Better Capital Planning via Exact Sensitivity Analysis Using the Influence Function

J.D. Opdyke, Senior Managing Director, DataMineit
JDOpdyke@DataMineit.com

Presented at American Bankers Association Operational Risk Modeling Forum
Washington, DC, July 18-20, 2012

***The views presented herein are the views of the sole author, J.D. Opdyke.**
All derivations, and all calculations and computations, were performed by J.D. Opdyke using SAS®.

American Bankers Association

DataMineIt

Contents

American Bankers Association

DataMineIt

1. Why Use the IF in OpRisk Severity Modeling?

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

Specific Objectives:

1) **Use the Influence Function (IF) to Develop and/or Select Estimators that yield an estimated capital distribution that is i) more robust to extreme tail events, ii) less variable, and iii) less biased vis-à-vis Jensen's inequality.**

2) **THEN, based on 1), Use IF to Generate Exact Capital Sensitivity Curves for Capital Planning. These show the EXACT impact on capital of additional (or dropped) losses.**

* For purposes of this presentation, potential dependence between the frequency and severity distributions is ignored. See Ergashev (2008).

2a. The Influence Function (IF) Defined

- **The IF can be used to demonstrate the "influence" that a data point of "contamination"* which deviates from the assumed severity distribution has on the estimated parameter:**

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

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** $T\left(F\right)=\int u\,dF\left(u\right)=\int uf\left(u\right)du$
- 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 "contamination," "statistical contamination," and "arbitrary deviation" 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.**

2a. The Influence Function (IF) Defined

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

- **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, χ (χ obviously can represent more than one data point as ε is a proportion of the data sample, but it is easier conceptually to view χ as a single data point whereby $\varepsilon = 1/n$: when this is combined with use of the empirical distribution, $\hat{F}$, this is, in fact, the Empirical Influence Function (EIF) – see below).**

- **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 χ, 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.**

2a. The Influence Function (IF) Defined

- **Note that 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).**

- **B-Robustness, arguably the most common definition of statistical robustness of an estimator, is based on the IF, and oftentimes the motivation for its derivation.**
- **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.**

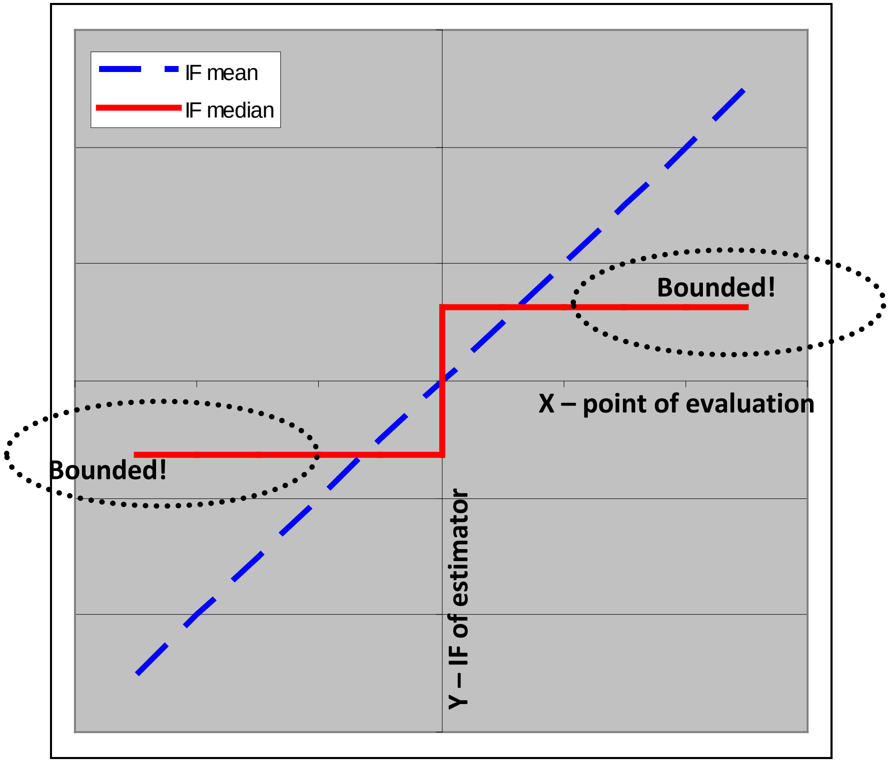
$$GES = \gamma^{*}\left(T,F\right) = \sup_{X}\left|IF\left(x;T,F\right)\right|$$

- **Comparing IFs of two estimators of location – the mean and the median – effectively demonstrates the concept of B-robustness.**

§ "B" comes from "bias," because if IF is bounded, so, too, must be the bias of the estimator is bounded (if any).

2a. The Influence Function (IF) Defined

Graph 1: Influence Functions of the Mean and the Median



- **Because the IF of the mean is <u>unbounded</u>, a single arbitrarily large (small) data point can render the mean meaninglessly large (small), 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).**

2a. The Influence Function (IF) Defined

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…**

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

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

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

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

2b. The Empirical Influence Function (EIF) Defined

- **<u>Empirical Influence Function</u>: The EIF naturally corresponds with the IF, and is given by**

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

- **EIF is simply the IF based on the empirical distribution.**

- **In practice, EIF is used as a plot of the difference between the values of the estimator based on the sample with and without the contaminated data point, *x*, as a function of *x*. The difference between the two estimator values is scaled by $\varepsilon = 1/n$. Even for relatively small sample sizes, EIF ≈ IF, so when samples of data are generated from a given distribution, *F*, the EIF can serve as an easily implemented verification that the calculations underlying the IF (which sometimes can be quite involved) are right. However, IF always is needed to establish definitively the relationship between the estimated parameter and *x*, for all relevant *x*.**

2c. The Influence Function Derived: MLE Examples

Below are derived the IFs of the parameters of Six Severity Distributions:

- **LogNormal**
- **LogGamma**
- **Generalized Pareto Distribution (GPD)**
- **Truncated LogNormal**
- **Truncated LogGamma**
- **Truncated GPD**

2c. The Influence Function Derived: MLE Examples

- **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**

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

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

So for MLE:

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

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

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

2c. The Influence Function Derived: MLE Examples

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

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

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

So we can write

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

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

And so the above becomes:

2c. The Influence Function Derived: MLE Examples

IF of MLEs for (left) truncated densities:

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

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

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

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

And so the IF is

2c. The Influence Function Derived: MLE Examples

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, and φ_{θ} is now vectorized. Parameter dependence exists when the off-diagonal terms are not zero.

2c. The Influence Function Derived: MLE Examples

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

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

for truncated and non-truncated distributions, respectively

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

2c. The Influence Function Derived: MLE Examples

LogNormal Derivatives:

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

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

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

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

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

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

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

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

2c. The Influence Function Derived: MLE Examples

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

into the Fisher Information for the LogNormal yields

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

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

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

2c. The Influence Function Derived: MLE Examples

which yields…

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

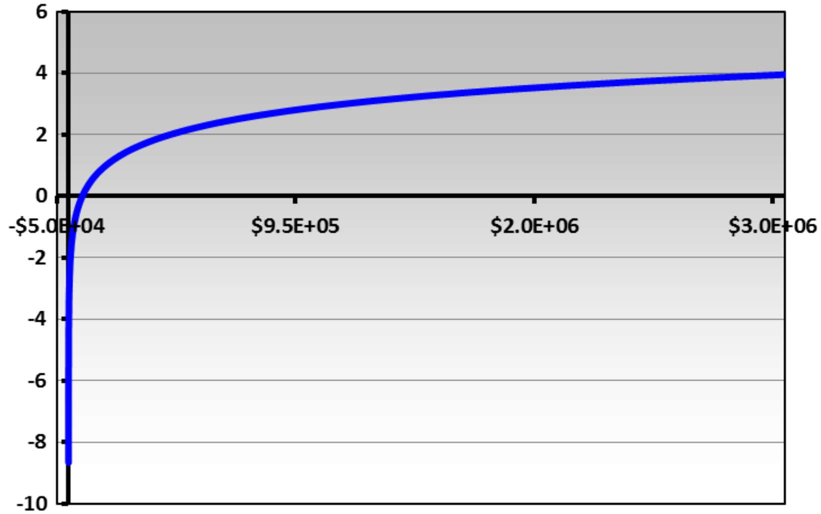
(zero off-diagonals indicate no parameter dependence)

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

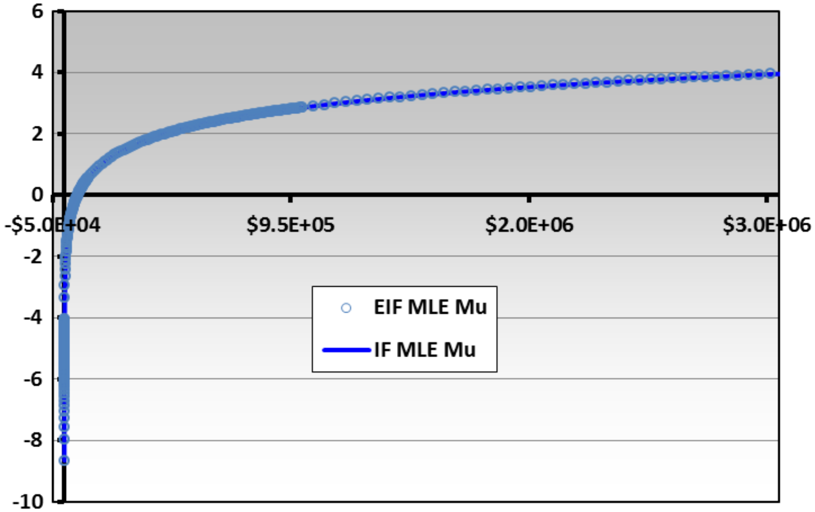
$$=\begin{bmatrix}-\sigma^2 & 0\\ 0 & -\sigma^2/2\end{bmatrix}\begin{bmatrix}\dfrac{\mu-\ln\left(x\right)}{\sigma^2}\\ \dfrac{1}{\sigma}-\dfrac{\left(\ln\left(x\right)-\mu\right)^2}{\sigma^3}\end{bmatrix}=\begin{bmatrix}\ln\left(x\right)-\mu\\ \dfrac{\left(\ln\left(x\right)-\mu\right)^2-\sigma^2}{2\sigma}\end{bmatrix}$$

2c. The Influence Function Derived: MLE Examples

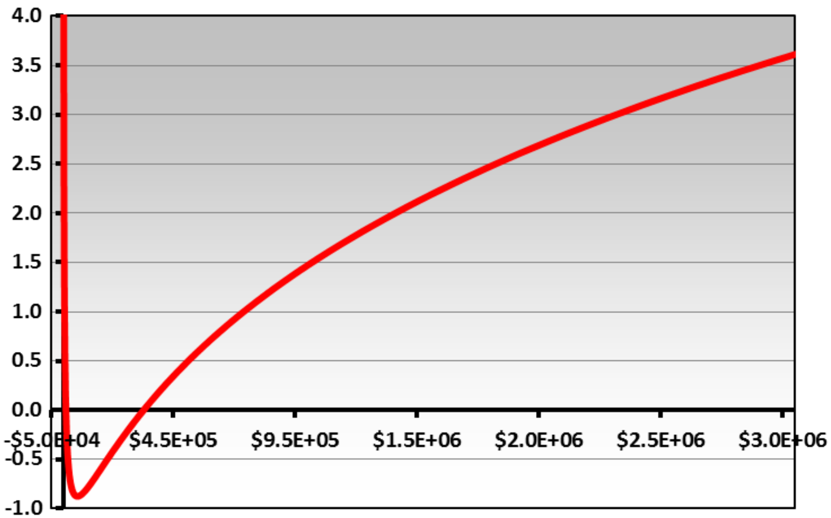
LogNormal: MLE IF



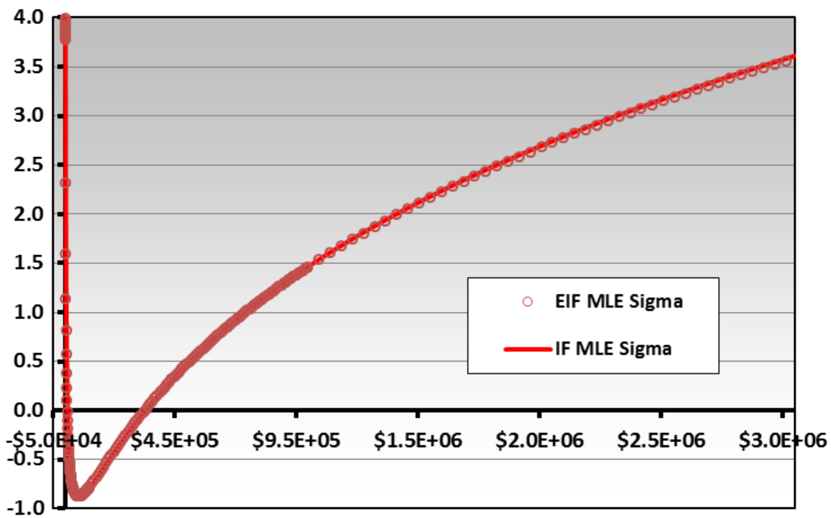
LogNormal: MLE IF vs. EIF



$\mu = 10.95$



$\sigma = 1.75$

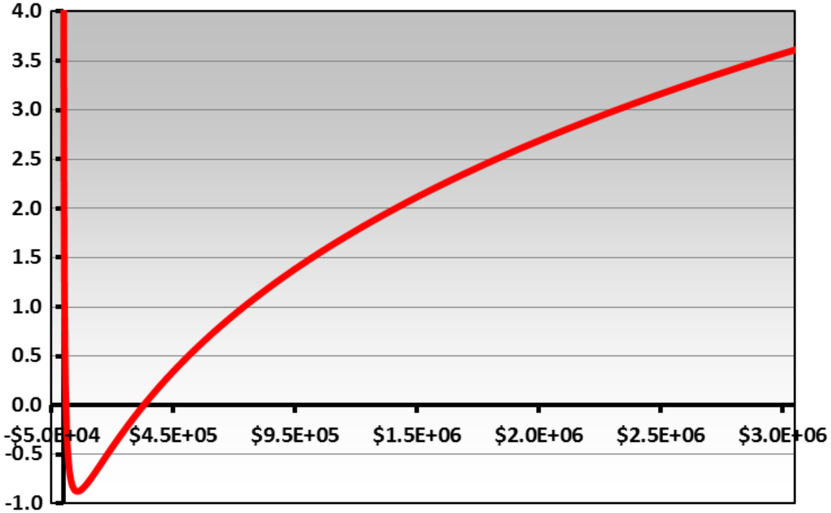


American Bankers Association

DataMineIt

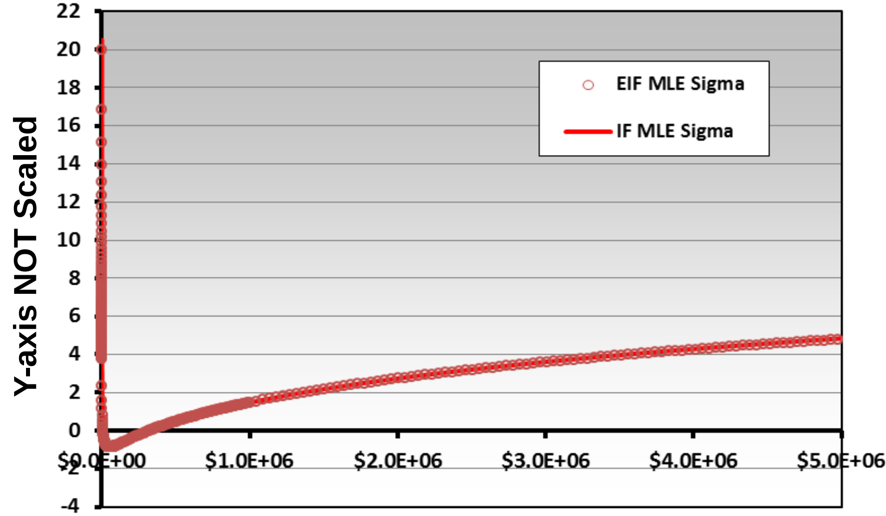
2c. The Influence Function Derived: MLE Examples

LogNormal: MLE IF



σ **= 1.75**

LogNormal: MLE IF vs. EIF



2c. The Influence Function Derived: MLE Examples

LogNormal Derivatives (for (left) Truncated case):

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

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

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

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

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

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

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

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

2c. The Influence Function Derived: MLE Examples

For the (left) Truncated LogNormal, inserting the derivations of

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

into the Fisher Information 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)

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

2c. The Influence Function Derived: MLE Examples

And inserting the derivatives into the φ_θ function yields:

$$\varphi_\theta = \begin{bmatrix} \varphi_\mu \\ \varphi_\sigma \end{bmatrix} = \begin{bmatrix} \partial\rho(x,\theta)/\partial\mu \\ \partial\rho(x,\theta)/\partial\sigma \end{bmatrix} = \begin{bmatrix} -\dfrac{\partial f(x,\theta)}{\partial\mu} \Big/ f(x,\theta) \\ -\dfrac{\partial f(x,\theta)}{\partial\sigma} \Big/ f(x,\theta) \end{bmatrix} = \begin{bmatrix} -\left[\dfrac{\ln(x)-\mu}{\sigma^2}\right] - \dfrac{\int_0^H \left[\dfrac{\ln(y)-\mu}{\sigma^2}\right] f(y;\mu,\sigma)\,dy}{1-F(H;\mu,\sigma)} \\ -\left[\dfrac{(\ln(x)-\mu)^2}{\sigma^3} - \dfrac{1}{\sigma}\right] - \dfrac{\int_0^H \left[\dfrac{(\ln(y)-\mu)^2}{\sigma^3} - \dfrac{1}{\sigma}\right] f(y;\mu,\sigma)\,dy}{1-F(H;\mu,\sigma)} \end{bmatrix}$$

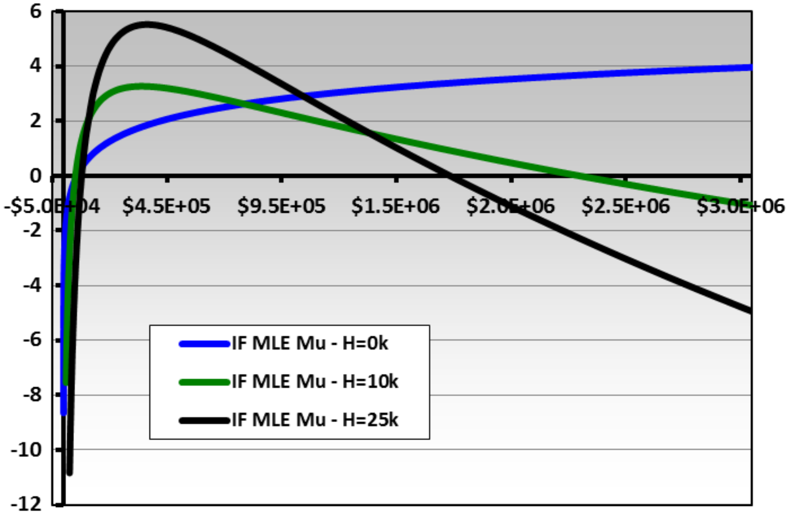
The Influence Function

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

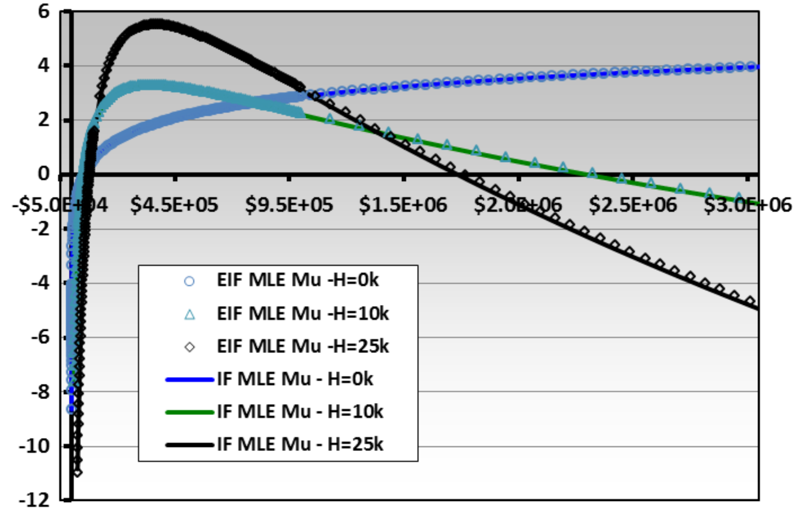
is then calculated numerically, as it is for all the remaining severity distributions except for the LogGamma.

2c. The Influence Function Derived: MLE Examples

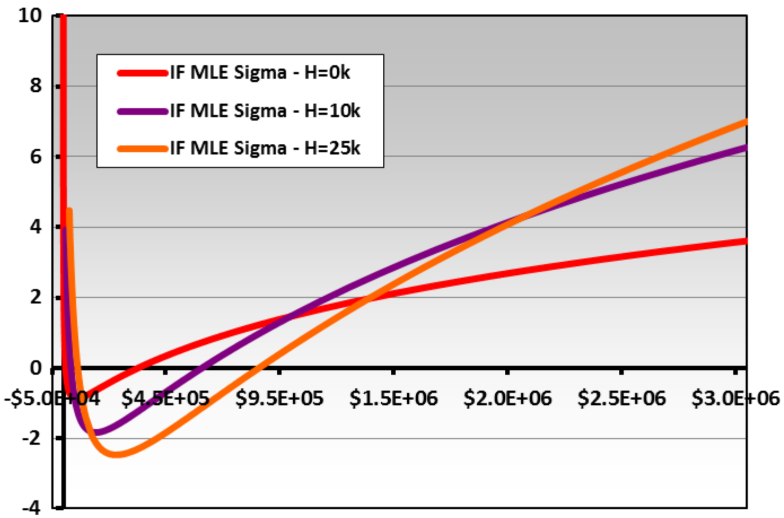
Truncated LogNormal: MLE IF



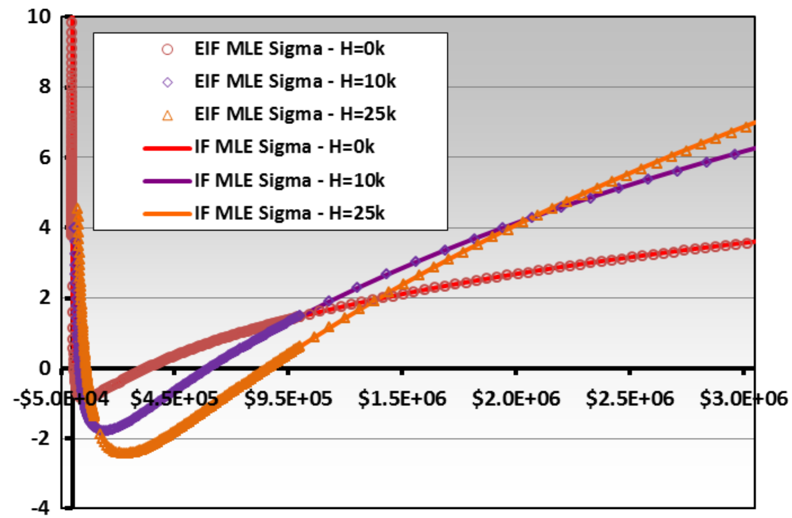
Truncated LogNormal: MLE IF vs. EIF



$\mu = 10.95$



$\sigma = 1.75$



American Bankers Association

DataMineIt

2c. The Influence Function Derived: MLE Examples

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

- **For the LogNormal, truncation causes not only a change in the shape, but also a change in the DIRECTION of $\hat{\mu}$(x) as *x* increases. Many would call this unexpected, if not counter-intuitive: when arbitrary deviations INCREASE, what many consider the location parameter, *μ*, actually DECREASES (exp(*μ*) is actually the scale parameter of the distribution).**

- **Note that this is not true for σ, which still increases as *x* increases, so truncation induces NEGATIVE covariance between the parameters.**

- **Many have thought this finding, when it shows up in simulations, to be numeric instability in the convergence algorithms used to obtain MLE estimators, but as the IF shows, this is the right result. And of course, neither the definition of the LogNormal density, nor that of the truncated LogNormal density, prohibits negative values for *μ*.**

- **This is but one example of the ways in which the IF can provide definitive answers to difficult statistical questions about which simulation-based approaches can provide only speculation and musing.**

2c. The Influence Function Derived: MLE Examples

LogGamma Distribution Derivatives:

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

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

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

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

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

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

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

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

2c. The Influence Function Derived: MLE Examples

Inserting the derivations of

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

into the Fisher Information for the LogGamma yields

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

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

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

2c. The Influence Function Derived: MLE Examples

which 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)

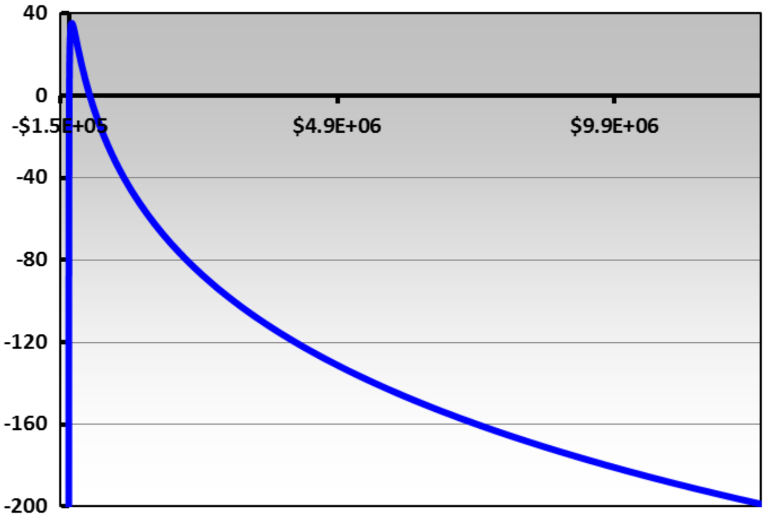
$$=\begin{bmatrix}-trigamma\left(a\right) & 1/b\\ 1/b & -a/b^2\end{bmatrix}^{-1}\begin{bmatrix}-\ln\left(b\right)-\ln\left(\ln\left(x\right)\right)+digamma\left(a\right)\\ -\frac{a}{b}+\ln\left(x\right)\end{bmatrix}=$$

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

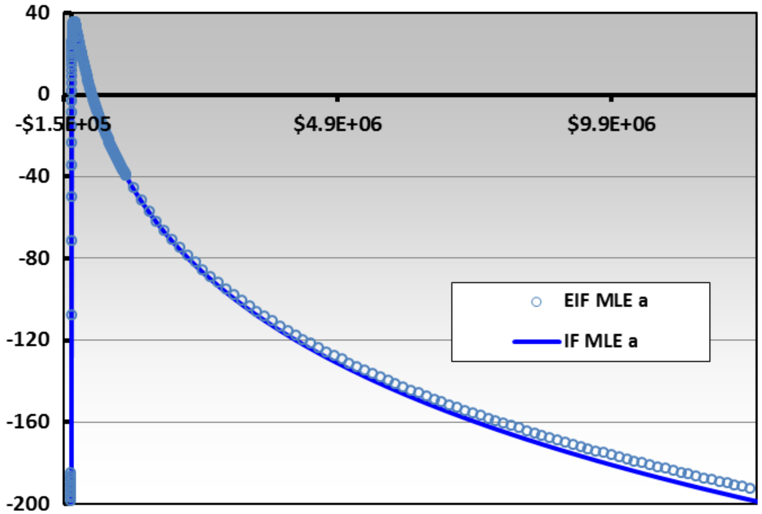
$$=\begin{bmatrix}\dfrac{\frac{a}{b^2}\left[\ln\left(b\right)+\ln\left(\ln\left(x\right)\right)-digamma\left(a\right)\right]-\frac{1}{b}\left[\ln\left(x\right)-\frac{a}{b}\right]}{trigamma\left(a\right)\left(\frac{a}{b^2}\right)-\frac{1}{b^2}}\\ \dfrac{\frac{1}{b}\left[\ln\left(b\right)+\ln\left(\ln\left(x\right)\right)-digamma\left(a\right)\right]-trigamma\left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

2c. The Influence Function Derived: MLE Examples

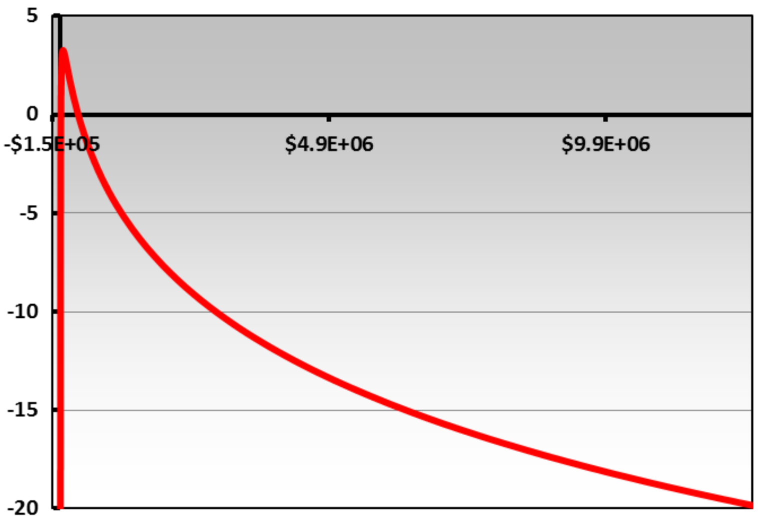
LogGamma: MLE IF

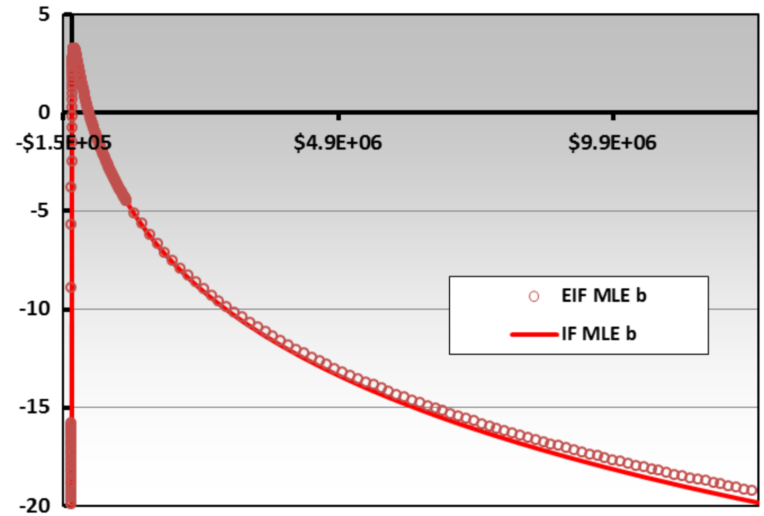


LogGamma: MLE IF vs. EIF



a = 35.5

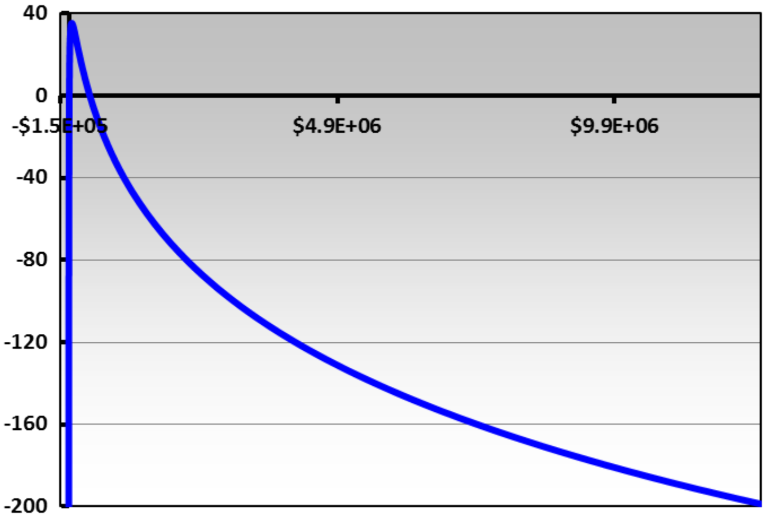




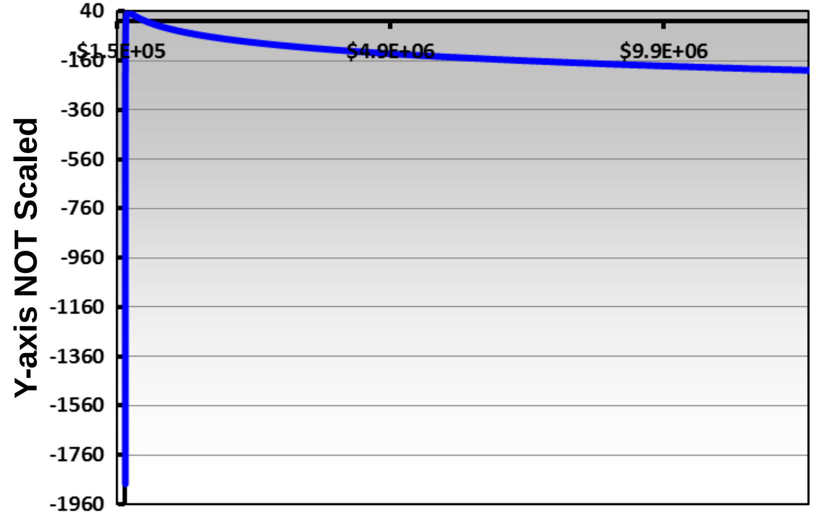
b = 3.25

2c. The Influence Function Derived: MLE Examples

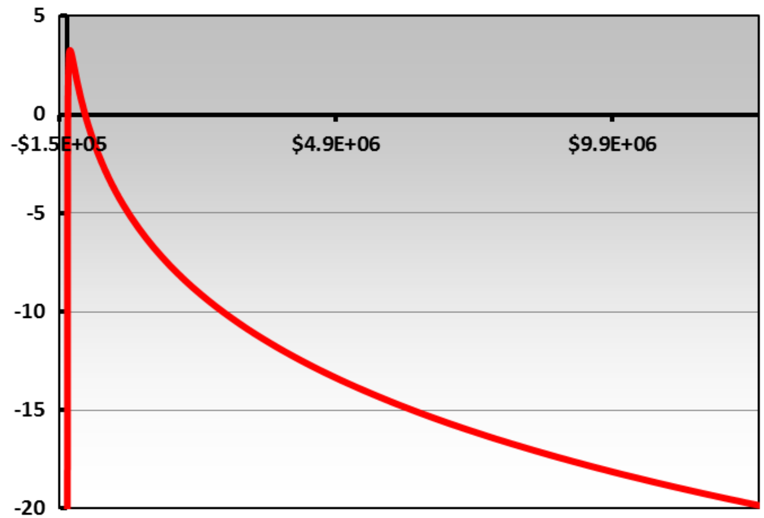
LogGamma: MLE IF



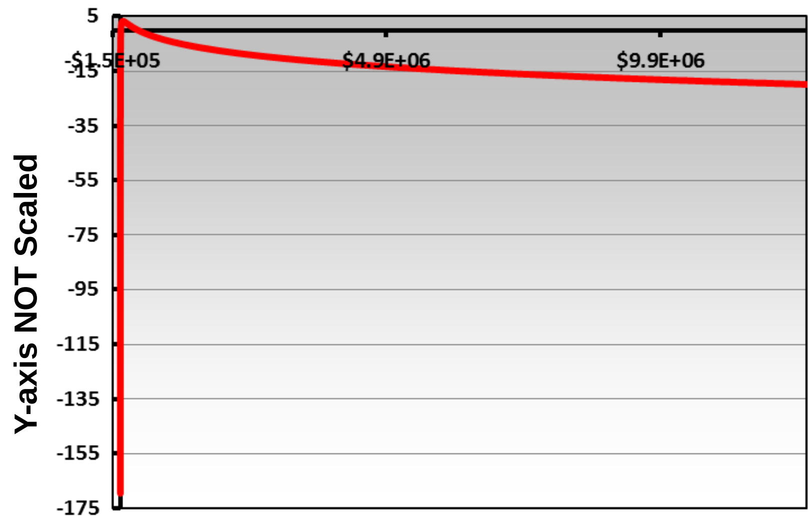
LogGamma: MLE IF



a = 35.5



b = 3.25



2c. The Influence Function Derived: MLE Examples

LogGamma Derivatives (for (left) Truncated Case):

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

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

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

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

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

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

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

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

2c. The Influence Function Derived: MLE Examples

For the (left) Truncated LogGamma, inserting the derivations of

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

into the Fisher Information yields:

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

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

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

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

(non-zero off-diagonals indicate parameter dependence)

2c. The Influence Function Derived: MLE Examples

And inserting the derivatives into the φ_θ function yields:

$$\varphi_\theta = \begin{bmatrix} \varphi_a \\ \varphi_b \end{bmatrix} = \begin{bmatrix} \partial\rho(x,\theta)/\partial a \\ \partial\rho(x,\theta)/\partial b \end{bmatrix} = \begin{bmatrix} -\dfrac{\partial f(x,\theta)}{\partial a} \Big/ f(x,\theta) \\ -\dfrac{\partial f(x,\theta)}{\partial b} \Big/ f(x,\theta) \end{bmatrix} = \begin{bmatrix} -\left[\ln(b)+\ln(\ln(y)) - digam(a)\right] - \dfrac{\int_1^H \left[\ln(b)+\ln(\ln(y)) - digam(a)\right] f(y;a,b)\,dy}{1-F(H;\mu,\sigma)} \\ -\left[\dfrac{a}{b} - \ln(y)\right] - \dfrac{\int_1^H \left[\dfrac{a}{b} - \ln(y)\right] f(y;a,b)\,dy}{1-F(H;\mu,\sigma)} \end{bmatrix}$$

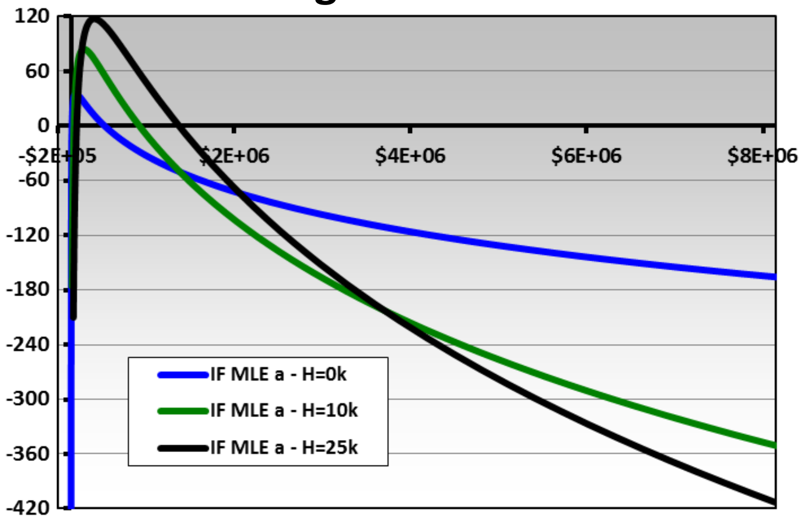
The Influence Function

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

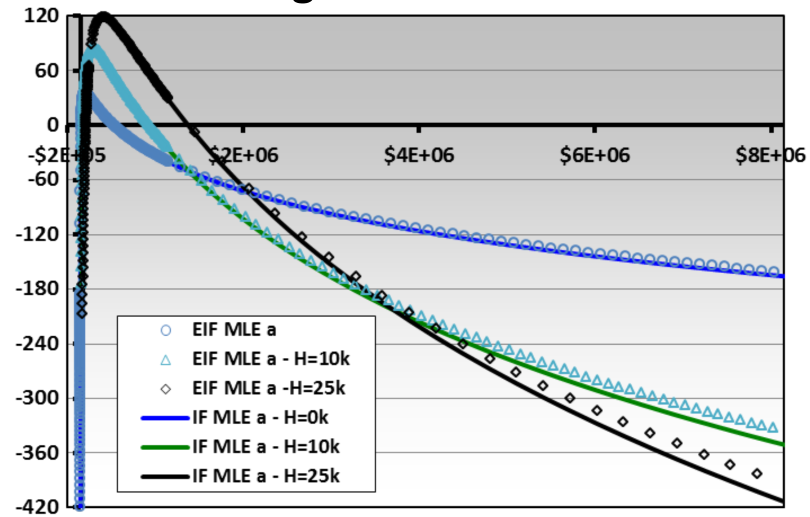
is then calculated numerically.

2c. The Influence Function Derived: MLE Examples

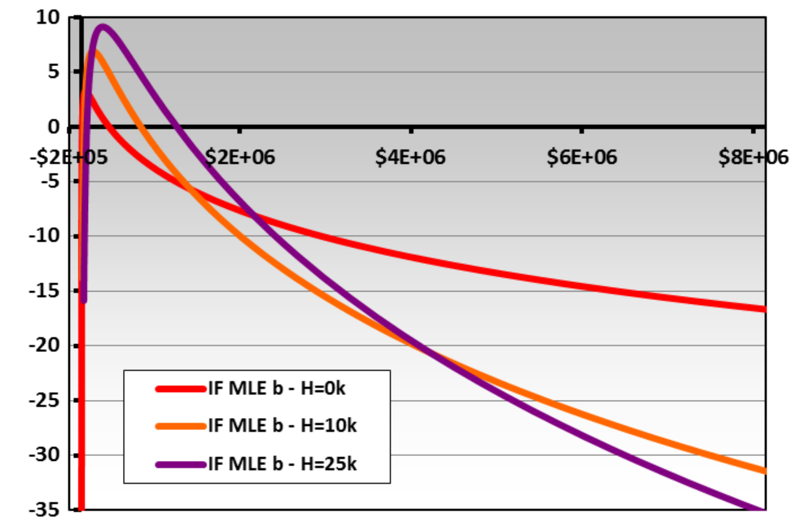
Truncated LogGamma: MLE IF



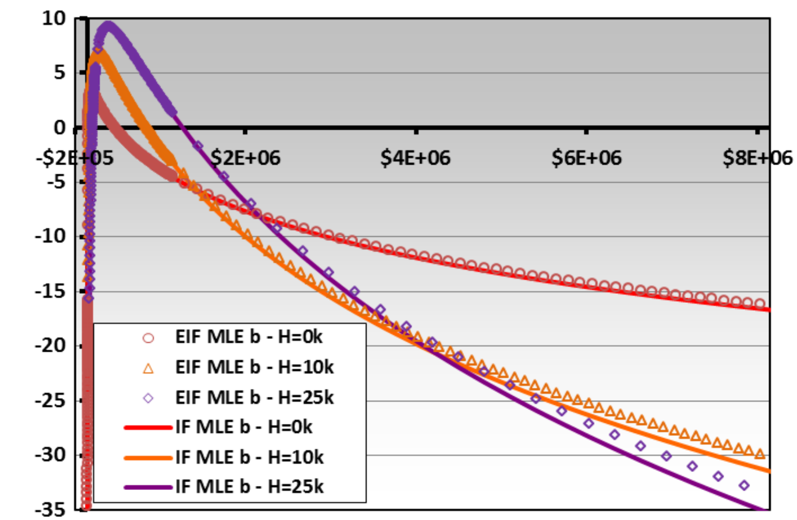
Truncated LogGamma: MLE IF vs. EIF



a = 35.5



b = 3.25



2c. The Influence Function Derived: MLE Examples

Generalized Pareto Distribution (GPD) Derivatives:

for $0 \le x < \infty;\ 0 < \beta;\ 0 \le \varepsilon$

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

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

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

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

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

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

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

2c. The Influence Function Derived: MLE Examples

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

into the Fisher Information 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)

2c. The Influence Function Derived: MLE Examples

And inserting the derivatives into the φ_θ function yields:

$$\varphi_\theta = \begin{bmatrix} \varphi_\beta \\ \varphi_\xi \end{bmatrix} = \begin{bmatrix} \partial\rho(x,\theta)/\partial\beta \\ \partial\rho(x,\theta)/\partial\xi \end{bmatrix} = \begin{bmatrix} -\dfrac{\partial f(x,\theta)}{\partial\beta} \Big/ f(x,\theta) \\ -\dfrac{\partial f(x,\theta)}{\partial\xi} \Big/ f(x,\theta) \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\beta}\left[\dfrac{\beta - x}{\beta + \varepsilon x}\right] \\ -\left[\left(\dfrac{-x(1+\varepsilon)}{\beta\varepsilon + \varepsilon^2 x}\right) + \dfrac{\ln\left(1 + \dfrac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right] \end{bmatrix}$$

The Influence Function

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

is then calculated numerically.

American Bankers Association

DataMineIt

2c. The Influence Function Derived: MLE Examples

Note that for the GPD specifically, Smith (1987)* conveniently simplifies the Fisher Information to yield

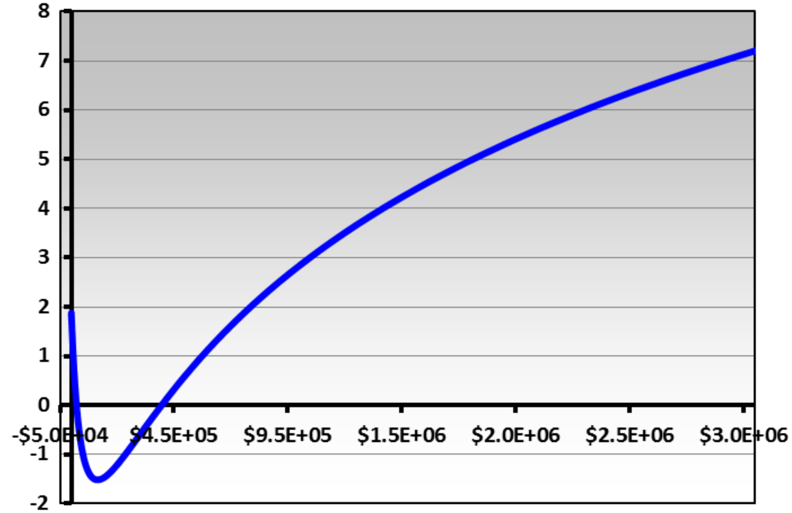
$$A(\theta)^{-1} = (1+\xi)\begin{bmatrix} 1+\xi & -\beta \\ -\beta & 2\beta^2 \end{bmatrix}$$

This gives the exact same result, as shown in the graphs below, as the numerical implementation above, and provides further independent validation of the more general framework presented herein (which can be used with all commonly used severity distributions).

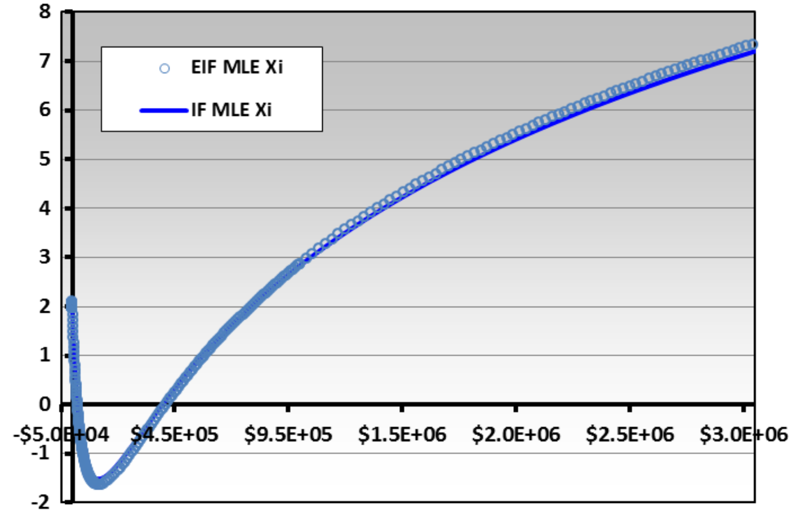
***NOTE: Smith (1987) is the oldest publication of this result that I have been able to find. Ruckdeschel & Horbenko (2010) re-present it in the context of Operational Risk.**

2c. The Influence Function Derived: MLE Examples

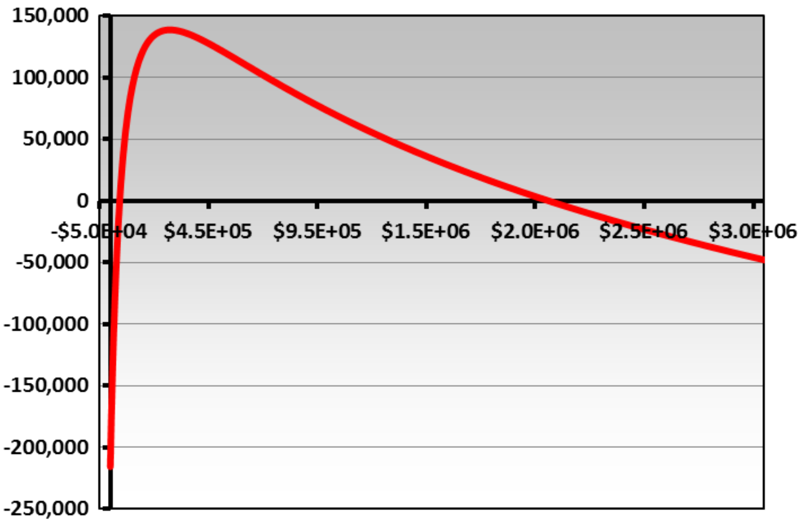
GPD: MLE IF



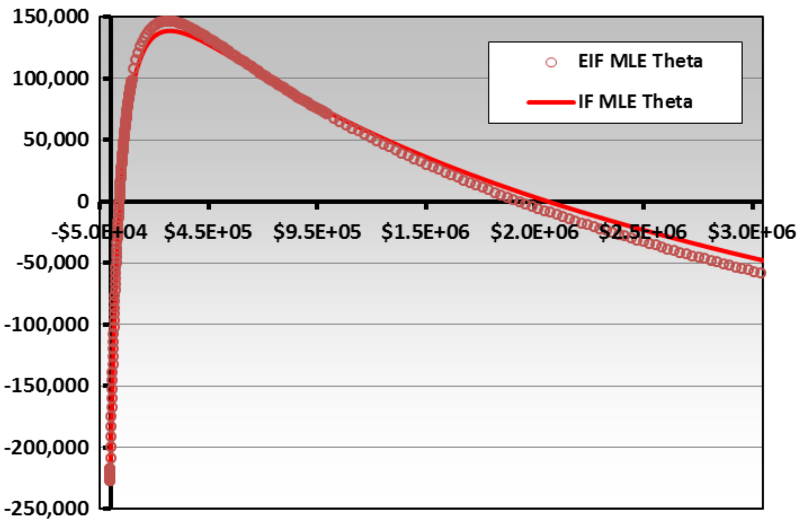
GPD: MLE IF vs. EIF



ξ = 0.875



β = 57,500



2c. The Influence Function Derived: MLE Examples

GPD Derivatives (for (left) Truncated Case):

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

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

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

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

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

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

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

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

2c. The Influence Function Derived: MLE Examples

For the (left) Truncated GPD, inserting the derivations of

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

into the Fisher Information yields:

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

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

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

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

2c. The Influence Function Derived: MLE Examples

(non-zero off-diagonals indicate parameter dependence)

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

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

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

2c. The Influence Function Derived: MLE Examples

And inserting the derivatives into the φ_θ function yields:

$$\varphi_\theta = \begin{bmatrix} \varphi_\beta \\ \varphi_\xi \end{bmatrix} = \begin{bmatrix} \partial\rho(x,\theta)/\partial\beta \\ \partial\rho(x,\theta)/\partial\xi \end{bmatrix} = \begin{bmatrix} -\dfrac{\partial f(x,\theta)}{\partial\beta} \Big/ f(x,\theta) \\ -\dfrac{\partial f(x,\theta)}{\partial\xi} \Big/ f(x,\theta) \end{bmatrix} =$$

$$= \begin{bmatrix} -\left[-\dfrac{1}{\beta}\left[\dfrac{\beta - x}{\beta + \varepsilon x}\right]\right] - \dfrac{\int_0^H -\dfrac{1}{\beta}\left[\dfrac{\beta - x}{\beta + \varepsilon x}\right] f(x;\beta,\varepsilon)\,dx}{1 - F(H;\mu,\sigma)} \\ -\left[\left(\dfrac{-x(1+\varepsilon)}{\beta\varepsilon + \varepsilon^2 x}\right) + \dfrac{\ln\left(1 + \dfrac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right] - \dfrac{\int_0^H \left[\left(\dfrac{-x(1+\varepsilon)}{\beta\varepsilon + \varepsilon^2 x}\right) + \dfrac{\ln\left(1 + \dfrac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right] f(x;\beta,\varepsilon)\,dx}{1 - F(H;\mu,\sigma)} \end{bmatrix}$$

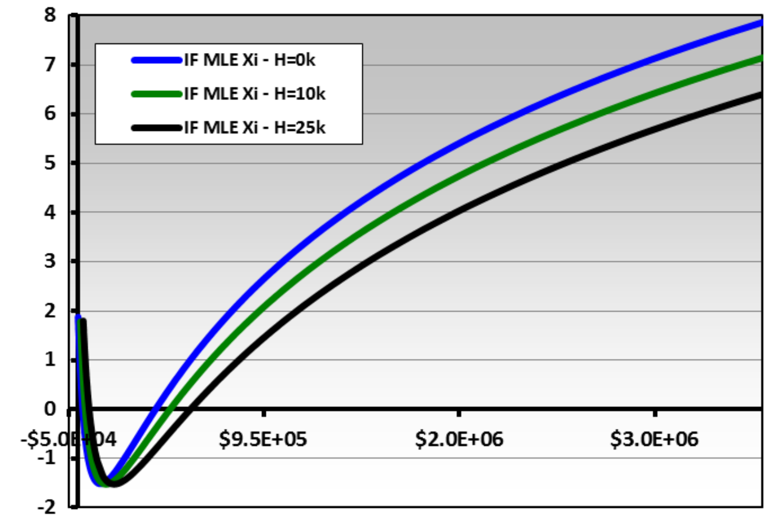
The Influence Function

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

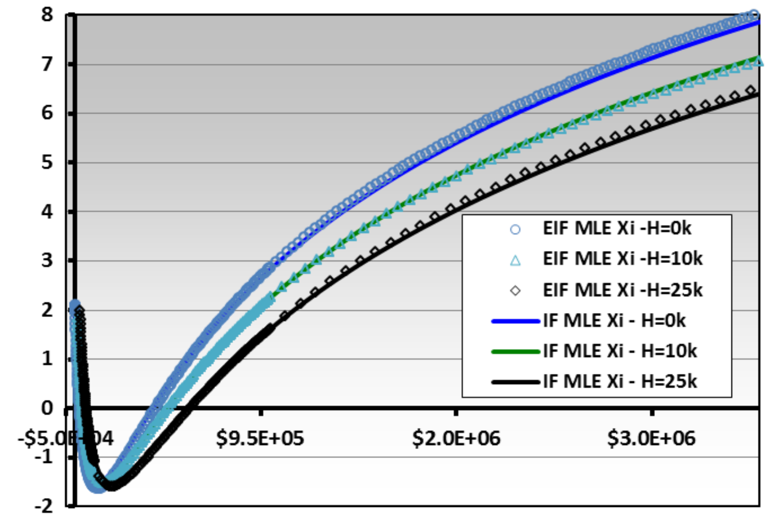
is then calculated numerically.

2c. The Influence Function Derived: MLE Examples

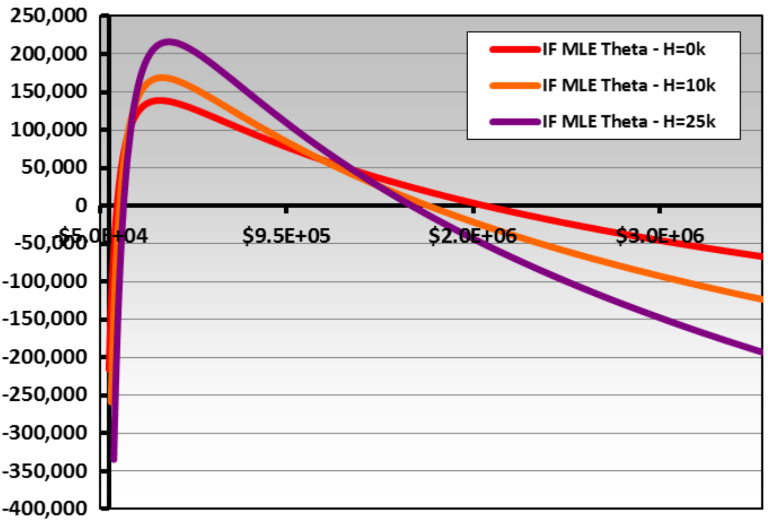
Truncated GPD: MLE IF



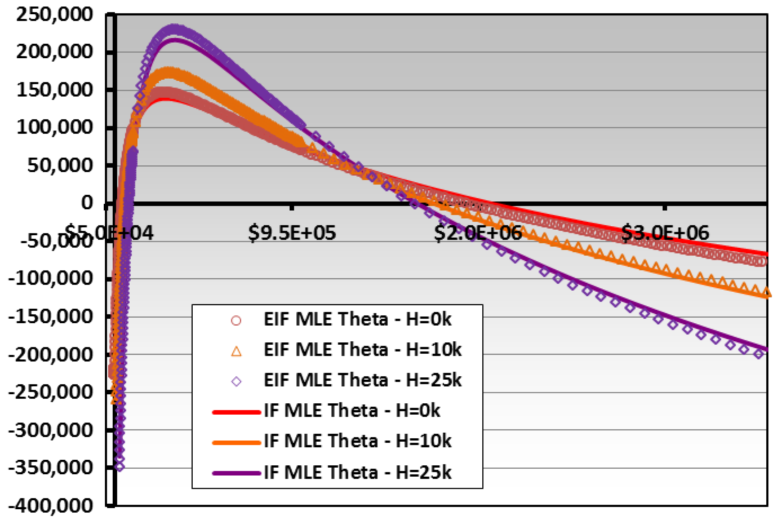
Truncated GPD: MLE IF vs. EIF



ε = 0.875



β = 57,500



American Bankers Association

DataMineIt

3. Using IF to Inform the Choice of Severity Estimator

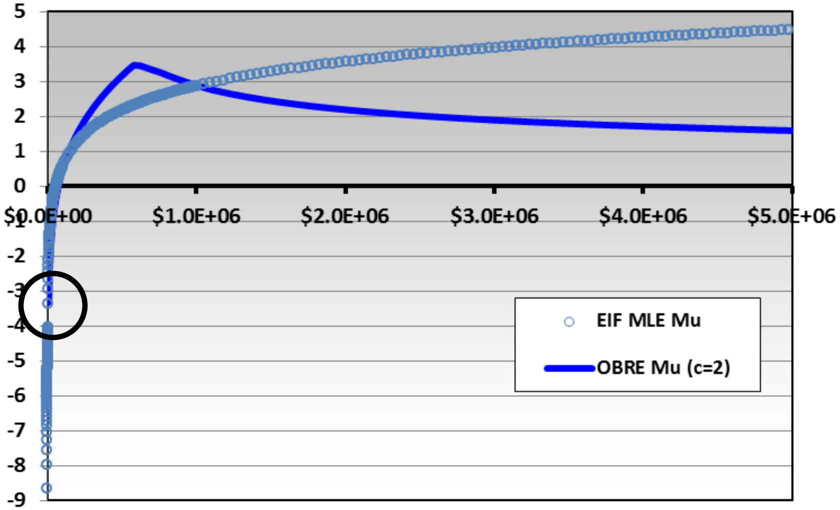
To restate, there are at least two major uses of the IF in this setting.

1.The first is to compare graphs and derivations like those generated above to choose and/or develop estimators that satisfy specified criteria most relevant to the particular setting in which they're being used. For example, non-iid data may be endemic to some settings (like the OpRisk setting), thus indicating the need for (B-)robust estimators (like OBRE – see Appendix 1). This is consistent with the ultimate goal of the OpRisk modeling exercise which is to generate a "better" estimated capital distribution, i.e. one that is i) more precise, ii) less biased, and iii) more robust to outlying loss events.

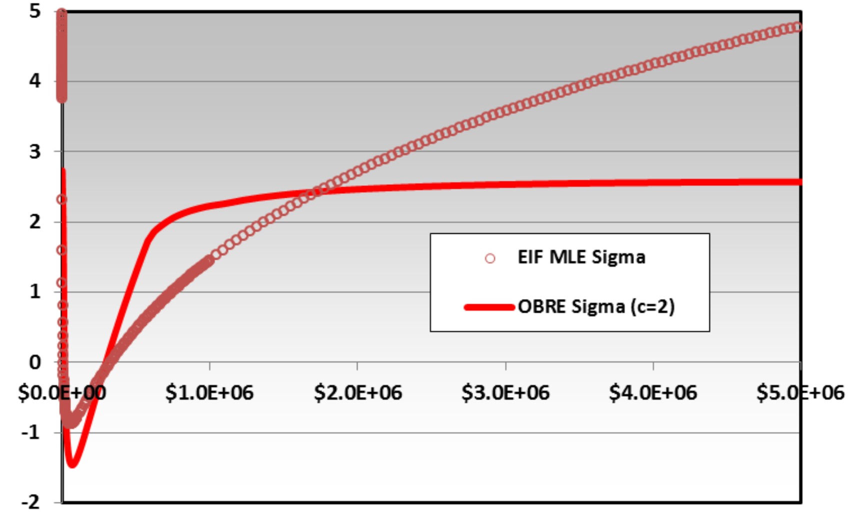
2.The second use of the IF, which builds on the first, is to use the EXACT CHANGES IN PARAMETER VALUES resulting from additional, dropped, or changed loss values ("*x*" on all the graphs above) to generate EXACT CAPITAL SENSITIVITY CURVES, and then use these curves for more effective and precise capital planning. This is treated in the sections below after a discussion of 1.

3a. Estimators for a "Better" Capital Distribution: More Stability over Time – More Robustness to Extreme Tail Events

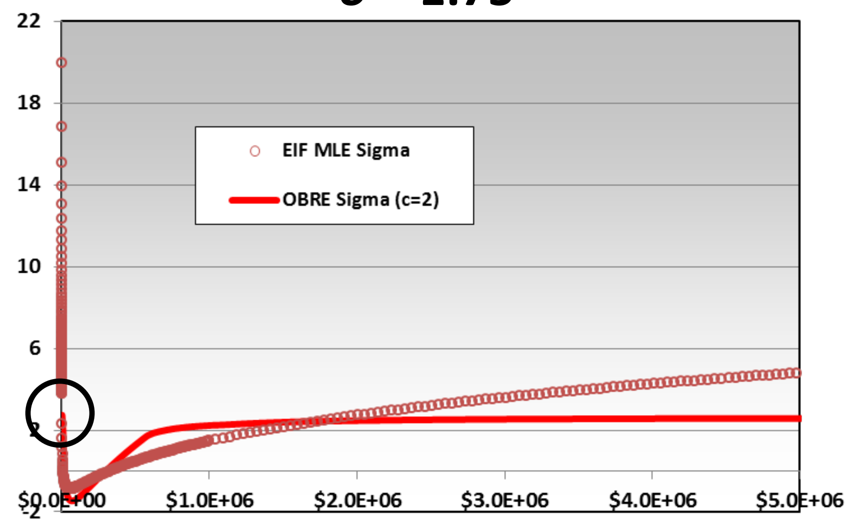
LogNormal EIFs: MLE vs. OBRE



μ = 10.95



σ = 1.75

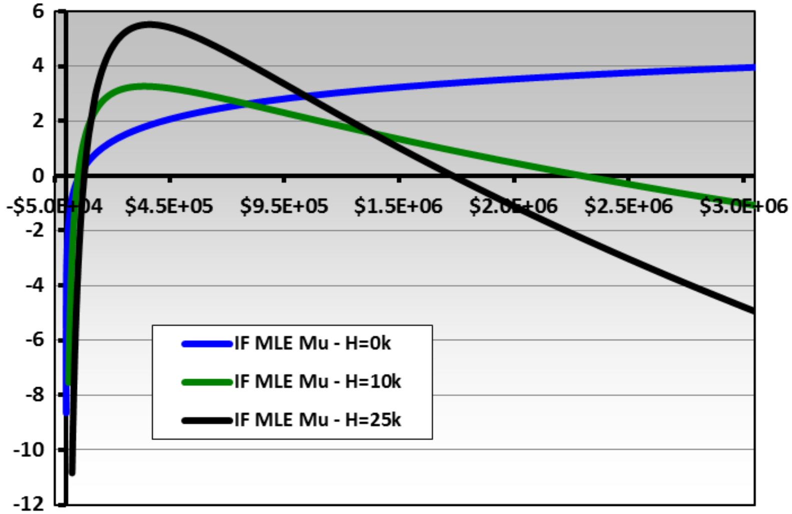


American Bankers Association

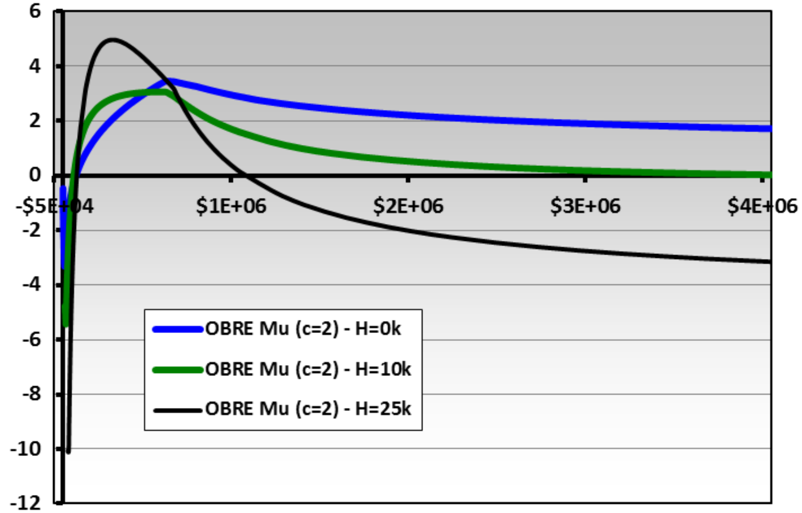
DataMineIt

3a. Estimators for a "Better" Capital Distribution: More Stability over Time – More Robustness to Extreme Tail Events

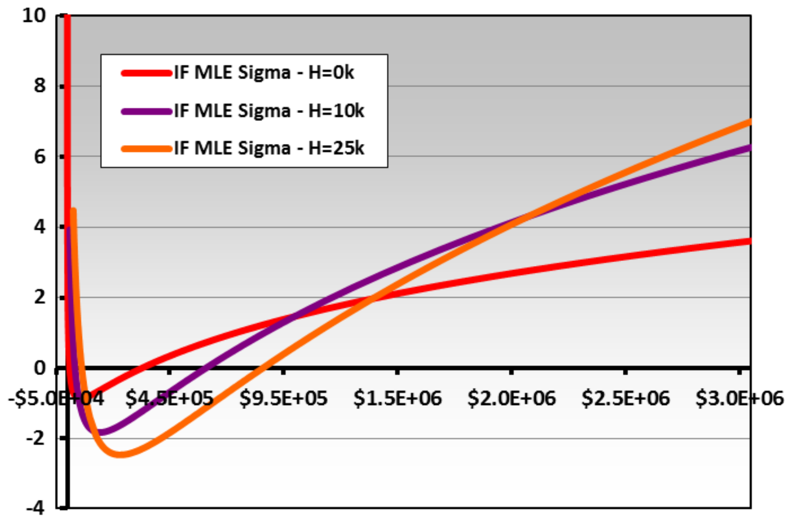
Truncated LogNormal: MLE EIF



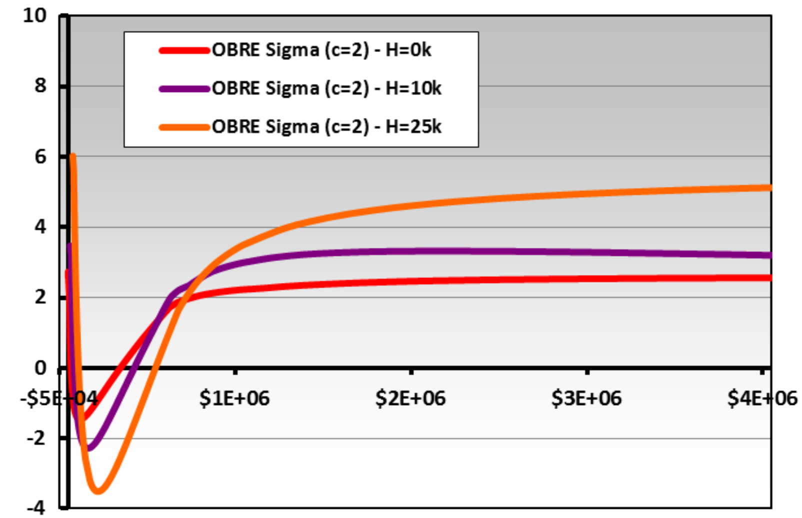
Truncated LogNormal: OBRE EIF



$\mu = 10.95$



$\sigma = 1.75$

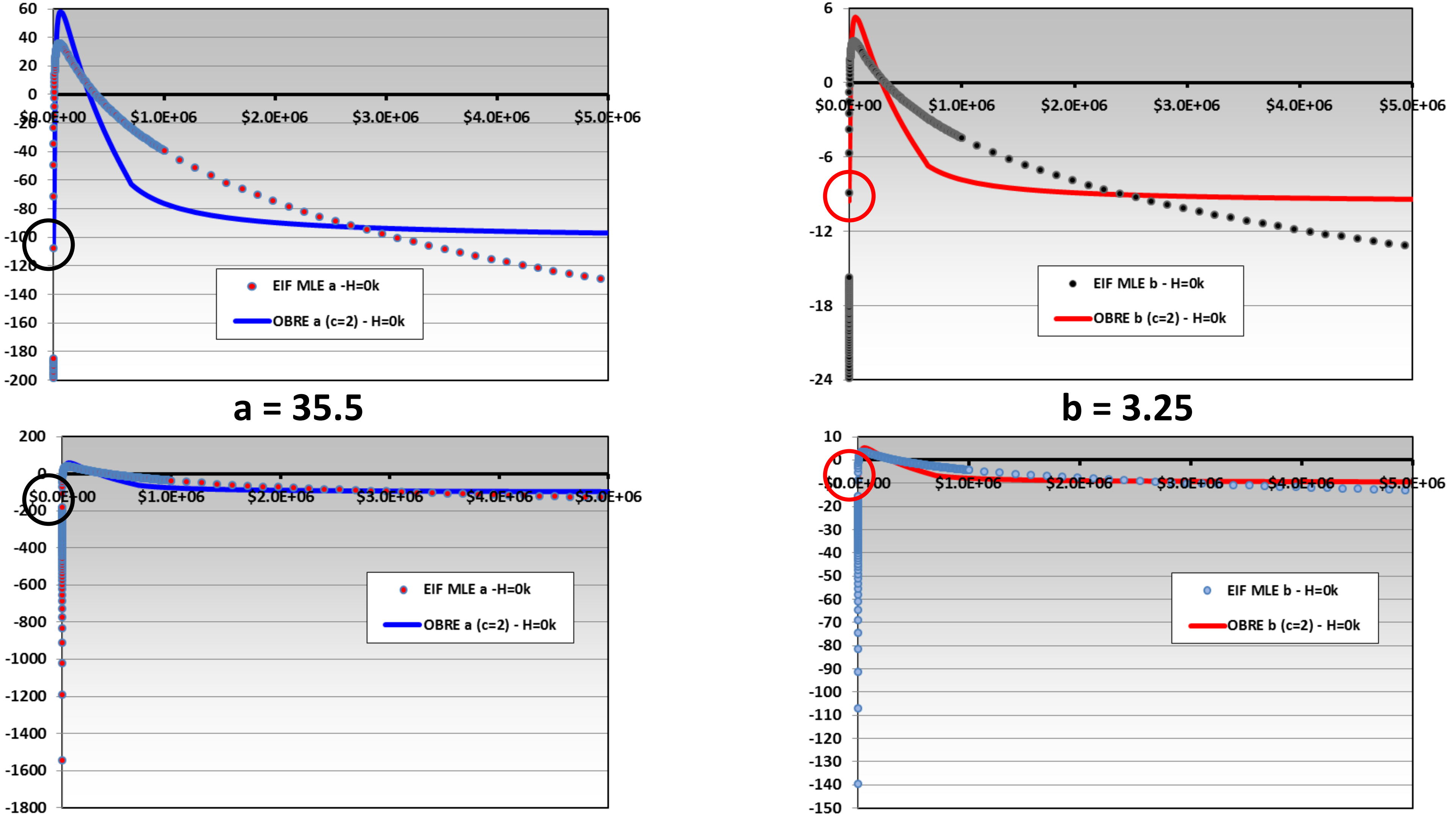


American Bankers Association

DataMineIt

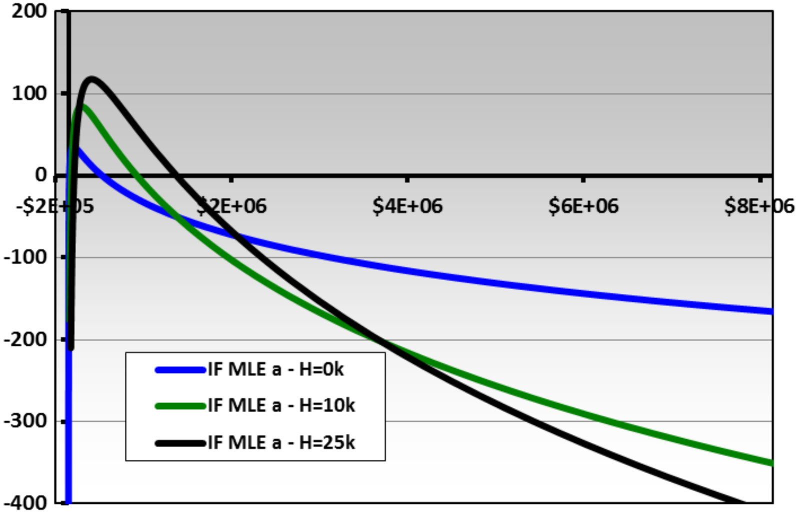
3a. Estimators for a "Better" Capital Distribution: More Stability over Time – More Robustness to Extreme Tail Events

LogGamma EIFs: MLE vs. OBRE

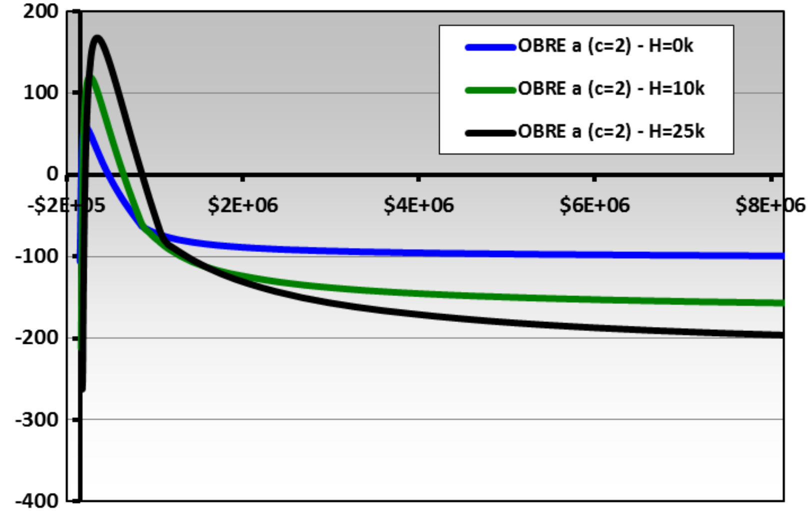


3a. Estimators for a “Better” Capital Distribution: More Stability over Time – More Robustness to Extreme Tail Events

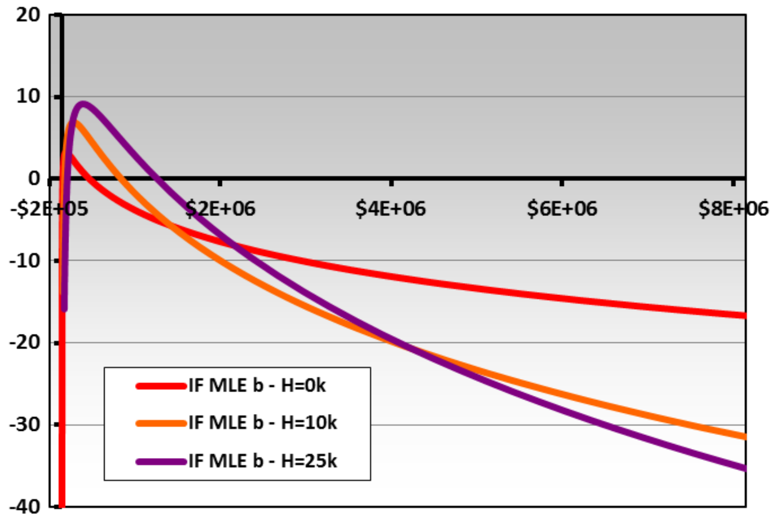
Truncated LogGamma: MLE EIF



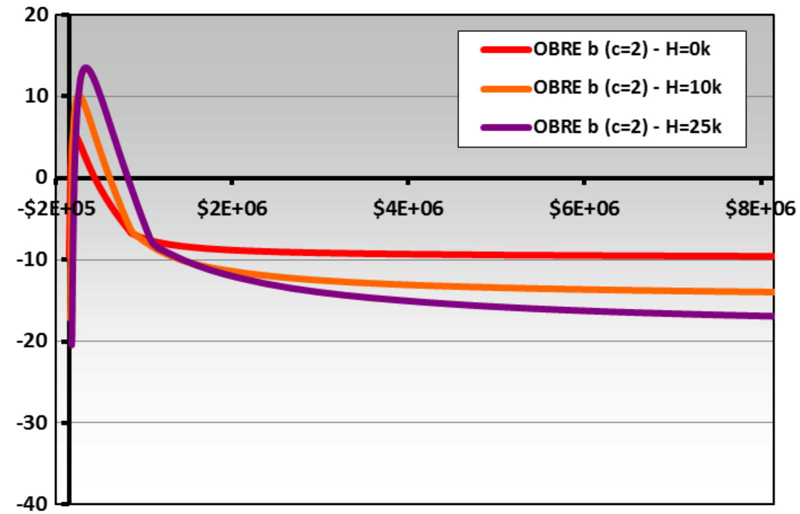
Truncated LogGamma: OBRE EIF



a = 35.5

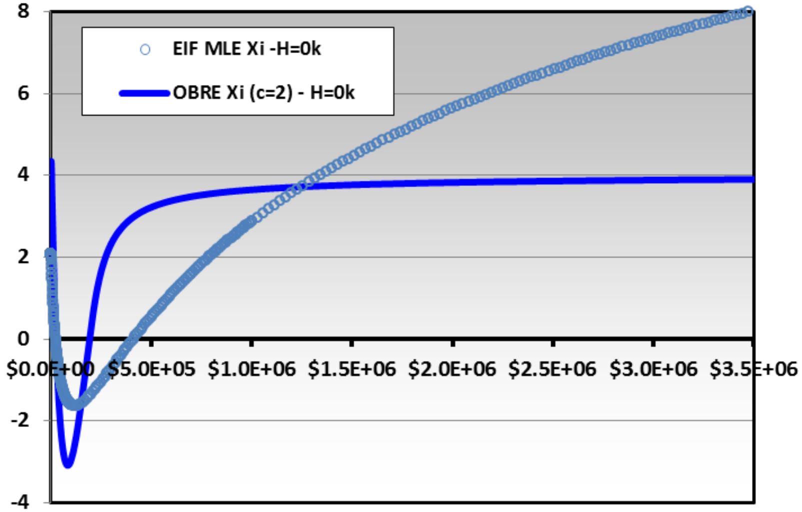


b = 3.25

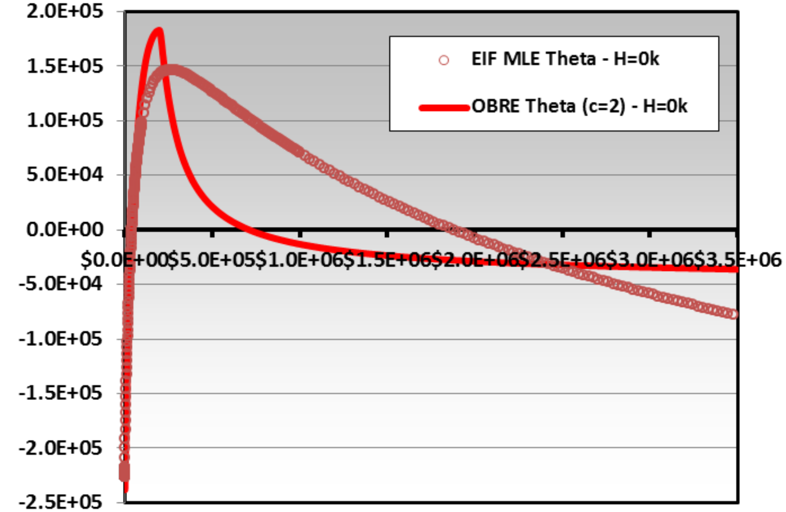


3a. Estimators for a "Better" Capital Distribution: More Stability over Time – More Robustness to Extreme Tail Events

GPD EIFs: MLE vs. OBRE



ε = 0.875



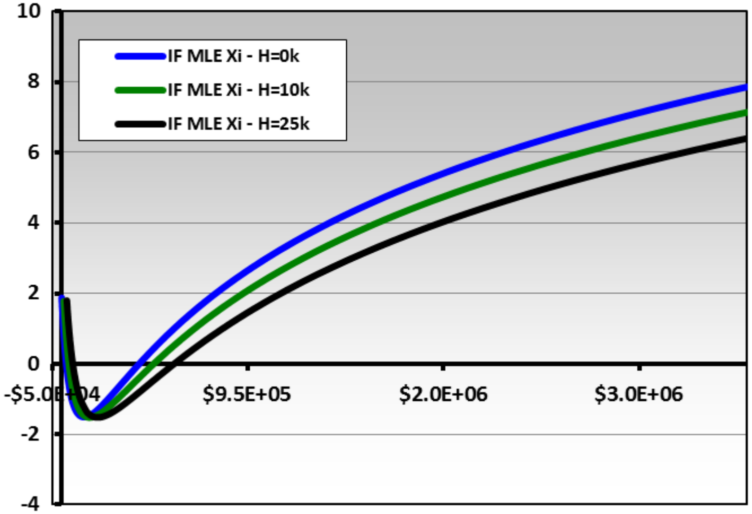
β = 57,500

American Bankers Association

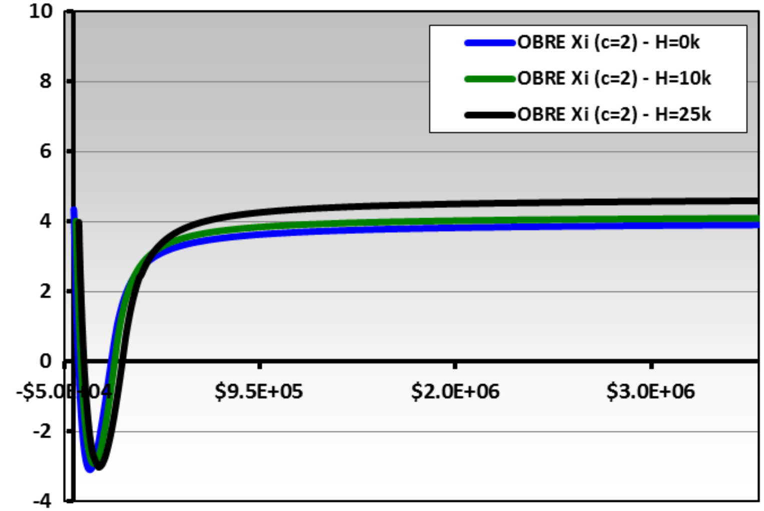
DataMineIt℠

3a. Estimators for a "Better" Capital Distribution: More Stability over Time – More Robustness to Extreme Tail Events

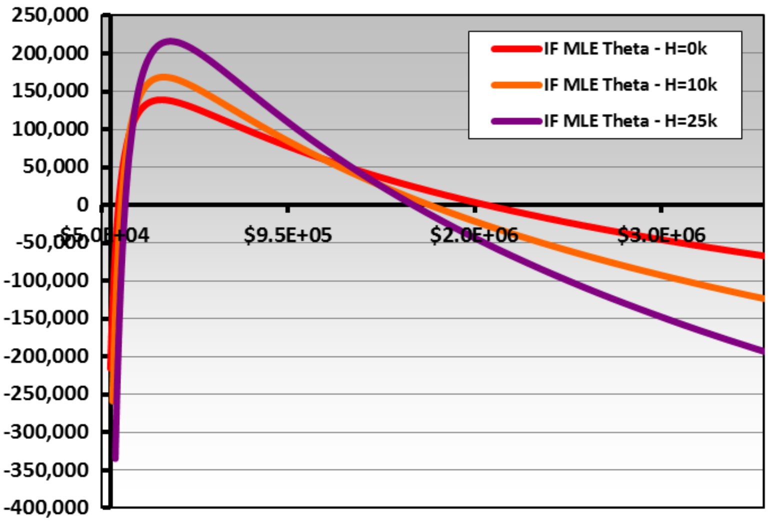
Truncated GPD: MLE EIF



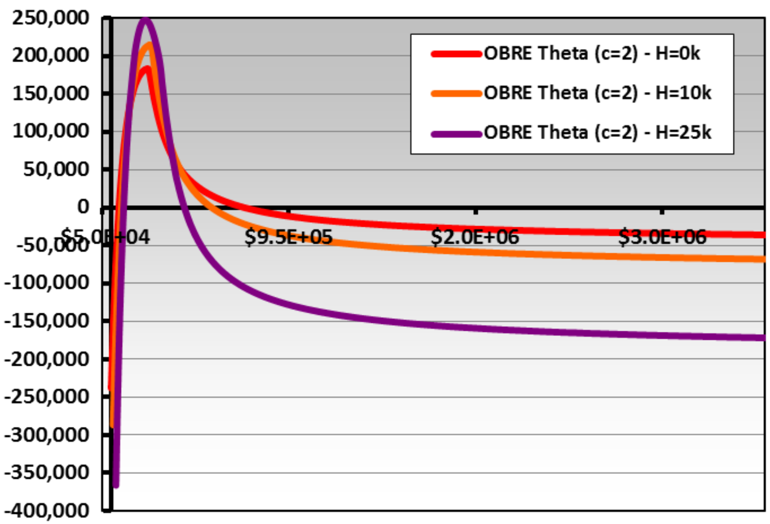
Truncated GPD: OBRE EIF



ξ = 0.875



β = 57,500



American Bankers Association

DataMineIt

3a. Estimators for a "Better" Capital Distribution: More Stability over Time – More Robustness to Extreme Tail Events

- **Note from the above derivations of their IFs, OBRE estimators have several potential advantages over their MLE counterparts. Not only are they B-robust, by definition, but they also avoid the truncation-induced/truncation-augmented covariance between parameters as *x* increases. The latter would appear to at least partially explain the extreme sensitivity of MLE estimators under truncation reported in the literature, which has perplexed some researchers.**

3b. Estimators for a "Better" Capital Distribution: Estimated with <u>Greater Efficiency</u>, i.e. More Precision, Less Variability

- **Estimators that are B-robust must give up some efficiency to obtain their robustness. However, this is only true under iid data. When data are NOT iid, as is the rule for OpRisk severity data, robust estimators can even be MORE efficient than MLE.**

- **The goal is to obtain an estimator that, under real world, non-iid conditions, is at least no less efficient than MLE, and hopefully <u>even more efficient in its capital estimates</u> (not just in the variability of its parameter values).**

- **The results of the simulation study shown below (see Appendix 2 for details), which compares the capital estimates of MLE vs. OBRE, show that we can have our cake and eat it too: OBRE can generate capital estimates that are less biased than those of MLE (which is discussed in the next section) while maintaining efficiency comparable to that of MLE. Given its superior robustness properties, a strong case can be made for its use in this setting over MLE because it is as good or better along all three major criteria – capital precision, capital accuracy, and capital robustness.**

- **The IF directly informs the issue of the robustness of an estimator, and even can be used to define the asymptotic variance of the estimators via** $Var = \int IF^2 dF$

3b. Estimators for a "Better" Capital Distribution: Estimated with Greater Efficiency, i.e. More Precision, Less Variability

TABLE 1			0% Deviation	6% Deviation	6% Deviation	6% Deviation
Distribution				**Both Tails (3% Each)**	**Left Tail**	**Right Tail**
LogN		**True SLA at 99.996%tile**	**$170,317,921**	**$173,118,560**	**$165,323,008**	**$180,654,136**
	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
	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%
	MLE	% within +/- 50%	80.0%	83.0%	80.0%	87.0%
	OBRE*	% within +/- 50%	80.0%	84.0%	82.0%	86.0%
	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
TLogN		**True SLA at 99.996%tile**	**$180,486,144**	**$183,180,240**	**$175,278,136**	**$190,682,320**
	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
	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%
	MLE	% within +/- 50%	71.0%	73.0%	72.0%	74.0%
	OBRE*	% within +/- 50%	72.0%	70.0%	71.0%	76.0%
	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^(11/8) ≈ 2.59**

3b. Estimators for a "Better" Capital Distribution: Estimated with Greater Efficiency, i.e. More Precision, Less Variability

TABLE 1			0% Deviation	6% Deviation	6% Deviation	6% Deviation
Distribution				**Both Tails (3% Each)**	**Left Tail**	**Right Tail**
LogG		**True SLA at 99.996%tile**	**$366,309,627**	**$370,407,112**	**$353,009,568**	**$387,304,656**
	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
	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%
	MLE	% within +/- 50%	63.0%	75.0%	70.0%	78.0%
	OBRE*	% within +/- 50%	59.0%	71.0%	72.0%	76.0%
	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
TLogG		**True SLA at 99.996%tile**	**$388,391,019**	**$392,310,056**	**$374,657,472**	**$409,562,640**
	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
	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%
	MLE	% within +/- 50%	63.0%	67.0%	66.0%	76.0%
	OBRE*	% within +/- 50%	56.0%	60.0%	66.0%	67.0%
	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

American Bankers Association

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

DataMineIt

3c. Estimators for a "Better" Capital Distribution: Estimated with <u>Less Bias</u> vis-à-vis the effects of Jensen's Inequality

- As M-Class estimators, both MLE and OBRE are asymptotically unbiased under iid data. However, for all right-skewed severity distributions, unbiased parameter estimates do NOT, UNDER THE LDA framework, yield unbiased capital estimates. In fact, if left unadjusted, they yield BIASED CAPITAL ESTIMATES.

- This is due to a 1906 analytic result by Jensen, known as "Jensen's inequality," which has been missed in the OpRisk literature to date (see Opdyke & Cavallo, 2012).

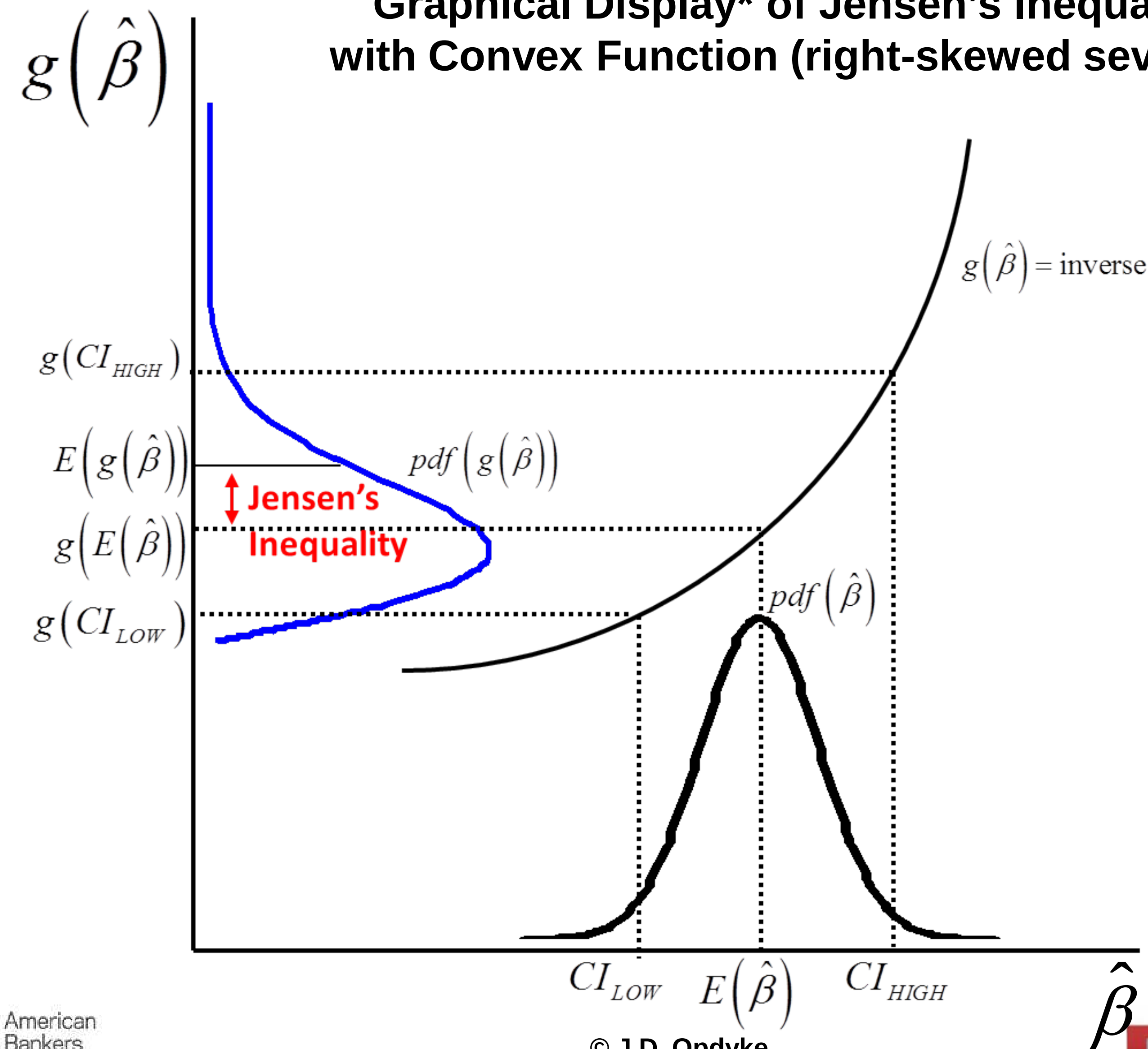
- Because the inverse cdf is convex (and not concave), the effect of this bias is always upwards, that is, estimating larger capital requirements than necessary, and can be very large. Its magnitude depends on three factors, all else equal:
    1. thickness of the tail of the severity distribution (heavier tail $\Rightarrow$ more bias)
    2. size of the quantile (higher quantile $\Rightarrow$ more bias)
    3. variance and skewness of the estimator (either larger $\Rightarrow$ more bias)

American Bankers Association

DataMineIt

3c. Estimators for a "Better" Capital Distribution: Estimated with <u>Less Bias</u> vis-à-vis the effects of Jensen's Inequality

Graphical Display* of Jensen's Inequality with Convex Function (right-skewed severity)



*From Kennedy (1992), p.37.

3c. Estimators for a "Better" Capital Distribution: Estimated with <u>Less Bias</u> vis-à-vis the effects of Jensen's Inequality

- **OBRE mitigates this bias to some degree, as seen in the mean capital estimates from Table 1 above, because its distribution generally is less skewed than that of MLE (even when its variance is comparable) <u>due to its robustness</u> (which we see in its IF).**

- **Completely eliminating this bias, while simultaneously maintaining efficiency and robustness, is the topic of continuing research.**

- **The main point here is to show how knowledge of the IFs of different estimators can help in the design and selection of estimators for "better" capital estimation (i.e. capital estimates that are less biased, more precise, and more robust to outlying events).**

Presenting "The Saga of the MLE Capital Scenarios", a (Divine) Comedy of (Statistical) Errors…?

(with apologies to Dante and Shakespeare)

- **Starring "the Absurd," and "the Improved but Still Crazy,"**
- **Featuring "That's Just Wrong,"**
- **with a Cameo Appearance from "Much More Reasonable"**

The "Absurd": Act 1, Scene 1

The "Absurd" enters as estimated capital exhibits counter-intuitive asymptotic behavior, increasing by orders of magnitude exactly as a new loss DECREASES by orders of magnitude.

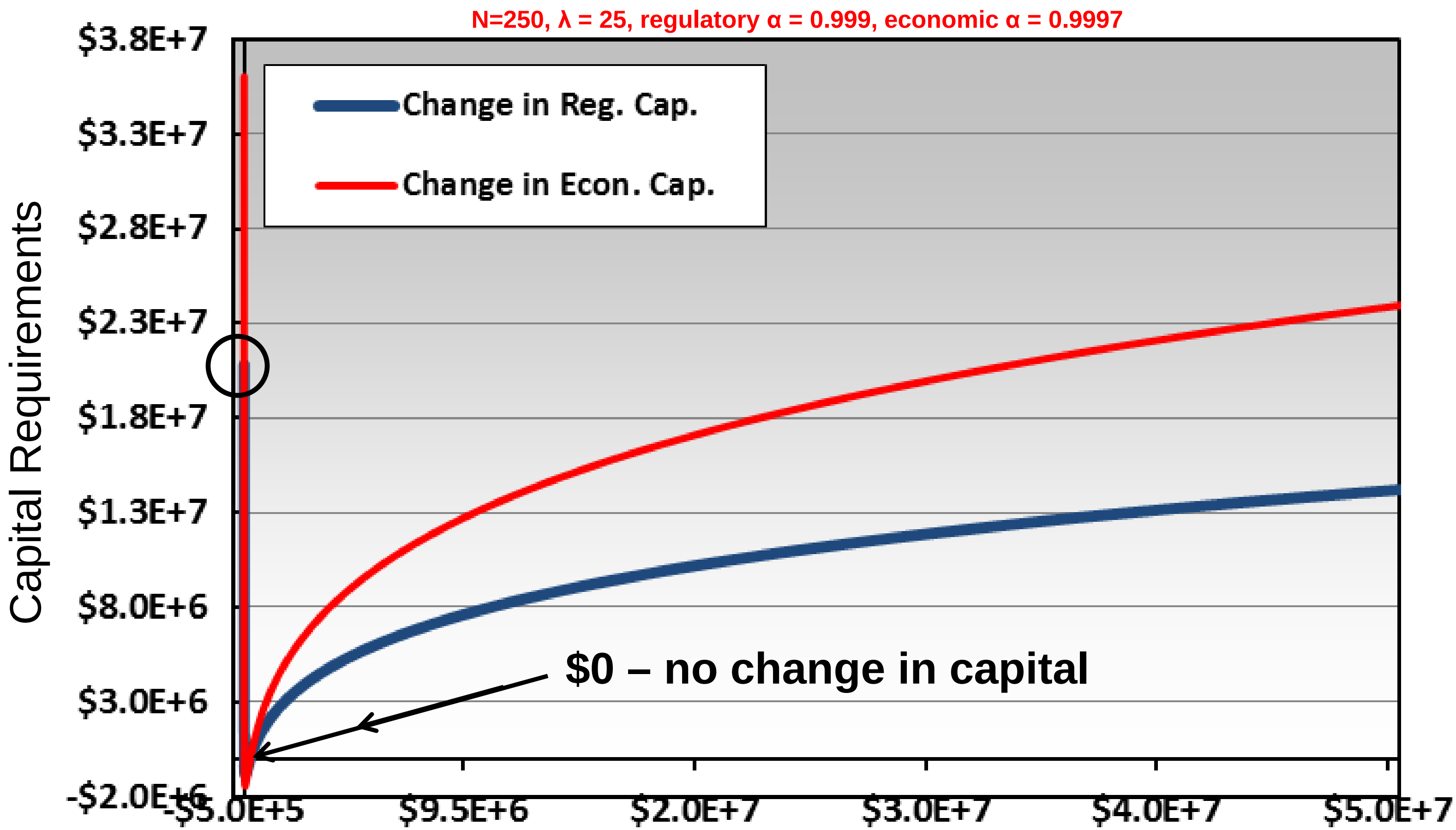
In other words, small left-tail losses – not "low frequency, high severity" losses – are possibly the greatest source of quarter-to-quarter instability and variability in MLE-based capital requirements.

How can this be??...

4. Using IF's Exact Capital Sensitivity Curves for Better Planning

The "Absurd":

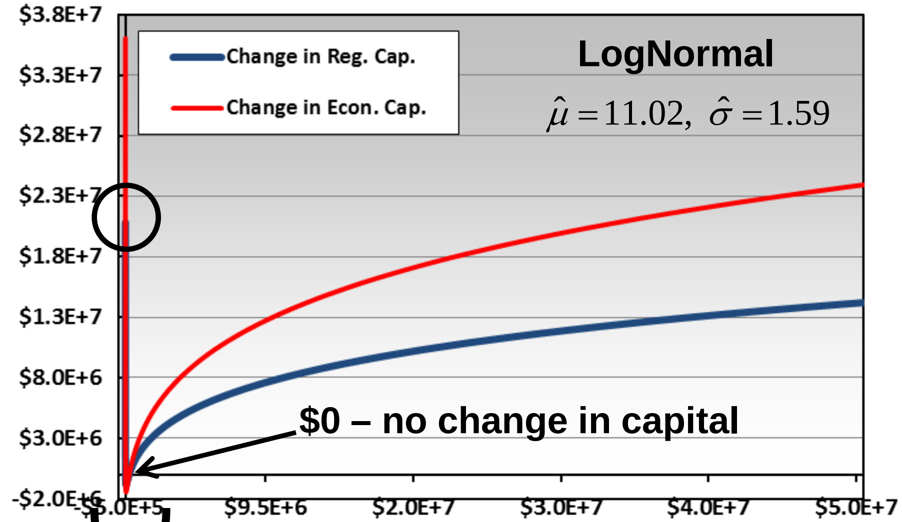
Based on a Random Draw from LogNormal (μ=10.95, σ=1.75) where MLE $\hat{\mu} = 11.02,\ \hat{\sigma} = 1.59$



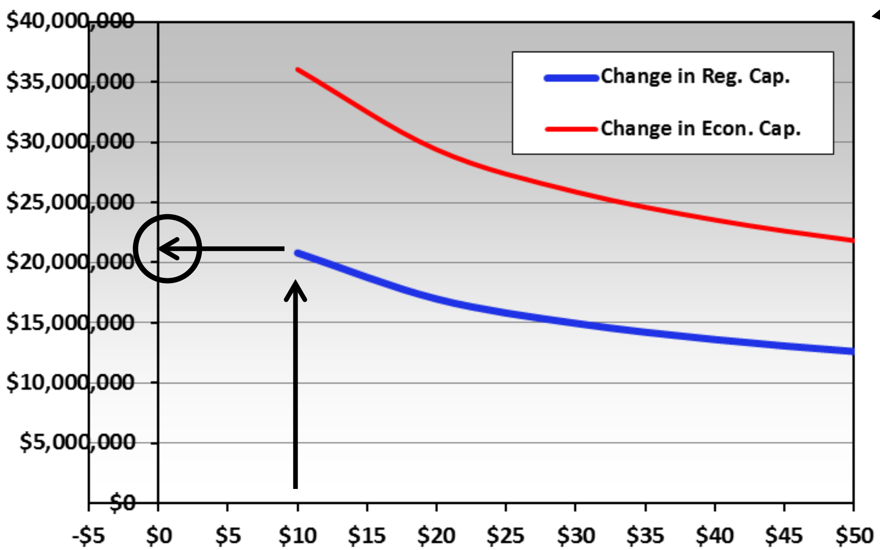
4. Using IF's Exact Capital Sensitivity Curves for Better Planning

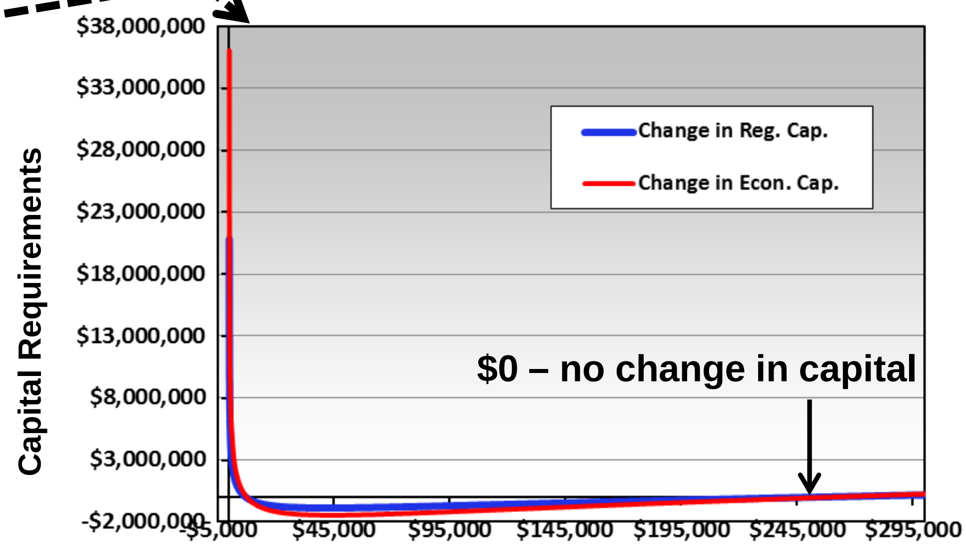
The "Absurd":

For MLE, a new loss of $10 increases regulatory capital by over $20m, and economic capital by over $36m. But a loss of about $250k increases capital by $0.



N=250, λ = 25, regulatory α = 0.999, economic α = 0.9997



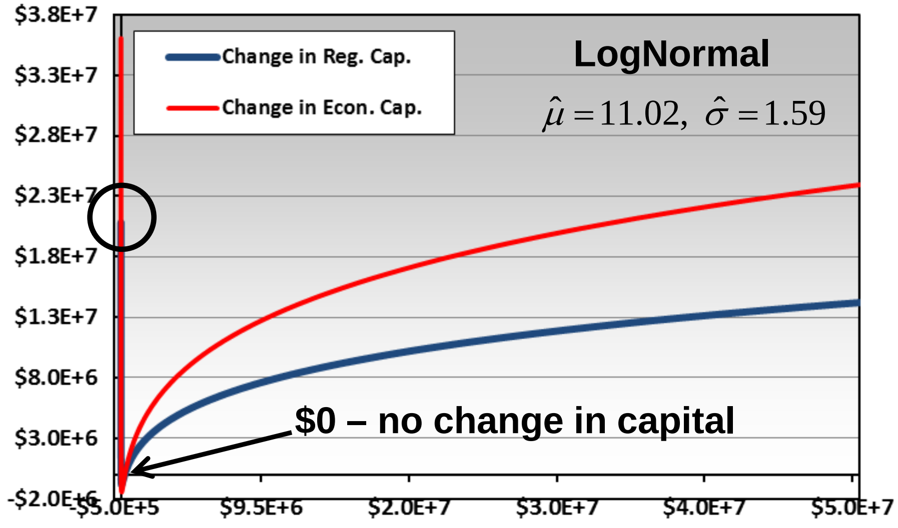


4. Using IF's Exact Capital Sensitivity Curves for Better Planning

The "Absurd":

WHY? **Check the MLE IF, which we derived previously as:**

$$IF_{\theta}\left(x;\theta,T\right)=\left[\begin{array}{c}\ln\left(x\right)-\mu \\ \dfrac{\left(\ln\left(x\right)-\mu\right)^{2}-\sigma^{2}}{2\sigma}\end{array}\right]$$

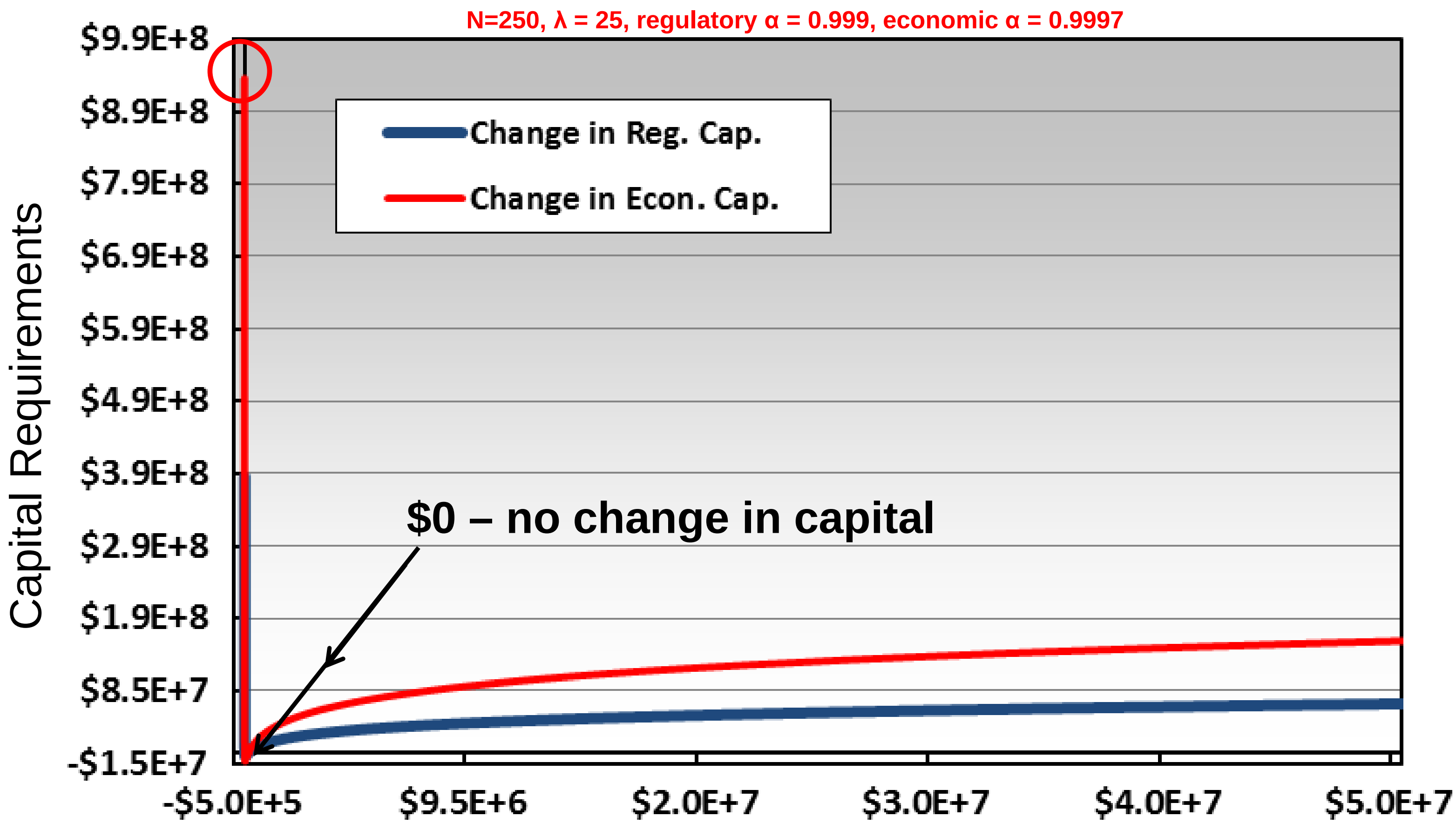


N=250, λ = 25, regulatory α = 0.999, economic α = 0.9997

The IF for the σ term becomes HUGE when *x* → 0^+, so required capital also is going to become HUGE as it is based directly on the HUGE parameter estimate for σ. Even though the IF indicates that the parameter estimate for *μ* decreases monotonically as *x* decreases, it does so at a much slower rate so the effect of σ will dominate the effect that *x* has on capital.

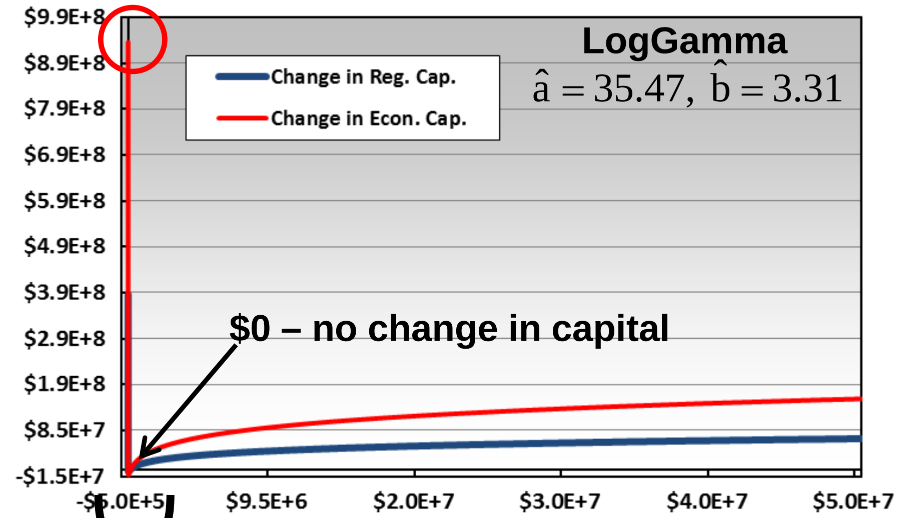
The "Absurd":

Based on a Random Draw from LogGamma (a=35.5, b=3.25) where MLE $\hat{a} = 35.47,\ \hat{b} = 3.31$

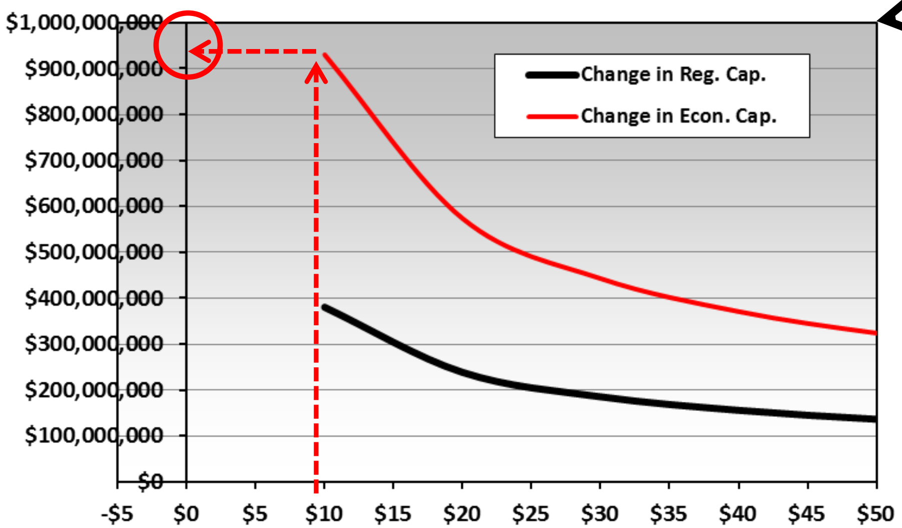


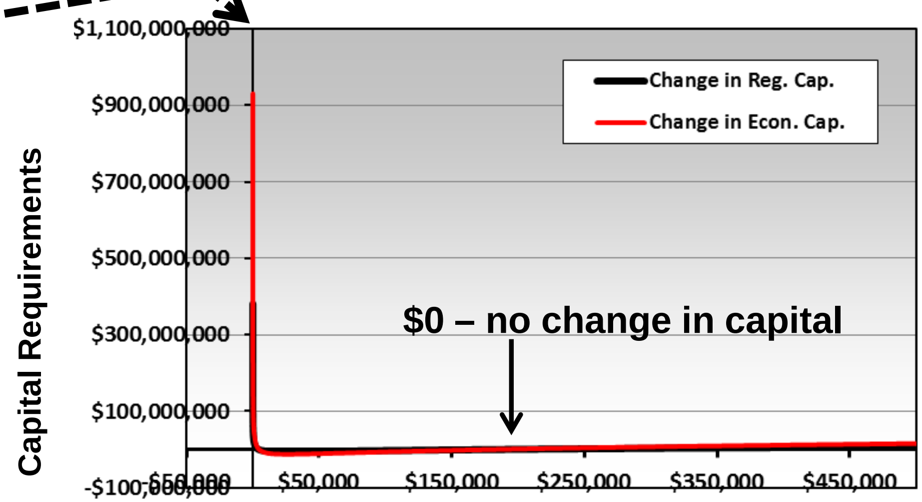
The “Absurd”:

For MLE, a new loss of $10 increases regulatory capital by over $380m, and economic capital by over $930m. But a loss of about $175k increases capital by $0.



N=250, λ = 25, regulatory α = 0.999, economic α = 0.9997





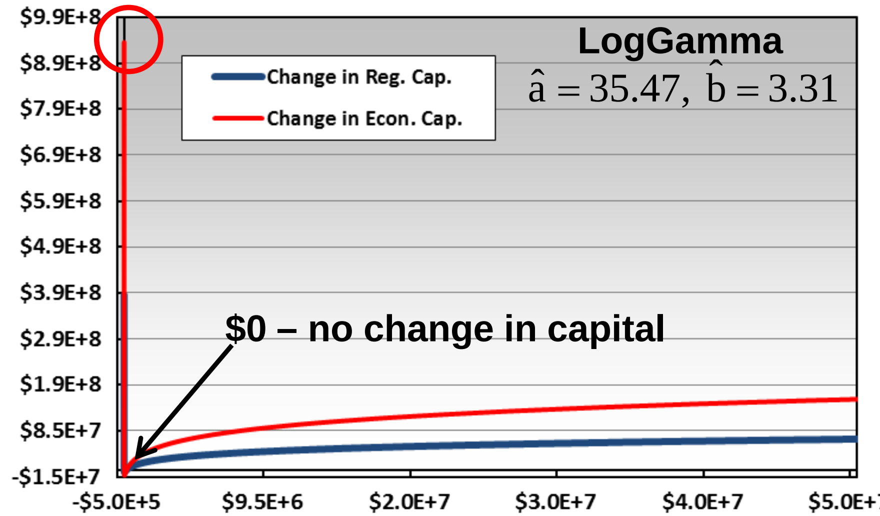
4. Using IF's Exact Capital Sensitivity Curves for Better Planning

The "Absurd":

WHY? Check the MLE IF, which we derived previously as:

$$IF_\theta\left(x;\theta,T\right)=$$

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



N=250, λ = 25, regulatory α = 0.999, economic α = 0.9997

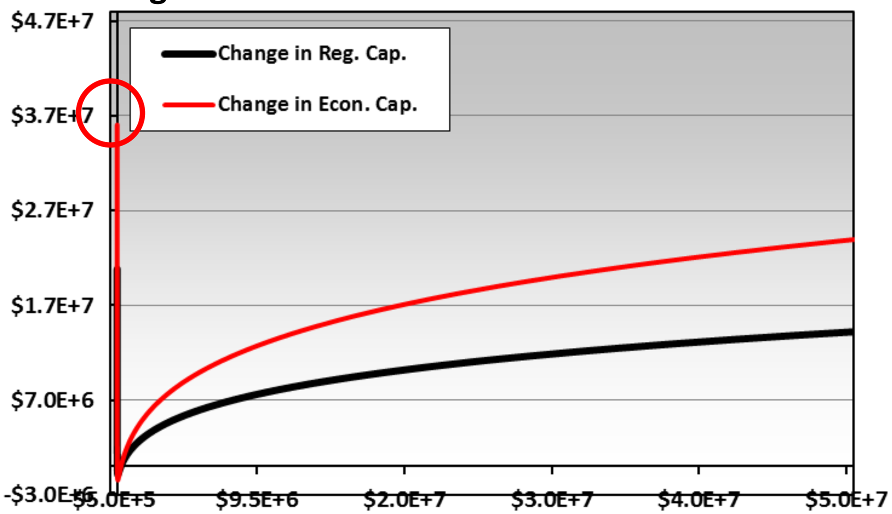
Here, -ln(x) in BOTH IF terms dominate the ln(ln(x)) terms, so ln(ln(x)) – ln(x), which inflects at x=exp(1), becomes a large negative number as $x \rightarrow 1^{+}$. However, for the LogGamma smaller b uniformly INCREASES the quantiles of the distribution, while smaller a DECREASES them. The b term dominates, however, because of the relative size of the constants in both numerators, so capital increases without bound as $x \rightarrow 1^{+}$.

The "Improved, but still Crazy": Act 1, Scene 2

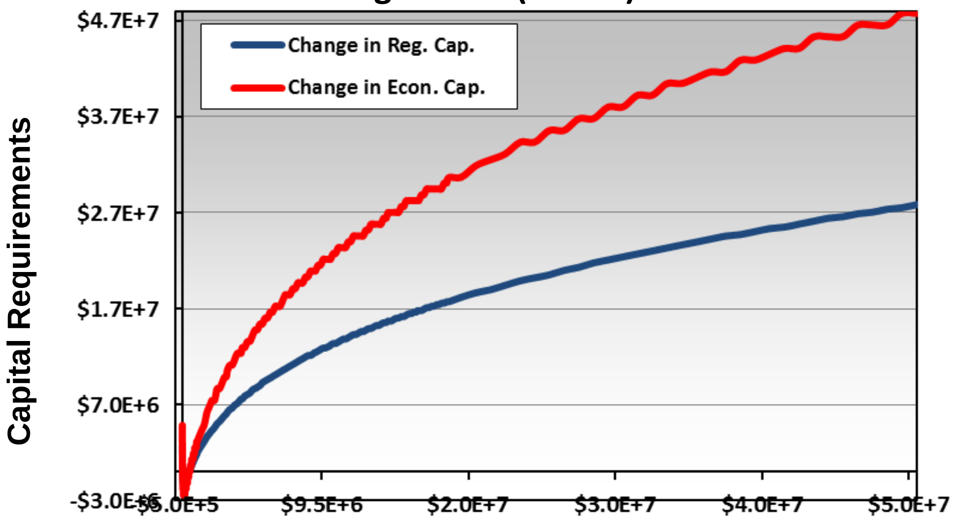
Truncation partially mitigates the "Absurd" asymptotic behavior of estimated capital, but note that even a relatively low threshold (e.g. $10k) makes a MUCH more heavy-tailed severity distribution, with much higher capital requirements, all else equal.

4. Using IF's Exact Capital Sensitivity Curves for Better Planning

LogNormal: MLE

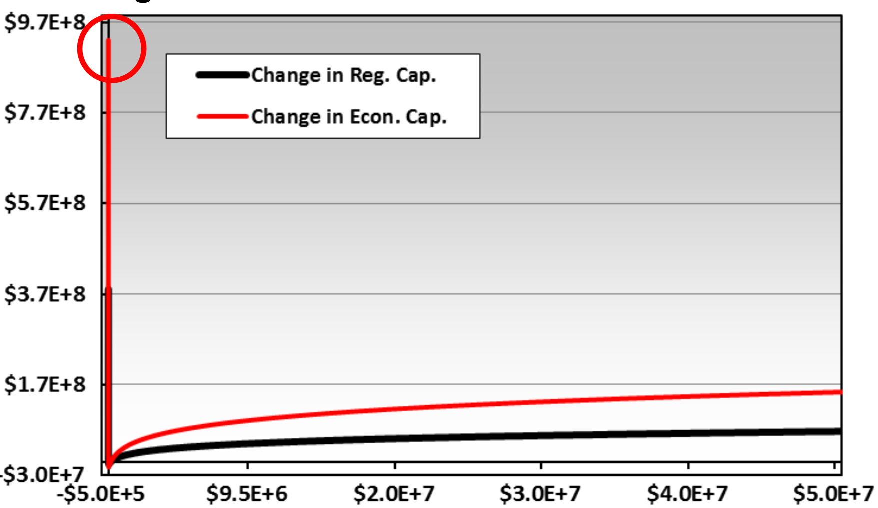


Truncated LogNormal (H=10k): MLE

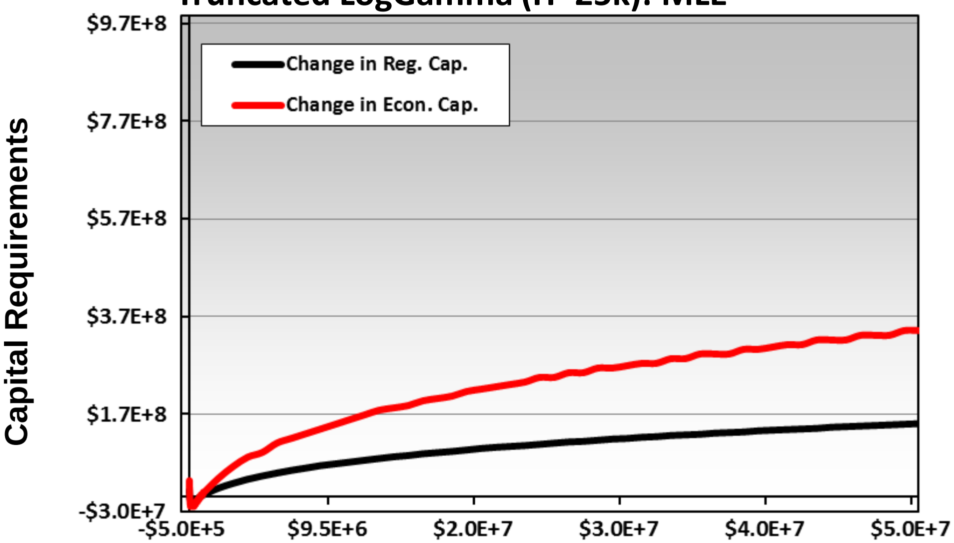


N=250, λ = 25, regulatory α = 0.999, economic α = 0.9997

LogGamma: MLE



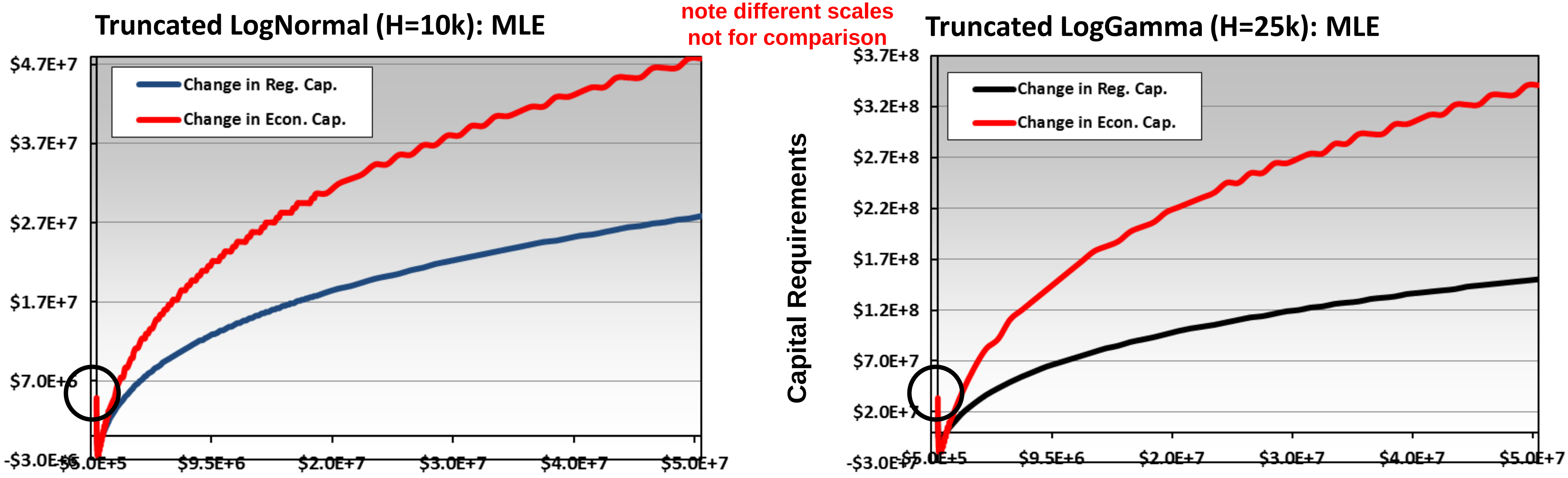
Truncated LogGamma (H=25k): MLE



American Bankers Association

DataMineIt

4. Using IF’s Exact Capital Sensitivity Curves for Better Planning



N=250, λ = 25, regulatory α = 0.999, economic α = 0.9997

For MLE, a new loss of $10,010 increases regulatory capital by over $2.7m, and economic capital by over $4.8m.

For MLE, a new loss of $25,010 increases regulatory capital by over $14.5m, and economic capital by over $33.5m.

American Bankers Association

DataMineIt

4. Using IF's Exact Capital Sensitivity Curves for Better Planning

- **These extreme capital responses to small, left-tail losses are not just mathematical curiosities: they are possibly the largest source of quarter-to-quarter instability of MLE-based capital requirements, because they are not as rare as "low frequency, high severity" losses. The effects are still extreme even for losses within $4k of the lower threshold, losses that every bank has in its severity modeling loss event datasets.**

					Change in Capital ($mill)					
Severity	Threshold	Parameter			H + $10 loss		H + $2k loss		H + $4k loss	
Dist.	H	Names	Parm 1	Parm 2	RC	EC	RC	EC	RC	EC
LogN	$0	μ, σ	10.953	1.749	$19.0	$33.3	$1.3	$2.4	$0.4	$0.8
LogN	$10,000	μ, σ	10.954	1.750	$2.6	$4.2	$2.0	$3.6	$1.5	$2.4
LogN	$25,000	μ, σ	10.917	1.749	$2.6	$4.8	$2.3	$4.2	$2.0	$3.6
LogG	$0	α, β	35.484	3.252	$590.9	$1,469.8	$14.1	$34.1	$3.6	$9.2
LogG	$10,000	α, β	35.513	3.263	$24.1	$62.2	$18.0	$43.1	$13.2	$33.5
LogG	$25,000	α, β	35.410	3.252	$26.4	$67.0	$22.8	$57.4	$19.2	$57.4
GPD	$0	ξ, β	0.8713	57,584	$27.9	$92.2	$24.0	$79.5	$20.4	$67.8
GPD	$10,000	ξ, β	0.8825	57,484	$31.2	$95.6	$26.4	$95.5	$24.0	$76.4
GPD	$25,000	ξ, β	0.8798	57,340	$38.4	$133.8	$36.0	$133.7	$31.2	$95.5

- **All it takes is a couple of new losses near the threshold, or changes in the values of such existing losses, to induce dramatic variability and instability in MLE-based capital requirements from quarter to quarter.**

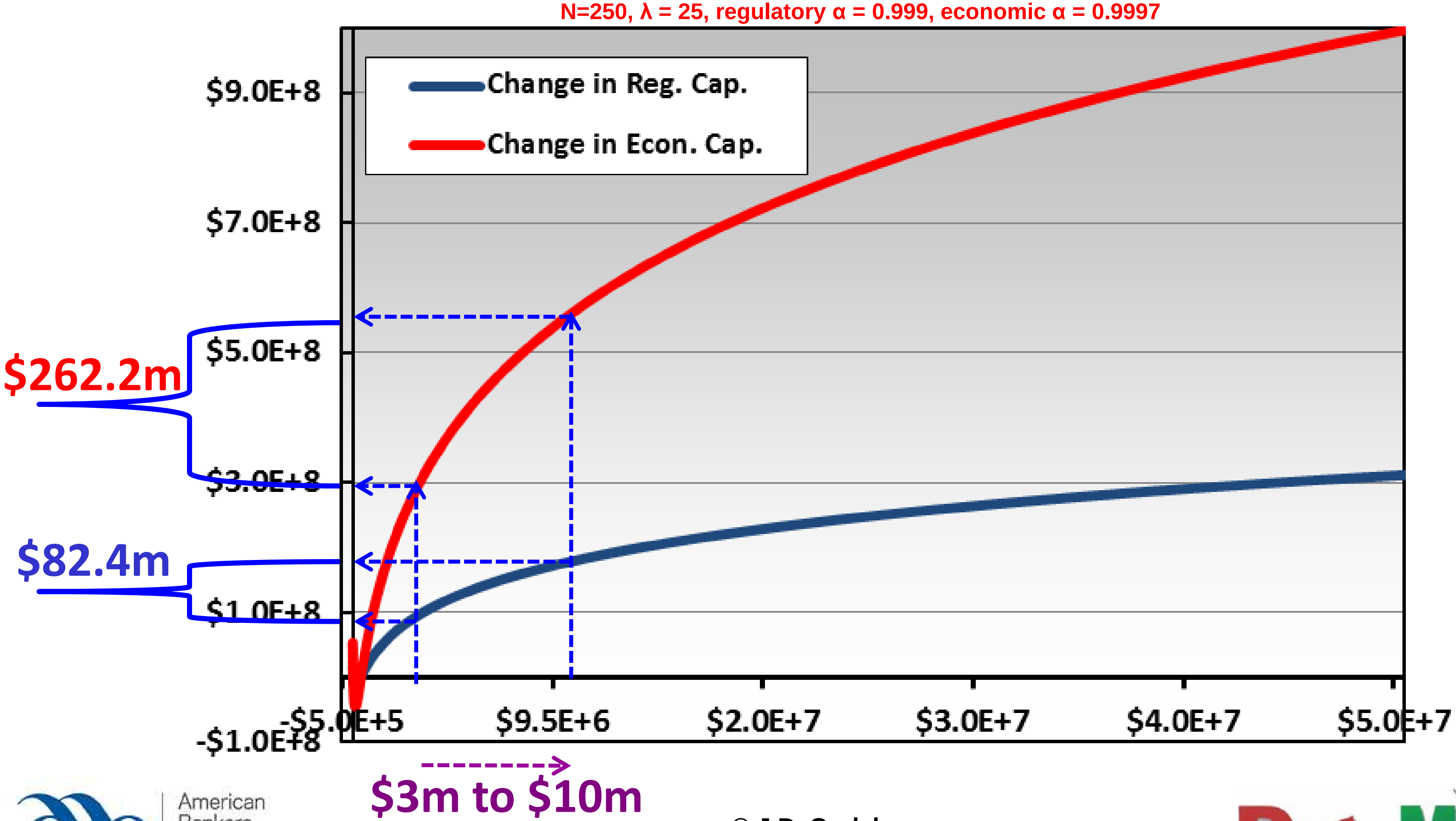
The "That's Just Wrong": Act 2, Scene 1

Under very heavy-tailed severity distributions (e.g. GPD, even withOUT infinite mean), MLE is simply too sensitive to changes in loss values to pass the "cest" – the capital estimate smell test.

4. Using IF's Exact Capital Sensitivity Curves for Better Planning

The "That's Just Wrong":

Based on a Random Draw from GPD (Ɛ = 0.875, β = 57,500) where MLE $\hat{\xi} = 0.833,\ \hat{\beta} = 60{,}895$



4. Using IF's Exact Capital Sensitivity Curves for Better Planning

- The above shows that under a GPD severity distribution ($\mathcal{E}$ = 0.833, β = 60,895), if an anticipated loss of $3m is actually realized as a $10m loss, the regulatory capital based on MLE estimators increases by over $82m, and the economic capital increases by over $262m.

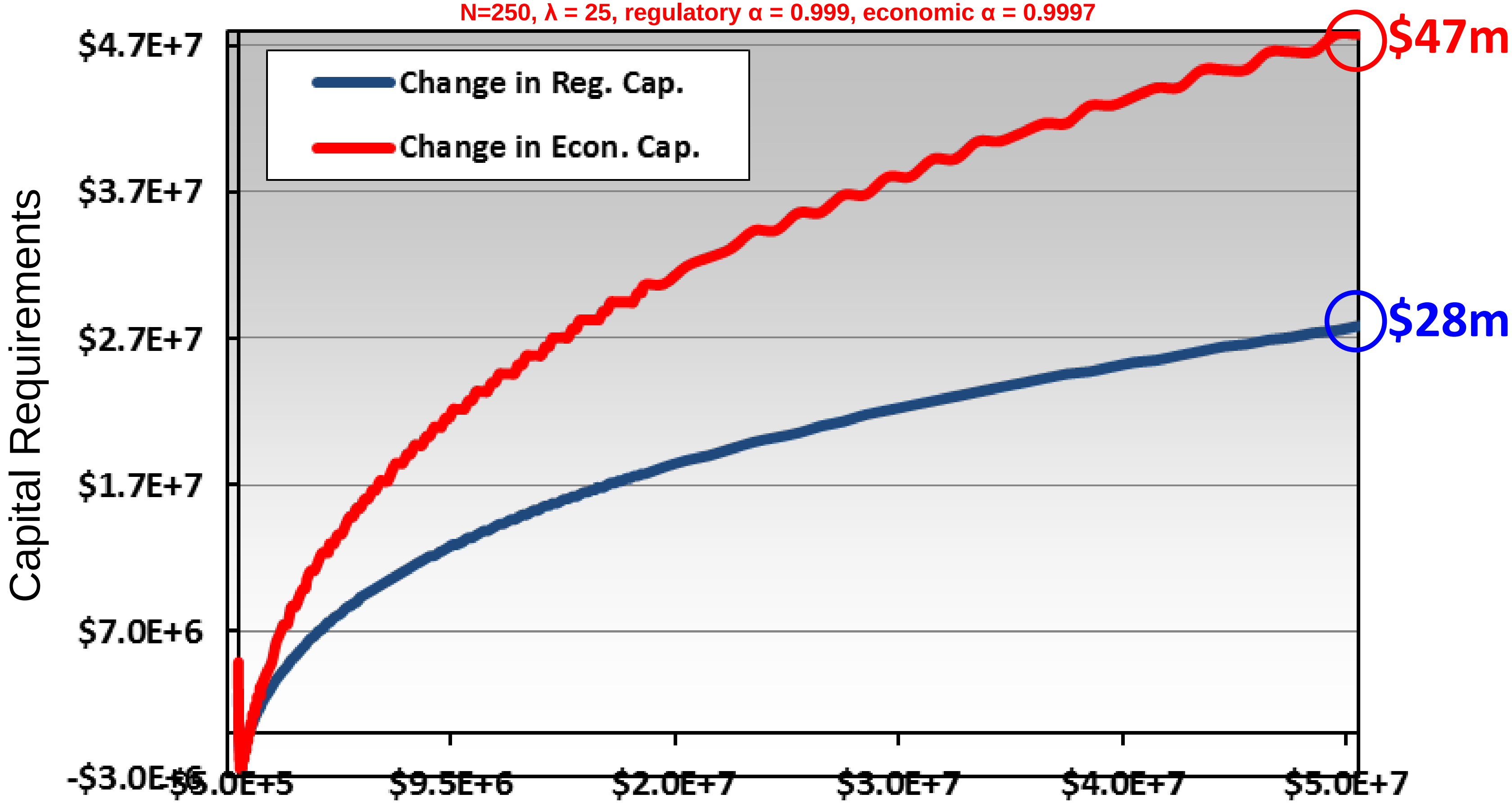
- It is not hyperbole to say that when a $7m increase in a single loss increases economic capital by hundreds of millions of dollars in an otherwise correctly specified MLE/LDA model, the LDA framework, and/or the use of MLE as a tool to implement LDA, are failing, by any measure, to provide reasonable, stable, data-based capital estimates.

The “Much More Reasonable”: Act 2, Scene 2

While the IF can utilize many realistic examples that easily expose misleading inadequacies of LDA/MLE, there are many scenarios that WOULD pass most capital sensitivity smell tests and that many would deem much more reasonable.

The "Much More Reasonable" :

Based on a Random Draw from Truncated LogNormal (μ=10.95, σ=1.75, H=10k) where MLE $\hat{\mu}=11.16,\ \hat{\sigma}=1.68$



4. Using IF's Exact Capital Sensitivity Curves for Better Planning

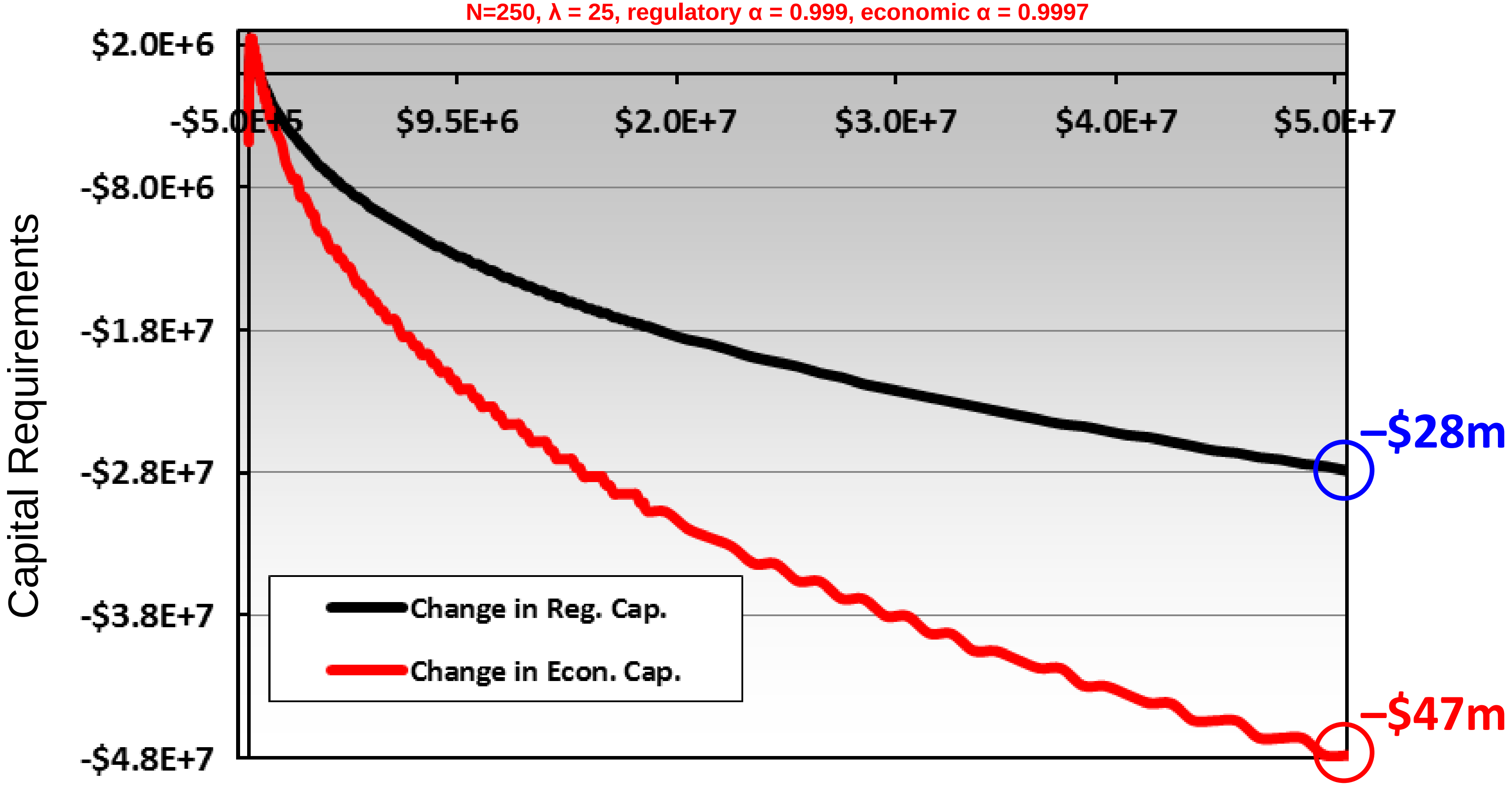
- **Under a Truncated LogNormal severity distribution (*μ*=11.16*, σ*=1.68*,* H=10k), a new $50m loss increases regulatory capital, based on MLE estimators, by $28m, and economic capital by $47m. This capital effect would not fall into most practitioners' "Absurd," "Still Crazy," or "That's Just Wrong" buckets.**

- **NOTE: If we completely drop a loss, say, due to a litigation that unexpectedly settled very favorably for the bank, we can modify the EIF to answer a slightly different question: how much does capital change if this loss was never included? The answer is just the negative EIF.**

$$-EIF\left(x\,|\,T,F\right)=\lim_{\varepsilon\to 0}\left[\frac{-\left[T\left(\hat{F}_{\varepsilon_{(n)}}\right)-T\left(\hat{F}_{(n-1)}\right)\right]}{\varepsilon}\right]$$

The "Much More Reasonable":

Based on a Random Draw from Truncated LogNormal (μ=10.95, σ=1.75, H=10k) where MLE $\hat{\mu}=11.16,\ \hat{\sigma}=1.68$



4. Using IF's Exact Capital Sensitivity Curves for Better Planning

- **NOTE: While all the above examples are prospective, focusing on current or possible future events, the IF can be used retrospectively as well for exact attribution analysis of capital changes due to specific losses in previous quarters. "But for" analyses can be constructed based on the exact affect on capital associated with each additional single loss event that occurred in a given quarter. This is an effective way to identify "the culprits:" specific losses that have caused grossly disproportionate changes in capital.**

- **NOTE: Preliminary results of OBRE-based capital estimates show fairly successful mitigation of MLE's extreme asymptotic behavior under new, small losses in the left tail, but TOO much robustness in the other direction, with estimates of capital requirements flattening off under very large right-tail losses. Effective utilization of OBRE's robustness tuning parameter may provide a solution, and this is currently being researched. But the point for this presentation is that the IF is the objective metric by which i) definitive assessments can be made not only of a single estimator across the entire domain of possible loss events, but also ii) comparative assessments can be made ACROSS estimators.**

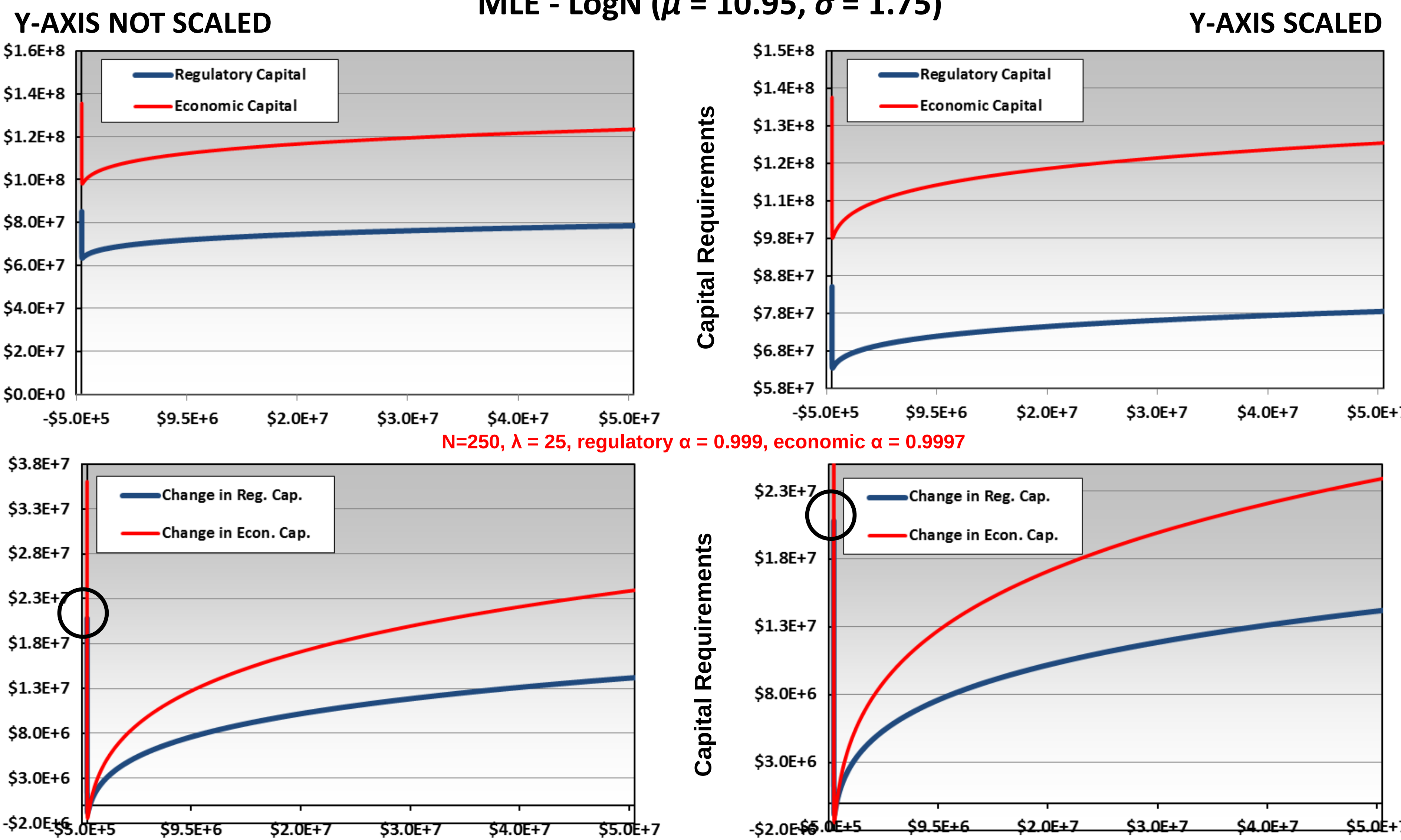
4. Using IF's Exact Capital Sensitivity Curves for Better Planning

- **<u>Bottom Line</u>: The capital estimate is essentially a high quantile estimate of the severity distribution. When using a fully parametric model to estimate high quantiles, the slightest deviation from parametric assumptions can change the quantile estimates in very dramatic and sometimes unanticipated ways. This is especially true when using non-robust estimators like MLE.**

- **<u>Moral of "The Saga of the Capital Scenarios"</u>: Given this bottom line, how could one NOT use the IF in capital planning?! Both to inform the choice of estimator given the characteristics of the data at hand, AND to gauge the EXACT impact of specific loss events that may be, or are, imminent?**

4. Using IF’s Exact Capital Sensitivity Curves for Better Planning

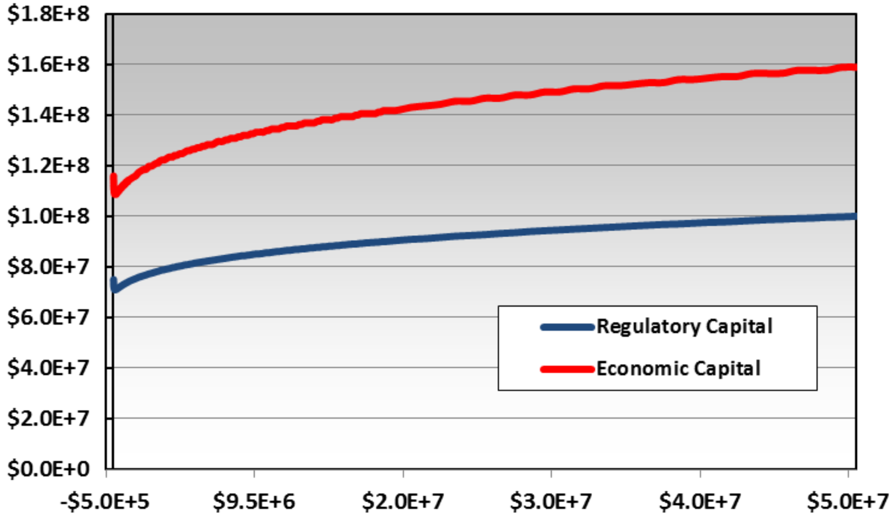
MLE - LogN (μ = 10.95, σ = 1.75)



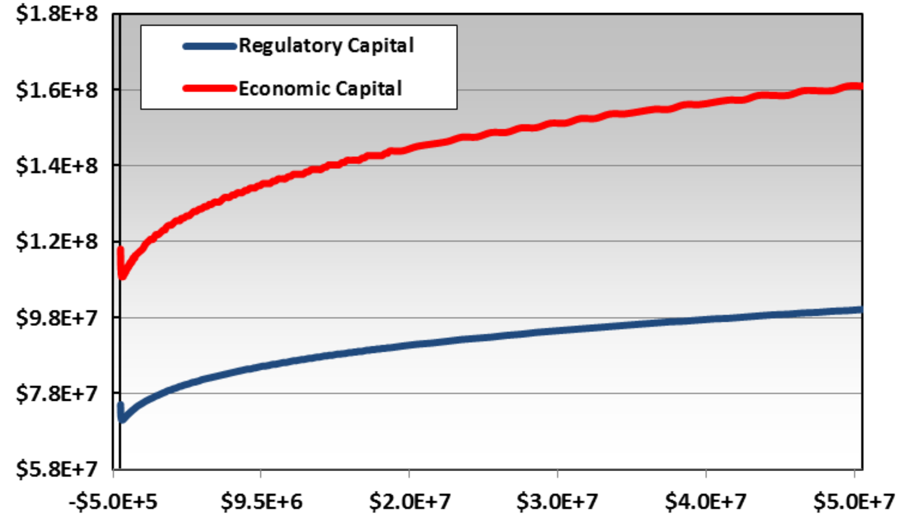
4. Using IF's Exact Capital Sensitivity Curves for Better Planning

MLE – Truncated LogN (μ = 10.95, σ = 1.75, H=10k)

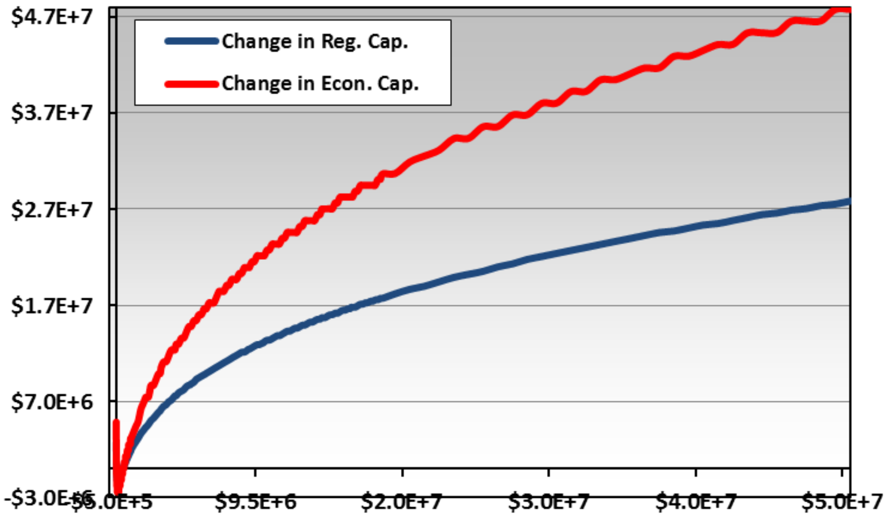
Y-AXIS NOT SCALED



Y-AXIS SCALED



N=250, λ = 25, regulatory α = 0.999, economic α = 0.9997



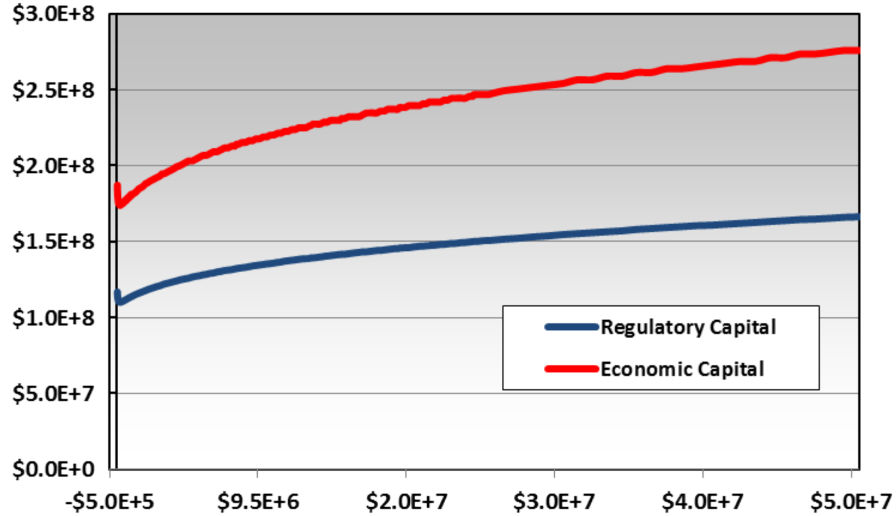
Capital Requirements

American Bankers Association

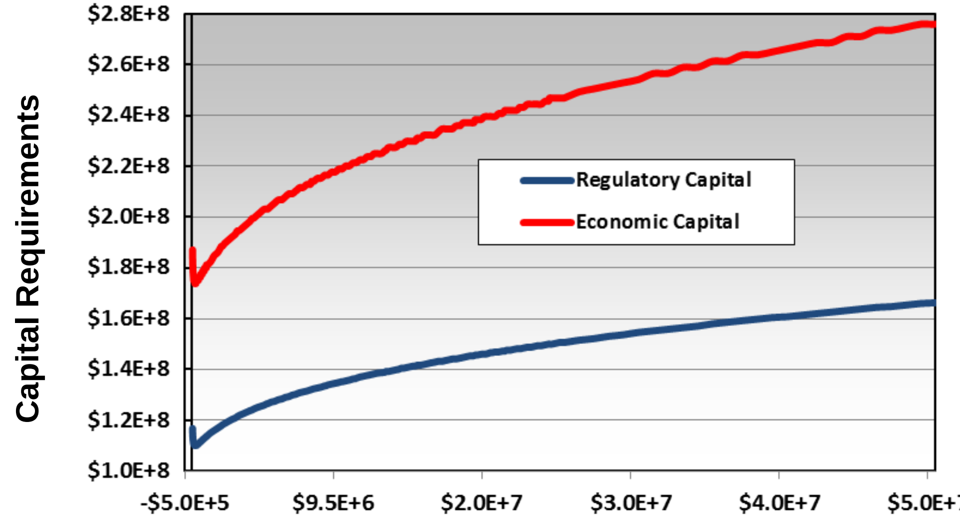
4. Using IF’s Exact Capital Sensitivity Curves for Better Planning

MLE – Truncated LogN (μ = 10.95, σ = 1.75, H=25k)

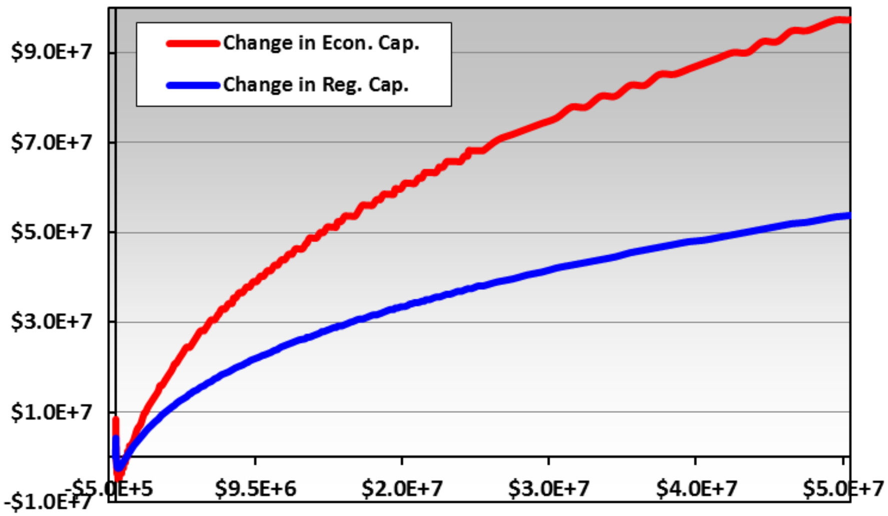
Y-AXIS NOT SCALED



Y-AXIS SCALED



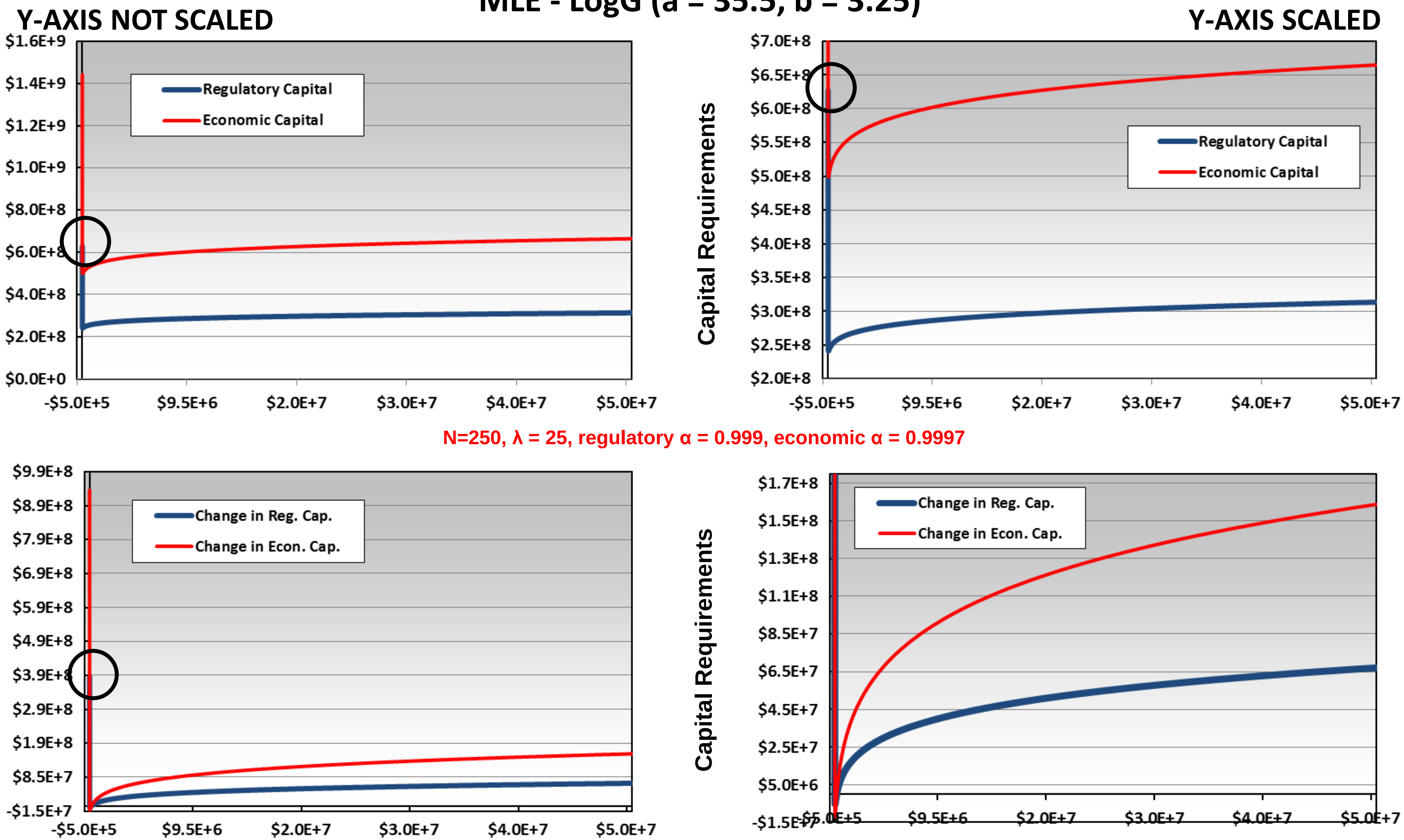
N=250, λ = 25, regulatory α = 0.999, economic α = 0.9997



Capital Requirements

4. Using IF's Exact Capital Sensitivity Curves for Better Planning

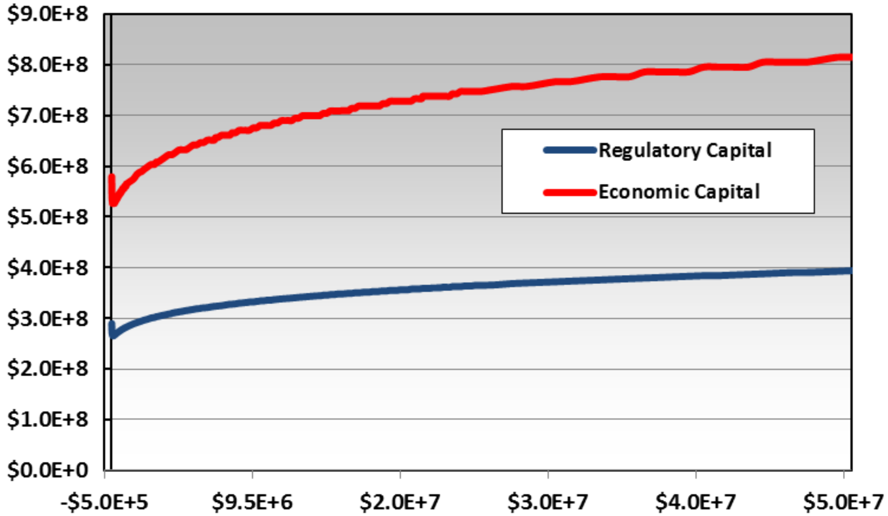
MLE - LogG (a = 35.5, b = 3.25)



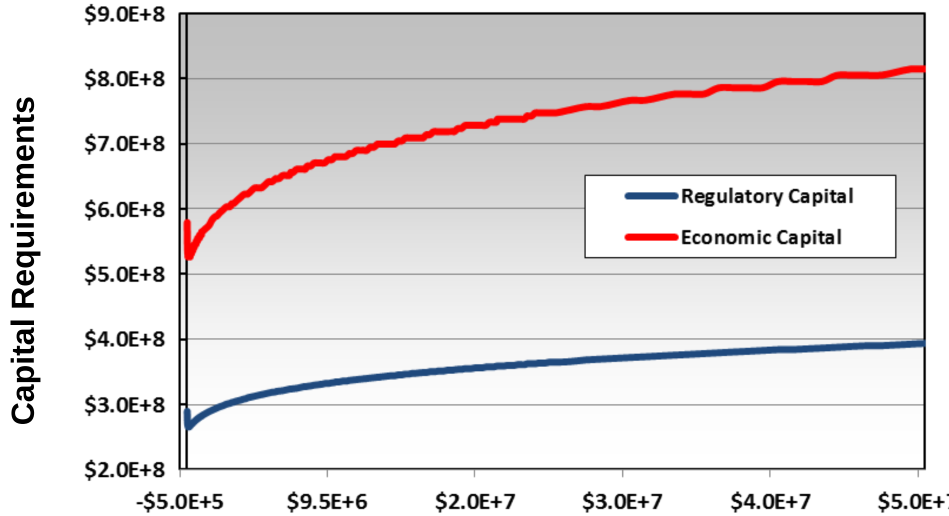
4. Using IF's Exact Capital Sensitivity Curves for Better Planning

MLE – Truncated LogG (a = 35.5, b = 3.25, H=10k)

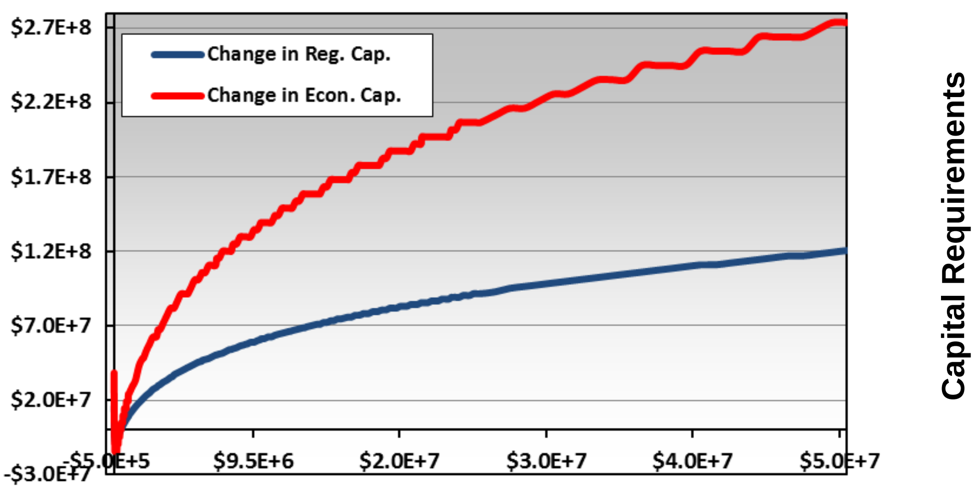
Y-AXIS NOT SCALED



Y-AXIS SCALED



N=250, λ = 25, regulatory α = 0.999, economic α = 0.9997



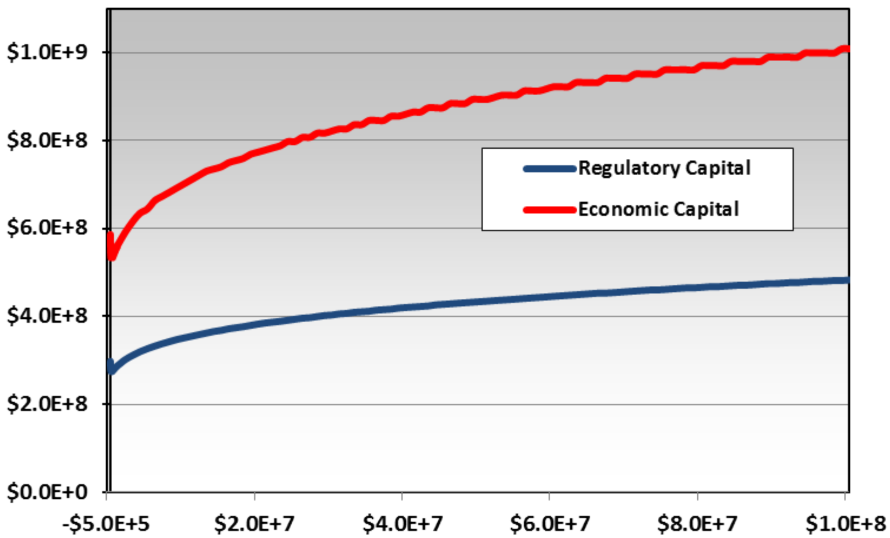
American Bankers Association

DataMineIt

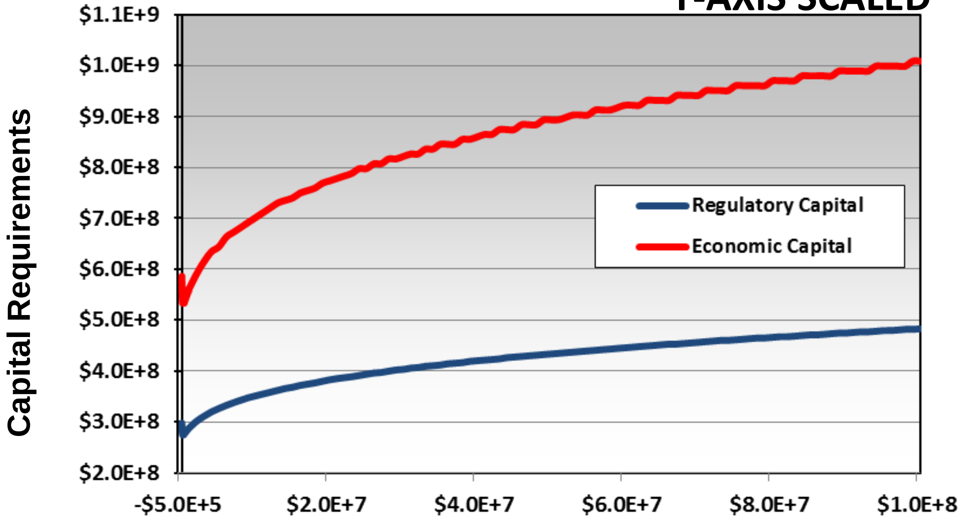
4. Using IF's Exact Capital Sensitivity Curves for Better Planning

MLE – Truncated LogG (a = 35.5, b = 3.25, H=25k)

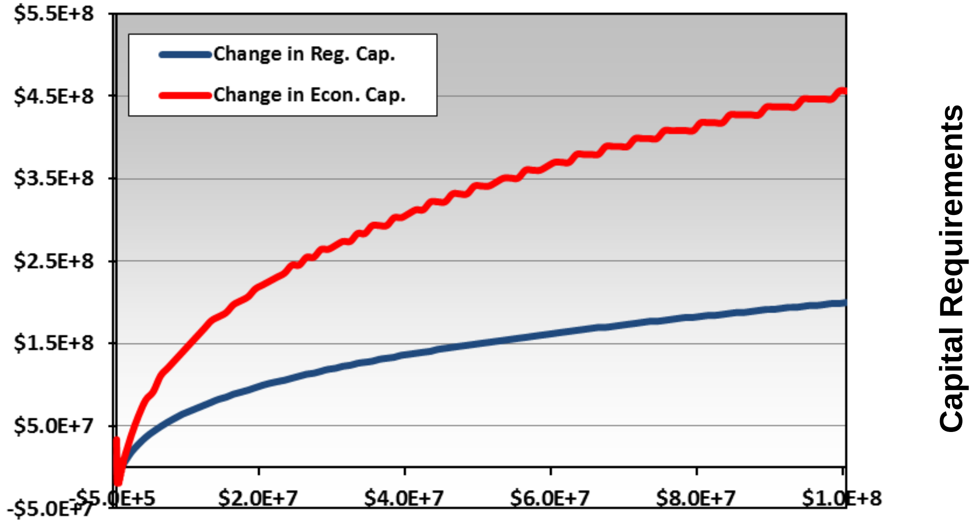
Y-AXIS NOT SCALED



Y-AXIS SCALED



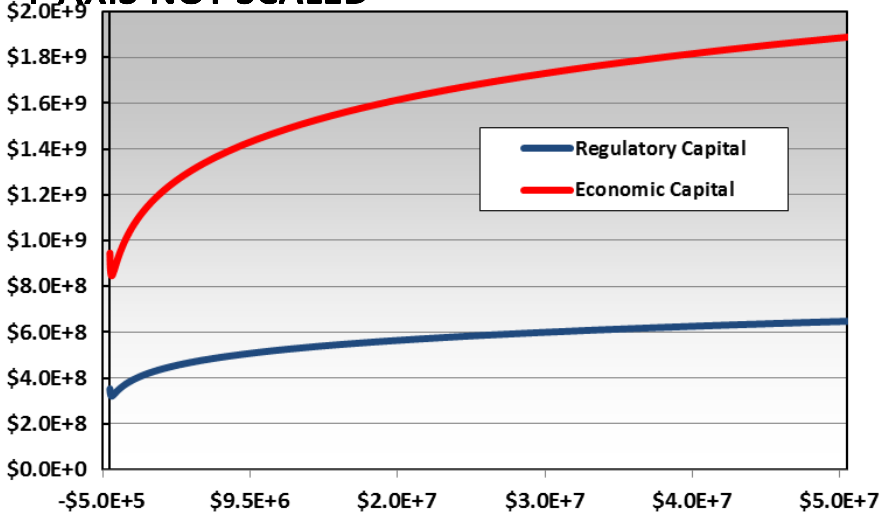
N=250, λ = 25, regulatory α = 0.999, economic α = 0.9997



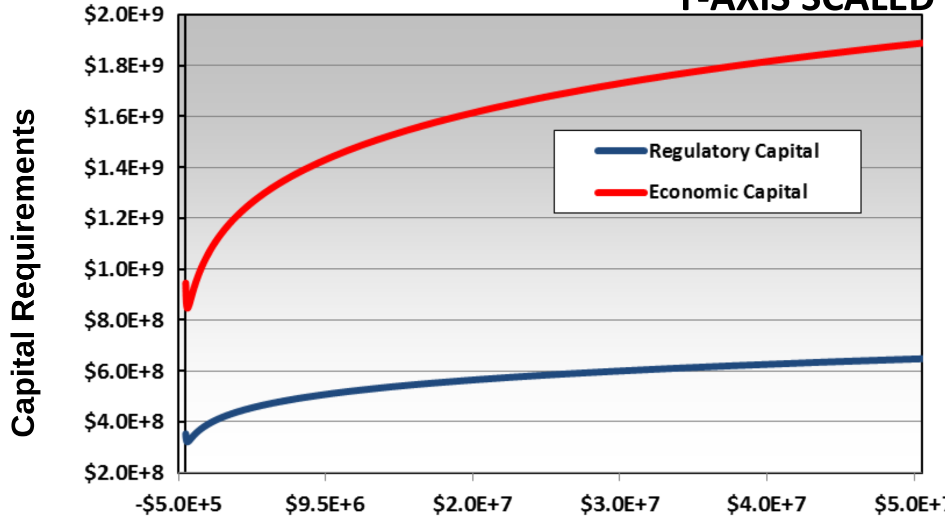
4. Using IF's Exact Capital Sensitivity Curves for Better Planning

MLE – GPD (Ɛ = 0.875, β = 57,500)

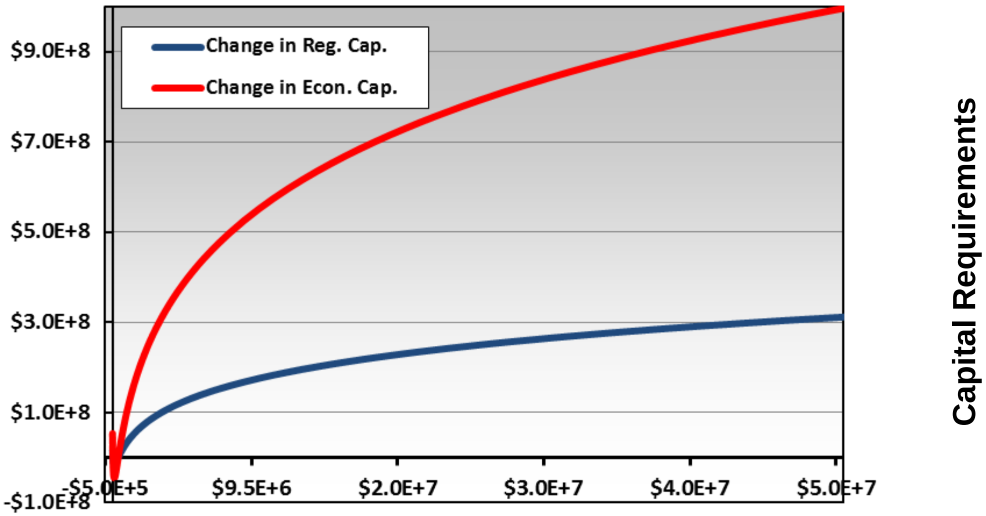
Y-AXIS NOT SCALED



Y-AXIS SCALED



N=250, λ = 25, regulatory α = 0.999, economic α = 0.9997



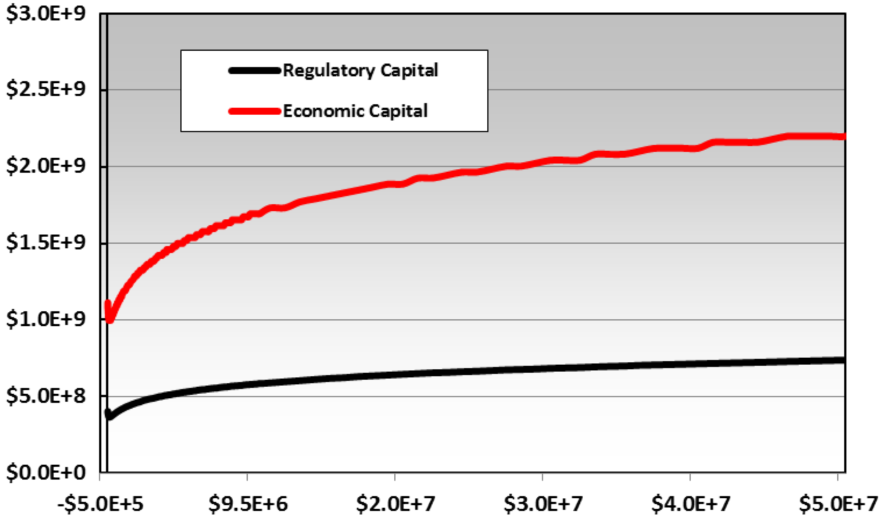
American Bankers Association

DataMineIt

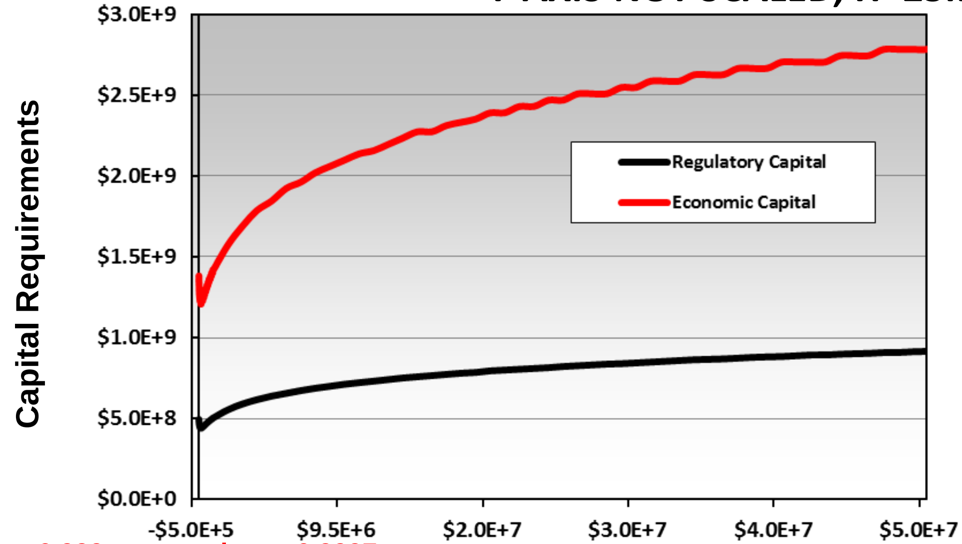
4. Using IF's Exact Capital Sensitivity Curves for Better Planning

MLE – Truncated GPD ($\mathcal{E}$ = 0.875, β = 57,500)

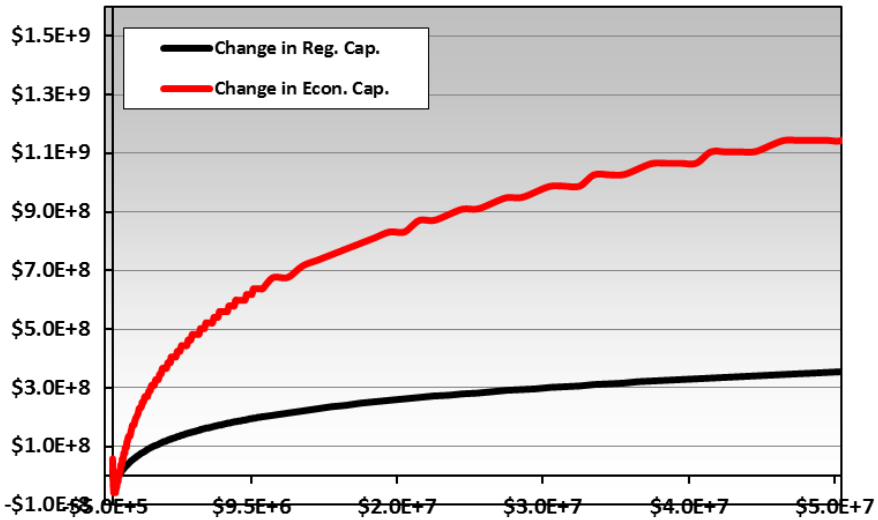
Y-AXIS NOT SCALED, H=10k

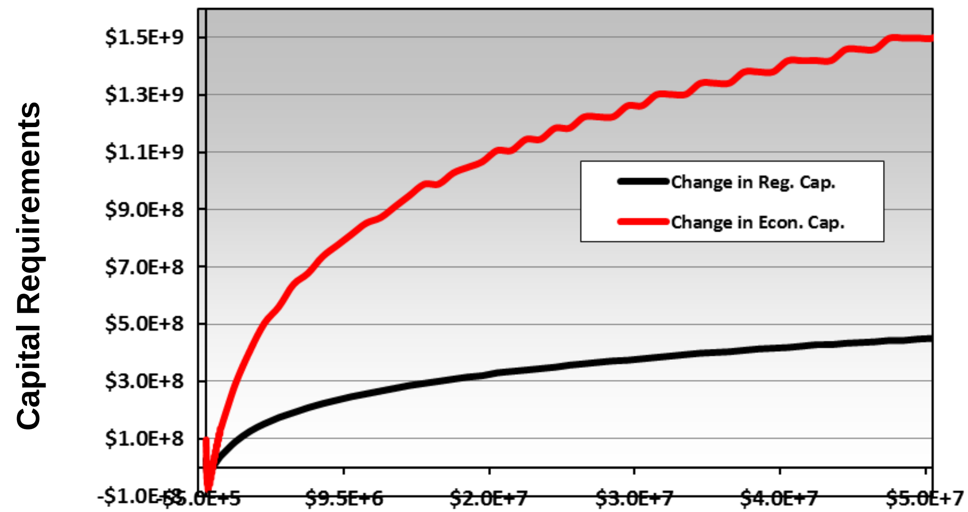


Y-AXIS NOT SCALED, H=25k



N=250, λ = 25, regulatory α = 0.999, economic α = 0.9997





5. Summary and Conclusions

- **The Influence Function (IF) is an extremely useful analytical tool in the Operational Risk severity modeling and capital estimation setting.**
- **The IF provides the EXACT behavior of virtually any estimator when losses are added, dropped, or changed.**
- **This provides great insight into severity estimator choice and development, which should be motivated almost exclusively by the need for an estimated capital distribution that is i) more precise, ii) less biased, and iii) more robust to extreme tail events over time.**
- **Once an estimator is selected, the IF's provision of EXACT CHANGES IN THE ESTIMATOR directly yields the EXACT CHANGES IN CAPITAL under new losses, with no (additional) estimation error (beyond that associated with severity and frequency parameter estimation).**
- **These EXACT CAPITAL SENSITIVITY CURVES allow for more accurate and more certain capital planning prospectively, under a wide range of hypothetical future scenarios, as well as retrospectively, for exact attribution and but-for analyses.**

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,** $s\left(x;\theta\right) = \left[\partial f\left(x;\theta\right)/\partial\theta\right]/ f\left(x;\theta\right)$ **, 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.**

American Bankers Association

DataMineIt

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.**

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):

STEP 1: 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.**

STEP 2: 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.

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

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

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.**

6. Appendix 2: SLA Capital Simulations

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

- **SLA (Single-Loss Approximation): Parameter estimates are used in Degen's (2010/2011) (similar to Böcker and Klüppelberg's (2005)) SLA formula to obtain capital estimates, and the distributions of these capital estimates are compared.**

$$\alpha = 0.999;\ \&\ \lambda = 25 \text{ arbitrarily;}\quad C_{\alpha} \approx F^{-1}\left(1-\frac{1-\alpha}{\lambda}\right)+\lambda\mu$$

- **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

6. Appendix 2: SLA Capital 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.

Time did not permit a full set of simulations to be run using GPD, but there are no methodological constraints against doing this, which preliminary runs confirm. Even when simulated random samples exhibit parameter values ($\hat{\xi} > 1$) yielding an infinite mean, which is especially common for the truncated GPD, utilization of Degen's (2010/2011) correct SLA approximation, which does not rely on the estimated mean of the severity distribution, is easily implemented and yields correct results.

$$C_\alpha \approx F^{-1}\left(1-\frac{1-\alpha}{\lambda}\right)-(1-\alpha)F^{-1}\left(1-\frac{1-\alpha}{\lambda}\right)\cdot\left(\frac{c_\xi}{1-1/\xi}\right)$$ where $c_\xi = (1-\xi)\dfrac{\Gamma^2(1-1/\xi)}{2\Gamma(1-2/\xi)}$ if $1<\xi<\infty$, and $c_\xi = 1$ if $\xi = 1$

This is not to say that severity distributions with infinite means are desirable or undesirable in this setting – only that the methodology contained herein is agnostic on the subject and is not adversely affected by it.

American Bankers Association

DataMineIt

6. Appendix 2: SLA Capital Simulations

- **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.**

- **OBRE value of *c*: 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. Developing fully data-driven algorithms to obtain these values is ongoing research.**

- **OBRE Starting Values: 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).**

6. 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.
- Degen, M., "The Calculation of Minimum Regulatory Capital Using Single-Loss Approximations," *The Journal of Operational Risk*, Vol. 5, No. 4, 3-17, Winter 2010/2011.
- Dupuis, D.J. (1998), "Exceedances Over High Thresholds: A Guide to Threshold Selection," *Extremes*, Vol. 1, No. 3, 251-261.
- Ergashev, B. (2008), ""Should Managers Rely on the Maximum Likelihood Estimation Method while Quantifying Operational Risk?" *The Journal of Operational Risk*, 3(2), pp. 63–86.
- 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.
- Jensen, J. L. W. V. (1906), "Sur les fonctions convexes et les inégalités entre les valeurs moyennes," *Acta Mathematica*, 30 (1), 175–193.
- Kennedy, P. (1992), *A Guide to Econometrics*, The MIT Press, Cambridge, MA.
- Ruckdeschel, P. and Horbenko, N. (2010) "Robustness Properties of Estimators in Generalized Pareto Models," Berichte des Fraunhofer ITWM, Nr. 182.
- 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. (1977), *Robust Statistical Procedures*, SIAM, Philadelphia.
- Opdyke, J.D. and Cavallo, A. (2012), "Estimating Operational Risk Capital: the Challenges of Truncation, the Hazards of MLE, and the Promise of Robust Statistics," *forthcoming*.
- 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.