiasedness, (relative) efficiency, and robustness – with an emphasis on how these affect (right) tail-fit for capital estimation.**

* Dependence between the frequency and serverity distributions under some circumstances is addressed later in the presentation.

** Technically, the term "efficient" can refer to an estimator that achieves the Cramér-Rao lower bound. Hereafter in this presentation, the terms "efficient" and "efficiency" are used in a relative sense, as in having a lower mean squared error relative to that of another estimator. See Appendix I.

2. MLE vs. Robust Statistics: Point-Counterpoint

<u>Maximum Likelihood Estimation (MLE)</u>:

"MLE does not inappropriately downweight extreme observations as do most/all robust statistics. And focus on extreme observations is the entire point of the OpRisk statistical modeling exercise! Why should we even partially ignore the (right) tail when that is where and how capital requirements are determined?! That's essentially ignoring data – the most important data – just because its hard to model!"

<u>Robust Statistics</u>:

"All statistical models are merely idealized approximations of reality, and OpRisk data clearly violate the fragile, textbook model assumptions required by MLE (e.g. iid data). And even under iid data, the expected value of high quantile estimates based on MLE parameter estimates is biased upwards for (right-skewed) heavy-tailed distributions (i.e. OpRisk severity distributions) due to Jensen's inequality (and this, of course, inflates OpRisk capital estimates). Robust Statistics explicitly and sytemmatically acknowledge and deal with non-iid data, sometimes using weights to avoid bias and/or inefficiency caused by unanticipated or unnoticed heterogeneity. And an ancillary benefit is mitigation of the bias in capital estimates due to Jensen's inequality. Consequently, under real-world, finite-sample, non-iid OpRisk loss data, Robust Statistics typically exhibit less bias, equal and sometimes even greater efficiency, and far more robustness than does MLE. These characteristics translate into a more reliable, stable estimation approach, regardless of the framework used by robust statistics (i.e. multivariate regression or otherwise) to obtain high quantile estimates of the severity distribution.

…to be revisited

2. MLE vs. Robust Statistics: Point-Counterpoint

- Due to the nature of estimating the far right tail of the OpRisk loss event distribution, and the relative paucity of data, some type of parametric statistical estimation is required.
- OpRisk data poses many serious challenges for such a statistical estimation, as described on slides 7-8.
- The validity of MLE, the "classical" approach, relies on assumptions clearly violated by the data.
- Are these violations are material in their effects on MLE? Are high quantile estimates based on MLE parameter estimates too volatile, biased, and/or non-robust for use in OpRisk severity distribution parameter estimation? To answer this, analytic results are derived (simulations are merely confirmatory) borrowing from the toolkit of robust statistics, which are examined as possible alternatives to MLE should the objections against it have merit.

2. MLE vs. Robust Statistics: Point-Counterpoint

Some Specific Questions to be Answered:

- **Does MLE become unusable under relatively modest deviations from i.i.d., especially for the heavy-tailed distributions used in this setting, or are these claims overblown?**
- **Is the bias of the expected value of MLE-based capital estimates large?**
- **Do analytical derivations of the MLE Influence Functions for severity distribution parameters support or contradict such claims? Are they consistent with simulation results? How does (possible) parameter dependence affect these results?**
- **Do these results hold under truncation? How much does truncation and the size of the collection threshold affect both MLE and Robust Statistics parameter estimates?**
- **Are widely used, well established Robust Statistics viable for severity distribution parameter estimation? Are they too inefficient relative to MLE for practical use? Do any implementation constraints (e.g. algorithmic/convergence issues) trip them up?**

3. OpRisk Empirical Challenges

The following characteristics of most Operational Risk loss event data make estimating severity distribution parameters very challenging, and are the source of the MLE vs. Alternatives debate:

1. **Relatively few actual data points on loss events**
2. **Extremely few actual data points on low frequency, high severity losses**
3. **The heavy-tailed nature of most relevant severity distributions**
4. **Heterogeneity, even within well-defined units of measure**
5. **The (left) truncated nature of most loss event data (since smaller losses below a threshold typically are ignored)**
6. **The changing nature, from quarter to quarter, of some of the data already in hand (e.g. financial restatements, dispute resolutions, etc.)**
7. **The real potential for a large quarter of new data to non-trivially change the severity distribution**
8. **The real potential for notable heterogeneity in the form of true, robustly defined statistical outliers (not just extreme events)**
9. **The ultimate need to estimate an extremely high quantile of the severity distribution**

- **Moreover, the _combined_ effect of 1-9 increases estimation difficulty far more than the sum of the individual challenges (for a nice descriptive summary, see Cpe et al., 2009).**
- **Bottom line: OpRisk loss data is most certainly not independent and identically distributed ("i.i.d."), which is a presumption of MLE; and even if it was iid, the expected value of high quantile estimates based on MLE estimates is biased due to Jensen's inequality. For the relevant heavy-tailed severity distributions, this bias is notable, as shown on pp.64-68 below.**

3. OpRisk Empirical Challenges

The practical consequences of 1-9 above for OpRisk modeling can include:

- A. Unusably large variances on the parameter estimates
- B. Extreme sensitivity in parameter values to data changes (i.e. financial restatements, dispute resolutions, etc.) and/or new and different quarters of loss data. This would translate into a lack of stability and reliability in capital estimates from quarter to quarter.
- C. Unreasonable sensitivity of parameter estimates to very large losses
- D. Unreasonable sensitivity of parameter estimates to very small losses (this counter-intuitive result is documented analytically below)
- E. Due to any of A-D, unusably large variance on estimated severity distribution (high) quantiles
- F. Due to any of A-E, unusably large variance on capital estimates
- G. A theoretical loss distribution that does not sync well with the empirical loss distribution: the quantiles of each simply do not match well. This would not bode well for future estimations from quarter to quarter even if key tail quantiles in the current estimation are reasonably close.
- H. Bias in MLE-based capital estimates

- So in the OpRisk setting, when estimating severity distribution parameters (using finite samples), the statistical criteria of unbiasedness, efficiency, and robustness are critical and directly determine the degree to which capital estimates from quarter to quarter are stable, reliable, precise, and robust.
- A quantitative definition of statistical "robustness" (more precisely, "B-robustness") is provided in the next several slides, after a brief definition of maximum likelihood estimation (MLE).

4. Maximum Likelihood Estimation (MLE)

- **Maximum Likelihood Estimation (MLE) is considered a "classical" approach to parameter estimation.**
- **MLE parameter estimates are the values that maximize the likelihood, under the assumed model, of observing the data sample at hand.**
- **When the assumed model is in fact the true generator of the data, <u>and those data are independent and identically distributed ("i.i.d.")</u>, MLE estimates are asymptotically unbiased ("consistent"), asymptotically normally distributed, and asymptotically efficient (i.e. they achieve the Cramér-Rao lower bound – see Appendix I).**
- **MLE values are obtained in practice by maximizing the log-likelihood function.**
- **As an example, derivations of MLE estimates of the parameters of the LogNormal distribution are shown below.**
- **<u>NOTE</u>: While MLE parameter estimates are asymptotically unbiased, the expected value of high quantiles (capital estimates) based on them actually IS biased due to Jensen's inequality. This is a well established analytical result (for right-skewed severity distributions) confirmed by the capital simulations shown on pp.64-68. The magnitude of this bias is notable and larger the thicker is the tail of the severity distribution.**

4. Maximum Likelihood Estimation (MLE)

For example, assuming an i.i.d. sample of n observations $x_1, x_2, \cdots x_n$ from the LogNormal distribution

$$f(x \mid \mu,\sigma) \sim \frac{1}{\sqrt{2\pi}\sigma x} \cdot e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} \qquad F(x \mid \mu,\sigma) = \frac{1}{2}\left[1 + erf\left(\frac{\ln(x)-\mu}{\sqrt{2\sigma^2}}\right)\right]$$

- **The likelihood function =** $L(\mu,\sigma \mid x) = \prod_{i=1}^{n} f(x_i \mid \mu,\sigma)$
- **The log-likelihood function =** $\hat{l}(\theta \mid x_1, x_2, \cdots x_n) = \ln\left[L(\mu,\sigma \mid x)\right] = \sum_{i=1}^{n} \ln\left[f(x_i \mid \mu,\sigma)\right]$
- **Then** $\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{\arg\max}\left[\hat{l}(\theta \mid x_1, x_2, \cdots x_n)\right]$
- **So simply maximize the objective function** $\hat{l}(\theta \mid x) = \sum_{i=1}^{n} \ln\left[f(x_i \mid \mu,\sigma)\right]$
- **By finding $\hat{\mu}$ such that** $\frac{\partial \hat{l}(\theta \mid x)}{\partial \mu} = 0$
- **And finding $\hat{\sigma}$ such that** $\frac{\partial \hat{l}(\theta \mid x)}{\partial \sigma} = 0$

4. Maximum Likelihood Estimation (MLE)

$$\hat{l}(\theta \mid x) = \sum_{i=1}^{n} \ln\left[f(x_i \mid \mu, \sigma) \right]$$

$$= \sum_{i=1}^{n} \ln(1) - \ln\left(\sqrt{2\pi}\sigma x_i\right) - \frac{1}{2}\left(\frac{\ln(x_i) - \mu}{\sigma}\right)^2$$

$$= \sum_{i=1}^{n} -\ln\left(\sqrt{2\pi}\right) - \ln(\sigma) - \ln(x_i) - \frac{\left[\ln(x_i) - \mu\right]^2}{2\sigma^2}$$

$$0 = \frac{\partial \hat{l}(\theta \mid x)}{\partial \mu} = \sum_{i=1}^{n} \frac{\partial}{\partial \mu}\left(-\frac{\left[\ln(x_i) - \mu\right]^2}{2\sigma^2} \right) = \sum_{i=1}^{n} \frac{2\left[\ln(x_i) - \mu\right]}{2\sigma^2} = \sum_{i=1}^{n} \frac{\left[\ln(x_i) - \mu\right]}{\sigma^2}$$

$$0 = -\frac{n\mu}{\sigma^2} + \frac{1}{\sigma^2}\sum_{i=1}^{n} \ln(x_i)$$

$n\mu = \sum_{i=1}^{n} \ln(x_i)$ **, so** $\hat{\mu}_{MLE} = \dfrac{\sum_{i=1}^{n} \ln(x_i)}{n}$

4. Maximum Likelihood Estimation (MLE)

$$\hat{l}(\theta \mid x) = \sum_{i=1}^{n} \ln\left[f(x_i \mid \mu, \sigma) \right]$$
$$= \sum_{i=1}^{n} \ln(1) - \ln\left(\sqrt{2\pi}\sigma x_i\right) - \frac{1}{2}\left(\frac{\ln(x_i) - \mu}{\sigma}\right)^2$$
$$= \sum_{i=1}^{n} -\ln\left(\sqrt{2\pi}\right) - \ln(\sigma) - \ln(x_i) - \frac{\left[\ln(x_i) - \mu\right]^2}{2\sigma^2}$$

$$0 = \frac{\partial \hat{l}(\theta \mid x)}{\partial \sigma} = \sum_{i=1}^{n} \frac{\partial}{\partial \sigma}\left(-\ln(\sigma) - \frac{\left[\ln(x_i) - \mu\right]^2}{2\sigma^2} \right) = \sum_{i=1}^{n} -\frac{1}{\sigma} - \frac{(-2)\left[\ln(x_i) - \mu\right]^2}{2\sigma^3}$$

$$= \sum_{i=1}^{n} \frac{\left[\ln(x_i) - \mu\right]^2}{\sigma^3} - \frac{n}{\sigma}; \quad \frac{n}{\sigma} = \frac{1}{\sigma^3} \sum_{i=1}^{n} \left[\ln(x_i) - \mu\right]^2; \quad n\sigma^2 = \sum_{i=1}^{n} \left[\ln(x_i) - \mu\right]^2;$$

so $\hat{\sigma}^2_{MLE} = \dfrac{\sum_{i=1}^{n} \left[\ln(x_i) - \hat{\mu}_{MLE}\right]^2}{n}$ **, which is asymptotically unbiased.**

ORX ASSOCIATION

DataMineIt℠

4. Maximum Likelihood Estimation (MLE)

- When the log-likelihood cannot be simplified to obtain closed-form algebraic solutions, numerical methods often can be used to obtain its maximum. For example, for the parameters of the Generalized Pareto Distribution (GDP), Grimshaw (1993) used a reparameterization to develop a numerical algorithm that obtains MLE estimates. Similarly, for the LogGamma distribution, Bowman & Shenton (1983, 1988) provide numerical methods to obtain MLE parameter estimates. For heavy-tailed severity distributions used in this setting (and generally), the use of numerical methods to obtain MLE estimates of distributional parameters is the rule rather than the exception (so MLE proponents cannot use this as an objection to other methods of estimation).

5a. Robust Statistics: Background and the IF

- The theory behind Robust Statistics is well developed and has been in use for nearly half a century (see Huber, 1964). Textbooks have institutionalized this sub-field of statistics for the past 30 years (see Huber, 1981, and Hampel et al., 1986).
- Robust Statistics is a general approach to estimation that recognizes all statistical models are merely idealized approximations of reality. Consequently, one of its main objectives is bounding the influence on the estimates of a small to moderate number of data points in the sample that deviate from the assumed statistical model.
- Why? So that in practice, when actual data samples generated by real-world processes do not exactly follow mathematically convenient textbook assumptions (e.g. all data points are not perfectly "i.i.d."), estimates generated by robust statistics do not "breakdown" and provide meaningless, or at least notably biased and inaccurate, values: their values remain "robust" to such violations.
- Based on the empirical challenges of modeling OpRisk loss data (which is most certainly not "i.i.d.") satisfying this robustness objective would appear to be central to the OpRisk severity distribution parameter estimation effort: robust statistics may be tailor-made for this problem!
- The tradeoff for obtaining robustness, however, is a loss of efficiency – a larger mean squared error (MSE – see Appendix I) – when the idealized model assumptions are true: if model assumptions are violated, robust statistics can be MORE efficient than MLE.

5a. Robust Statistics: Background and the IF

The Influence Function (IF)

- **Perhaps the most useful analytical tool for assessing whether, and the degree to which, a statistic is "robust" in the sense that it bounds or limits the influence of arbitrary deviations* from the assumed model is the Influence Function (IF), defined below:**

$$IF\left(x \mid T,F\right) = \lim_{\varepsilon \to 0}\left[\frac{T\left\{\left(1-\varepsilon\right)F + \varepsilon\delta_x\right\} - T\left(F\right)}{\varepsilon}\right] = \lim_{\varepsilon \to 0}\left[\frac{T\left(F_\varepsilon\right) - T\left(F\right)}{\varepsilon}\right]$$

where

- F **is the distribution that is the assumed source of the data sample**
- T **is a statistical functional, that is, a statistic defined by the distribution that is the (assumed) source of the data sample. For example, the statistical functional for the mean is** $T\left(F\right) = \int u dF\left(u\right) = \int u f\left(u\right) du$
- x **is a particular point of evaluation, and the points being evaulated are those that deviate from the assumed** F**.**
- δ_x **is the probability measure that puts mass 1 at the point** x**.**

*** The terms "arbitrary deviation" and "contamination" or "statistical contamination" are used synonymously to mean data points that come from a distribution other than that assumed by the statistical model. They are not necessarily related to issues of data quality per se.**

5a. Robust Statistics: Background and the IF

$$IF(x|T,F) = \lim_{\varepsilon \to 0}\left[\frac{T\{(1-\varepsilon)F + \varepsilon\delta_x\} - T(F)}{\varepsilon}\right] = \lim_{\varepsilon \to 0}\left[\frac{T(F_\varepsilon) - T(F)}{\varepsilon}\right]$$

- F_ε is simply the distribution that includes some proportion of the data, ε, that is an arbitrary deviation away from the assumed distribution, F. So the Influence Function is simply the difference between the value of the statistical functional INCLUDING this arbitrary deviation in the data, vs. EXCLUDING the arbitrary deviation (the difference is then scaled by ε).
- So the IF is defined by three things: an estimator T, an assumed distribution/model F, and a deviation from this distribution, x (x obviously can represent more than one data point as ε is a proportion of the data sample, but it is easier conceptually to view x as a single data point whereby $\varepsilon = 1/n$: this is, in fact, the Empirical Influence Function (EIF) – see Appendix III).
- Simply put, the IF shows how, in the limit (asymptotically as $\varepsilon \to 0$, so as $n \to \infty$), an estimator's value changes as a function of x, the value of arbitrary deviations away from the assumed statistical model, F. In other words, the IF is the functional derivative of the estimator with respect to the distribution.

5a. Robust Statistics: Background and the IF

- **IF is a special case of the Gâteaux derivative, but its existence requires even weaker conditions (see Hampel et al., 1986, and Huber, 1977), so its use is valid under a very wide range of application (including the relevant OpRisk severity distributions).**

5a. Robust Statistics: Background and the IF

B-Robustness as Bounded IF

- If IF is bounded as x becomes arbitrarily large/small, the estimator is said to be "B-robust"*; if IF is not bounded and the estimator's values become arbitrarily large as deviations from the model become arbitrarily large/small, the estimator is NOT B-robust.
- The Gross Error Sensitivity (GES) measures the worst case (approximate) influence that an arbitrary deviation can have on the value of an estimator. If GES is finite, an estimator is B-robust; if it is infinite, it is not B-robust.

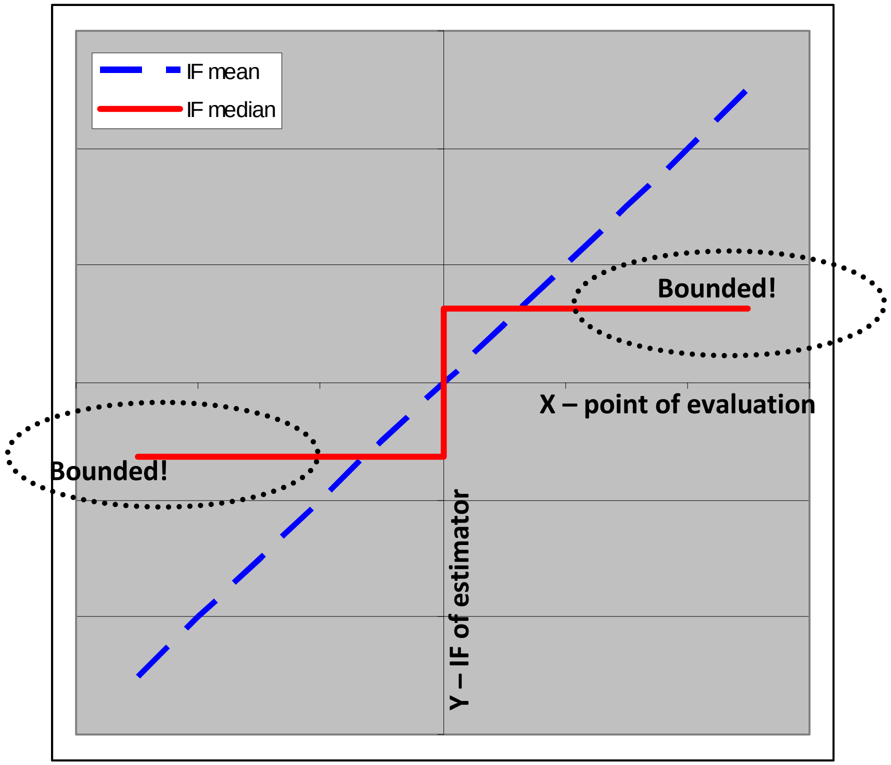
$$GES = \gamma^{*}\left(T,F\right) = \sup_{x}\left|IF\left(x;T,F\right)\right|$$

- A useful example demonstrating the concept of B-robustness is the comparison of the IFs of two common location estimators: the mean and the median. The former is unbounded with an infinite GES, and thus is not B-robust, while the latter is bounded, with a finite GES, and thus is B-robust.

* "B" comes from "bias," because if IF is bounded, the bias of the estimator is bounded.

5a. Robust Statistics: Background and the IF

Graph 1: Influence Functions of the Mean and the Median



- **Because the IF of the mean is <u>unbounded</u>, a single arbitrarily large data point can render the mean meaninglessly large, but that is not true of the median.**
- **The IF of the mean is derived mathematically below (see Hampel et al., 1986, pp.108-109 for a similar derivation for the median, also presented in Appendix II for convenience).**

5a. Robust Statistics: Background and the IF

Derivation of IF of the Mean:
Assuming $F = \Phi$ **, the standard normal distribution:**

$$IF(x|T,F) = \lim_{\varepsilon \to 0}\left[\frac{T(F_\varepsilon) - T(F)}{\varepsilon}\right]$$

$$= \lim_{\varepsilon \to 0}\left[\frac{T\{(1-\varepsilon)F + \varepsilon\delta_x\} - T(F)}{\varepsilon}\right]$$

The statistical functional of the mean is defined by $T(F) = \int u dF(u) = \int uf(u) du$ **, so…**

$$= \lim_{\varepsilon \to 0}\left[\frac{\int ud\{(1-\varepsilon)\Phi + \varepsilon\delta_x\}(u) - \int ud\Phi(u)}{\varepsilon}\right]$$

$$= \lim_{\varepsilon \to 0}\left[\frac{(1-\varepsilon)\int ud\Phi(u) + \varepsilon\int ud\delta_x(u) - \int ud\Phi(u)}{\varepsilon}\right]$$

$= \lim_{\varepsilon \to 0}\left[\frac{\varepsilon x}{\varepsilon}\right]$, because $\int ud\Phi(u) = 0$ so $IF(x;T,F) = x$

Or if $F \neq \Phi$ and $\int udF(u) \neq 0,$ then $IF(x|T,F) = \lim_{\varepsilon \to 0}\left[\frac{-\varepsilon\mu + \varepsilon x}{\varepsilon}\right] = x - \mu$

5a. Robust Statistics: Background and the IF

Many important robustness measures are based directly on the IF: brief definitions are presented below, with complete definitions listed in Appendix III.

- **Gross Error Sensitivity (GES):** Measures the worst case (approximate) influence that a small amount of contamination of a fixed size can have on the value of the estimator. If finite, the IF is bounded, and the estimator is "B-robust."
- **Rejection Point:** The point beyond which IF = zero and data points have no effect on the estimate.
- **Empirical Influence Function:** The non-asymptotic, finite-sample influence function.
- **Sensitivity Curves:** The scaled, non-asymptotic, finite-sample influence function (the difference between two empirical functionals, one based on a sample with contamination, one without, multiplied by *n*.)
- **Asymptotic Variance and ARE:** The variance of the estimator, and the ratio of the variances of two estimators.
- **Change-in-Variance Sensitivity:** For M-estimators, the derivative of the asymptotic variance when contaminated, divided by the asymptotic variance. Assesses how sensitive is the estimator to changes in its asymptotic variance due to contamination at *F*. If finite, then estimator is "V-robust," which is stronger than B-robustness.
- **Local Shift Sensitivity:** Assesses how sensitive the estimator is to small changes in the values of the observations; what is the worst effect on an estimator caused by shifting an observation slightly from point x to point y?
- **Breakdown Point:** A measure of global robustness, not local robustness like IF. The percentage of data points that can be contaminated with the estimator still providing useful information, that is, not "breaking down."

5a. Robust Statistics: Background and the IF

- **As may now be apparent, the robust statistics approach, and the analytical toolkit on which it relies, can be used to assess the performance of a very wide range of estimators, regardless of how they are classified; it is not limited to a small group of estimators. Hence, it has very wide ranging application and general utility.**
- **And a major objective of a robust statistics approach, as described above, is to bound the influence function of an estimator so that the estimator remains robust to deviations from the assumed statistical model (distribution). This approach would appear to be tailor-made to tackle many of the empirical challenges resident in OpRisk loss data.**
- **And as noted above, even under textbook iid data conditions, the expected value of capital estimates (high quantile estimates) based on MLE parameter estimates will be biased upwards (due to Jensen's inequality), sometimes dramatically (see pp. 64-68). Mitigation of this bias is an ancillary benefit of at least one of the robust statistics studied herein.**

5b. IF Derived: MLE Estimators of Severity Parameters

- **The goal of this section is to derive the IFs of the MLE estimators of the parameters of the relevant severity distributions. For this presentation-format of this paper, these distributions include: LogNormal, Truncated LogNormal, LogGamma, and Truncated LogGamma. I have made similar derivations for additional severity distributions, but include only the above for the sake of brevity. Additional distributions are included in the journal-format version of this paper.**

- **The point is to demonstrate analytically the non-robustness of MLE for the relevant estimations in the OpRisk setting, and hence the utility of IF as a heuristic and applied tool for assessing estimator performance. For example, deriving the IF for the mean (the MLE estimator of the specified model) gave an analytical result above of $IF\left(x \mid \mu, T\right) = x - \mu$ We know this is not B-robust because as x becomes arbitrarily large, so too does the IF: it is not bounded. Graphs comparing the IFs of these MLE estimators to the corresponding IFs of robust estimators will be shown in Section 7 (technically, the EIFs are compared, but the EIFs converge asymptotically to the IFs, and for the sample sizes used (n=250), the MLE IFs and MLE EIFs are virtually identical).**

- **In addition to determining whether any of the MLE estimators are B-robust, the IFs demonstrate the ranges of contamination (x) under which the estimators are the most volatile, show the relationships between a distribution's parameters, and how those relationships may change under different conditions (such as truncation).**

5b. IF Derived: MLE Estimators of Severity Parameters

- **New Results and Points of Note:**
  - **Derivations of the IFs, MLE or otherwise, must account for dependence between the parameters of the severity distribution: this is something that sometimes has been overlooked in the relevant OpRisk severity modeling literature.**
  - **IFs for the MLE estimators for the (left) truncated* distributions have not been reported in the literature: they are new results.**
  - **OBRE previously has not been applied to truncated data (with one exception that does not use the standard implementation algorithm): so these, too, are new results.**
  - **Truncation Induces Dependence/Extreme Sensitivity: Truncation induces dependence between the parameters of the severity distribution, if not there already (in which case truncation appears to augment it). This is shown in the formulae and graphs of the IFs, and appears to be the source of the extreme "sensitivity" of MLE estimators of truncated distributions reported in the literature, based on simulations. This is the first paper to present the analytic results under truncation.**

*** Unless otherwise noted, all truncation herein refers to left truncation, that is, truncation of the lower (left) tail, because data collection thresholds for losses ignore losses below a specified threshold. Under reasonable assumptions, truncation does induce dependence between the frequency and severity distributions, but this is ignored (as is often convention in this setting) for the purposes of this presentation.**

5b. IF Derived: MLE Estimators of Severity Parameters

- **MLEs belong to the class of "M-estimators," so called because they generalize "M"aximum likelihood estimation. Broad classes of estimators have the same form of IF (see Hampel et al. ,1986), so all M-estimators conveniently share the same form of IF.**

- **M-estimators are consistent and asymptotically normal.**

- **M-estimators are defined as any estimator** $T_n = T_n\left(X_1, \cdots, X_n\right)$ **that satisfies**

$$\sum_{i=1}^{n} \rho\left(X_i, T_n\right) = \min_{T_n}! \quad \text{or} \quad \sum_{i=1}^{n} \varphi\left(X_i, T_n\right) = 0 \quad \text{where} \quad \varphi\left(x,\theta\right) = \frac{\partial \rho\left(x,\theta\right)}{\partial \theta}$$

if the derivative of ρ **exists, and** ρ **is defined on** $\wp \times \Theta$ **.**

So for MLE:

$$\rho\left(x,\theta\right) = -\ln\left[f\left(x,\theta\right)\right]$$

$$\varphi_\theta\left(x,\theta\right) = \frac{\partial \rho\left(x,\theta\right)}{\partial \theta} = -\frac{\partial f\left(x,\theta\right)}{\partial \theta} \bigg/ f\left(x,\theta\right)$$ **(note that this is simply the score function)**

$$\varphi_\theta'\left(x,\theta\right) = \frac{\partial \varphi_\theta\left(x,\theta\right)}{\partial \theta} = \frac{\partial \rho^2\left(x,\theta\right)}{\partial \theta^2} = \frac{-\frac{\partial f^2\left(x,\theta\right)}{\partial \theta^2} \cdot f\left(x,\theta\right) + \left[\frac{\partial f\left(x,\theta\right)}{\partial \theta}\right]^2}{\left[f\left(x,\theta\right)\right]^2}$$

5b. IF Derived: MLE Estimators of Severity Parameters

- **And for M-estimators, IF is defined as (assuming a nonzero denominator):**

$$IF_{\theta}\left(x \mid \theta, T\right) = \frac{\varphi_{\theta}\left(y,\theta\right)}{-\int_{a}^{b} \varphi'_{\theta}\left(y,\theta\right) dF\left(y\right)}$$

where a and b define the domain of the density (in this setting, typically a = 0 and b = ∞).

So we can write

$$IF_{\theta}\left(x \mid \theta, T\right) = \frac{-\dfrac{\dfrac{\partial f\left(y,\theta\right)}{\partial\theta}}{f\left(y,\theta\right)}}{-\int_{a}^{b} \dfrac{-\dfrac{\partial f^{2}\left(y,\theta\right)}{\partial\theta^{2}} \cdot f\left(y,\theta\right) + \left[\dfrac{\partial f\left(y,\theta\right)}{\partial\theta}\right]^{2}}{\left[f\left(y,\theta\right)\right]^{2}} dF\left(y\right)} = \frac{\dfrac{\dfrac{\partial f\left(y,\theta\right)}{\partial\theta}}{f\left(y,\theta\right)}}{\int_{a}^{b} \dfrac{\left[\dfrac{\partial f\left(y,\theta\right)}{\partial\theta}\right]^{2} - \dfrac{\partial f^{2}\left(y,\theta\right)}{\partial\theta^{2}} \cdot f\left(y,\theta\right)}{f\left(y,\theta\right)} dy}$$

For the (left) truncated densities, $g\left(x,\theta\right) = \dfrac{f\left(x,\theta\right)}{1-F\left(H,\theta\right)}$ **where H is the truncation threshold.**

And so the above becomes:

5b. IF Derived: MLE Estimators of Severity Parameters

IF of MLEs for (left) truncated densities:

$$\rho(x;\theta) = -\ln\left(g(x;\theta)\right) = -\ln\left(\frac{f(x;\theta)}{1-F(H;\theta)}\right) = -\ln\left(f(x;\theta)\right) + \ln\left(1-F(H;\theta)\right)$$

$$\varphi_\theta(x,H;\theta) = \frac{\partial \rho(x;\theta)}{\partial \theta} = -\frac{\frac{\partial f(x;\theta)}{\partial \theta}}{f(x;\theta)} - \frac{\frac{\partial F(H;\theta)}{\partial \theta}}{1-F(H;\theta)}$$

$$\varphi'_\theta(x,H;\theta) = \frac{\partial \varphi_\theta(x,H;\theta)}{\partial \theta} = \frac{\partial^2 \rho(x;\theta)}{\partial \theta^2} =$$

$$= \frac{-\frac{\partial^2 f(x;\theta)}{\partial \theta^2}\cdot f(x;\theta) + \left[\frac{\partial f(x;\theta)}{\partial \theta}\right]^2}{\left[f(x;\theta)\right]^2} + \frac{-\frac{\partial^2 F(H;\theta)}{\partial \theta^2}\cdot\left[1-F(H;\theta)\right] - \left[\frac{\partial F(H;\theta)}{\partial \theta}\right]^2}{\left[1-F(H;\theta)\right]^2}$$

And so the IF is

5b. IF Derived: MLE Estimators of Severity Parameters

IF of MLEs for (left) truncated densities:

$$IF_\theta\left(x;\theta,T\right)=\frac{-\dfrac{\dfrac{\partial f\left(x;\theta\right)}{\partial\theta}}{f\left(x;\theta\right)}-\dfrac{\dfrac{\partial F\left(H;\theta\right)}{\partial\theta}}{1-F\left(H;\theta\right)}}{-\dfrac{1}{1-F\left(H;\theta\right)}\displaystyle\int_a^b\frac{\left[\dfrac{\partial f\left(y;\theta\right)}{\partial\theta}\right]^2-\dfrac{\partial^2 f\left(y;\theta\right)}{\partial\theta^2}\cdot f\left(y;\theta\right)}{f\left(y;\theta\right)}dy+\dfrac{\left[\dfrac{\partial F\left(H;\theta\right)}{\partial\theta}\right]^2+\dfrac{\partial^2 F\left(H;\theta\right)}{\partial\theta^2}\cdot\left[1-F\left(H;\theta\right)\right]}{\left[1-F\left(H;\theta\right)\right]^2}}$$

Note that a and b are now H and (typically) ∞ , respectively.

As noted previously, we must account for (possible) dependence between the parameter estimates, and so we must use the matrix form of the IF defined below (see Stefanski & Boos (2002) and D.J. Dupuis (1998)):

$$IF_\theta\left(x;\theta,T\right)=A\left(\theta\right)^{-1}\varphi_\theta=\begin{bmatrix}-\displaystyle\int_a^b\frac{\partial\varphi_{\theta_1}}{\partial\theta_1}dK\left(y\right) & -\displaystyle\int_a^b\frac{\partial\varphi_{\theta_1}}{\partial\theta_2}dK\left(y\right)\\ -\displaystyle\int_a^b\frac{\partial\varphi_{\theta_2}}{\partial\theta_1}dK\left(y\right) & -\displaystyle\int_a^b\frac{\partial\varphi_{\theta_2}}{\partial\theta_2}dK\left(y\right)\end{bmatrix}^{-1}\begin{bmatrix}\varphi_{\theta_1}\\ \varphi_{\theta_2}\end{bmatrix}$$

Where K is either F or G , $A\left(\theta\right)$ is simply the Fisher Information (if the data follow the assumed model), and φ_θ is now vectorized. Parameter dependence exists when the off-diagonal terms are not zero.

5b. IF Derived: MLE Estimators of Severity Parameters

Note that the off-diagonal cross-terms are the second-order partial derivatives:

$$-\int_a^b \frac{\partial \varphi_{\theta_1}}{\partial \theta_2} dG(y) = -\frac{1}{1-F(H;\theta)} \int_a^b \frac{\left[\frac{\partial f(y;\theta)}{\partial \theta_1}\right]\left[\frac{\partial f(y;\theta)}{\partial \theta_2}\right] - \frac{\partial^2 f(y;\theta)}{\partial \theta_1 \partial \theta_2} \cdot f(y;\theta)}{f(y;\theta)} dy + \frac{\left[\frac{\partial F(H;\theta)}{\partial \theta_1}\right]\left[\frac{\partial F(H;\theta)}{\partial \theta_2}\right] + \frac{\partial^2 F(H;\theta)}{\partial \theta_1 \partial \theta_2} \cdot \left[1-F(H;\theta)\right]}{\left[1-F(H;\theta)\right]^2}$$

and

$$-\int_a^b \frac{\partial \varphi_{\theta_1}}{\partial \theta_2} dF(y) = \int_a^b \frac{\left[\frac{\partial f(y,\theta)}{\partial \theta_1}\right]\left[\frac{\partial f(y,\theta)}{\partial \theta_2}\right] - \frac{\partial f^2(y,\theta)}{\partial \theta_1 \partial \theta_2} \cdot f(y,\theta)}{f(y,\theta)} dy$$

With the above defintion, all that needs be done to derive IF for each severity distribution is the calculation of the first and second order derivatives of each density, as well as, for the (left) truncated cases, the first and second order derivatives of the cumulative distribution functions: that is, derive

$$\frac{\partial f(y;\theta)}{\partial \theta_1}, \frac{\partial f(y;\theta)}{\partial \theta_2}, \frac{\partial^2 f(y;\theta)}{\partial \theta_1 \partial \theta_2}, \frac{\partial^2 f(y;\theta)}{\partial \theta_1^2}, \frac{\partial^2 f(y;\theta)}{\partial \theta_2^2}, \frac{\partial F(H;\theta)}{\partial \theta_1}, \frac{\partial F(H;\theta)}{\partial \theta_2}, \frac{\partial^2 F(H;\theta)}{\partial \theta_1 \partial \theta_2}, \frac{\partial F^2(H;\theta)}{\partial \theta_1^2}, \text{ and } \frac{\partial F^2(H;\theta)}{\partial \theta_2^2}$$

This is done in Appendix IV for the four severity distributions examined herein.

This “plug-n-play” approach makes derivation and use of the IFs corresponding to each severity distribution’s parameters considerably more convenient.

5b. IF Derived: MLE Estimators of Severity Parameters

Below, I "plug-n-play" to obtain $A(\theta)$ for the four severity distributions. Note that for the LogNormal, (left) truncation induces parameter dependence, and for the LogGamma, it augments dependence that was there even before truncation. For the truncated cases and the LogGamma, after the cells of $A(\theta)$ are obtained, IF is calculated numerically.

From Appendix IV, inserting the derivations of $\frac{\partial f(y;\theta)}{\partial\theta_1}, \frac{\partial f(y;\theta)}{\partial\theta_2}, \frac{\partial^2 f(y;\theta)}{\partial\theta_1\partial\theta_2}, \frac{\partial^2 f(y;\theta)}{\partial\theta_1^2}$, and $\frac{\partial^2 f(y;\theta)}{\partial\theta_2^2}$

for the LogNormal yields

$$-\int_0^\infty \frac{\partial\varphi_\mu}{\partial\mu} dF(y) = -\int_0^\infty \left[\frac{\ln(y)-\mu}{\sigma^2}\right]^2 - \left[\frac{(\ln(y)-\mu)^2}{\sigma^4} - \frac{1}{\sigma^2}\right] f(y)dy = -\int_0^\infty \frac{1}{\sigma^2} f(y)dy = -\frac{1}{\sigma^2}$$

$$-\int_0^\infty \frac{\partial\varphi_\sigma}{\partial\sigma} dF(y) = -\int_0^\infty \left(\frac{3(\ln(y)-\mu)^2}{\sigma^4} - \frac{1}{\sigma^2}\right) f(y)dy = \frac{-3}{\sigma^4}\int_0^\infty (\ln(y)-\mu)^2 f(y)dy + \frac{1}{\sigma^2} = \frac{-3\sigma^2}{\sigma^4} + \frac{1}{\sigma^2} = -\frac{2}{\sigma^2}$$

$$-\int_0^\infty \frac{\partial\varphi_\mu}{\partial\sigma} dF(y) = -\int_0^\infty \frac{\partial\varphi_\sigma}{\partial\mu} dF(y) = \int_0^\infty \left(\left[\frac{\ln(y)-\mu}{\sigma^2}\right]\left[\frac{(\ln(y)-\mu)^2}{\sigma^3} - \frac{1}{\sigma}\right] - \left[\frac{\ln(y)-\mu}{\sigma^2}\right]\left[\frac{(\ln(y)-\mu)^2}{\sigma^3} - \frac{1}{\sigma}\right]\right) f(y)dy = 0$$

5b. IF Derived: MLE Estimators of Severity Parameters

Inserting Appendix IV derivations of for the LogNormal yields…

$$IF_{\theta}\left(x;\theta,T\right)=A\left(\theta\right)^{-1}\varphi_{\theta}=\begin{bmatrix}-\int_a^b \frac{\partial\varphi_{\theta_1}}{\partial\theta_1}dK\left(y\right) & -\int_a^b \frac{\partial\varphi_{\theta_1}}{\partial\theta_2}dK\left(y\right)\\ -\int_a^b \frac{\partial\varphi_{\theta_2}}{\partial\theta_1}dK\left(y\right) & -\int_a^b \frac{\partial\varphi_{\theta_2}}{\partial\theta_2}dK\left(y\right)\end{bmatrix}^{-1}\begin{bmatrix}\varphi_{\theta_1}\\ \varphi_{\theta_2}\end{bmatrix}=$$

(zero off-diagonals indicate no parameter dependence)

$$=\begin{bmatrix}-1/\sigma^2 & 0\\ 0 & -2/\sigma^2\end{bmatrix}^{-1}\begin{bmatrix}\dfrac{\mu-\ln\left(x\right)}{\sigma^2}\\ \dfrac{1}{\sigma}-\dfrac{\left(\ln\left(x\right)-\mu\right)^2}{\sigma^3}\end{bmatrix}=$$

$$=\begin{bmatrix}-\sigma^2 & 0\\ 0 & -\sigma^2/2\end{bmatrix}\begin{bmatrix}\dfrac{\mu-\ln\left(x\right)}{\sigma^2}\\ \dfrac{1}{\sigma}-\dfrac{\left(\ln\left(x\right)-\mu\right)^2}{\sigma^3}\end{bmatrix}=\begin{bmatrix}\ln\left(x\right)-\mu\\ \dfrac{\left(\ln\left(x\right)-\mu\right)^2-\sigma^2}{2\sigma}\end{bmatrix}$$

5b. IF Derived: MLE Estimators of Severity Parameters

From Appendix IV, inserting the derivations of

$$\frac{\partial f(y;\theta)}{\partial\theta_1}, \frac{\partial f(y;\theta)}{\partial\theta_2}, \frac{\partial^2 f(y;\theta)}{\partial\theta_1\partial\theta_2}, \frac{\partial^2 f(y;\theta)}{\partial\theta_1^2}, \frac{\partial^2 f(y;\theta)}{\partial\theta_2^2}, \frac{\partial F(H;\theta)}{\partial\theta_1}, \frac{\partial F(H;\theta)}{\partial\theta_2}, \frac{\partial^2 F(H;\theta)}{\partial\theta_1\partial\theta_2}, \frac{\partial F^2(H;\theta)}{\partial\theta_1^2}, \text{ and } \frac{\partial F^2(H;\theta)}{\partial\theta_2^2}$$

for the (left) Truncated LogNormal yields

$$-\int_H^\infty \frac{\partial\varphi_\mu}{\partial\mu} dG(y) = -\frac{1}{\sigma^2} + \frac{\left[\int_0^H \frac{\ln(y)-\mu}{\sigma^2} f(y)dy\right]^2 + \int_0^H \frac{(\ln(y)-\mu)^2}{\sigma^4} - \frac{1}{\sigma^2} f(y)dy \cdot \left[1-F(H;\mu,\sigma)\right]}{\left[1-F(H;\mu,\sigma)\right]^2}$$

$$-\int_H^\infty \frac{\partial\varphi_\sigma}{\partial\sigma} dG(y) = -\frac{1}{\left[1-F(H;\mu,\sigma)\right]} \cdot \int_H^\infty \frac{3(\ln(y)-\mu)^2}{\sigma^4} f(y)dy + \frac{1}{\sigma^2} +$$

$$+\frac{\left[\int_0^H \frac{(\ln(y)-\mu)^2}{\sigma^3} - \frac{1}{\sigma} f(y)dy\right]^2 + \int_0^H \left[\frac{1}{\sigma^2} - \frac{3(\ln(y)-\mu)^2}{\sigma^4}\right] + \left[\frac{(\ln(y)-\mu)^2}{\sigma^3} - \frac{1}{\sigma}\right]^2 f(y)dy \cdot \left[1-F(H;\mu,\sigma)\right]}{\left[1-F(H;\mu,\sigma)\right]^2}$$

$$-\int_H^\infty \frac{\partial\varphi_\mu}{\partial\sigma} dG(y) = -\int_0^\infty \frac{\partial\varphi_\sigma}{\partial\mu} dF(y) = -\frac{1}{\left[1-F(H;\mu,\sigma)\right]} \cdot \int_H^\infty \frac{-2(\ln(y)-\mu)}{\sigma^3} f(y)dy +$$

(non-zero off-diagonals indicate parameter dependence)

$$+\frac{\left[\int_0^H \frac{\ln(y)-\mu}{\sigma^2} f(y)dy\right] \times \left[\int_0^H \frac{(\ln(y)-\mu)^2}{\sigma^3} - \frac{1}{\sigma} f(y)dy\right] + \left(\int_0^H \frac{-2(\ln(y)-\mu)}{\sigma^3} f(y)dy + \int_0^H \left[\frac{\ln(y)-\mu}{\sigma^2}\right] \cdot \left[\frac{(\ln(y)-\mu)^2}{\sigma^3} - \frac{1}{\sigma}\right] f(y)dy\right) \cdot \left[1-F(H;\mu,\sigma)\right]}{\left[1-F(H;\mu,\sigma)\right]^2}$$

5b. IF Derived: MLE Estimators of Severity Parameters

From Appendix IV, inserting the derivations of $\frac{\partial f(y;\theta)}{\partial \theta_1}, \frac{\partial f(y;\theta)}{\partial \theta_2}, \frac{\partial^2 f(y;\theta)}{\partial \theta_1 \partial \theta_2}, \frac{\partial^2 f(y;\theta)}{\partial \theta_1^2}$, and $\frac{\partial^2 f(y;\theta)}{\partial \theta_2^2}$ **for the LogGamma yields**

$$-\int_1^\infty \frac{\partial \varphi_a}{\partial a} dF(y) = -\int_1^\infty \frac{\partial\left(-\ln(b) - \ln(\ln(y)) + digamma(a)\right)}{\partial a} f(y)dy = -\int_1^\infty trigamma(a) f(y)dy = -trigamma(a)$$

$$-\int_1^\infty \frac{\partial \varphi_b}{\partial b} dF(y) = -\int_1^\infty \frac{\partial\left(-\frac{a}{b} + \ln(y)\right)}{\partial b} f(y)dy = -\int_1^\infty \frac{a}{b^2} f(y)dy = -\frac{a}{b^2}$$

$$-\int_1^\infty \frac{\partial \varphi_a}{\partial b} dF(y) = -\int_1^\infty \frac{\partial \varphi_b}{\partial a} dF(y) = -\int_1^\infty \frac{\partial\left(-\ln(b) - \ln(\ln(y)) + digamma(a)\right)}{\partial b} f(y)dy = -\int_1^\infty \frac{\partial\left(-\frac{a}{b} + \ln(y)\right)}{\partial a} f(y)dy = -\int_1^\infty -\frac{1}{b} dy = \frac{1}{b}$$

5b. IF Derived: MLE Estimators of Severity Parameters

Inserting Appendix IV derivations of for the LogGamma yields…

$$IF_\theta\left(x;\theta,T\right)=A\left(\theta\right)^{-1}\varphi_\theta=\begin{bmatrix}-\int_a^b\frac{\partial\varphi_{\theta_1}}{\partial\theta_1}dK\left(y\right) & -\int_a^b\frac{\partial\varphi_{\theta_1}}{\partial\theta_2}dK\left(y\right)\\ -\int_a^b\frac{\partial\varphi_{\theta_2}}{\partial\theta_1}dK\left(y\right) & -\int_a^b\frac{\partial\varphi_{\theta_2}}{\partial\theta_2}dK\left(y\right)\end{bmatrix}^{-1}\begin{bmatrix}\varphi_{\theta_1}\\ \varphi_{\theta_2}\end{bmatrix}=$$

(non-zero off-diagonals indicate parameter dependence)

$$=\begin{bmatrix}-trigamma\left(a\right) & 1/b\\ 1/b & -a/b^2\end{bmatrix}^{-1}\begin{bmatrix}-\ln\left(b\right)-\ln\left(\ln\left(x\right)\right)+digamma\left(a\right)\\ -\frac{a}{b}+\ln\left(x\right)\end{bmatrix}=$$

$$=\frac{1}{\left(-a/b^2\right)\cdot trigamma\left(a\right)-1/b^2}\begin{bmatrix}-a/b^2 & -1/b\\ -1/b & -trigamma\left(a\right)\end{bmatrix}\begin{bmatrix}-\ln\left(b\right)-\ln\left(\ln\left(x\right)\right)+digamma\left(a\right)\\ -\frac{a}{b}+\ln\left(x\right)\end{bmatrix}=$$

$$=\begin{bmatrix}\dfrac{\frac{a}{b^2}\left[\ln\left(b\right)+\ln\left(\ln\left(x\right)\right)-digamma\left(a\right)\right]-\frac{1}{b}\left[\ln\left(x\right)-\frac{a}{b}\right]}{trigamma\left(a\right)\left(\frac{a}{b^2}\right)-\frac{1}{b^2}}\\ \dfrac{\frac{1}{b}\left[\ln\left(b\right)+\ln\left(\ln\left(x\right)\right)-digamma\left(a\right)\right]-trigamma\left(a\right)\left[\ln\left(x\right)-\frac{a}{b}\right]}{trigamma\left(a\right)\left(\frac{a}{b^2}\right)-\frac{1}{b^2}}\end{bmatrix}$$

5b. IF Derived: MLE Estimators of Severity Parameters

From Appendix IV, inserting the derivations of

$$\frac{\partial f(y;\theta)}{\partial\theta_1},\frac{\partial f(y;\theta)}{\partial\theta_2},\frac{\partial^2 f(y;\theta)}{\partial\theta_1\partial\theta_2},\frac{\partial^2 f(y;\theta)}{\partial\theta_1^2},\frac{\partial^2 f(y;\theta)}{\partial\theta_2^2},\frac{\partial F(H;\theta)}{\partial\theta_1},\frac{\partial F(H;\theta)}{\partial\theta_2},\frac{\partial^2 F(H;\theta)}{\partial\theta_1\partial\theta_2},\frac{\partial F^2(H;\theta)}{\partial\theta_1^2},\text{ and }\frac{\partial F^2(H;\theta)}{\partial\theta_2^2}$$

for the (left) Truncated LogGamma yields

$$-\int_H^\infty \frac{\partial\varphi_a}{\partial a}dG(x) = -trigamma(a) + \frac{\left[\int_1^H \ln(b)+\ln(\ln(x)) - digamma(a)\, f(x)dx\right]^2 + \left[1-F(H;a,b)\right]\cdot\int_1^H\left[\ln(b)+\ln(\ln(x))-digamma(a)\right]^2 - trigamma(a)\, f(x)dx}{\left[1-F(H;a,b)\right]^2}$$

$$-\int_H^\infty \frac{\partial\varphi_b}{\partial b}dG(x) = -\frac{a}{b^2} + \frac{\left[\int_1^H\left(\frac{a}{b}-\ln(y)\right)f(x)dx\right]^2 + \left[1-F(H;a,b)\right]\cdot\int_1^H \frac{a(a-1)}{b^2} - \frac{2a\ln(y)}{b} + \left[\ln(y)\right]^2 f(x)dx}{\left[1-F(H;a,b)\right]^2}$$

$$-\int_H^\infty \frac{\partial\varphi_a}{\partial b}dG(x) = -\int_H^\infty \frac{\partial\varphi_b}{\partial a}dG(x) = \frac{1}{b} + \frac{\left[1-F(H;a,b)\right]\cdot\frac{1}{b}\cdot F(H;a,b) + \left[1-F(H;a,b)\right]\cdot\int_1^H\left[\ln(b)+\ln(\ln(x))-digamma(a)\right]\cdot\left[\frac{a}{b}-\ln(x)\right]f(x)dx}{\left[1-F(H;a,b)\right]^2}$$

$$+\frac{\int_1^H \ln(b)+\ln(\ln(x))-digamma(a)\, f(x)dx\cdot\int_1^H\left(\frac{a}{b}-\ln(x)\right)f(x)dx}{\left[1-F(H;a,b)\right]^2}$$

(non-zero off-diagonals indicate parameter dependence)

OBRE Defined:

The Optimally Bias-Robust Estimator (OBRE) is provided for a given sample of data as the value $\hat{\theta}$ of θ that solves (1):

(1) $$\sum_{i=1}^{n} \varphi_c^{A,a}\left(x_i;\theta\right)=0$$ **where** **(1.a)** $$\varphi_c^{A,a}\left(x;\theta\right)=A\left(\theta\right)\cdot\left[s\left(x;\theta\right)-a\left(\theta\right)\right]\cdot W_c\left(x;\theta\right)$$

and

(1.b) $$W_c\left(x;\theta\right)=\min\left\{1;\frac{c}{\left\|A\left(\theta\right)\cdot\left[s\left(x;\theta\right)-a\left(\theta\right)\right]\right\|}\right\}$$

and *A* and *a* respectively are a dim(θ) x dim(θ) matrix and a dim(θ)-dimensional vector determined by the equations:

$$E\left[\varphi_c^{A,a}\left(x;\theta\right)\cdot\varphi_c^{A,a}\left(x;\theta\right)^T\right]=I$$ **((2) – ensures bounded IF)**

$$E\left[\varphi_c^{A,a}\left(x;\theta\right)\right]=0$$ **((3) – ensures Fisher consistency)**

$s\left(x;\theta\right)$ **is simply the score function,** $s\left(x;\theta\right)=\left[\partial f\left(x;\theta\right)/\partial\theta\right]\big/ f\left(x;\theta\right)$ **, so OBRE is defined in terms of a weighted standardized scores function, where** $W_c\left(x;\theta\right)$ **are the weights. *c* is a tuning parameter,** $\sqrt{\dim\left(\theta\right)}\le c\le\infty$ **, regulating from very robust to MLE, respectively.**

OBRE Defined:

- The weights make OBRE robust, but it maintains efficiency as close as possible to MLE (subject to its constraints) because it is based on the scores function. Hence, its name: "Optimal" B-Robust Estimator. The constraints – bounded IF and Fisher consistency – are implemented with *A* and *a*, respectively, which can be viewed as Lagrange multipliers. And *c* regulates the robustness-efficiency tradeoff: a lower *c* gives a more robust estimator, and $c = \infty$ is MLE. Bottom line: by minimizing the trace of the asymptotic covariance matrix, OBRE is maximally efficient for a given level of robustness, which is controlled by the analyst with *c*. Many choose *c* to achieve 95% efficiency relative to MLE, but this actual value for *c* depends on the model being implemented.

- Several versions of the OBRE exist with minor variations on exactly how they bound the IF. The OBRE defined above is the so-called "standardized" OBRE "which has proved to be numerically more stable" (see Alaiz and Victori-Feser, 1996). The "standardized" OBRE is used in this study.

5c. Robust Estimators: OBRE and CvM

OBRE Computed:

To compute OBRE, (1) must be solved under conditions (2) and (3), for a given tuning parameter value *c*, via Newton-Raphson (see D.J. Dupuis, 1998):

<u>STEP 1</u>: Decide on a precision threshold, *η*, an initial value for *θ*, and initial values *a* = 0 and $A=\sqrt{\left[J(\theta)^{-1}\right]^{T}}$ **where** $J(\theta)=\int s(x;\theta)\cdot s(x;\theta)^{T}\,dF_{\theta}(x)$ **is the Fisher Information.**

<u>STEP 2</u>: Solve for *a* and *A* in the following equations:

$$A^{T}A=M_{2}^{-1} \quad \text{and} \quad a=\int s(x,\theta)W_{c}(x,\theta)dF_{\theta}(x)\Big/\int W_{c}(x,\theta)dF_{\theta}(x)$$

where $M_{k}=\int\left[s(x;\theta)-a\right]\cdot\left[s(x;\theta)-a\right]^{T}\cdot W_{c}(x,\theta)^{k}\,dF_{\theta}(x),\ \text{k=1,2}$

which gives the "current values" of *θ*, *a*, and *A* used to solve the given equations.

<u>STEP 3</u>: Now compute M_1 **and** $\Delta\theta=M_{1}^{-1}\cdot\left\{\frac{1}{n}\cdot\sum_{i=0}^{n}\left[s(x_i;\theta)-a\right]\cdot W_{c}(x_i,\theta)\right\}$

<u>STEP 4</u>: If $\max_{j}\left|\frac{\Delta\theta_j}{\theta_j}\right|>\eta\ \ (j=1,2)$ **then** $\theta\rightarrow\theta+\Delta\theta$ **and return to <u>STEP 2</u>, otherwise stop.**

OBRE Computed:

- **The idea of the above algorithm is to first compute *A* and *a* for a given *θ* by solving (2) and (3). This is followed by a Newton-Raphson step given these two new matrics, and these steps are iterated until convergence is achieved.**

- **The above algorithm follows D.J. Dupuis (1998), who cautions on two points of implementation in an earlier paper by Alaiz and Victoria-Feser (1996):**

  - **Alaiz and Victoria-Feser (1996) state that integration can be avoided in the calculation of *a* in STEP 2 and M_1 in STEP 3, but Dupuis (1998) cautions that the former calculation of *a* requires integration, rather than a weighted average from plugging in the empirical density, or else (1.a) will be satisfied by all estimates.**

  - **Also, perhaps mainly as a point of clarification, Dupuis (1998) clearly specifies $\max_j \left| \frac{\Delta\theta_j}{\theta_j} \right| > \eta \;\left(j = 1,2 \right)$ in STEP 4 rather than just $\Delta\theta > \eta$ as in Alaiz and Victoria-Feser (1996).**

- **The initial values for *A* and *a* in STEP 1 correspond to the MLE.**

5c. Robust Estimators: OBRE and CvM

OBRE Computed:

- **The algorithm converges if initial values for *θ* are reasonably close to the ultimate solution. Initial values can be MLE, or a more robust estimate from another estimator, or even an OBRE estimate obtained with *c* = large and initial values as MLE, which would then be used as a starting point to obtain a second and final OBRE estimate with *c* = smaller. In this study, MLE estimates were used as initial values, and no convergence problems were encountered, even when the loss dataset contained 6% arbitrary deviations from the assumed model.**

- **Note that the weights generated and used by OBRE, W_c , can be extremely useful for another important objective of robust statistics – outlier detection. Within the OpRisk setting, this can be especially useful for determining appropriate "units of measure" (uom), the grouping of loss events by some combinations of business unit and event type, each uom with the same (or close) loss distribution. As discussed below, the extreme quantiles that need to be estimated for regulatory capital and economic capital purposes are extremely sensitive to even slight changes in the variability of the parameter estimates. This, along with the a) unavoidable tradeoff between statistical power (sample size) and homogeneity; b) loss-type definitional issues; and c) remaining heterogeneity within units of measure even under ideal conditions, all make defining units of measure an extremely challenging and crucial task; good statistical methods can and should be utilized to successfully execute on this challenge.**

5c. Robust Estimators: OBRE and CvM

CvM Defined:

The Cramér von Mises estimator is a "minimum distance" estimator (MDE), yielding the parameter value of the assumed distribution that minimizes its distance from the empirical distribution. Given the CvM statistic $W^2(\theta)$ in its common form,

$$W^2(\theta) = \frac{1}{n} \cdot \sum_{i=1}^{n} \left[F_n(x_i) - F_\theta(x_i) \right]^2$$

where F_n is the empirical distribution and F_θ is the assumed distribution, the minimum CvM estimator (MCVME) is that value $\hat{\theta}$ of θ, for the given sample, that minimizes $W^2(\theta)$:

$$\hat{\theta}_{MCVME} = \underset{\theta}{\arg\min} \left\{ n \cdot \int \left[F_n(x) - F_\theta(x) \right]^2 dF_\theta(x) \right\}$$

5c. Robust Estimators: OBRE and CvM

CvM Computed:

The computational formula typically used to calculate the MCVME is:

$$W^2(\theta) = \frac{1}{12n} \cdot \sum_{s=1}^{n} \left[F_\theta\left(x_{(s)}\right) - \frac{2s-1}{2n} \right]^2$$

where $x_{(s)}$ **is the ordered (s)'th value of x.**

- **MCVME is an M-class estimator, and as such it is consistent and asymptotically normal.**
- **MDE's are very similar conceptually, and typically differ in how they weight the data points. For example, Anderson-Darling, another MDE, weights the tail more than does CvM. CvM is very widely used, perhaps the most widely used MDE, hence its inclusion.**
- **Before presenting results comparing MLE to OBRE and CvM, I talk briefly about (left) truncation, and reemphasize its analytic and empirical importance in this setting.**

6. Truncation Matters, the Threshold Matters

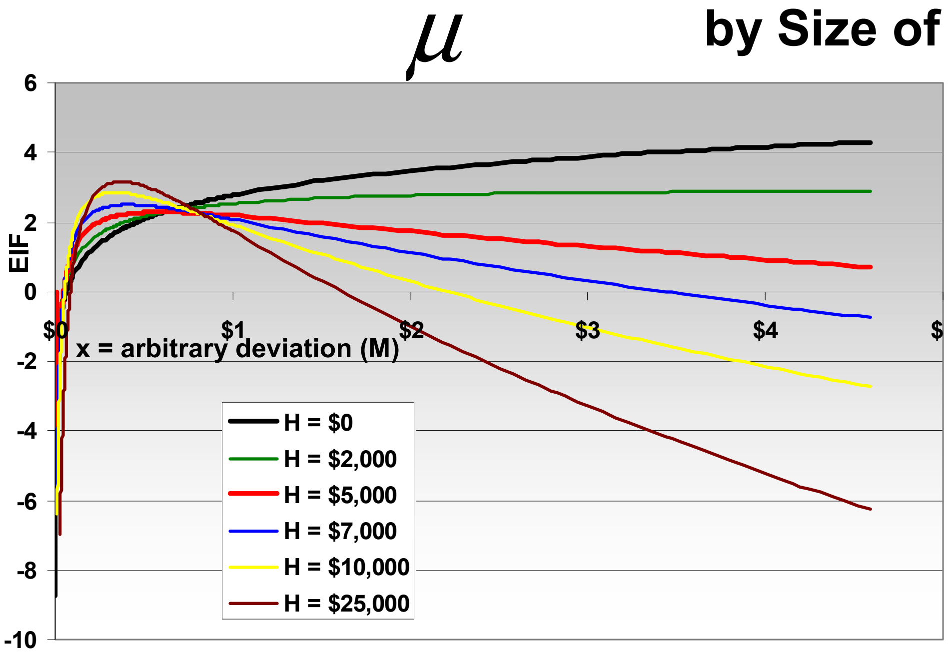
- **The effects of a collection threshold on parameter estimation can be unexpected, even counterintuitive, both in the magnitude of the effect, and its direction.**
- **Note first that given the size of the economic and regulatory capital estimates generated from severity distribution modeling (into the hundreds of millions and even billions of dollars), the size of the thresholds appear tiny, and the % of the non-truncated distributions that fall below the thresholds do not appear shockingly large, either (assuming, of course, that the loss distribution below the threshold is the same as that above it, which is solely a heuristic assumption here).**
- **However, the effects of (left) truncation on MLE severity distribution parameter estimates can be dramatic, even for low thresholds.**
- **Not only are the effects dramatic, but arguably very unexpected. The entire shape AND DIRECTION of some of the IFs change as does the threshold, over relatively small changes in the threshold value.**
- **Note that this is not merely a sensitivity to simulation assumptions, but rather, an analytical result.**

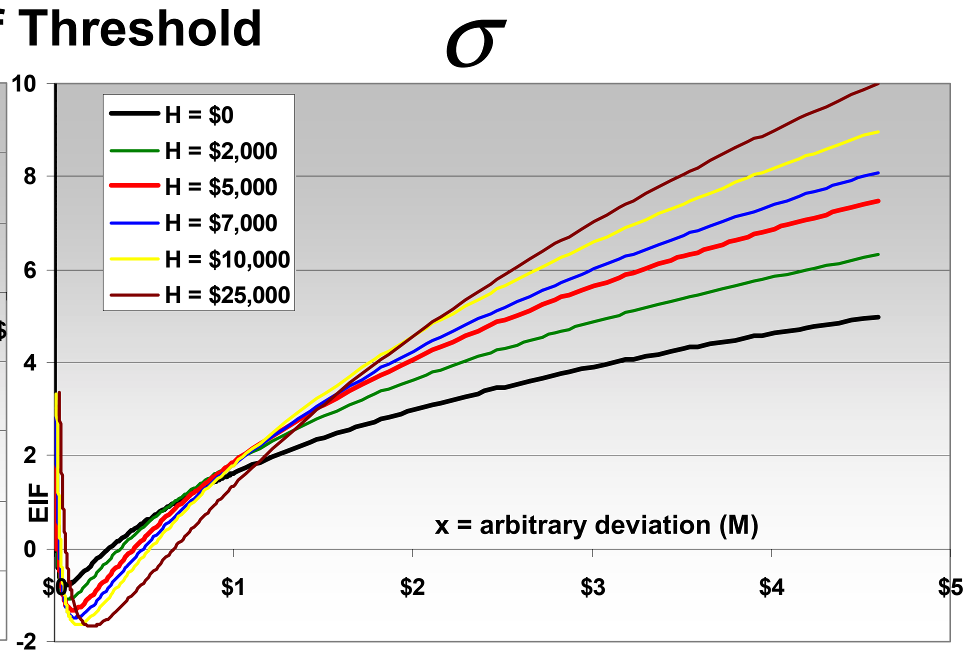
Collection Threshold	LogNormal (μ=11, σ=2) % Below	LogGamma (a=35.5, b=3.25) % Below
$1,000	0.7%	2.0%
$2,000	2.4%	4.5%
$3,000	4.4%	6.7%
$4,000	6.5%	8.8%
$5,000	8.6%	10.7%
$10,000	17.6%	18.5%
$15,000	24.6%	24.4%
$20,000	30.2%	29.2%
$25,000	34.9%	33.1%

6. Truncation Matters, the Threshold Matters

- **The effects of a collection threshold on parameter estimation can be unexpected, even counterintuitive, both in the magnitude of the effect, and its direction.**

EIF of Truncated LogNormal (μ = 11, σ = 2) MLE Parameter Estimates: by Size of Threshold





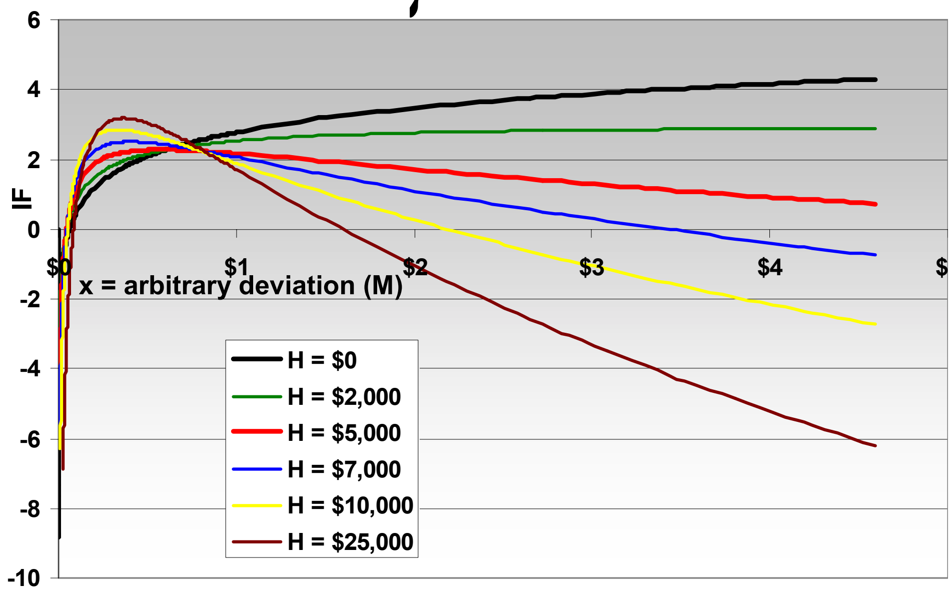
- **Note the NEGATIVE covariance between parameters induced by (left) truncation. Many would call this unexpected, if not counter-intuitive: the location parameter, *μ*, DECREASES under larger and larger arbitrary deviations.**

6. Truncation Matters, the Threshold Matters

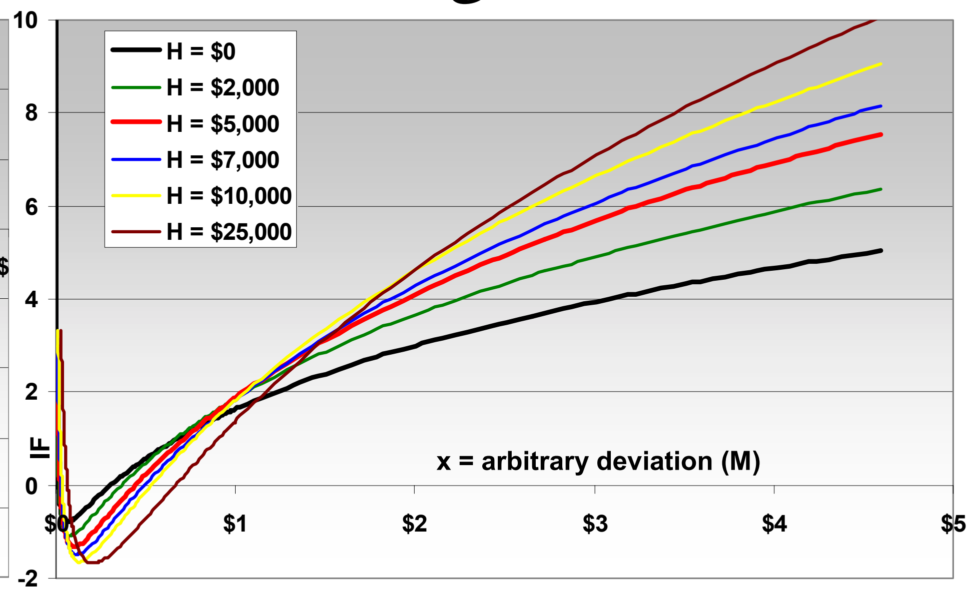
- The effects of a collection threshold on parameter estimation can be unexpected, even counterintuitive, both in the magnitude of the effect, and its direction.

IF of Truncated LogNormal (μ = 11, σ = 2) MLE Parameter Estimates: by Size of Threshold

μ



σ



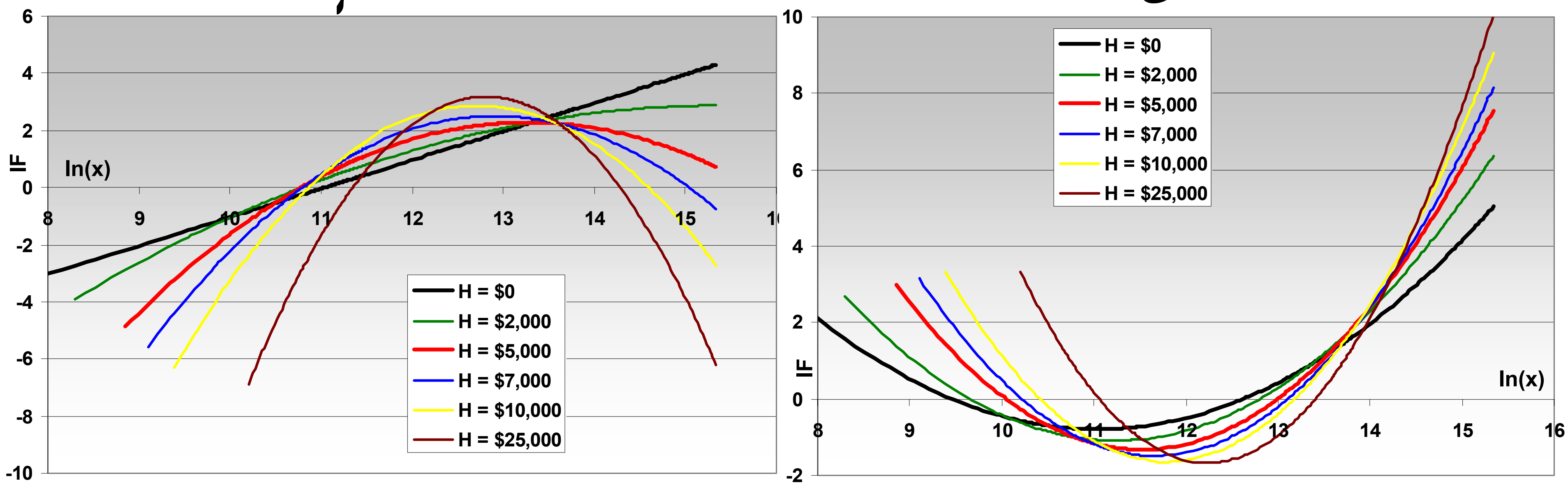
- Note the NEGATIVE covariance between parameters induced by (left) truncation. Many would call this unexpected, if not counter-intuitive: the location parameter, *μ*, DECREASES under larger and larger arbitrary deviations.

6. Truncation Matters, the Threshold Matters

- **The effects of a collection threshold on parameter estimation can be unexpected, even counterintuitive, both in the magnitude of the effect, and its direction.**

IF of Truncated LogNormal (μ = 11, σ = 2) MLE Parameter Estimates: by Size of Threshold

μ σ



Log Scale

- **Note the log-linear** $IF_{\mu}\left(x;\mu,\sigma;MLE\right)=\ln\left(x\right)-\mu$ **under no truncation is analogous to the** $IF_{\mu}\left(x;\mu,\sigma;MLE\right)=x-\mu$ **obtained earlier under the normal distribution.**

ORX ASSOCIATION

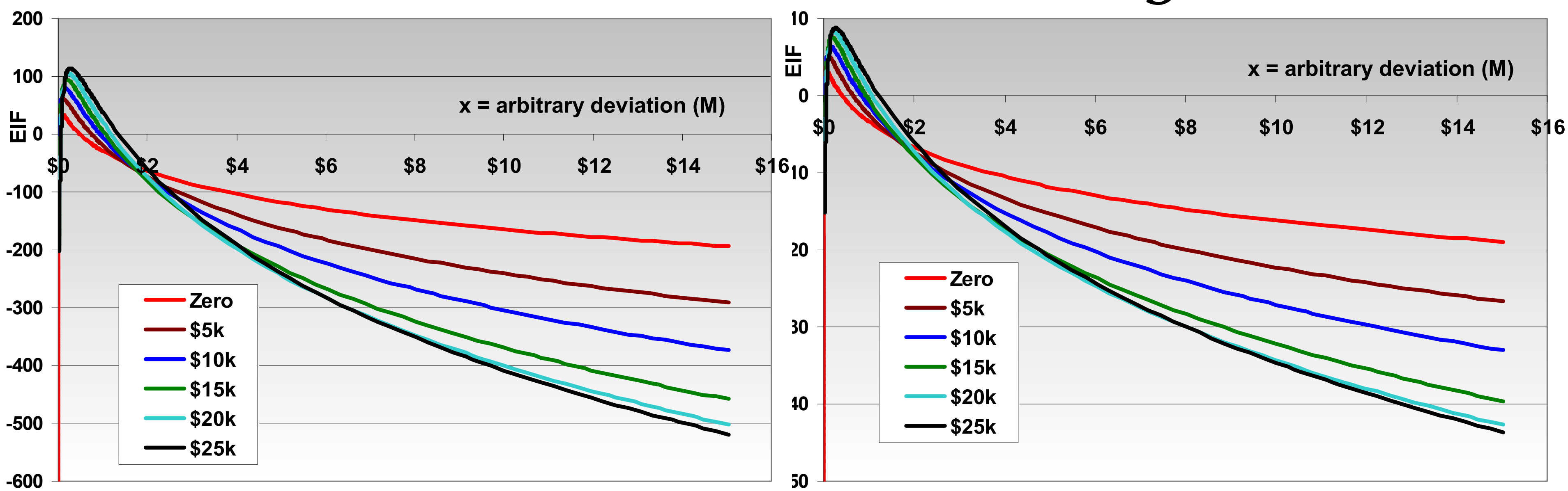
DataMineIt℠

6. Truncation Matters, the Threshold Matters

- **The effects of a collection threshold on parameter estimation can be unexpected, even counterintuitive, both in the magnitude of the effect, and its direction.**

EIF of Truncated LogGamma (a = 35.5, b = 3.25) MLE Parameter Estimates: by Size of Threshold

a *b*



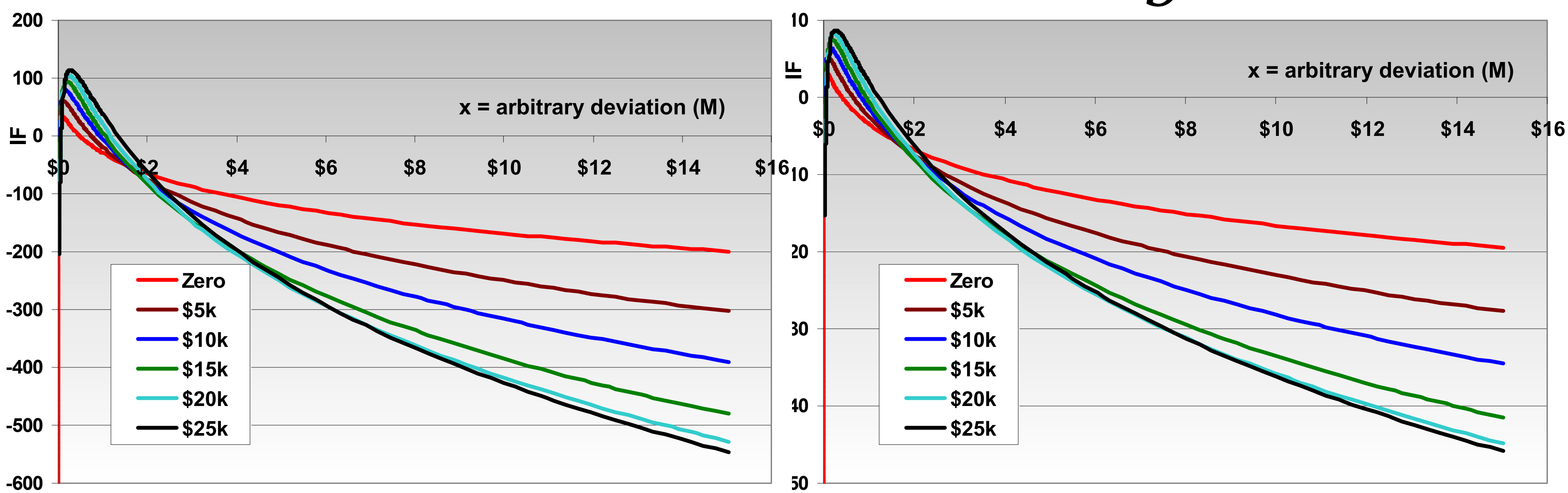
- **Note that for the LogGamma, (left) truncation augments the already POSITIVE covariance between parameters.**

6. Truncation Matters, the Threshold Matters

- **The effects of a collection threshold on parameter estimation can be unexpected, even counterintuitive, both in the magnitude of the effect, and its direction.**

IF of Truncated LogGamma (a = 35.5, b = 3.25) MLE Parameter Estimates:

a **by Size of Threshold** *b*



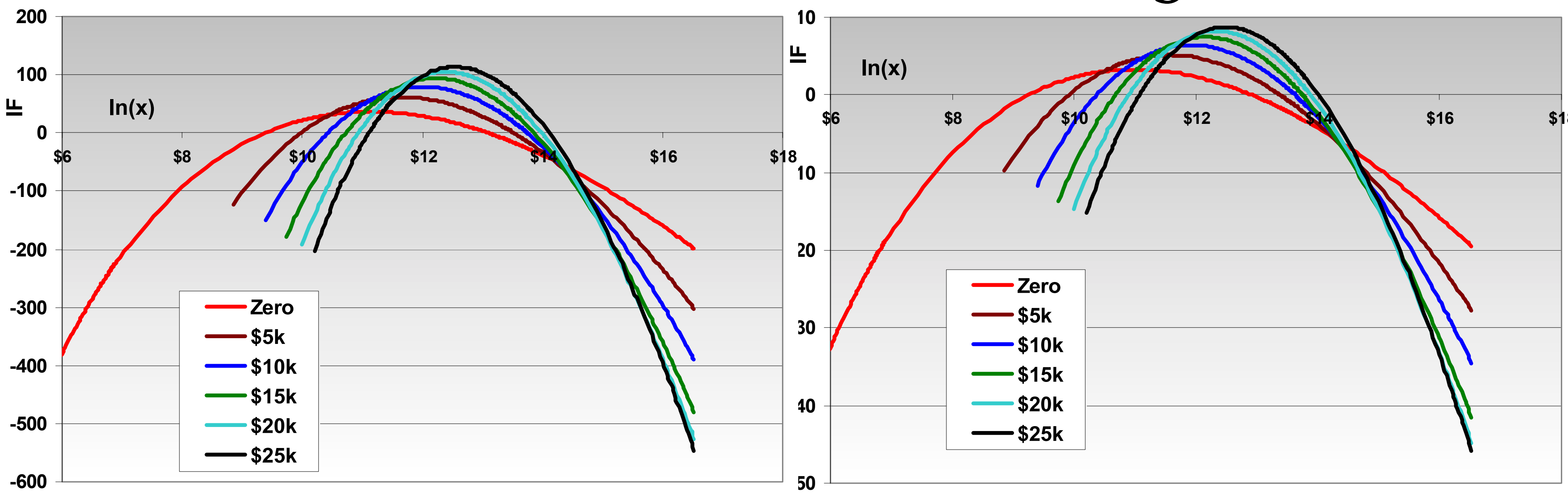
- **Note that for the LogGamma, (left) truncation augments the already POSITIVE covariance between parameters.**

6. Truncation Matters, the Threshold Matters

- **The effects of a collection threshold on parameter estimation can be unexpected, even counterintuitive, both in the magnitude of the effect, and its direction.**

IF of Truncated LogGamma (a = 35.5, b = 3.25) MLE Parameter Estimates:

a **by Size of Threshold** *b*



Log Scale

- **Note that for the LogGamma, (left) truncation augments the already POSITIVE covariance between parameters.**

6. Truncation Matters, the Threshold Matters

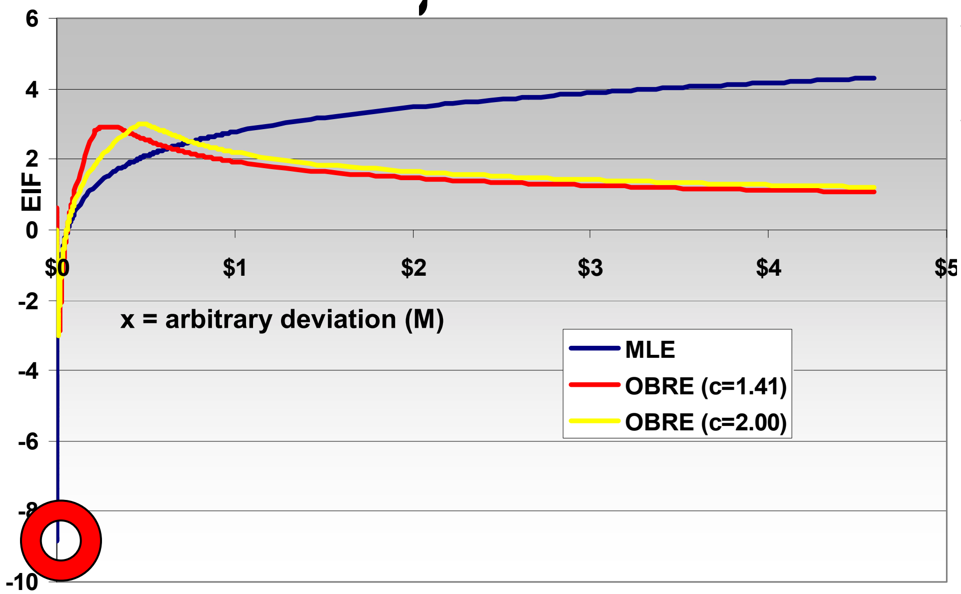
- **These arguably unexpected, and even counterintuitive results, both in the magnitude of the effect of (left) truncation and sometimes its direction, not to mention the potential for dramatic change in the relationship between parameters of the same distribution, would appear to explain the extreme sensitivity of MLE estimators under truncation reported in the literature, which has perplexed some researchers.**

7a. Results: Disproportionate Impact of Left Tail

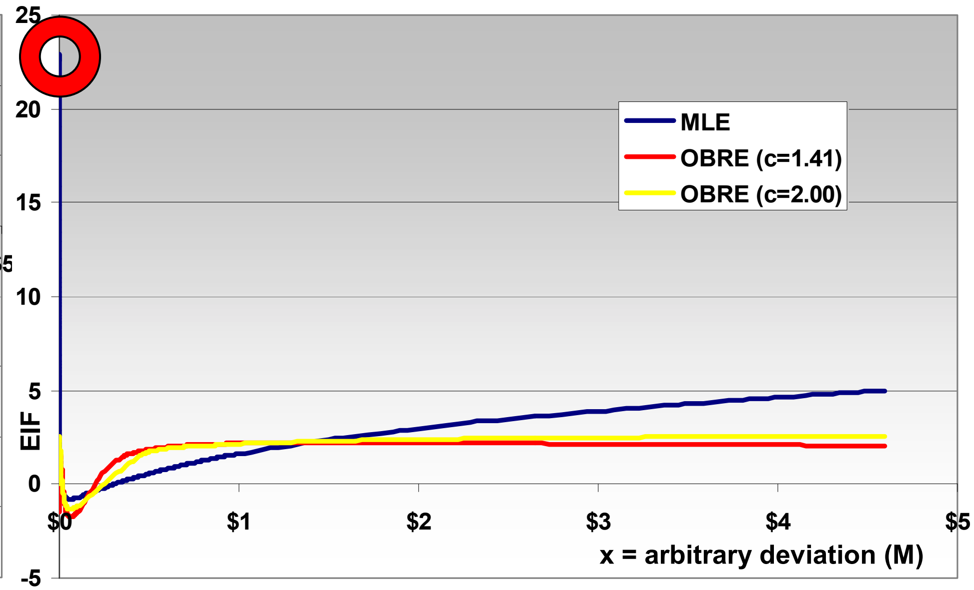
- **NOTE: Arbitrary deviations from the assumed model do not have to be large in absolute value to have a large impact on MLE estimates. The IF is a useful tool for spotting such counter-intuitive and important effects that are potentially devastating to the estimation process.**

EIF of LogNormal (n=250, μ = 11, σ = 2) Parameter Estimates: OBRE v. MLE

μ



σ

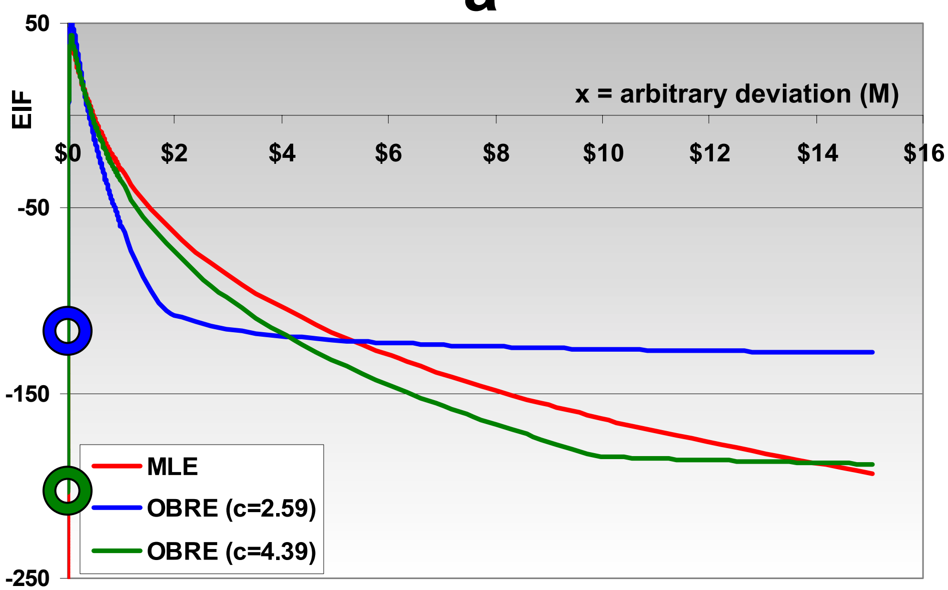


7a. Results: Disproportionate Impact of Left Tail

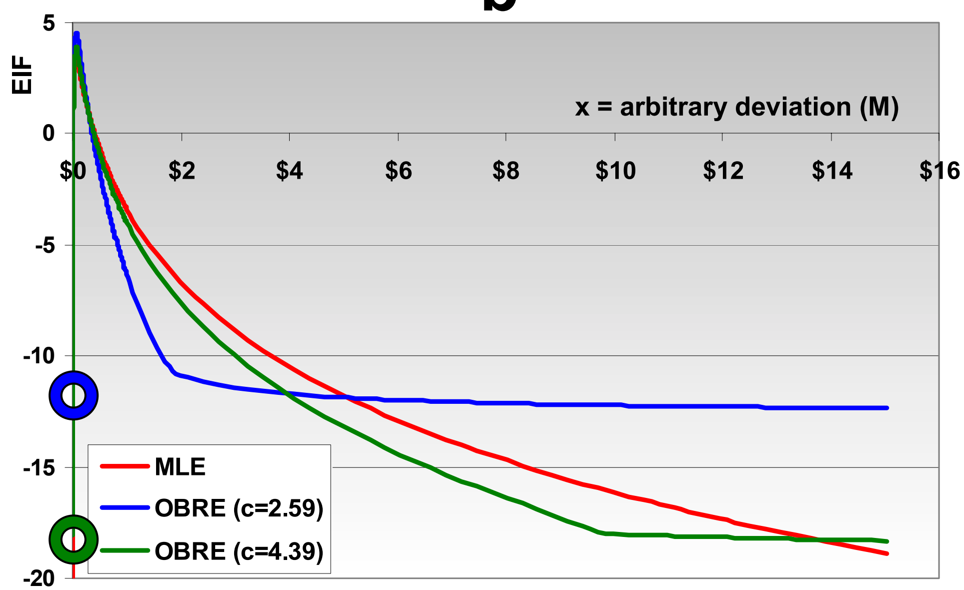
- **NOTE: Arbitrary deviations from the assumed model do not have to be large in absolute value to have a large impact on MLE estimates. The IF is a useful tool for spotting such counter-intuitive and important effects that are potentially devastating to the estimation process.**

EIF of LogGamma (n=250, a = 35.5, b = 3.25) Parameter Estimates: OBRE v. MLE

a



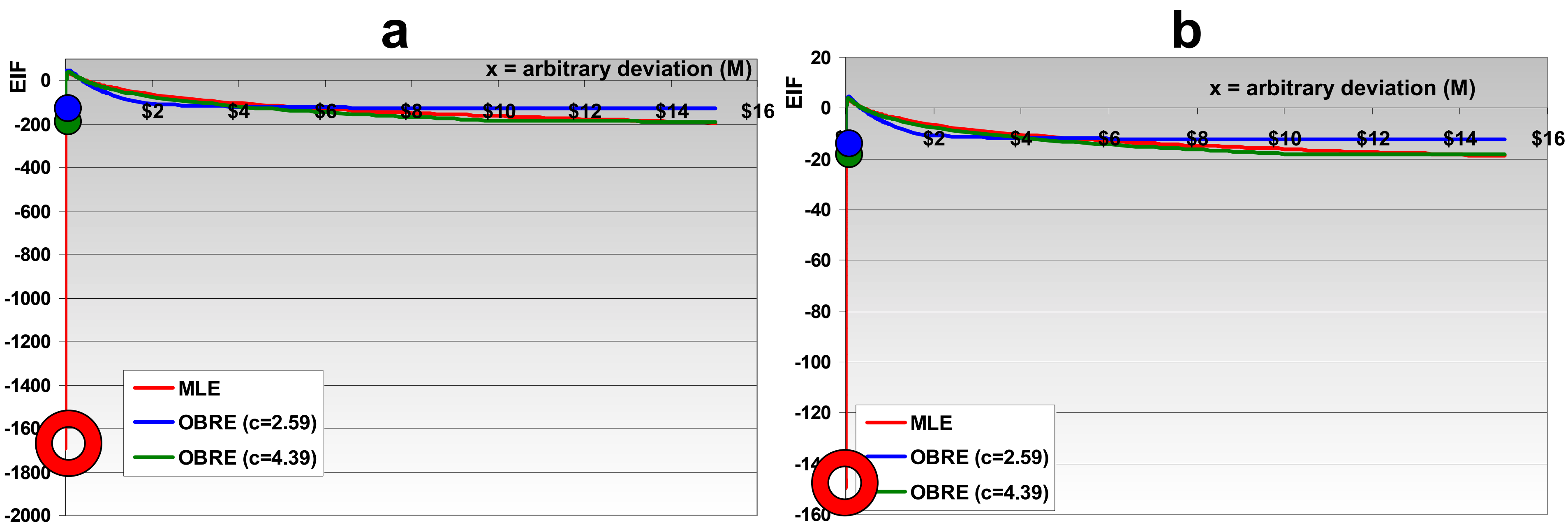
b



7a. Results: Disproportionate Impact of Left Tail

- **NOTE: Arbitrary deviations from the assumed model do not have to be large in absolute value to have a large impact on MLE estimates. The IF is a useful tool for spotting such counter-intuitive and important effects that are potentially devastating to the estimation process.**

EIF of LogGamma (n=250, a = 35.5, b = 3.25) Parameter Estimates: OBRE v. MLE

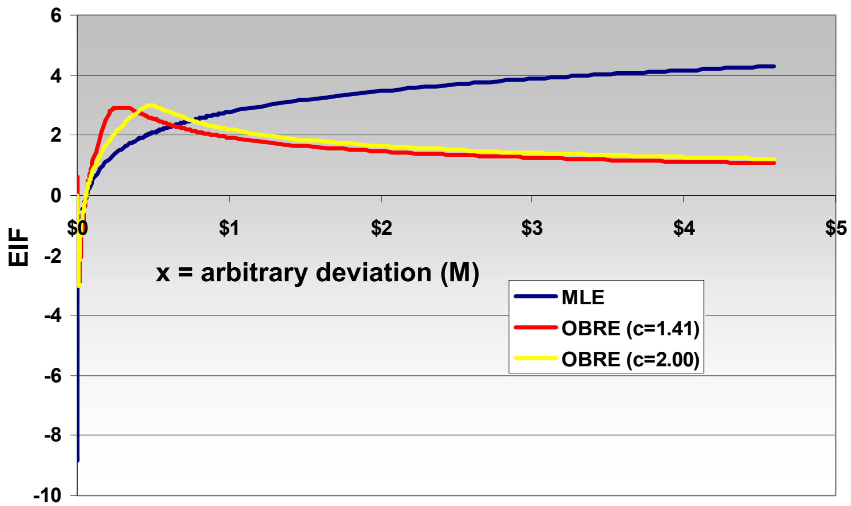


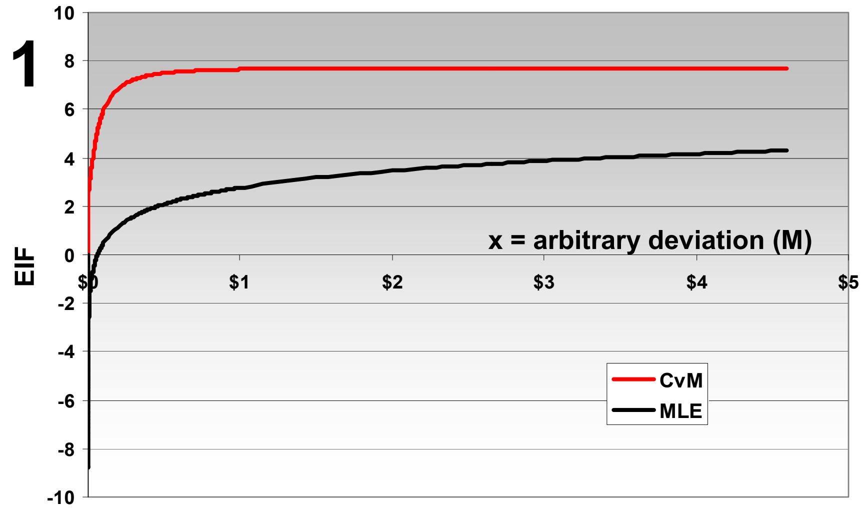
7a. Results: LogNormal Distribution (n=250)

EIF's: OBRE v. MLE by Deviation

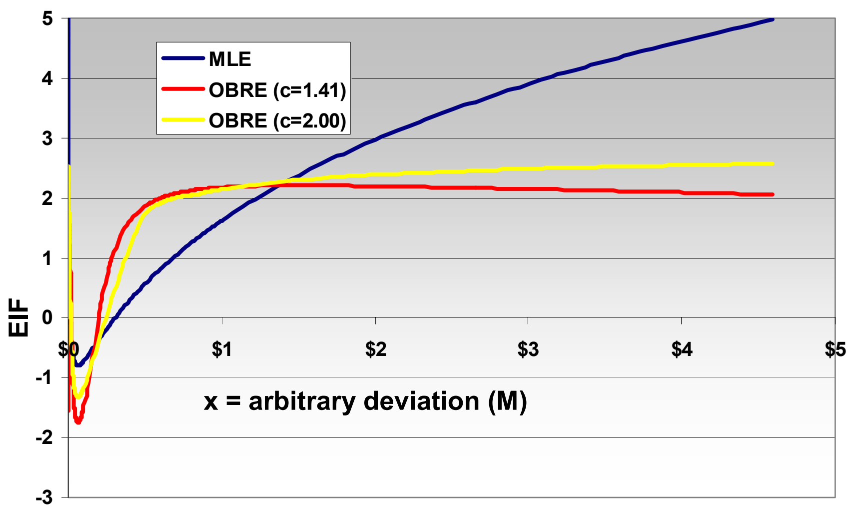
EIF's: CvM vs. MLE by Deviation

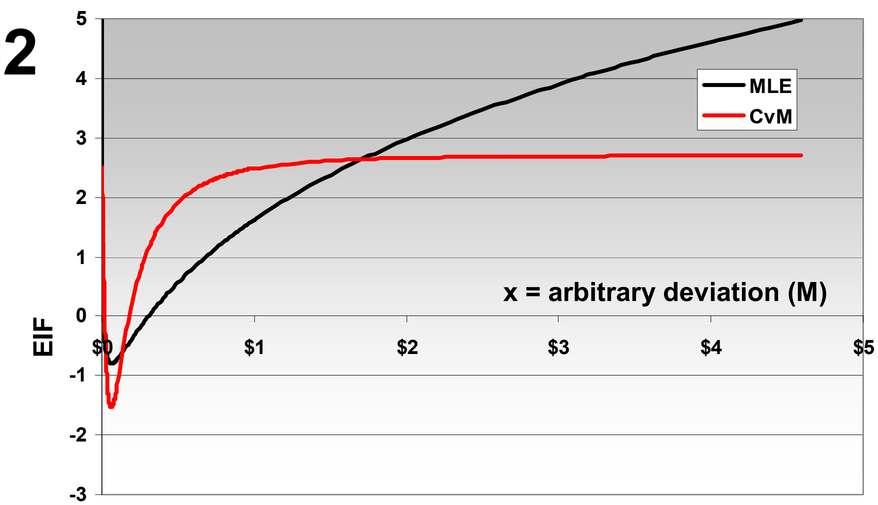
μ=11





σ=2



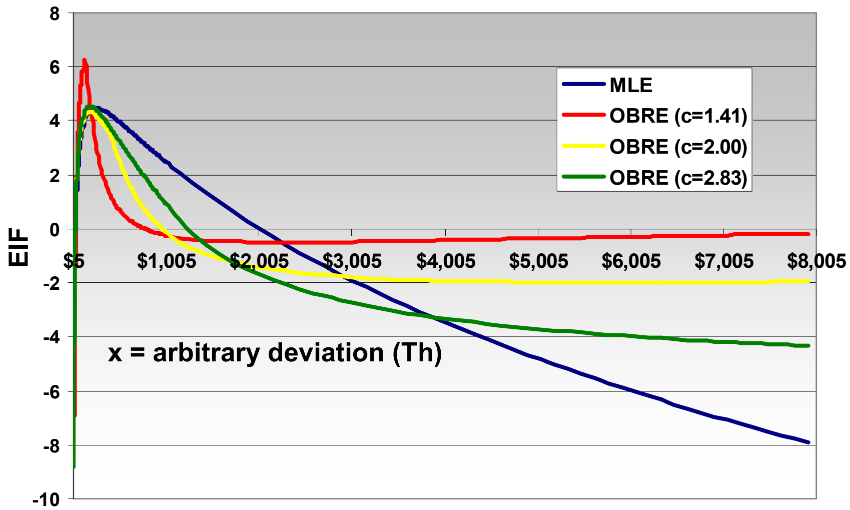


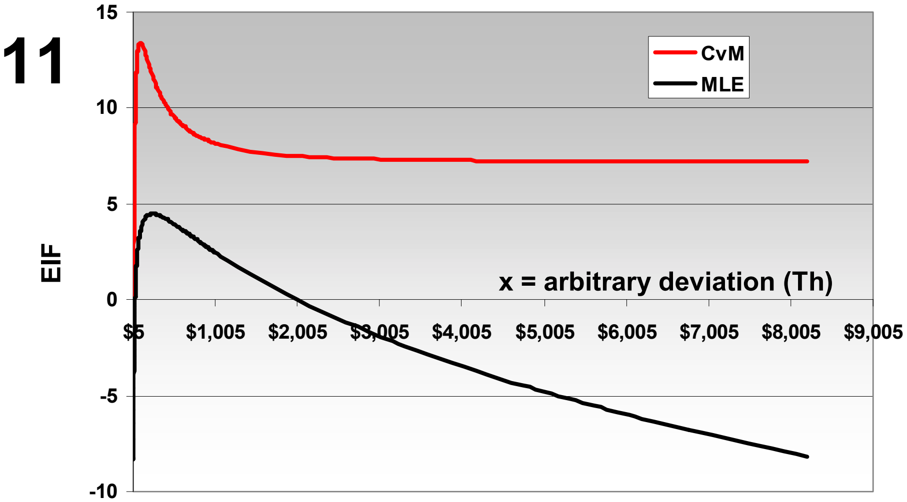
7a. Results: Truncated LogNormal (n=250, H=$5,000)

EIF's: OBRE v. MLE by Deviation

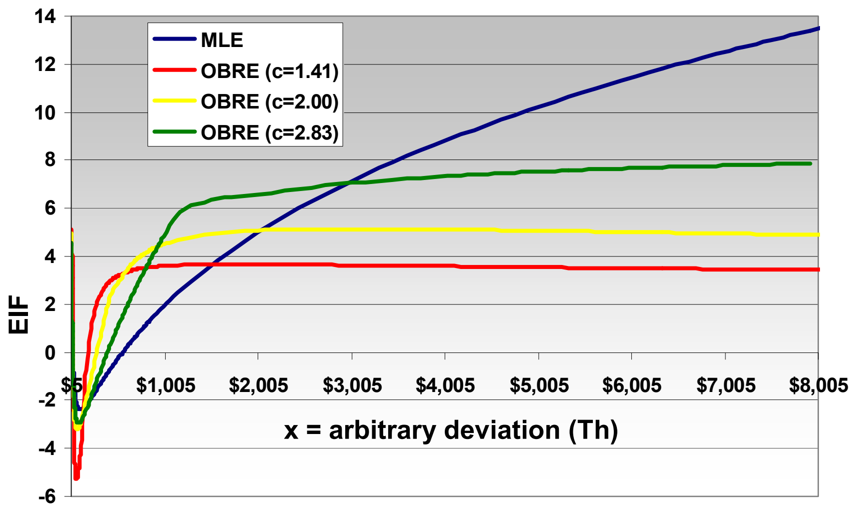
EIF's: CvM vs. MLE by Deviation

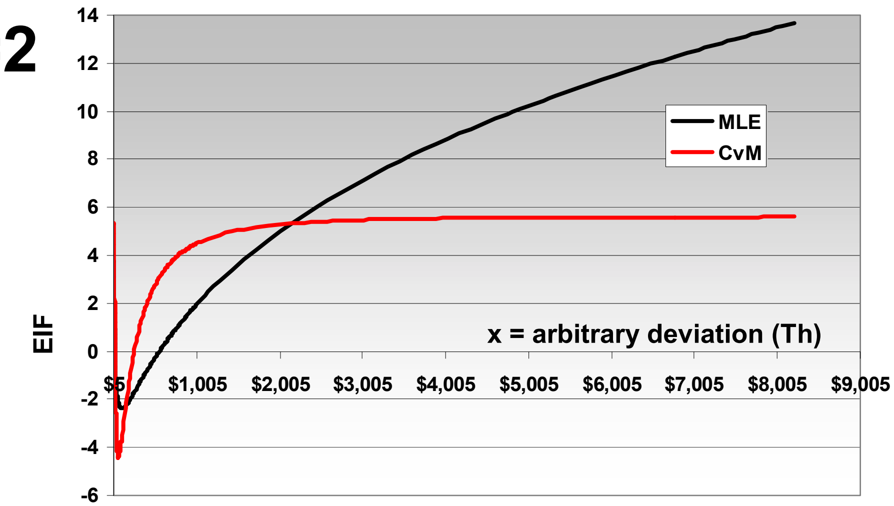
μ=11





σ=2





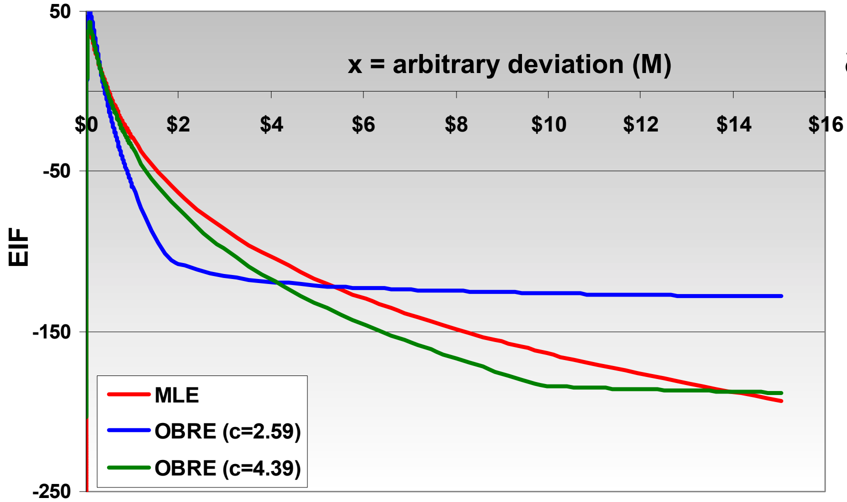
ORX ASSOCIATION

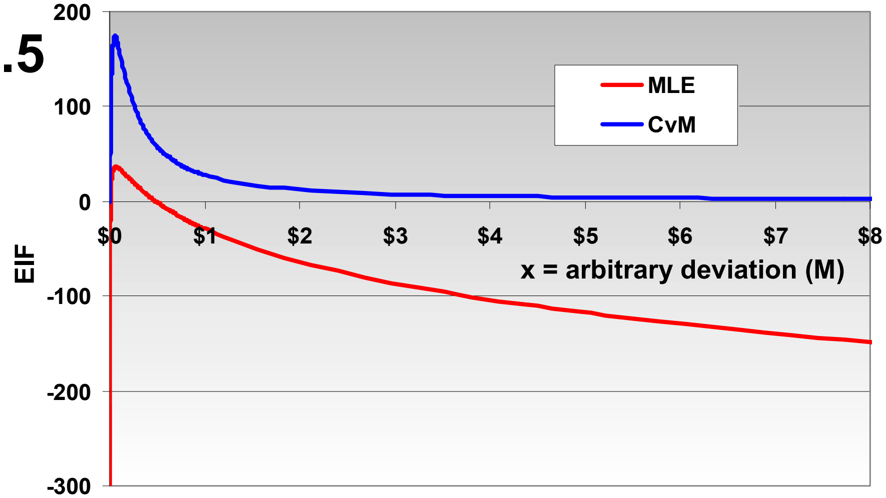
DataMineIt℠

7a. Results: LogGamma Distribution (n=250)

EIF's: OBRE v. MLE by Deviation

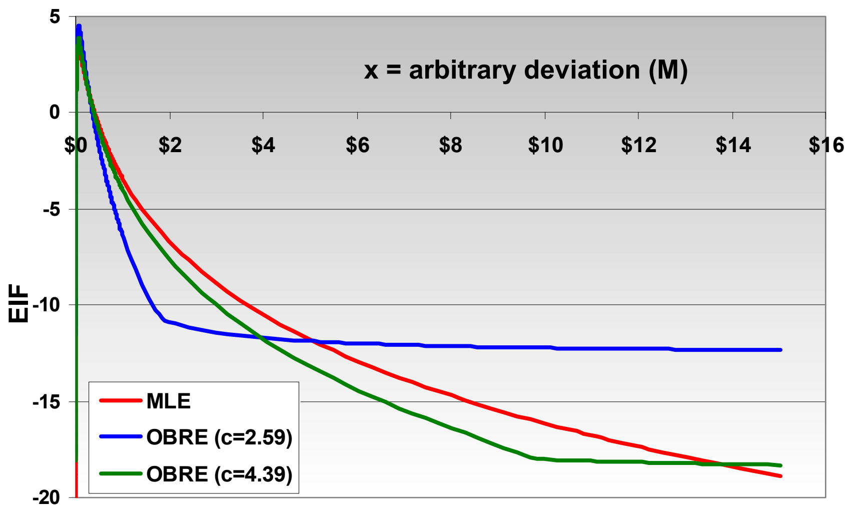
EIF's: CvM vs. MLE by Deviation

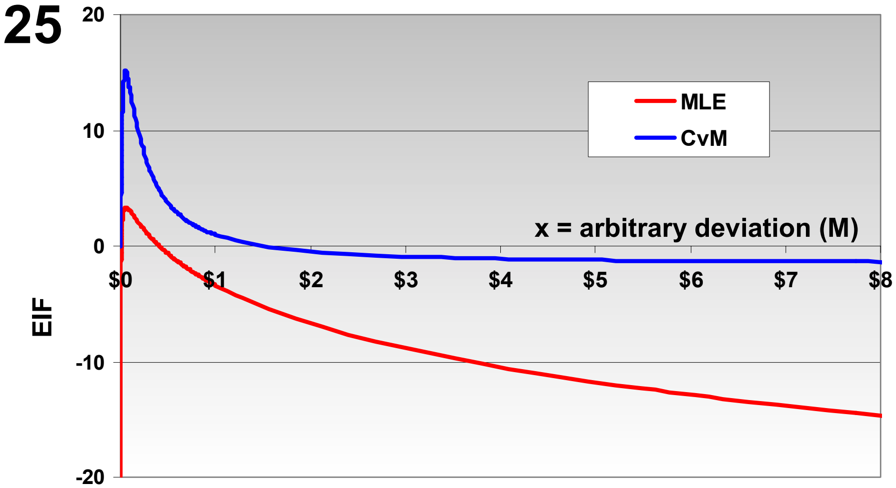




a=35.5

b=3.25



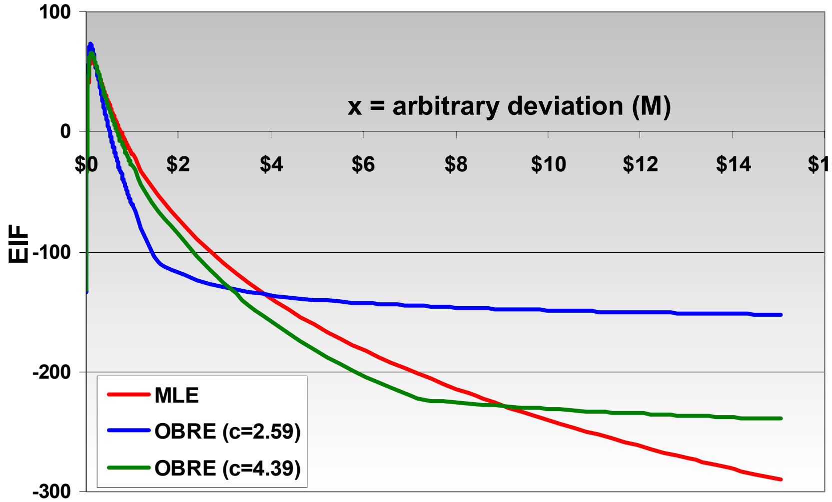


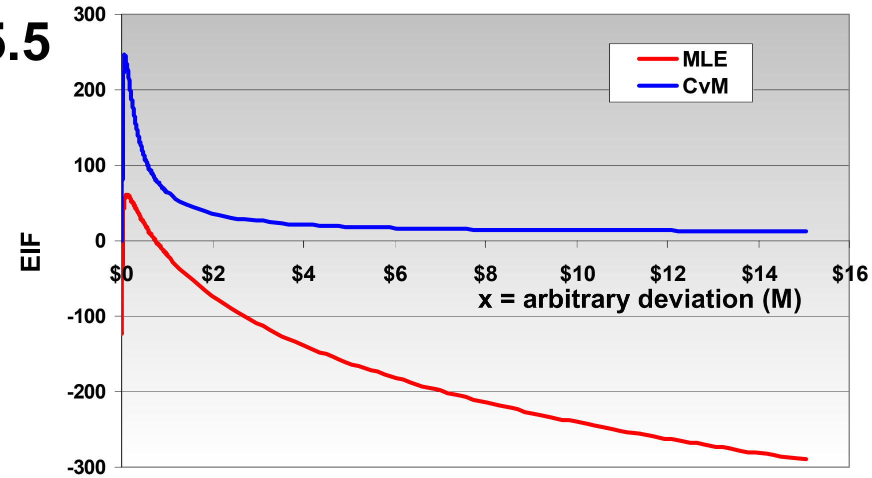
7a. Results: Truncated LogGamma (n=250, H=$5,000)

EIF's: OBRE v. MLE by Deviation

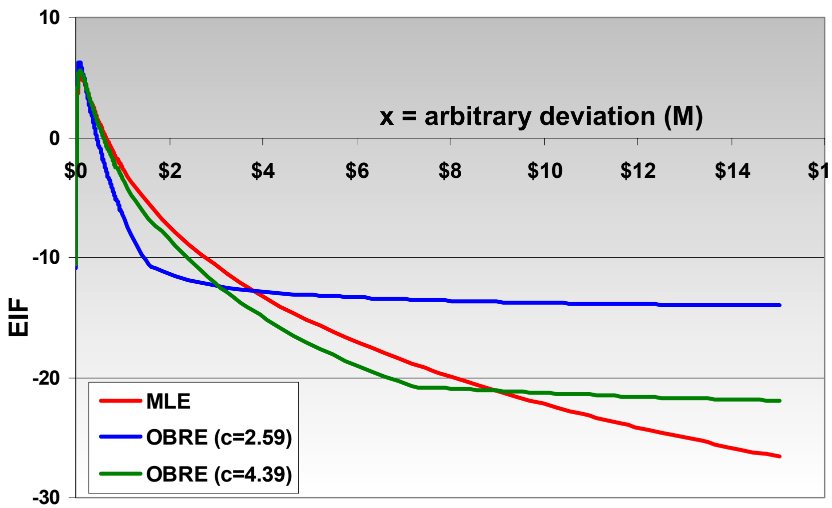
EIF's: CvM vs. MLE by Deviation

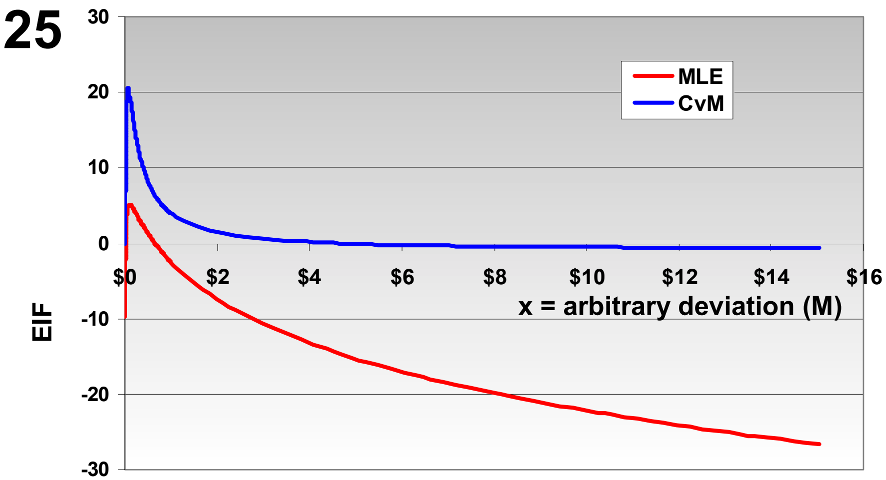
a=35.5





b=3.25





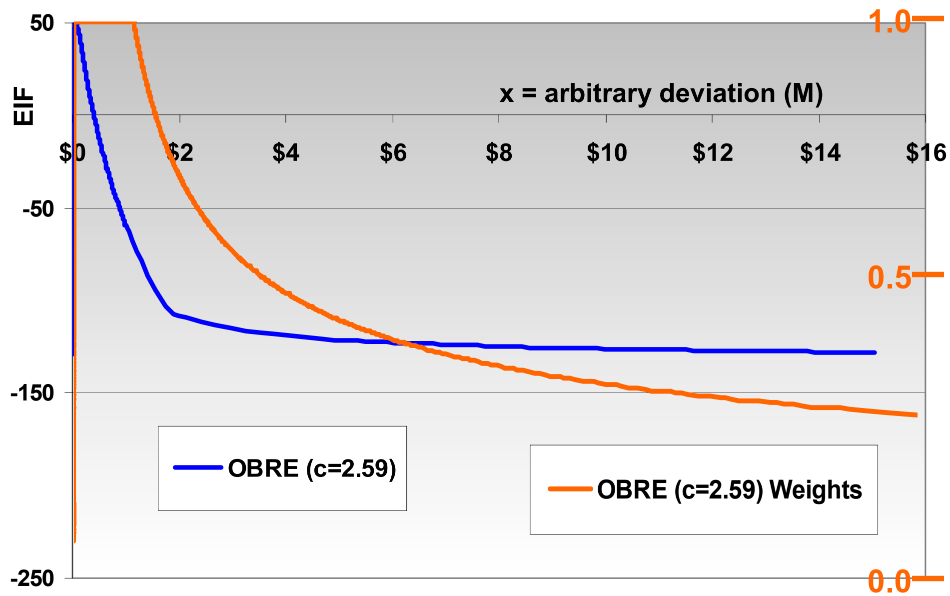
ORX ASSOCIATION

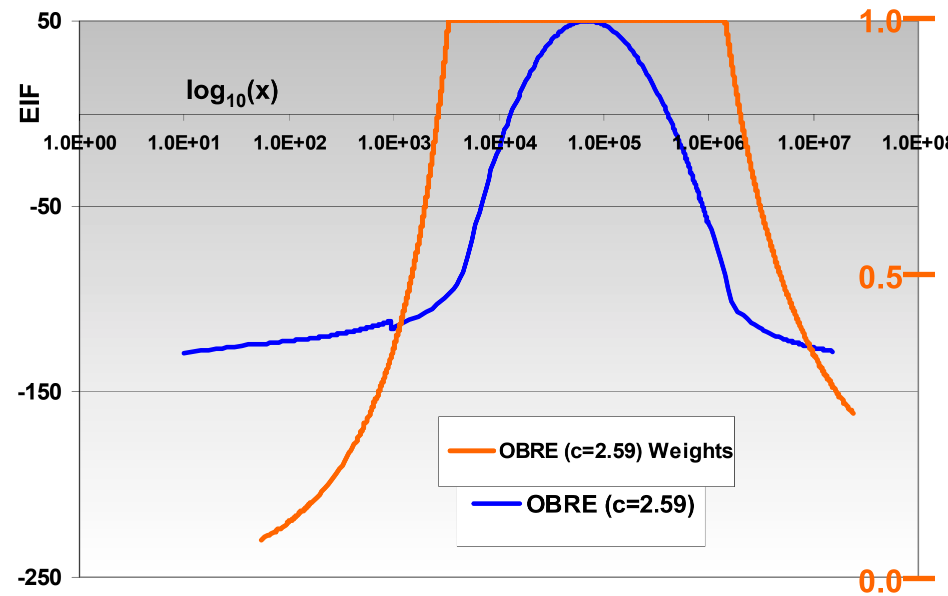
DataMineIt℠

7a. Results: OBRE Weights

- **OBRE Weights, one for each data point, range from one to zero, approaching the latter as values deviate from the assumed distribution.**

LogGamma (n=250, a=35.5, b=3.25)
OBRE EIF of *a* vs. OBRE Weights by Arbitrary Deviation





Log Scale

7a. Results: OBRE Weights

- **OBRE Weights contain very valuable information: they are indicators of the degree to which a particular data point (<u>within the context of the data sample at hand!</u>) deviates from the assumed statistical model.**
- **As such they can be used for outlier detection, unit-of-measure construction, and possibly in the parameter estimation process itself.**
- **For the latter, they are arguably superior to "trimming" (observation deletion) based on sample quantiles, maximum/minimum k observations, absolute deviations, or other relatively arbitrary and inflexible metrics.**

7b. Results: SLA Simulations

The simulations generate MLE parameter estimates vs. OBRE and CvM parameter estimates. Each is used to generate a distribution of capital estimates based on SLA.

- **SLA (Single-Loss Approximation): Parameter estimates are used in Böcker & Klüppelberg's (2005) SLA formula to obtain capital estimates, and the distributions of these capital estimates are compared.** $C_{\alpha} \approx F^{-1}\left(1-\frac{1-\alpha}{\lambda}\right)+(\lambda-1)\mu$

  $\alpha = 0.999$; and $\lambda = 25$ arbitrarily.

- **Sample Size: n = 250 was chosen as a reasonable size for many units-of-measure. Depending on the bank, some will have larger n, some smaller, but if the results were not useful for this n = 250, then sample size would have been a real issue with these methods going forward, so that is why n = 250 was selected.**

- **Severity Distributions: the LogNormal and the LogGamma. Both are commonly used in this setting, but they are very distinct distributions, with the latter being more heavy-tailed (see table). Results obtained from other distributions will be included in journal-format version of this paper.**

X%Tile	LogNormal (μ=11, σ=2)	LogGamma (a=35.5, b=3.25)
50.0000%	$59,874	$50,045
75.0000%	$230,724	$179,422
90.0000%	$776,928	$614,477
95.0000%	$1,606,723	$1,333,228
99.0000%	$6,278,840	$6,162,960
99.9000%	$28,932,168	$38,778,432
99.9700%	$57,266,640	$92,087,922
99.9960%	$159,698,811	$355,104,952
99.9988%	$279,358,818	$760,642,911

7b. Results: SLA Simulations

- **Truncation: The Truncated LogNormal and Truncated LogGamma, with a collection threshold of \$5k, are included.**
- **Parameter values: These were choosen (both LogNormal and Truncated LogNormal, μ = 11, σ = 2, and both LogGamma and Truncated LogGamma a = 35.5, b = 3.25) so as to reflect a) fairly large differences between the Lognormal and the LogGamma; b) general empirical realities based on OpRisk work I've done (but not proprietary results); c) yet, some "stretching" vis-à-vis fairly large (but still realistic) parameter values (the base distributions have means of about \$442k and \$467k, respectively). Obviously, for any given setting, all estimation methods should be tested extensively for parameter value ranges relevant to the specific estimation effort.**
- **Arbitrary Deviations: Mixture distributions are used to test the robustness of the estimators to deviations from iid data. Three scenarios are studied: 6% Left tail contamination, 6% Right tail contamination, and 3% Left tail + 3% Right tail contamination. For the LogNormal, the left and right tail contamination is drawn from LogNormal(μ = 9.5, σ = 2) and LogNormal(μ = 11.576, σ = 2), respectively, and for the LogGamma, the left and right tail contamination is drawn from LogGamma(a = 31.8, b = 3.25) and LogGamma(a = 37, b = 3.25), respectively. Each of these has a mean that deviates just under \$350,000 from the respective base distributions.**

7b. Results: SLA Simulations

- **<u>OBRE value of *c*</u>: For OBRE, different values for *c*, the tuning parameter, were used with the given parameter values, and those which provided the most obviously appropriate tradeoff between accuracy and precision of the corresponding SLA capital estimates were used. Algorithms that may be useful to obtain these values are discussed below.**

- **<u>OBRE Starting Values</u>: MLE estimates were used as starting point for the OBRE algorithm, and for this study, no convergence problems were encountered. That said, values of η, *c*, n, and the distribution parameters all are very interrelated, and like any convergence algorithm, must be carefully monitored. For example, values of $\eta = 0.01$ were sufficient for LogNormal parameter estimation, but for LogGamma estimation, $\eta = 0.005$ and even $\eta = 0.0001$ were sometimes required due to its longer tail and the need for greater precision. Such variation is typical of convergence algorithms, so their responsible use requires an awareness of these issues. While starting values are sometimes noted in the literature as being important for the convergence of OBRE algorithms, this emphasis may be due to the relatively small sample sizes (as low as n = 40) being used in some of those studies (see Horbenko, Ruckdeschel, & Bae, 2011).**

- **<u>CvM Starting values</u>: A wide range of parameter values were provided for the Gaussian quadrature optimization algorithm. No convergence issues were encountered with the LogNormal and Truncated LogNormal distributions, but that was not the case in fully a third of the LogGamma and Truncated LogGamma distributions where second-order optimality conditions were violated.**

7b. Results: SLA Simulations

- First, before addressing the issue of developing inferential algorithms, one might ask whether by using robust statistics, we can even "back into" SLA results that are more accurate, all else equal, compared to MLE. That is, do there even exist, and can we find, values of the tuning parameter that will provide SLA estimates with less bias than MLE while not appreciably increasing variance? Given the difficulty of high quantile estimation in general, let alone in the OpRisk setting, this certainly is not a given, and it is the focus of the next several slides.
- The SLA#s in Table 1(a-d) were obtained by "backing into" optimal values of the tuning parameter knowing the true SLA value ex ante. Informal robustness tests were then conducted ex post to provide an initial assessment as to the feasibility of developing a process for statistical inference. One such possible process is sketched in the pages following Tables 1(a-d).
- NOTE: In Tables 1(a-d), note the large bias in the expected value of MLE-based capital estimates, under iid data with no contamination, due to Jensen's inequality. This bias grows with the heaviness of the tail of the severity distribution.

7b. Results: SLA Simulations

Table 1: Summary of SLA Estimates "Backed Into" with Optimal Tuning Parameter and Weight Usage for OBRE

	0% Deviation	3% Each Tail	6% Left Tail	6% Right Tail
LogNormal				
True SLA at 99.996%tile	$170,317,921	$173,118,560	$165,323,008	$180,654,136
OBRE Closer v. MLE	**$6,832,168**	**$7,748,825**	**$7,360,382**	**$2,773,099**
Truncated LogNormal				
True SLA at 99.996%tile	$180,486,144	$183,180,240	$175,278,136	$190,682,320
OBRE Closer v. MLE	**$20,759,747**	**$15,740,849**	**$15,538,087**	**$18,370,891**
LogGamma				
True SLA at 99.996%tile	$366,309,627	$370,407,112	$353,009,568	$387,304,656
OBRE Closer v. MLE	**$43,389,280**	**$46,872,690**	**$46,172,221**	**$45,206,643**
Truncated LogGamma				
True SLA at 99.996%tile	$388,391,019	$392,310,056	$374,657,472	$409,562,640
OBRE Closer v. MLE	**$63,221,137**	**$71,691,291**	**$73,131,423**	**$68,666,193**

7b. Results: SLA Simulations

	TABLE 1a: **LogNormal**	0% Deviation	6% Deviation Both Tails (3% Each)	6% Deviation Left Tail	6% Deviation Right Tail
	True SLA at 99.996%tile	$170,317,921	$173,118,560	$165,323,008	$180,654,136
CvM	Mean	$185,211,363	$187,672,888	$182,490,921	$189,357,264
MLE	Mean	$177,821,938	$184,864,199	$181,071,343	$186,460,684
OBRE*	Mean	$170,989,770	$177,115,375	$173,710,961	$177,620,687
	OBRE Closer v. MLE	**$6,832,168**	**$7,748,825**	**$7,360,382**	**$2,773,099**
CvM	Mean %Difference from True	8.7%	8.4%	10.4%	4.8%
MLE	Mean %Difference from True	4.4%	6.8%	9.5%	3.2%
OBRE*	Mean %Difference from True	0.4%	2.3%	5.1%	-1.7%
CvM	% within +/- 50%	74.0%	70.0%	75.0%	77.0%
MLE	% within +/- 50%	80.0%	83.0%	80.0%	87.0%
OBRE*	% within +/- 50%	80.0%	84.0%	82.0%	86.0%
CvM	RMSE	$102,185,795	$93,890,094	$85,334,514	$88,837,541
MLE	RMSE	$79,516,780	$68,157,312	$66,129,189	$66,662,079
OBRE*	RMSE	$79,571,542	$76,325,792	$70,325,414	$73,332,644

***NOTE: c = 2^(11/8) ≈ 2.59**

7b. Results: SLA Simulations

	TABLE 1b: Truncated LogNormal	0% Deviation	6% Deviation Both Tails (3% Each)	6% Deviation Left Tail	6% Deviation Right Tail
	True SLA at 99.996%tile	$180,486,144	$183,180,240	$175,278,136	$190,682,320
CvM	Mean	$205,843,384	$215,660,029	$213,128,501	$222,407,560
MLE	Mean	$201,471,561	$207,653,389	$203,560,697	$214,920,757
OBRE*	Mean	$180,711,814	$191,912,540	$188,022,611	$196,549,866
	OBRE Closer v. MLE	**$20,759,747**	**$15,740,849**	**$15,538,087**	**$18,370,891**
CvM	Mean %Difference from True	14.0%	17.7%	21.6%	16.6%
MLE	Mean %Difference from True	11.6%	13.4%	16.1%	12.7%
OBRE*	Mean %Difference from True	0.1%	4.8%	7.3%	3.1%
CvM	% within +/- 50%	59.0%	64.0%	60.0%	64.0%
MLE	% within +/- 50%	71.0%	73.0%	72.0%	74.0%
OBRE*	% within +/- 50%	72.0%	70.0%	71.0%	76.0%
CvM	RMSE	$318,935,475	$148,248,416	$151,366,218	$176,135,956
MLE	RMSE	$140,551,905	$109,436,060	$111,794,444	$118,952,011
OBRE*	RMSE	$133,209,674	$110,730,346	$116,252,565	$129,840,945

***NOTE: c = 2^(9/8) ≈ 2.18**

7b. Results: SLA Simulations

	TABLE 1c: **LogGamma**	0% Deviation	6% Deviation Both Tails (3% Each)	6% Deviation Left Tail	6% Deviation Right Tail
	True SLA at 99.996%tile	$366,309,627	$370,407,112	$353,009,568	$387,304,656
CvM	Mean	$436,699,482	$460,516,168	$449,553,624	$465,199,876
MLE	Mean	$415,025,578	$430,550,666	$420,202,603	$434,679,718
OBRE*	Mean	$360,982,956	$383,677,976	$374,030,382	$385,136,237
	OBRE Closer v. MLE	**$43,389,280**	**$46,872,690**	**$46,172,221**	**$45,206,643**
CvM	Mean %Difference from True	19.2%	24.3%	27.3%	20.1%
MLE	Mean %Difference from True	13.3%	16.2%	19.0%	12.2%
OBRE*	Mean %Difference from True	-1.5%	3.6%	6.0%	-0.6%
CvM	% within +/- 50%	54.0%	62.0%	59.0%	63.0%
MLE	% within +/- 50%	63.0%	75.0%	70.0%	78.0%
OBRE*	% within +/- 50%	59.0%	71.0%	72.0%	76.0%
CvM	RMSE	$331,448,466	$332,027,462	$310,337,566	$332,386,275
MLE	RMSE	$271,095,454	$243,734,467	$233,682,773	$244,208,780
OBRE*	RMSE	$222,205,047	$258,303,584	$252,743,932	$252,990,317

***NOTE: c = 2^(19/8) ≈ 5.187**
W ≥ 0.85

7b. Results: SLA Simulations

	TABLE 1d:	0% Deviation	6% Deviation	6% Deviation	6% Deviation
	Truncated LogGamma		Both Tails (3% Each)	Left Tail	Right Tail
	True SLA at 99.996%tile	$388,391,019	$392,310,056	$374,657,472	$409,562,640
CvM	Mean	$524,605,463	$519,079,398	$509,418,297	$524,493,597
MLE	Mean	$470,229,619	$470,391,969	$463,087,826	$479,560,215
OBRE*	Mean	$407,008,482	$398,700,677	$389,956,403	$410,894,022
	OBRE Closer v. MLE	**$63,221,137**	**$71,691,291**	**$73,131,423**	**$68,666,193**
CvM	Mean %Difference from True	35.1%	32.3%	36.0%	28.1%
MLE	Mean %Difference from True	21.1%	19.9%	23.6%	17.1%
OBRE*	Mean %Difference from True	4.8%	1.6%	4.1%	0.3%
CvM	% within +/- 50%	50.0%	51.0%	54.0%	62.0%
MLE	% within +/- 50%	63.0%	67.0%	66.0%	76.0%
OBRE*	% within +/- 50%	56.0%	60.0%	66.0%	67.0%
CvM	RMSE	$584,908,158	$393,702,817	$341,259,642	$440,805,965
MLE	RMSE	$360,712,711	$237,737,636	$270,317,853	$311,345,233
OBRE*	RMSE	$273,966,583	$237,477,157	$237,181,395	$272,922,481

***NOTE: $c = 2^{(19/8)} \approx 5.187$**
$W \geq 0.85$

7b. Results: SLA Simulations

Steps to obtain OBRE Tuning Parameter:

1. For a given sample, obtain MLE parameter estimates for the appropriate distribution

2. Using those parameter estimates, simulate some number of samples (say, B=500) from that distribution and obtain B MLE parameter estimates and B corresponding SLA capital esimates (the mean of these SLA estimates will most likely overshoot the "true" SLA).

3. For a given tuning parameter value c, calculate B OBRE parameter estimates and B corresponding SLA capital estimates based on the B samples.

4. Repeat 3. for different values of c (say, 2^(8/8), 2^(9/8), 2^(10/8), 2^(11/8), 2^(12/8), depending on the distribution) and choose the value of c that most closely approximates the "true" SLE (based on 2.) without dramatically increasing the RMSE of the OBRE-based SLA (the RMSE calculated based on the B samples).

- The above is viable only if the value of c ultimately chosen is robust to initial parameter misspecification. Preliminary tests indicate that it is.

7b. Results: SLA Simulations

- For the LogNormal and Truncated LogNormal, the values of c chosen for Table 1 were subsequently tested on data samples generated from distributions with both parameters a full standard deviation away from the original parameters – in the same direction! (for independent parameters, Pr<0.03) For the LogGamma and Truncated LogGamma, values a half a standard deviation, in the same direction, were used (Pr<0.10). For the LogNormal, this created distributions with means -$115K / +$160K smaller/larger, and for the LogGamma, means -$191K / +$318K smaller/ larger, respectively. In all four cases, the original value of c was chosen as the best c.

- This robustness to parameter misspecification may be related to the heaviness of the tail of the distribution, with less robustness under heavier tails. And preliminary tests using parameter misspecifications that were even larger indicated this, while also yielding "borderline" results under which different values of c COULD have been chosen as "better." So if this approach is shown to be practically useable, it would have to be well tested on a given set of data / distributions / ranges of parameter values.

- To test this approach, a simulation study that repeats Steps 1.- 4. on a large number of samples needs to be carried out. This would be computationally expensive, unless shortcuts can be derived. Time has not permitted this to date, but it is a required next step to demonstrate viability and useability across a sufficiently wide range of conditions.

7b. Results: SLA Simulations

- Finding the "best" value of c directly yielded the SLA#s in Table 1 for the LogNormal and Truncated LogNormal distributions. Unfortunately, this was not the case for the LogGamma and Truncated LogGamma distributions: even after finding the best value of c, the high quantile estimates based on OBRE parameter estimates still notably overshot the "true" high quantile (although not quite as much as did MLE's estimates). So something else is needed.
- The information-laden OBRE weights are a natural place to turn to attempt to estimate SLA capital estimates with greater precision.
- One possibility is to first obtain OBRE weights on the data points, and then reestimate OBRE excluding observations with weights below a certain value, i.e. excluding those observations that deviate dramatically from the assumed distribution. Recall that weight values will change from sample to sample, because they are based on deviations from the presumed distribution (which is different for each sample), not on an arbitrary absolute value, or an arbitrary trimming requirement. Some samples will exclude no observations based on the criteria, and others will exclude several.
- To obtain the SLA #s for LogGamma and Truncated LogGamma in Table 1, a process similar to that used with the tuning parameter was followed: the optimal weight-exclusion value was found, and then tested for robustness to initial parameter misspecification ex post. A procedure for statistical inference might look something like the below:

7b. Results: SLA Simulations

If after Step 4. the OBRE-based SLA estimate is still unacceptably high relative to the "true" SLA from the sampling exercise, proceed to Step 5.:

5. **Based on the OBRE parameter estimates obtained in Step 4., generate some number of new samples (D=500) and for each sample, generate OBRE weights. Then exclude observations with weights below a certain value, and estimate OBRE parameter estimates for all D samples. For example, W<0.5 may correspond to about 0.3% of all observations, on average; W<0.7 may correspond to about 0.6% of all observations, on average; and W<0.9 may correspond to about 0.9% of all observations, on average (but this will, and should, vary from sample to sample).**

6. **Repeat Step 5. for different values of W (e.g. W<0.6, W<0.7, W<0.8, W<0.9). Select the value of W that is closest to the "true" SLA.**

7. **Use the value of W obtained in 6., along with the value of c obtained in Step 4., to estimate OBRE on the original sample.**

7b. Results: SLA Simulations

- Of course, testing such a process requires a computationally demanding simulation, unless computational shortcuts are derived.
- The informal robustness tests on the "W" values demonstrated less robustness that did the tests on the tuning parameter: for the LogGamma and Truncated LogGamma, deviations from the simulated OBRE parameter estimates (from Step 5.) one quarter of a standard deviation from the "true" (simulation) parameter values (in the same direction) yielded the same values of W, but larger deviations quickly yielded very different values of W, which yielded very different capital estimates. It would appear that if an inferential procedure can be developed that utilizes OBRE weight values, it will probably need to have far more statistical power than that related to choosing the value of the tuning parameter.

7b. Results: SLA Simulations

- **<u>BOTTOM LINE</u>:**
  The main point of this exercise was to establish that, at least mechanically, the tuning parameter and the weight values COULD be used to obtain more accurate (less biased) SLA estimates without notably increasing the already large variance on the MLE distribution of SLA estimates. This has been done, and the non-trivial magnitude of the values "left on the table" have been established in Table 1. But of course the next and more important step is to develop and test a complete procedure for statistical inference.

- **The proposed Steps 1. – 4. may be quite sufficient for more medium- to somewhat heavy tailed distributions like the LogNormal and Truncated LogNormal, but Steps 5.-7., or something more complex, may be required for distributions with very heavy tails, like the LogGamma.**

2. MLE vs. Robust Statistics: Point-Counterpoint

Maximum Likelihood Estimation (MLE):

"MLE does not inappropriately downweight extreme observations as do most/all robust statistics. And focus on extreme observations is the entire point of the OpRisk statistical modeling exercise! Why should we even partially ignore the (right) tail when that is where and how capital requirements are determined?! That's essentially ignoring data – the most important data – just because its hard to model!"

Robust Statistics:

"All statistical models are merely idealized approximations of reality, and OpRisk data clearly violate the fragile, textbook model assumptions required by MLE (e.g. iid data). And even under iid data, the expected value of high quantile estimates based on MLE parameter estimates is biased upwards for (right-skewed) heavy-tailed distributions (i.e. OpRisk severity distributions) due to Jensen's inequality (and this, of course, inflates OpRisk capital estimates). Robust Statistics explicitly and sytemmatically acknowledge and deal with non-iid data, sometimes using weights to avoid bias and/or inefficiency caused by unanticipated or unnoticed heterogeneity. And an ancillary benefit is mitigation of the bias in capital estimates due to Jensen's inequality. Consequently, under real-world, finite-sample, non-iid OpRisk loss data, Robust Statistics typically exhibit less bias, equal and sometimes even greater efficiency, and far more robustness than does MLE. These characteristics translate into a more reliable, stable estimation approach, regardless of the framework used by robust statistics (i.e. multivariate regression or otherwise) to obtain high quantile estimates of the severity distribution.

8. Point-Counterpoint Revisted: Who Wins?

Some Specific Questions to be Answered:

- **Does MLE become unusable under relatively modest deviations from i.i.d., especially for the heavy-tailed distributions used in this setting YES, or are these claims overblown? NO**
- **Is the bias of the expected value of MLE-based capital estimates large? YES (ESP. FOR VERY HEAVY TAILS)**
- **Do analytical derivations of the MLE Influence Functions for severity distribution parameters support or contradict such claims? NO, THEY SUPPORT THEM Are they consistent with simulation results? YES How does (possible) parameter dependence affect these results? NOTABLY (ESP. FOR VERY HEAVY TAILS)**
- **Do these results hold under truncation? YES How much does truncation and the size of the collection threshold affect both MLE and Robust Statistics parameter estimates? RESPECTIVELY: VERY BADLY, NOT MUCH/ROBUST**
- **Are widely used, well established Robust Statistics viable for severity distribution parameter estimation? FOR SLA, OBRE IS Are they too inefficient relative to MLE for practical use? NO, SOMETIMES MORE EFFICIENT Do any implementation constraints (e.g. algorithmic/convergence issues) trip them up? NOT OBRE, BUT CvM ON VERY HEAVY TAILED DISTRIBUTIONS.**

ORX ASSOCIATION

DataMineIt℠

8. Point-Counterpoint Revisted: Confirmation

"Estimation of operational risk is badly influenced by the quality of data, as not all external data is relevant, some losses (i.e. 'outliers') may not be captured by the ideal model, and induce bias, and some data may not be reported at all. This can result in systematic over- or under-estimation of operational risk. … robust estimation of the regulatory capital for the operational risk hence provides a useful technique to avoid bias when working with data influenced by outliers and possible deviations from the ideal models." (Horbenko, Ruckdeschel, & Bae, 2010)

"…recent empirical findings suggest that classical methods will frequently fit neither the bulk of the operational loss data nor the outliers well… Classical estimators that assign equal importance to all available data are highly sensitive to outliers and in the presence of just a few extreme losses can produce arbitrarily large estimates of mean, variance and other vital statistics. …On the contrary, robust methods take into account the underlying structure of the data and "separate" the bulk of the data from outlying events, [in – sic] this way avoiding upward bias in the vital statistics and forecasts." (Chernobai & Rachev, 2006)

"Since we can assume that deviation from the model assumptions almost always occurs in finance and insurance data, it is useful to complement the analysis with procedures that are still reliable and reasonably efficient under small deviations from the assumed parametric model and highlight which observations (e.g. outliers) or deviating substructures have most influence on the statistical quantity under observation. Robust statistics achieves this by a set of different statistical frameworks that generalize classical statistical procedures such as maximum likelihood or OLS." (Embrechts & Dell'Aquila, 2006)

9. Findings Summary & Next Steps

1) **Counter-Intuitive Disproportionate Impact of Small (Left Tailed) Losses/Deviations:** Small aribtrary deviations away from the presumed model (that is, deviations in the left tail) can have very large, disproportionate biasing effects on MLE estimates. This is an analytically derived result of the IFs of the LogNormal, Truncated LogNormal, LogGamma, and Truncated LogGamma (and other distributions), not an artifact of sensitivity to simulation assumptions. It is important for its magnitude, and the fact that it is overlooked.

2) **The Threshold Matters … a lot! Truncation Induces or Augments Parameter Dependence, Sometimes in Very Counterintuitive Ways, with Dramatic Effects:** This is an analytically derived result of the IFs of the LogNormal, Truncated LogNormal, LogGamma, and Truncated LogGamma (and other distributions), not an artifact of sensitivity to simulation assumptions. This is an important finding as it would appear to explain the extreme sensitivity, and sometimes counterintuitive behavior, of MLE estimates to truncation that is often cited in the literature (based on simulations alone).

3) **All Analytically Derived IFs Virtually Exactly Match EIFs**

4) **OBRE v. CvM:** The flexibility provided by OBRE's tuning parameter appears to be quite necessary in this setting, that is, for SLA-based capital estimation. This gives it a strong advantage over CvM, which did not perform well here. The latter also encountered convergence issues in this study, while the former did not.

9. Findings Summary & Next Steps

5) **<u>OBRE is Robust to Arbitrary Deviations from the Presumed Model</u>:** As expected, OBRE-based SLA estimates are more robust than their MLE counterparts to deviations from the presumed severity distribution (even without the "optimal" tuning parameter values). This was true across distributions and types of deviations.

6) **<u>MLE Overshoots High Quantiles</u>:**
As expected, and as is well documented in the literature, even under ideal iid conditions MLE distribution parameter estimates overshoot, on average, when used to estimate very high quantiles (due to Jensen's inequality). The heavier the tail, the larger the bias.

7) **<u>Robust Statistics Overshoot, Too! (for very heavy tails)</u> :**
Unfortunately, for the very heavy tailed distributions (e.g. LogGamma, not LogNormal), even robust estimates of distribution parameters overshoot, on average, when used to estimate very high quantiles. Judicious use of OBRE weights in an innovative statistical inference procedure may be able to adequately address this.

8) **<u>A Lot Left on the Table</u>:**
Due to 6), a less biased finite sample quantile estimator than one based on MLE estimates, all else equal, would not only provide more accurate capital estimates, but also estimates that uniformly and non-trivially lessen banks' capital requirements. It appears that the heavier the tail of the severity distribution, the greater the absolute (and possibly relative) value of these "savings."

9. Findings Summary & Next Steps

9) **<u>Meaninful Variance Reduction (Only) from Additional Information / Methodology</u>:** The severity distribution quantiles requiring estimation are so large, and the extant data so (relatively) scarce, that even using the absolute "best" estimator will not provide sufficient variance reduction to obtain capital estimates that are not "all over the map." Additional information / methodology is required to obtain meaningful variance reduction in the capital estimate distribution, and one excellent potential source / method is that of estimating these distribution parameters with regression. On internal losses, rich covariate information exists, and the inferential technology exists not only to estimate these parameters with multivariate inference, but better still, with OBRE-based multivariate inference. The ultimate solution here may be an OBRE regression (there are many examples in the applied literature), which would be at once robust, as well as more efficient in its use of currently unused information to achieve much needed variance reduction. Such an approach could, and probably should, also guide the creation/definition of units of measure, as well as directly address the issue of how to appropriately deal with time-varying thresholds (e.g. using the appropriate index as a covariate with real (not nominal) loss data).

ORX ASSOCIATION

DataMineIt SM

10. Conclusions

- **Keep an Open Mind**:
  The application of Robust Statistics to OpRisk severity distribution parameter estimation is relatively new (at least to obtain capital estimates). Because this challenging problem is far from being definitively "solved" by the industry, and it is not a theoretical problem, applied researchers need to keep open minds to different approaches. Many of the methods currently gaining some acceptance would have been considered by most practitioners to be unacceptably "heroic" just a half decade ago.
- **We can do better than MLE**:
  The point of using robust statistics in this setting is not to underweight certain data points per se, but rather, to use weights, directly or indirectly, to avoid the well-documented non-robustness of MLE to what we are sure are many violations of presumed model assumptions (e.g. iid data). Striking the right balance between over- and underweighting obviously is key, and something which judicious and creative use of the OBRE tuning parameter, along with OBRE weights, may be able to achieve. The dollar amounts "left on the table" due to MLE's overshooting bias (as per Jensen's inequality) of the high quantile required for capital estimation, not to mention that due to non-iid loss data, make this pursuit well worth it.
- Many of the challenges of OpRisk loss event data appear to be tailor-made for a robust statistics approach, and the results presented herein appear promising for its application in this setting.

11. Appendix I

Mean Squared Error: This is the average of the squared devations of sample-based estimate values from the true population value of the parameter being estimated, as shown below:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\theta - \hat{\theta}\right)^2 = Variance\left(\theta\right) + \left[\,Bias\left(\theta\right)\right]^2$$

If an estimator is unbiased, bias = 0 and MSE = Variance. "Efficiency" can be defined in slightly different ways, but it is always inversely related to MSE.

The Cramér-Rao Lower Bound: is the inverse of the information matrix – the negative of the expected value of the second-order derivative of the log-likelihood. Because this is the lower bound for the variance of any unbiased estimator, efficiency is usually defined in reference to it, if not in reference to another estimator (in which case it is usually called relative efficiency).

11. Appendix II

For the median, we must use additional results from Hampel et al. (1986) related to L-estimators (of location), which are of the form $T_n(X_1,\ldots,X_n)=\sum_{i=1}^{n} a_i X_{i:n}$, where $X_{1:n},\ldots,X_{n:n}$ is the ordered sample and the a_i are coefficients.

"L" of "L-estimators" comes from "linear" combinations of order statistics. A natural sequence of location estimators is obtained if the weights a_i are generated by $a_i = \int_{[(i-1)/n,\, i/n]} h\,d\lambda \Big/ \int_{[0,1]} h\,d\lambda$, where $h:[0,1]\rightarrow \Re$ satisfies $\int_{[0,1]} h\,d\lambda \neq 0$

Under regularity conditions, these estimators are asymptotically normal and the corresponding functional is

$T(G)=\dfrac{\int xh(G(x))\,dG(x)}{\int h(F(y))\,dF(y)}$, which is Fisher consistent with influence function

$$IF(x;T,F)=\frac{\int_{[0,x]} h(F(y))\,d\lambda(y)-\int\left[\int_{[0,t]} h(F(y))\,d\lambda(y)\right]dF(t)}{\int h(F(y))\,dF(y)}$$

where the denominator is nonzero because it equals $\int_{[0,1]} h\,d\lambda$

Now the median corresponds to $h=\delta_{(1/2)}$, so $\int_{[0,1]} h\,d\lambda = 1$ and $T(G)=\int_{[0,1]} G^{-1}(y)h(y)\,d\lambda(y)=G^{-1}(1/2)$

So its influence function is $IF(x;T,F)=\dfrac{1}{2f\left(F^{-1}(1/2)\right)}\,\text{sign}\left(x-F^{-1}(1/2)\right)$

and for standard normal, median=0, $F^{-1}(1/2)=0$, so

$$IF(x;T,F)=\frac{\text{sign}(x-0)}{2f(0)}=\frac{\text{sign}(x)}{2\frac{1}{\sqrt{2\pi}}\exp(-0/2)}=\frac{\text{sign}(x)}{2\frac{1}{\sqrt{2\pi}}\exp(0)}=\frac{\text{sign}(x)}{\frac{2}{\sqrt{2\pi}}}=\text{sign}(x)\sqrt{\frac{\pi}{2}}$$

11. Appendix III

Many important robustness measures are based directly on the IF:

<u>Gross Error Sensitivity</u> (GES) is the supremum being taken over all x where IF exists:

$$\gamma^{*}(T,F)=\sup_{x}\left|IF(x;T,F)\right|$$

This measures the worst case (approximate) influence that a small amount of contamination of a fixed size can have on the value of the estimator. If GES is finite, that is, if IF is bounded, the estimator is B-robust ("B" comes from "bias," because GES can be regarded as an upper bound on the (standardized) asymptotic bias of the estimator). Robustifying an estimator typically makes it less efficient, so the conflict between robustness and efficiency is often best solved with Op timal B-robust estimators (OBRE) – estimates which cannot be improved with respect to both GES and asymptotic variance (shown below). So GES is very useful, alongside IF, for comparing two estimators. If, for example, a comparison of the IFs of two estimators leads to ambiguous conclusions, that is, if one estimator's IF has tighter bounds over one range but the other's is tighter over another range, then GES is a useful tool describing which is better under the worst case scenario.

11. Appendix III

Rejection Point: If IF does not exist in some area and is equal to zero, then contamination of points in that area do not have any influence on the estimator at all. The rejection point, then, is defined as

$$\rho^* = \inf\left\{r > 0; IF\left(x;T,F\right) = 0 \text{ when } |x| > r\right\}$$

Observations farther away than ρ^* are rejected completely, so it is generally desirable if ρ^* is finite. In other words, for estimators with finite rejection point, there will be some point beyond which extreme outlying data points will have no influence on the value of the estimator (because the value of the influence function is zero), and in general, this is a desirable characteristic of an estimator, adding to its robustness against data that deviates notably from the model's assumptions.

11. Appendix III

Empirical Influence Function: The empirical influence function (EIF) naturally corresponds with the IF, and is given by

$$IF\left(x;T,\hat{F}\right)=\lim_{\varepsilon\to 0}\left[\frac{T\left\{\left(1-\varepsilon\right)\hat{F}+\varepsilon\delta_{x}\right\}-T\left(\hat{F}\right)}{\varepsilon}\right]$$

To implement this in practice, EIF is simply a plot of
as a function of *x*, where *x* is the added contamination data point inserted in place of observation . The EIF can be described as an estimation using the original sample, but with only *n* – 1 of the observations, compared to one using the same sample with one additional data point, *x*, the contamination. This also is closely related to the jackknife (the finite sample approximation of the asymptotic variance, treated below, is the jackknife estimator of the variance).

11. Appendix III

Sensitivity Curve: A concept very closely related to the empirical influence function, that is, the non-asymptotic, finite sample IF, is Sensitivity Curves. Analogous to the EIF, these answer the question: how sensitive is the estimator, based on the finite empirical sample at hand, to single-point contaminations at each data point? From Hampel et al. (1986), the sensitivity curve is simply $SC_n(x) = n\left[T_n(x_1,\ldots,x_{n-1},x) - T_{n-1}(x_1,\ldots,x_{n-1})\right]$, which is just a translated and rescaled version of EIF. The functional is applied to two empirical samples (both with one original data point removed): one with a point of contamination, and one without. The difference between the values of the empirical functional, multiplied by *n*, is the sensitivity curve.

Analogously, when the estimator is a functional, then

$$SC_n(x) = \frac{1}{n}\left[T\left(\left(1-\frac{1}{n}\right)F_{n-1} + \frac{1}{n}\delta_x\right) - T\left(F_{n-1}\right)\right]$$

, where F_n is the empirical distribution $(x_1,\ldots,x_{n-1})$. In fact, based on the above, $SC_n(x)$ will in many cases converge to $IF(x;T,F)$ asymptotically.

11. Appendix III

Asymptotic Variance and ARE: Based on the IF, an important measure of efficiency is the asymptotic variance, from which the asymptotic relative efficiency (ARE) directly can be calculated. The ARE is simply a measure of the relative size of the variances of two estimators, telling us which is more efficient.

$$V(T,F)=\int IF(x;T,F)^2\,dF(x)$$

$$ARE_{T,S}=V(S,F)/V(T,F)$$

Understanding the (relative) efficiency of an estimator is especially important within the framework of robust statistics, because some degree of efficiency typically is lost when estimators are made robust. Knowing the extent of efficiency loss is important, because we want estimators that are both robust and efficient, and these are competing criteria by which we need to compare estimators, under different distributions and against each other. Designing estimators to be OBRE (optimally *B*-robust estimators), for example, requires finding estimators that simultaneously can be made no more efficient, and no more robust, and to do this requires knowing how efficient and robust an estimator is.

11. Appendix III

Change-in-Variance Sensitivity: The "change-in-variance" sensitivity assesses how sensitive is the estimator to changes in its asymptotic variance due to contamination at *F*. The denominator of CVS is the asymptotic variance (see section on *M*-estimators above for a definition of ψ), and the numerator is the derivative of the asymptotic variance when contaminated.

$$CVS(\varphi, F) := \sup\left\{ \frac{CVF(x;\varphi,F)}{V(\varphi,F)} ; x \in \Re \setminus C(\varphi) \cup D(\varphi) \right\}$$ where the

change-in-variance function is

$$CVF(x;\varphi,F) = \frac{\partial}{\partial\varepsilon}\left[V\left(\varphi, (1-\varepsilon)F + \varepsilon\left(\frac{1}{2}\delta_x + \frac{1}{2}\delta_{-x} \right)\right)\right]_{\varepsilon=0}$$ for continuous

ψ, for which no delta functions arise. The above is valid for all *M*-estimators. If CVS is finite, *T* is *V*-robust ("V" is for Variance). *V*-robustness is stronger than *B*-robustness: if an estimator is V-robust, it must also be *B*-robust (and if an estimator is not *B*-robust, then it is not *V*-robust). Note that unlike IF, only large positive values for *CVF*, not large negative values, point to nonrobustness.

11. Appendix III

Local Shift Sensitivity: The point of “local shift sensitivity” is to summarize how sensitive the estimator is to small changes in the values of the observations; in other words, how much is the estimator affected by shifting an observation slightly from point x to point y? When the “worst” effect of this “wiggling” is obtained, and it is standardized, the local shift sensitivity is defined as

$$\lambda^* = \sup_{x \neq y} \left| IF(y;T,F) - IF(x;T,F) \right| / \left| y - x \right|$$

This helps to evaluate how sensitive an estimator is to changes in the data, all else equal. And this is relevant in this setting because loss data does change from quarter to quarter, if financials are restated, litigation is settled, etc. So this is an important tool for assessing the robustness of a particular estimator, and can be used in simulation studies to compare the behavior of multiple estimators under such data changes.

11. Appendix III

<u>Breakdown Point</u>: While the IF and its related summary values are related to local robustness, describing the effects of a(n infinitesimal) contamination at point *x*, the "breakdown point" is a measure of global robustness – it describes the global reliability of an estimator by asking, up to what percentage of the data can be contaminated before the estimator stops providing valuable information? The asymptotic contamination breakdown point of the estimate *T* at *F*, denoted ε^*, is the largest $\varepsilon^* \in (0,1)$ such that for $\varepsilon < \varepsilon^*$, $T\left((1-\varepsilon)F + \varepsilon H\right)$ remains bounded as a function of H and also bounded away from the boundary of θ.

Analogously, the finite sample breakdown point ε^* of the estimator T_n at the sample $(x_1,\ldots,x_n)$ is given by

$$\varepsilon_n^*\left(T_n;x_1,\ldots,x_n\right) := \frac{1}{n}\max\left\{m;\max_{i_1,\ldots,i_n}\ \sup_{y_1,\ldots,y_n}\left|T_n\left(z_1,\ldots,z_n\right)\right| < \infty\right\},$$

where the sample $(z_1,\ldots,z_n)$ is obtained from the sample $(x_1,\ldots,x_n)$ by replacing the *m* data points $\left(x_{i_1},\ldots,x_{i_m}\right)$ by arbitrary values $(y_1,\ldots,y_n)$.

The mean, for example, has asymptotic breakdown point and finite sample breakdown point, respectfully, of $\varepsilon^* = 0$ and $\varepsilon^* = 1/n$, because a single observation with value = arbitrarily large (i.e. ∞) renders its values meaningless. In constrast, those of the median are $\varepsilon^* = 0.5$, and $\varepsilon^* = 1/2$ for an even *n* and $\varepsilon^* = (n-1)/2n$ for odd *n*, respectfully, which is far more robust than the mean.

11. Appendix IV

LogNormal Derivatives:

for $0 < x < \infty;\ 0 < \mu < \infty;\ 0 < \sigma < \infty$

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}$$

$$F(x;\mu,\sigma) = \frac{1}{2}\left[1 + erf\left(\frac{\ln(x)-\mu}{\sqrt{2\sigma^2}}\right)\right]$$

$$\frac{\partial}{\partial\mu} f(x;\mu,\sigma) = \left[\frac{\ln(x)-\mu}{\sigma^2}\right] f(x;\mu,\sigma)$$

$$\frac{\partial}{\partial\sigma} f(x;\mu,\sigma) = \left[\frac{(\ln(x)-\mu)^2}{\sigma^3} - \frac{1}{\sigma}\right] f(x;\mu,\sigma)$$

$$\frac{\partial^2}{\partial\mu^2} f(x;\mu,\sigma) = \left[\frac{(\ln(x)-\mu)^2}{\sigma^4} - \frac{1}{\sigma^2}\right] f(x;\mu,\sigma)$$

$$\frac{\partial^2}{\partial\sigma^2} f(x;\mu,\sigma) = \left(\left[\frac{1}{\sigma^2} - \frac{3(\ln(x)-\mu)^2}{\sigma^4}\right] + \left[\frac{(\ln(x)-\mu)^2}{\sigma^3} - \frac{1}{\sigma}\right]^2\right) f(x;\mu,\sigma)$$

$$\frac{\partial}{\partial\mu\partial\sigma} f(x;\mu,\sigma) = \left[\frac{\ln(x)-\mu}{\sigma^2}\right]\left[\frac{(\ln(x)-\mu)^2}{\sigma^3} - \frac{3}{\sigma}\right] f(x;\mu,\sigma)$$

11. Appendix IV

LogNormal Derivatives (for (left) Truncated case):

Due to Leibniz's Rule, these derivatives can be moved inside these integrals.

$$g(x;\mu,\sigma)=\frac{f(x;\mu,\sigma)}{1-F(H;\mu,\sigma)}$$

$$G(x;\mu,\sigma)=1-\frac{1-F(x;\mu,\sigma)}{1-F(H;\mu,\sigma)}$$

$$\frac{\partial F(H;\mu,\sigma)}{\partial\mu}=\frac{\partial}{\partial\mu}\int_0^H f(y;\mu,\sigma)dy=\int_0^H\frac{\partial}{\partial\mu}f(y;\mu,\sigma)dy=\int_0^H\left[\frac{\ln(y)-\mu}{\sigma^2}\right]f(y;\mu,\sigma)dy$$

$$\frac{\partial F(H;\mu,\sigma)}{\partial\sigma}=\frac{\partial}{\partial\sigma}\int_0^H f(y;\mu,\sigma)dy=\int_0^H\frac{\partial}{\partial\sigma}f(y;\mu,\sigma)dy=\int_0^H\left[\frac{(\ln(y)-\mu)^2}{\sigma^3}-\frac{1}{\sigma}\right]f(y;\mu,\sigma)dy$$

$$\frac{\partial^2 F(H;\mu,\sigma)}{\partial\mu^2}=\frac{\partial^2}{\partial\mu^2}\int_0^H f(y;\mu,\sigma)dy=\int_0^H\frac{\partial^2}{\partial\mu^2}f(y;\mu,\sigma)dy=\int_0^H\left[\frac{(\ln(y)-\mu)^2}{\sigma^4}-\frac{1}{\sigma^2}\right]f(y;\mu,\sigma)dy$$

$$\frac{\partial^2 F(H;\mu,\sigma)}{\partial\sigma^2}=\frac{\partial^2}{\partial\sigma^2}\int_0^H f(y;\mu,\sigma)dy=\int_0^H\frac{\partial^2}{\partial\sigma^2}f(y;\mu,\sigma)dy=\int_0^H\left[\frac{1}{\sigma^2}-\frac{3(\ln(y)-\mu)^2}{\sigma^4}\right]+\left[\frac{(\ln(y)-\mu)^2}{\sigma^3}-\frac{1}{\sigma}\right]^2 f(y;\mu,\sigma)dy$$

$$\frac{\partial F(H;\mu,\sigma)}{\partial\mu\partial\sigma}=\frac{\partial}{\partial\mu\partial\sigma}\int_0^H f(y;\mu,\sigma)dy=\int_0^H\frac{\partial}{\partial\mu\partial\sigma}f(y;\mu,\sigma)dy=\int_0^H\left[\frac{\ln(y)-\mu}{\sigma^2}\right]\left[\frac{(\ln(y)-\mu)^2}{\sigma^3}-\frac{3}{\sigma}\right]f(y;\mu,\sigma)dy$$

11. Appendix IV

LogGamma Distribution Derivatives:

assuming $1 \le x < \infty;\ 0 < a;\ 0 < b$

$$f(x;a,b) = \frac{b^a (\log(x))^{(a-1)}}{\Gamma(a) x^{b+1}}$$

$$F(x;a,b) = \frac{b^a}{\Gamma(a)} \int_{\ln(1)}^{\ln(x)} y^{(a-1)} \exp(-yb)\,dy$$

$$\frac{\partial}{\partial a} f(x;a,b) = \left[\ln(b) + \ln(\ln(x)) - digam(a)\right] f(x;a,b)$$

$$\frac{\partial}{\partial b} f(x;a,b) = \left[\frac{a}{b} - \ln(x)\right] f(x;a,b)$$

$$\frac{\partial^2}{\partial a^2} f(x;a,b) = \left(\left[\ln(b) + \ln(\ln(x)) - digam(a)\right]^2 - trigamma(a)\right) \cdot f(x;a,b)$$

$$\frac{\partial^2}{\partial b^2} f(x;a,b) = \left[\frac{a(a-1)}{b^2} - \frac{2a(\ln(x))}{b} + (\ln(x))^2\right] \cdot f(x;a,b)$$

$$\frac{\partial}{\partial a \partial b} f(x;a,b) = \left(\frac{1}{b} + \left[\ln(b) + \ln(\ln(x)) - digam(a)\right] \times \left[\frac{a}{b} - \ln(x)\right]\right) f(x;a,b)$$

ORX ASSOCIATION

DataMineIt℠

11. Appendix IV

LogGamma Derivatives (for (left) Truncated Case):

Due to Leibniz's Rule, these derivatives can be moved inside these integrals.

$$g(x;\mu,\sigma)=\frac{f(x;\mu,\sigma)}{1-F(H;\mu,\sigma)}$$

$$G(x;\mu,\sigma)=1-\frac{1-F(x;\mu,\sigma)}{1-F(H;\mu,\sigma)}$$

$$\frac{\partial F(H;a,b)}{\partial a}=\int_1^H\left[\ln(b)+\ln(\ln(y))-digam(a)\right]f(y;a,b)\,dy$$

$$\frac{\partial F(H;a,b)}{\partial b}=\int_1^H\left[\frac{a}{b}-\ln(y)\right]f(y;a,b)\,dy$$

$$\frac{\partial^2 F(H;a,b)}{\partial a^2}=\int_1^H\left(\left[\ln(b)+\ln(\ln(y))-digam(a)\right]^2-trigamma(a)\right)\cdot f(y;a,b)\,dy$$

$$\frac{\partial^2 F(H;a,b)}{\partial b^2}=\int_1^H\left[\frac{a(a-1)}{b^2}-\frac{2a(\ln(y))}{b}+(\ln(y))^2\right]\cdot f(y;a,b)\cdot dy$$

$$\frac{\partial F(H;a,b)}{\partial a\partial b}=\int_1^H\left(\frac{1}{b}+\left[\ln(b)+\ln(\ln(y))-digam(a)\right]\times\left[\frac{a}{b}-\ln(y)\right]\right)f(y;a,b)\,dy$$

11. Appendix IV

Generalized Pareto Distribution Derivatives:

assuming $\varepsilon \geq 0$, for $0 \leq x < \infty;\ 0 < \beta < \infty;\ 0 \leq \varepsilon < \infty$

$$f(x;\varepsilon,\beta) = \frac{1}{\beta}\left[1+\varepsilon\frac{x}{\beta}\right]^{\left[-\frac{1}{\varepsilon}-1\right]}$$

$$F(x;\varepsilon,\beta) = 1-\left[1+\varepsilon\frac{x}{\beta}\right]^{\left[-\frac{1}{\varepsilon}\right]}$$

$$\frac{\partial}{\partial\beta}f(x;\beta,\varepsilon) = -\frac{1}{\beta}\left[\frac{\beta-x}{\beta+\varepsilon x}\right]f(x;\beta,\varepsilon)$$

$$\frac{\partial}{\partial\varepsilon}f(x;\beta,\varepsilon) = \left[\left(\frac{-x(1+\varepsilon)}{\beta\varepsilon+\varepsilon^2 x}\right)+\frac{\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right]f(x;\beta,\varepsilon)$$

$$\frac{\partial^2}{\partial\beta^2}f(x;\beta,\varepsilon) = \left(\left[\frac{1}{\beta^2}-\frac{x(1+\varepsilon)(2\beta+\varepsilon x)}{\left(\beta^2+\beta\varepsilon x\right)^2}\right]+\frac{1}{\beta^2}\left[\frac{\beta-x}{\beta+\varepsilon x}\right]^2\right)f(x;\beta,\varepsilon)$$

$$\frac{\partial^2}{\partial\varepsilon^2}f(x;\beta,\varepsilon) = \left(\left[\frac{x\beta+2\varepsilon x^2+\varepsilon^2 x^2}{\left(\beta\varepsilon+\varepsilon^2 x\right)^2}+\frac{x}{(\beta+\varepsilon x)\varepsilon^2}-\frac{2\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^3}\right]+\left[\left(\frac{-x(1+\varepsilon)}{\beta\varepsilon+\varepsilon^2 x}\right)+\frac{\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right]^2\right)f(x;\beta,\varepsilon)$$

$$\frac{\partial}{\partial\varepsilon\partial\beta}f(x;\beta,\varepsilon) = \left(\left[-\frac{1}{\beta}\left(\frac{\beta-x}{\beta+\varepsilon x}\right)\right]\left[\left(\frac{-x(1+\varepsilon)}{\beta\varepsilon+\varepsilon^2 x}\right)+\frac{\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right]+\left[\frac{\varepsilon x(1+\varepsilon)}{\left(\beta\varepsilon+\varepsilon^2 x\right)^2}-\frac{x}{\beta\varepsilon(\beta+\varepsilon x)}\right]\right)f(x;\beta,\varepsilon)$$

11. Appendix IV

Generalized Pareto Distribution Derivatives (for (left) Truncated Case):

Due to Leibniz's Rule, these derivatives can be moved inside these integrals.

$$g(x;\mu,\sigma)=\frac{f(x;\mu,\sigma)}{1-F(H;\mu,\sigma)}$$

$$G(x;\mu,\sigma)=1-\frac{1-F(x;\mu,\sigma)}{1-F(H;\mu,\sigma)}$$

$$\frac{\partial F(H;\beta,\varepsilon)}{\partial\beta}=\int_0^H -\frac{1}{\beta}\left[\frac{\beta-x}{\beta+\varepsilon x}\right]f(x;\beta,\varepsilon)\,dx$$

$$\frac{\partial F(H;\beta,\varepsilon)}{\partial\varepsilon}=\int_0^H \left[\left(\frac{-x(1+\varepsilon)}{\beta\varepsilon+\varepsilon^2 x}\right)+\frac{\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right]f(x;\beta,\varepsilon)\,dx$$

$$\frac{\partial^2 F(H;\beta,\varepsilon)}{\partial\beta^2}=\int_0^H \left(\left[\frac{1}{\beta^2}-\frac{x(1+\varepsilon)(2\beta+\varepsilon x)}{\left(\beta^2+\beta\varepsilon x\right)^2}\right]+\frac{1}{\beta^2}\left[\frac{\beta-x}{\beta+\varepsilon x}\right]^2\right)f(x;\beta,\varepsilon)\,dx$$

$$\frac{\partial^2 F(H;\beta,\varepsilon)}{\partial\varepsilon^2}=\int_0^H \left(\left[\frac{x\beta+2\varepsilon x^2+\varepsilon^2 x^2}{\left(\beta\varepsilon+\varepsilon^2 x\right)^2}+\frac{x}{(\beta+\varepsilon x)\varepsilon^2}-\frac{2\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^3}\right]+\left[\left(\frac{-x(1+\varepsilon)}{\beta\varepsilon+\varepsilon^2 x}\right)+\frac{\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right]^2\right)f(x;\beta,\varepsilon)\,dx$$

$$\frac{\partial F(H;\beta,\varepsilon)}{\partial\varepsilon\partial\beta}=\int_0^H \left(\left[-\frac{1}{\beta}\left(\frac{\beta-x}{\beta+\varepsilon x}\right)\right]\left[\left(\frac{-x(1+\varepsilon)}{\beta\varepsilon+\varepsilon^2 x}\right)+\frac{\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right]+\left[\frac{\varepsilon x(1+\varepsilon)}{\left(\beta\varepsilon+\varepsilon^2 x\right)^2}-\frac{x}{\beta\varepsilon(\beta+\varepsilon x)}\right]\right)f(x;\beta,\varepsilon)\,dx$$

11. Appendix IV

Inserting the derivations of $\frac{\partial f(y;\theta)}{\partial\theta_1}, \frac{\partial f(y;\theta)}{\partial\theta_2}, \frac{\partial^2 f(y;\theta)}{\partial\theta_1\partial\theta_2}, \frac{\partial^2 f(y;\theta)}{\partial\theta_1^2}$, and $\frac{\partial^2 f(y;\theta)}{\partial\theta_2^2}$

for the GPD yields

$$-\int_0^\infty \frac{\partial\varphi_\varepsilon}{\partial\varepsilon} dF(x) = -\int_0^\infty \left[\frac{x\beta + 2\varepsilon x^2 + \varepsilon^2 x^2}{\left(\beta\varepsilon + \varepsilon^2 x\right)^2} + \frac{x}{(\beta + \varepsilon x)\varepsilon^2} - \frac{2\ln\left(1 + \frac{\varepsilon x}{\beta}\right)}{\varepsilon^3} \right] f(x)\,dx$$

$$-\int_0^\infty \frac{\partial\varphi_\beta}{\partial\beta} dF(x) = -\int_0^\infty \left[\frac{1}{\beta^2} - \frac{x(1+\varepsilon)(2\beta + \varepsilon x)}{\left(\beta^2 + \beta\varepsilon x\right)^2} \right] f(x)\,dx$$

$$-\int_0^\infty \frac{\partial\varphi_\varepsilon}{\partial\beta} dF(x) = -\int_0^\infty \frac{\partial\varphi_\beta}{\partial\varepsilon} dF(x) = -\int_0^\infty \left[\frac{x}{\beta\varepsilon(\beta + \varepsilon x)} - \frac{\varepsilon x(1+\varepsilon)}{\left(\beta\varepsilon + \varepsilon^2 x\right)^2} \right] f(x)\,dx$$

(non-zero off-diagonals indicate parameter dependence)

11. Appendix IV

Inserting the derivations of

$$\frac{\partial f(y;\theta)}{\partial\theta_1},\frac{\partial f(y;\theta)}{\partial\theta_2},\frac{\partial^2 f(y;\theta)}{\partial\theta_1\partial\theta_2},\frac{\partial^2 f(y;\theta)}{\partial\theta_1^2},\frac{\partial^2 f(y;\theta)}{\partial\theta_2^2},\frac{\partial F(H;\theta)}{\partial\theta_1},\frac{\partial F(H;\theta)}{\partial\theta_2},\frac{\partial^2 F(H;\theta)}{\partial\theta_1\partial\theta_2},\frac{\partial F^2(H;\theta)}{\partial\theta_1^2},\text{ and }\frac{\partial F^2(H;\theta)}{\partial\theta_2^2}$$

for the (left) Truncated GPD yields

$$-\int_0^\infty \frac{\partial\varphi_\varepsilon}{\partial\varepsilon}dG(x) = -\frac{1}{\left[1-F(H;\beta,\varepsilon)\right]}\cdot\int_H^\infty\left[\frac{x\beta+2\varepsilon x^2+\varepsilon^2x^2}{\left(\beta\varepsilon+\varepsilon^2x\right)^2}+\frac{x}{(\beta+\varepsilon x)\varepsilon^2}-\frac{2\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^3}\right]f(x)dx$$

$$+\frac{\left(\int_0^H\left[\left(\frac{-x(1+\varepsilon)}{\beta\varepsilon+\varepsilon^2x}\right)+\frac{\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right]f(x;\beta,\varepsilon)dx\right)^2+\left[1-F(H;\beta,\varepsilon)\right]\cdot\int_0^H\left(\left[\frac{x\beta+2\varepsilon x^2+\varepsilon^2x^2}{\left(\beta\varepsilon+\varepsilon^2x\right)^2}+\frac{x}{(\beta+\varepsilon x)\varepsilon^2}-\frac{2\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^3}\right]+\left[\left(\frac{-x(1+\varepsilon)}{\beta\varepsilon+\varepsilon^2x}\right)+\frac{\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right]^2\right)f(x;\beta,\varepsilon)dx}{\left[1-F(H;\beta,\varepsilon)\right]^2}$$

$$-\int_0^\infty \frac{\partial\varphi_\beta}{\partial\beta}dG(x) = -\frac{1}{\left[1-F(H;\beta,\varepsilon)\right]}\cdot\int_H^\infty\left[\frac{1}{\beta^2}-\frac{x(1+\varepsilon)(2\beta+\varepsilon x)}{\left(\beta^2+\beta\varepsilon x\right)^2}\right]f(x)dx$$

$$+\frac{\left(\int_0^H-\frac{1}{\beta}\left[\frac{\beta-x}{\beta+\varepsilon x}\right]f(x;\beta,\varepsilon)dx\right)^2+\left[1-F(H;\beta,\varepsilon)\right]\cdot\int_0^H\left(\left[\frac{1}{\beta^2}-\frac{x(1+\varepsilon)(2\beta+\varepsilon x)}{\left(\beta^2+\beta\varepsilon x\right)^2}\right]+\frac{1}{\beta^2}\left[\frac{\beta-x}{\beta+\varepsilon x}\right]^2\right)f(x;\beta,\varepsilon)dx}{\left[1-F(H;\beta,\varepsilon)\right]^2}$$

11. Appendix IV

Inserting the derivations of

$$\frac{\partial f(y;\theta)}{\partial\theta_1},\frac{\partial f(y;\theta)}{\partial\theta_2},\frac{\partial^2 f(y;\theta)}{\partial\theta_1\partial\theta_2},\frac{\partial^2 f(y;\theta)}{\partial\theta_1^2},\frac{\partial^2 f(y;\theta)}{\partial\theta_2^2},\frac{\partial F(H;\theta)}{\partial\theta_1},\frac{\partial F(H;\theta)}{\partial\theta_2},\frac{\partial^2 F(H;\theta)}{\partial\theta_1\partial\theta_2},\frac{\partial F^2(H;\theta)}{\partial\theta_1^2},\text{ and }\frac{\partial F^2(H;\theta)}{\partial\theta_2^2}$$

for the (left) Truncated GPD yields

(non-zero off-diagonals indicate parameter dependence)

$$-\int_0^\infty \frac{\partial\varphi_\varepsilon}{\partial\beta}dG(x) = -\int_0^\infty \frac{\partial\varphi_\beta}{\partial\varepsilon}dG(x) = -\frac{1}{\left[1-F(H;\beta,\varepsilon)\right]}\cdot\int_H^\infty\left[\frac{x}{\beta\varepsilon(\beta+\varepsilon x)}-\frac{\varepsilon x(1+\varepsilon)}{\left(\beta\varepsilon+\varepsilon^2 x\right)^2}\right]f(x)dx$$

$$+\frac{\left(\int_0^H\left[\left(\frac{-x(1+\varepsilon)}{\beta\varepsilon+\varepsilon^2 x}\right)+\frac{\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right]f(x;\beta,\varepsilon)dx\right)\times\left(\int_0^H-\frac{1}{\beta}\left[\frac{\beta-x}{\beta+\varepsilon x}\right]f(x;\beta,\varepsilon)dx\right)}{\left[1-F(H;\beta,\varepsilon)\right]^2}$$

$$+\frac{\left[1-F(H;\beta,\varepsilon)\right]\cdot\int_0^H\left(\left[\frac{x\beta+2\varepsilon x^2+\varepsilon^2 x^2}{\left(\beta\varepsilon+\varepsilon^2 x\right)^2}+\frac{x}{(\beta+\varepsilon x)\varepsilon^2}-\frac{2\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^3}\right]+\left[\left(\frac{-x(1+\varepsilon)}{\beta\varepsilon+\varepsilon^2 x}\right)+\frac{\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right]^2\right)f(x;\beta,\varepsilon)dx}{\left[1-F(H;\beta,\varepsilon)\right]^2}$$

11. References

- Alaiz, M., and Victoria-Feser, M. (1996), "Modelling Income Distribution in Spain: A Robust Parametric Approach," *TDARP Discussion Paper No. 20*, STICERD, London School of Economics.
- Böcker, K., and Klüppelberg, C. (2005), "Operational VaR: A Closed-Form Approximation," *RISK Magazine*, 12, 90-93.
- Bowman, K.O., and Shenton, L.R. (1983), "Maximum Likelihood Estimators for the Gamma Distribution Revisited," Communications in Statistics-Simulation and Computation, 12, 697-710.
- Bowman, K.O., and Shenton, L.R. (1988), Properties of Estimators for the Gamma Distribution, Marcel Dekker, New York.
- Cpe., E, G. Mignola, G. Antonini, and R. Ugoccioni, "Chllenges and Pitfalls in Measuring Operational Risk from Loss Data," *The Journal of Operational Risk*, Vol. 4, No. 4, 3-27.
- Dupuis, D.J. (1998), "Exceedances Over High Thresholds: A Guide to Threshold Selection," *Extremes*, Vol. 1, No. 3, 251-261.
- Ergashev B., (2008), "Should Risk Managers Rely on the Maximum Likelihood Estimation Method While Quantifying Operational Risk," *The Journal of Operational Risk*, Vol. 3, No. 2, 63-86.
- Grimshaw, S. (1993), "Computing Maximum Likelihood Estimates for the Generalized Pareto Distribution," *Technometrics*, Vol. 35, No. 2, 185-191.
- Hampel, F.R., E. Ronchetti, P. Rousseeuw, and W. Stahel, (1986), *Robust Statistics: The Approach Based on Influence Functions*, John Wiley and Sons, New York.
- Horbenko, N., Ruckdeschel, P. and Bae, T. (2011), "Robust Estimation of Operational Risk," *The Journal of Operational Risk*, Vol.6, No.2, 3-30.
- Huber, P.J. (1964), "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35, 73-101.
- Huber, P.J. (1977), *Robust Statistical Procedures*, SIAM, Philadelphia.
- Huber, P.J. (1981), *Robust Statistics*, John Wiley and Sons, Inc.
- Stefanski, L., and Boos, D. (2002), *The American Statistician*, Vol. 56, No. 1, pp.29-38.
- Victoria-Feser, M., and Ronchetti, E. (1994), "Robust Methods for Personal-Income Distribution Models," *The Canadian Journal of Statistics*, Vol.22, No.2, pp.247-258.

J.D. Opdyke
President, DataMineit
JDOpdyke@DataMineit.com
www.DataMineit.com

Providing risk analytics and statistical consulting to the banking, credit, and consulting sectors.