Getting Extreme VaR Right: Eliminating Convexity & Approximation Biases Under Heavy-tailed, Moderately-Sized Samples

J.D. Opdyke

QuantMinds/RiskMinds September 10, 2019

Boston, MA

QuantMinds Americas Co-located with RiskMinds Americas

RiskMinds Americas Co-located with QuantMinds Americas

Disclaimer

J.D. Opdyke is VP-Financial Risk and Measurement, Enterprise Risk and Return Management at Allstate.

J.D. Opdyke is the sole author of all work contained herein, which was completed while he was a Senior Managing Director at DataMineit, LLC, before he began his position at Allstate.

The views presented herein are solely those of the author and do not necessarily reflect the views of DataMineit, LLC, or Allstate or any of its subsidiaries or legal entities.

Absent Defensible Risk Measurement, There is no Risk Management

"**Measurement is the first step** that leads to control and eventually to improvement. If you can't measure something, you can't understand it. If you can't understand it, you can't control it. If you can't control it, you can't improve it." (emphasis added)

- H.J. Harrington

This presentation is all about increasing the accuracy, precision, and robustness of measuring the magnitude of extremely high severity, low probability events to guide decision-making for more effective and efficient risk management. Absent useable and scientifically defensible measurement, there simply is no meaningful risk management.

Contents

I. The Model

- **Credit Risk, Operational Risk, Insurance Risk, Actuarial, Catastrophic Loss, and many other models all have used The Compound Loss Distribution (CLD) approach.**
- **The CLD convolves a frequency distribution (representing the # of losses that can occur during a specified time period (e.g. 1 year)) and a severity distribution (representing the magnitudes of these losses) to generate a compound loss distribution.**
- **We then estimate risk metrics (e.g. Value-at-Risk (VaR)) and related 'capital' based on the CLD.**

$$\text{(annual) CLD} \sim S = \sum_{i=1}^{N} X_i$$

$$\text{where } X \sim f_X(x;\theta),\ N \sim p_N(n;\lambda) \sim \text{(typically) } \lambda^n e^{-\lambda} / n!,^{**}$$

$$\text{and } g_{X,N}(x,n) = f_X(x;\theta) \cdot p_N(n;\lambda)\ \left[\text{frequency/severity independence}\right]$$

$$\text{estimated capital} \sim VaR_\alpha(S) = \inf\left\{s \in \mathbb{R} : F_S(s) \geq \alpha\right\} \text{ for } \alpha \in (0,1),$$

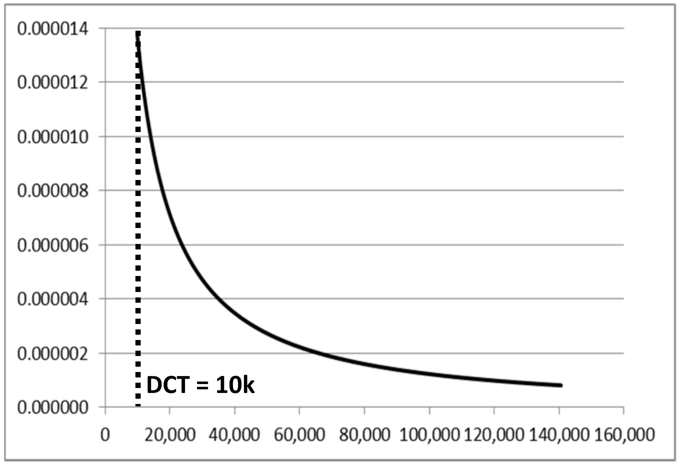
$$\text{where } F_S(s) \text{ is the cdf of } S\ \left(\alpha = 0.999 \text{ for } 99.9\%\text{tile}\right)$$

**NOTE: As explained below, the choice of frequency distribution affects the ultimate VaR estimate very little – orders of magnitude less – than the choice the severity distribution. The Poisson is the most commonly used frequency distribution, but only minor changes are required to analytic results related to the CLD if other frequency distributions (e.g. the Negative Binomial) are used.

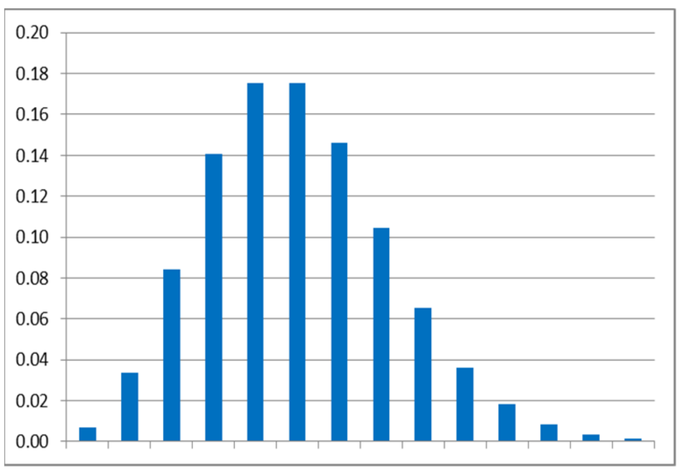
I. The Model

Compound Loss Distribution Approach (for one unit-of-measure):

***f* () ~ Estimated Severity PDF – Truncated LogNormal (μ=10, σ=2.8, H=10k)**

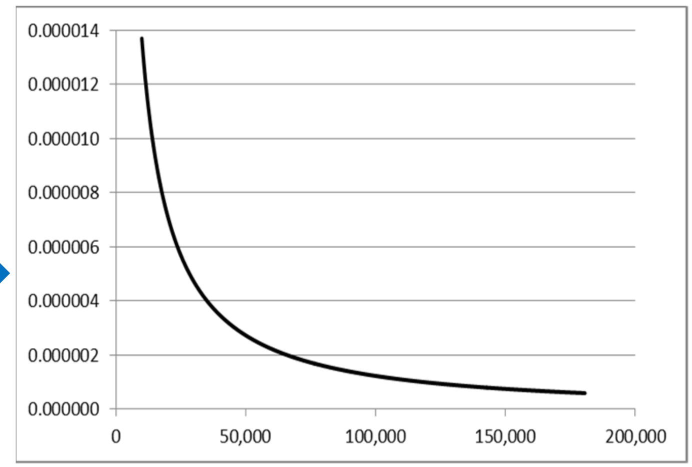


***p* () ~ Estimated Frequency PMF – Poisson (annual λ=25)**



Convolution via simulation (in practice, rarely a closed form solution … but for the VaR there are good & widely accepted analytical approximations much faster than Monte Carlo simulation)

Compound Loss Distribution (CLD)



(Regulatory) Capital = VaR at 99.9%tile

I. The Model

1. **Per the above, most applications of CLD assume independence between the frequency and severity distributions, although this is widely recognized as an unrealistic simplifying assumption. Initial research (see Stahl, 2016) tells us that risk metrics like VaR increase in magnitude and variance when this assumption is relaxed.**

2. **Note that this estimate of VaR is based on estimates of the parameters of the severity and frequency distributions: this is a different estimator than the Empirical VaR estimator, which is shown below:**

$$VaR_{\alpha-empirical} \sim N\left(VaR_{\alpha}, SD\right) \text{ where } SD = \frac{\sqrt{\alpha\left(1-\alpha\right)}}{f\left(VaR_{\alpha}\right)\sqrt{n}}$$

$\alpha =$ specified percentile, a.k.a. 'confidence level' (here, $\alpha = 0.999+$)

$VaR_{\alpha} =$ quantile at specified percentile

$f\left(VaR_{\alpha}\right) =$ value of probability density at the quantile of the specified percentile

$\left[\text{note that when the quantile itself is being estimated, this becomes } f\left(\widehat{VaR_{\alpha}}\right)\right]$

$n =$ number of loss event data points in the sample

   a. **As we shall see below, VaR-CLD is most certainly not normally distributed, at least not under finite sample conditions, under which it is highly skewed (see Opdyke, 2014, Opdyke, 2017, and (only relevant for VaR-SLD) Guégan, et al., 2017).**

I. The Model

b. However, inverting this relationship gives us the number of losses needed to obtain a specified precision on our quantile (capital) estimate, and this is instructive to see that the variance for extreme VaR is extremely large, which matches our intuition: our certainty in a 1-in-1000 year estimate (i.e. VaR99.9) should be very low compared with those of smaller VaRs. More on this later.

$$n = \frac{4\alpha(1-\alpha)}{\varepsilon^2\left[f\left(VaR_\alpha\right)\cdot VaR_\alpha\right]^2}$$

$\varepsilon = 2 \cdot SD/VaR_\alpha =$ relative error of the quantile estimate

$n =$ number of data points required to achieve precision of $\varepsilon\%$

c. Shevchenko (2011) uses this to show that even for a relatively light-tailed severity distribution (LogNormal, μ=0, σ=2), at least 50,000 to 100,000 years worth of losses would be required to attain quantile (capital) estimates within 10% of the true value (which is another reason we cannot use this estimator).

3. Also very important to note that all results contained herein for VaR-CLD, under the conditions discussed herein in the next section, also apply to VaR based on a Single Loss Distribution (VaR-SLD) where frequency (sample size) is held constant. In other words, when we have only a severity distribution.

This is notable: the only difference between VaR-CLD and VaR-SLD is a larger variance and a larger bias associated with the former over the latter, which again matches our intuition, especially given that VaR-CLD has been shown to be notably skewed under finite sample conditions (see Opdyke, 2014 and (only relevant for VaR-SLD) Guégan et al., 2017).

II. Relevant Conditions of Use/Application

The algorithm/model presented herein (JAEQE – Jensen-Adjusted Extreme Quantile Estimates) fills a hole in the literature when the conditions below all exist concurrently (which is often):

i. **VaR is Extreme: At the very least, we're referencing VaR99.5 (representing a 1-in-200 year loss, on average), although most define "extreme" here as VaR99.9 or greater (representing a 1-in-1000 year loss, on average).**

   a. **Importantly, for the CLD, VaR99.9 requires estimating much higher VaR for the severity distribution specifically: for, say, $\lambda = 30$, and $\alpha = 0.999$ and $\alpha = 0.9997$, (which correspond to typical values of Regulatory and Economic Capital, respectively) $p = \left[1-\left(1-\alpha\right)/\lambda\right] =$ 0.999967 and 0.99999, respectively. These are EXTREMELY high VaR!**

   b. **This contributes significantly to very large increases in the size of the variance and bias associated with the estimate of VaR-CLD.**

ii. **Sample sizes are Finite: As shown below, estimation bias due to Jensen's inequality disappears asymptotically, so we are concerned with sample sizes of, say, somewhere north of 150 and south of 1,000.**

iii. **Severity Loss distributions are Heavy-tailed: Convexity in VaR-CLD as a function of one or more severity parameters (specifically, those associated with the tail index, as shown below) only exists for heavy-tailed severity distributions: this is a non-issue for, say, Gaussian data.**

The extant literature treats different combinations of these 3 conditions, but rarely, if ever, all 3 simultaneously.

Notably, the degree to which the above 3 conditions are true directly affects the size of the variance, bias, and RMSE of the VaR-CLD (and VaR-SLD) estimate, as discussed further below.

III. Approximation, Estimation, and Model Error

Three sources of error exist when estimating VaR-CLD (see Opdyke, 2017; Abdimomunov et al., 2019):

1. **Approximation Error (not relevant to VaR-SLD) – formulaic error in approximating VaR.**
2. **Estimation Error – NOT choosing the right parameter values of the right distributions.**
3. **Model Error – NOT choosing the right frequency and severity distributions.**

1. Very rarely do frequency-severity pairs provide exact, closed-form solutions to VaR-CLD. The Single-Loss Approximation (SLA) of Degen (2010) is the only closed-form approximation that does NOT (always) require numeric integration (only when solving for the point b) below). This is shown below.

a) if $\xi < 1$, $C_\alpha \approx F^{-1}\left(1-\frac{1-\alpha}{\lambda}\right)+\lambda\mu$ where μ is the mean of F

b) if $\xi = 1$, $C_\alpha \approx F^{-1}\left(1-\frac{1-\alpha}{\lambda}\right)+c_\xi\lambda\mu_F\left[F^{-1}\left(1-\frac{1-\alpha}{\lambda}\right)\right]$ where $c_\xi = 1,\ \mu_F(x)=\int_0^x\left[1-F(s)\right]ds$

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

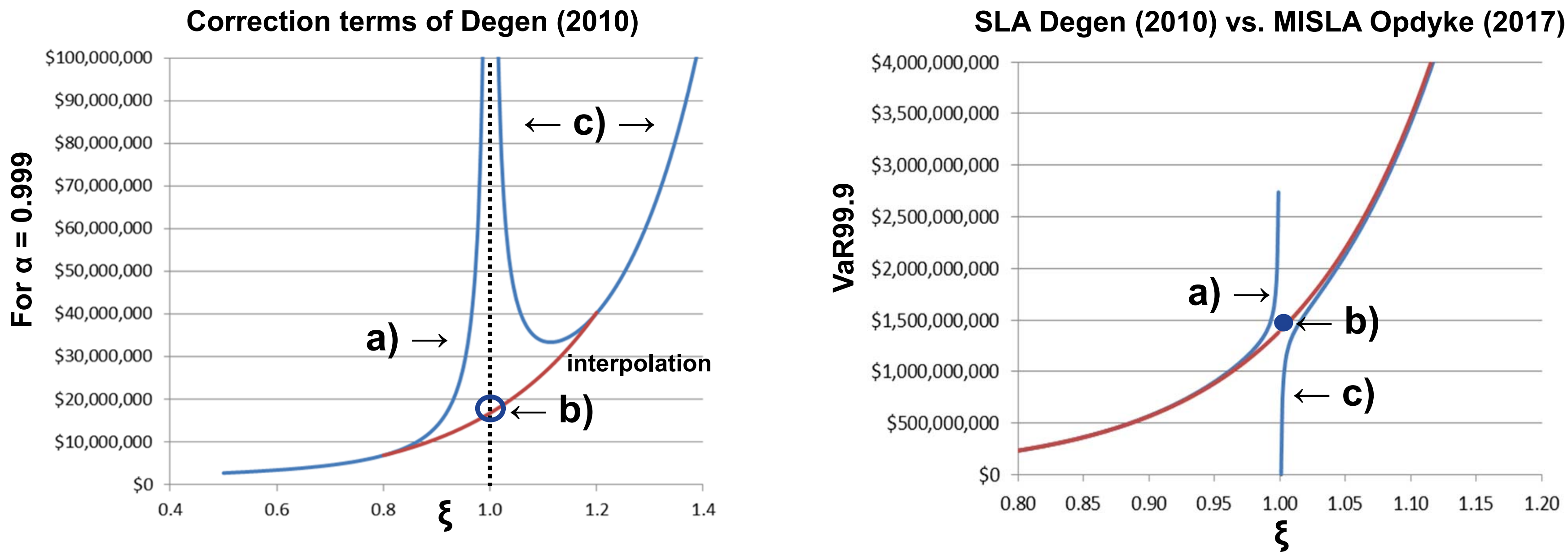
($\xi \geq 2$ is so extreme as to not be relevant in this setting)

(the above assumes a Poisson-distributed frequency distribution and can be modified if this assumption does not hold)

Opdyke (2017) showed SLA to have a discontinuity when the tail index approaches a value of 1.0, and he solved the problem by implementing a straightforward nonlinear interpolation across the discontinuity.

III. Approximation, Estimation, and Model Error

Figure 1: SLA vs. MISLA for GPD Severity (θ = 55,000; ξ = tail index)

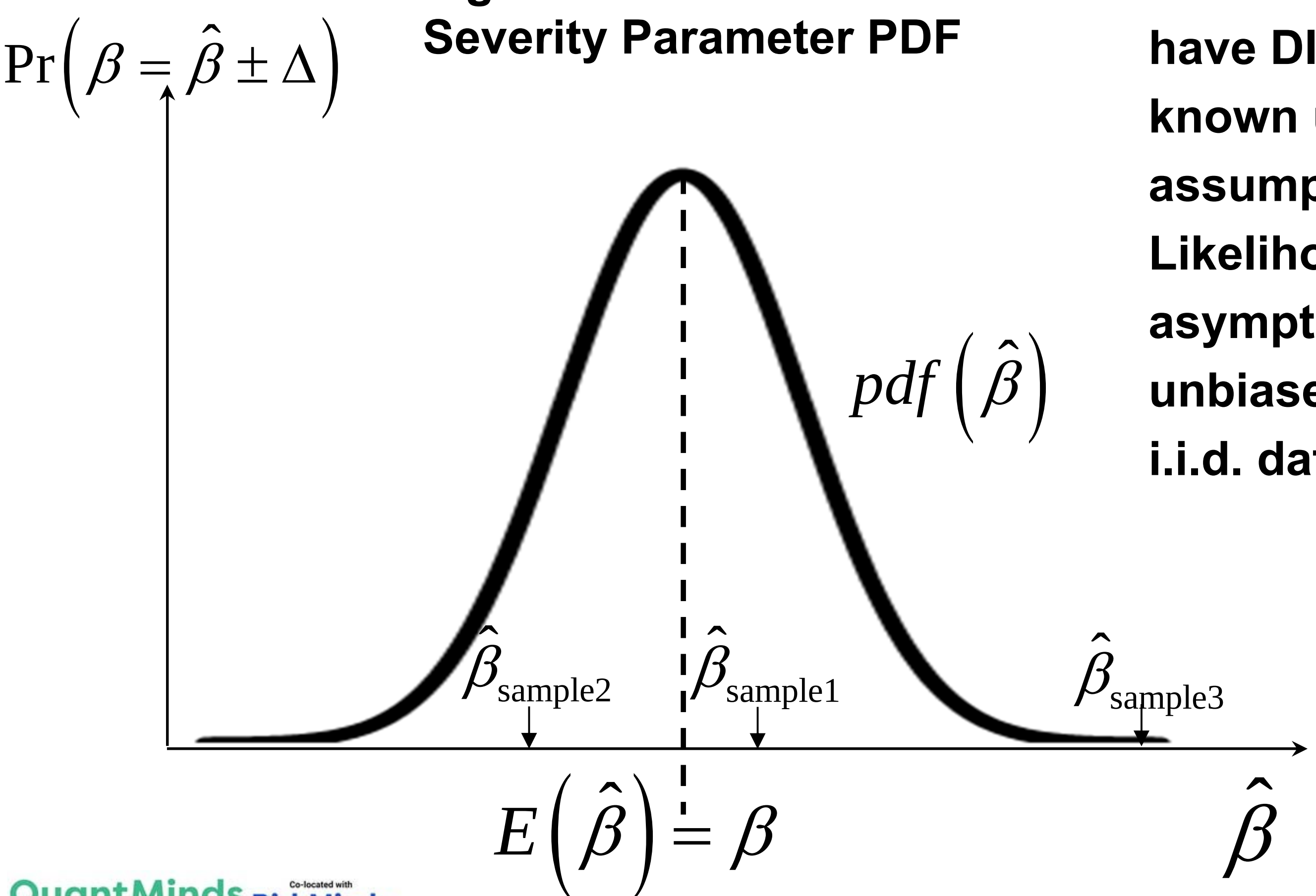


- **Opdyke (2017) conducted an extensive simulation study of methods for approximating VaR-CLD, and found his MISLA and the PE2 of Hernandez et al. (2014) to be the fastest and most accurate, with the former slightly faster and the latter slightly more accurate. However, when both are applied across the range of most or all possible severity parameter values as herein, MISLA is much faster than PE2 on average as it does not require numeric integration, and PE2 always does. Therefore MISLA is used throughout this study as the best method for eliminating approximation error.**
- **Note that VaR-CLD (y-axis above) is a highly convex function the tail index (which is a direct function of one of the severity parameters!!).**

2. Estimation Error

- **Loss/Returns Data = a Sample, NOT a Population**
- **Therefore, true severity parameters, β , will never be known.**

Figure 2:
Severity Parameter PDF



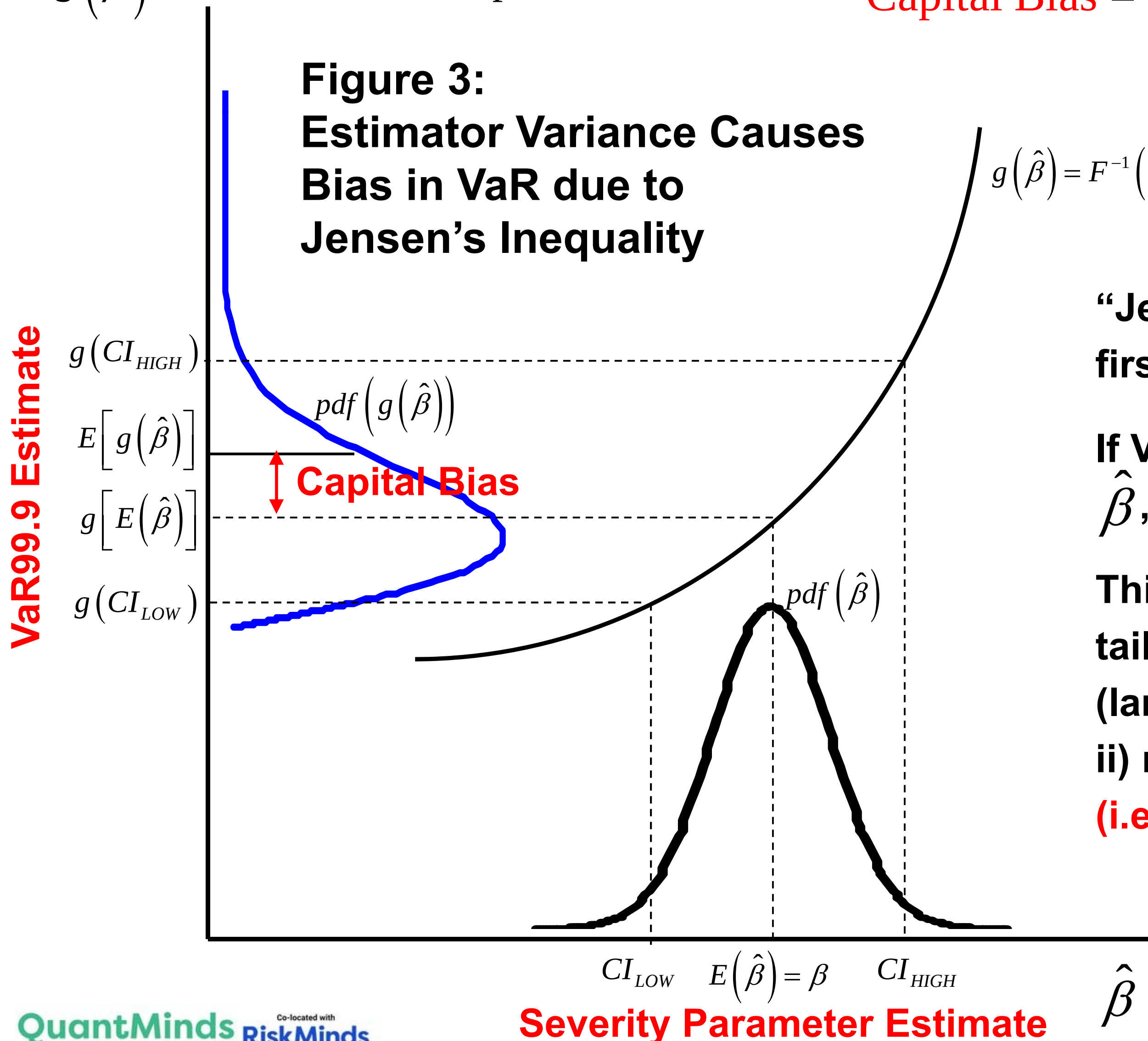
BUT Estimated Parameters, $\hat{\beta}$, have DISTRIBUTIONS that are known under specified assumptions (e.g. Maximum Likelihood Estimators (MLE) are asymptotically normal, unbiased, and efficient under i.i.d. data)

III. Approximation, Estimation, and Model Error

$g\left(\hat{\beta}\right) = \hat{C} = \text{Estimated Capital}$

Capital Bias $= \left(E\left[g\left(\hat{\beta}\right)\right] - g\left[E\left(\hat{\beta}\right)\right]\right) > 0$

Figure 3: Estimator Variance Causes Bias in VaR due to Jensen's Inequality



"Jensen's inequality" first proved in 1906.

If VaR, g(), is strictly convex in $\hat{\beta}$, capital always will be inflated.

This is true for i) relevant (heavy-tailed) severities under relevant (large) parameter values, WHEN ii) n=not large, AND iii) p is large (i.e. p≥0.995).

* Graph based on Kennedy (1992), p.37.

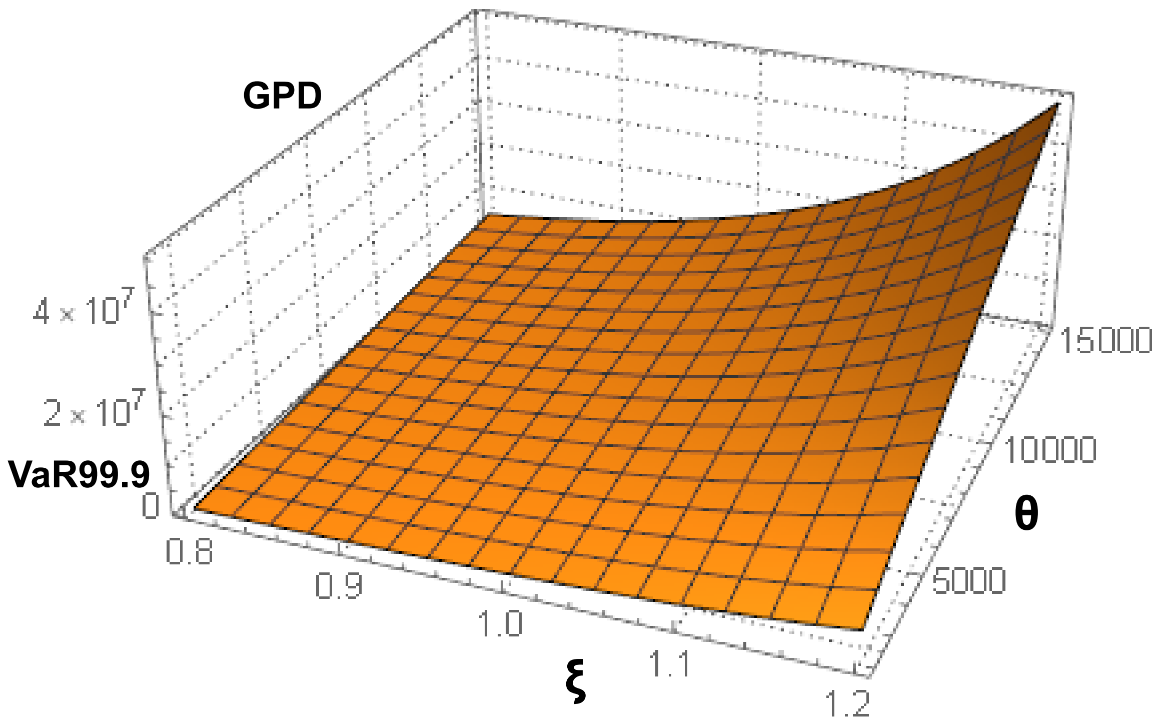
III. Approximation, Estimation, and Model Error

- **For straightforward severities with closed-form VaR functions, Figure 3 (for large p) – the convexity of VaR as a function of severity parameter(s) – can be shown analytically.**

- **For example, for the LogNormal (which has no dependence between parameters):**

$$VaR_{LN} = \exp\left(\mu + \sigma\Phi^{-1}(p)\right), \text{ so } \partial^2 VaR_{LN} / \partial\mu^2 = VaR_{LN} > 0 \text{ and } \partial^2 VaR_{LN} / \partial\sigma^2 = VaR_{LN} \cdot \left[\Phi^{-1}(p)\right]^2 > 0$$

- **For more complicated distributions, especially when truncated, these marginal checks of each parameter are easy to do graphically (such results are shown for GPD in Figures 4a/4b), as are bivariate graphs of the response surface.**



$$VaR_{GPD} = \left[\theta((1-p)^{-\xi} - 1)\right] / \xi$$

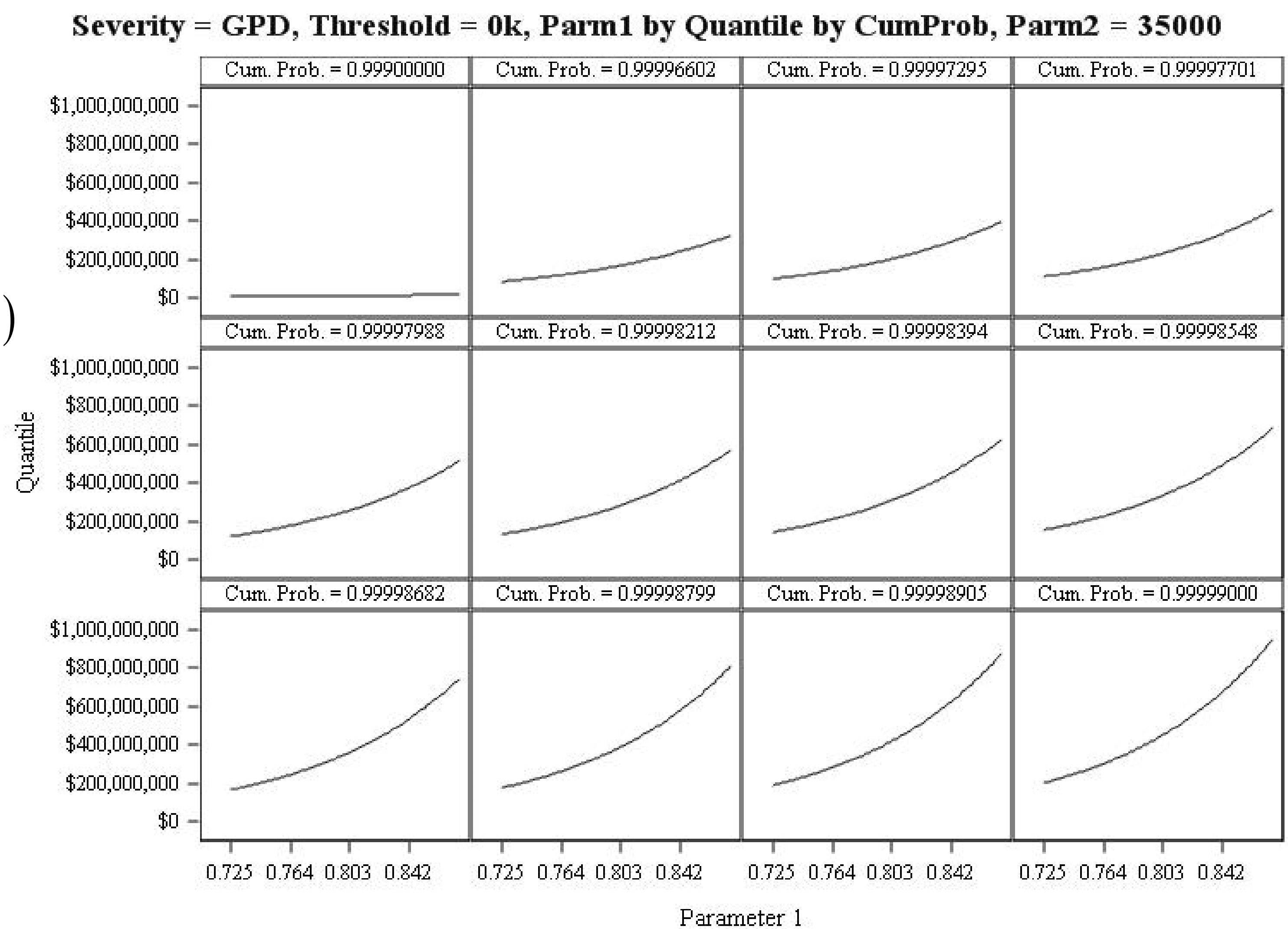
$$\partial^2 VaR_{GPD} / \partial\theta^2 = 0$$

$$\partial^2 VaR_{GPD} / \partial\xi^2 = \left[(1-p)^{-\xi}\theta(2 - 2(1-p)^{\xi} + \xi Log[1-p](2 + \xi Log[1-p]))\right] / \xi^3 > 0$$

III. Approximation, Estimation, and Model Error

Figure 4a: For Tail Index Parameter ξ, Convexity of VaR-CLD Increases Rapidly as a Function of *p*

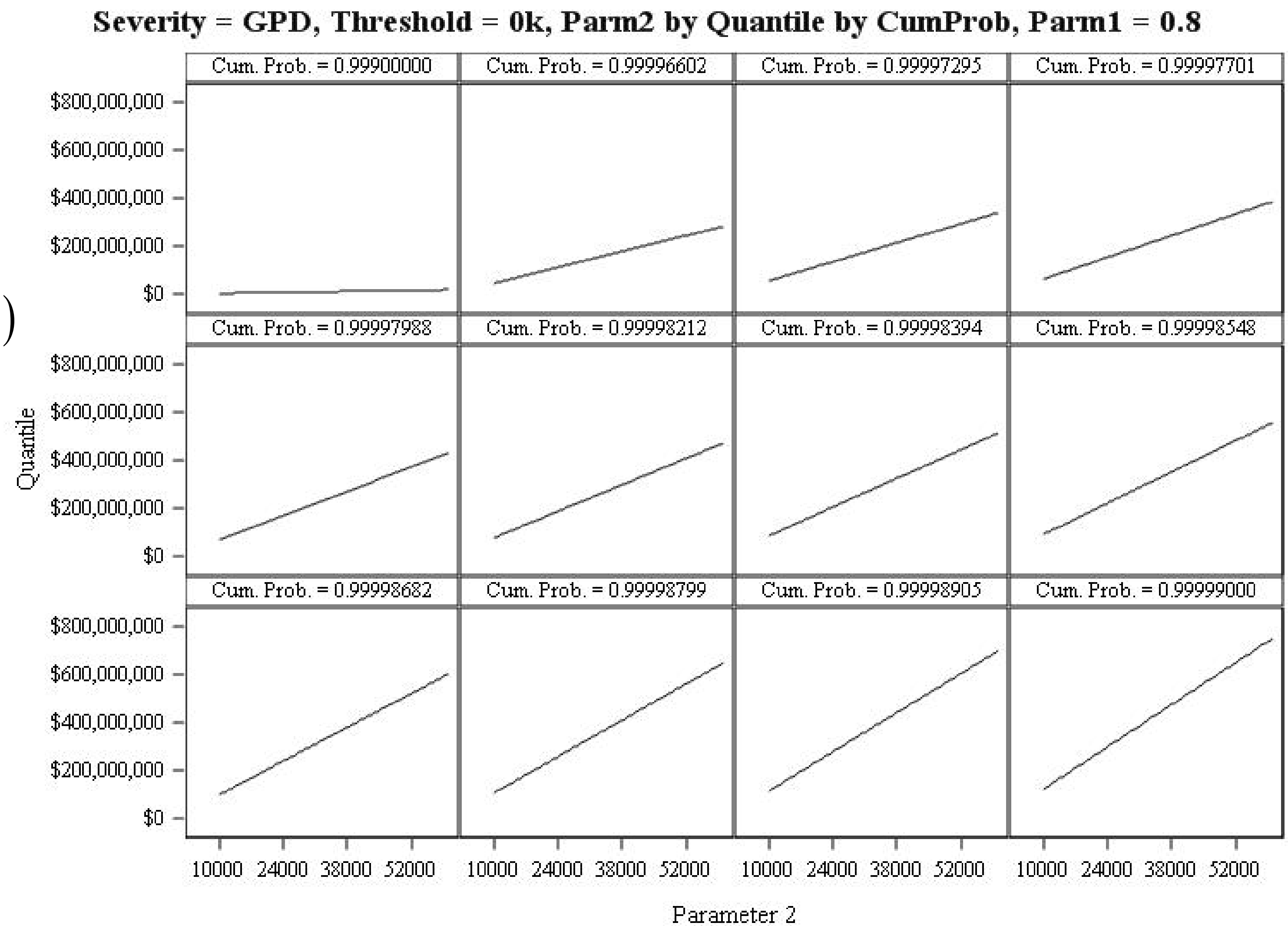
$F^{-1}(p;\xi)$



III. Approximation, Estimation, and Model Error

Figure 4b: VaR-CLD is a LINEAR function of *θ*, the NON-tail Index Parameter

$F^{-1}(p;\theta)$



III. Approximation, Estimation, and Model Error

TABLE 1: VaR as a function of Severity Parameters in the Tail (i.e. p ≥ 0.995) by Parameter by Severity
(see Appendix 4 for Distributional Characteristics of 1, 3, 4, 6, 8, 9)

Severity Distribution	VaR is Convex/Linear as Function of...			Relationship between Parameters
	Parameter 1	Parameter 2	Parameter 3	
1) LogNormal (μ, σ)	Convex	Convex		Independent
2) LogLogistic (α, β)	Linear	Convex		Independent
3) LogGamma (a, b)	Convex	Convex		Dependent
4) GPD (ξ, θ)	Convex	Linear		Dependent
5) Burr (type XII) (ϒ, α, β)	Convex	Convex	Linear	Dependent
6) Truncated 1)	Convex	Convex		Dependent
7) Truncated 2)	Linear	Convex		Dependent
8) Truncated 3)	Convex	Convex		Dependent
9) Truncated 4)	Convex	Linear		Dependent
10) Truncated 5)	Convex	Convex	Linear	Dependent

Note that, as shown in Figure 4a, VaR's convexity in *p* (when *p* is large) INCREASES in *p*: larger quantiles are associated with greater convexity.

III. Approximation, Estimation, and Model Error

Before proceeding, it is important to define not only the MECHANISM by which VaR's estimate (as a function of severity parameter(s)) becomes biased, but also the mechanisms defining the DEGREE of this bias (and increased variance), and these are 3:

1. **The larger the sample size (and smaller the variance of the severity parameter estimate(s)) the smaller that VaR bias. This can be seen in Figure 3: smaller variance on the X-axis shrinks the skewness and bias of VaR on the Y-axis.**
2. **The farther out in the tail we go, i.e. the larger the *p*, the larger the VaR bias. This is shown above in analytics derivations and graphically.**
3. **Finally, the heavier the tail, the larger the VaR bias (so truncated distributions, ceteris paribus, exhibit more VaR bias, as do the same severity distributions with 'larger' (tail index) parameter values)**

As shown empirically in Opdyke (2014), VaR is a slightly concave function of *λ*, the frequency parameter. This makes sense intuitively as it is essentially 1. above: *λ* defines the sample size, and bias shrinks as sample size increases because parameter estimate variance decreases. And even when VaR is a linear function of one of the severity parameters, it is a highly convex function of the other, i.e. the one associated with the tail index* (see Figure 1). So on net, VaR as a function of all 3 (or more) CLD parameters is convex (and thus, POSITIVELY biased – this bias only goes in one direction!).

Note that for many applications (e.g. Operational Risk modeling), the above conditions put us squarely in the bias goldilocks zone: we have few observations (several hundred at most), heavy-tailed loss distributions (e.g. GPD, LogGamma, and LogNormal with large σ≥2), and must estimate quantiles associated with large percentiles (e.g. p ≥ 0.995 or 0.999).

* Degen (2010) defines the tail index as follows: A positive measurable function, f, is regularly varying with parameter β (written as $f \in RV_\beta$) if f satisfies $\lim_{t\to\infty} f(tx)/f(t) = x^\beta$ for all x>0. In the case f = a probability density, one in particular has that $f \in RV_{-1/\xi-1}$ implies $1-F \in RV_{-1/\xi}$, where ξ is the tail index.

III. Approximation, Estimation, and Model Error

Also before proceeding, several important mentions below:

A. **Because CVaR / ES is a (provably) convex function of severity parameter estimates (see Brown, 2007, Bardou et al., 2010, & Ben-Tal, 2005), switching from VaR to CVaR does not avoid this problem (and in fact, appears to make it worse).**

B. **Jensen Inequality-induced VaR bias is unaffected by whether moments are in/finite per se.**

C. **The extensive focus in this setting on PARAMETER estimation ($\hat{\theta}$), as opposed to VaR estimation, has been largely misguided BECAUSE PARAMETER ESTIMATION MISSES THE MECHANISM OF VaR BIAS (see Figure 3)!**

$$\left[E\left(VaR\left[f\left(x;\hat{\theta}\right)\right]\right)-VaR\left(E\left[f\left(x;\hat{\theta}\right)\right]\right)\right]>0$$

Over the past dozen years, especially in the operational risk space, many dozens of journal papers have focused on manipulating parameter estimates to achieve better VaR estimates, with at best mixed success: only robust parameter estimation has had some limited success here, but this is only because more extreme values of the distribution of parameter estimates (X-axis in Figure 3) are downweighted: this approach STILL does not directly address the convexity of VaR as a function of the (severity) parameter values, which is the only way to improve the accuracy, precision, and robustness of these VaR estimates.

D. **Other bias reduction strategies (e.g. see Kim and Hardy, 2007), do not appear to work under these conditions because most involve shifting the distribution of the estimator, often using some type of bootstrap, which can easily result in negative VaR estimates and greater instability in the estimator. Also, approaches that rely on the derivative(s) of VaR(s) such as (Taylor) series expansions, can easily run into numeric precision issues for some severities given the very high quantiles being estimated. So even when such solutions exist in theoretical form, practical challenges may derail their application under these conditions.**

IV. How can Jensen's Inequality (1906!) Have Been Missed?

- **Jensen's Inequality WAS PROVEN IN 1906: how can it have been missed in this setting?!**
- **Aside from Opdyke and Cavallo (2012a, 2012b), Opdyke (2014 & 2017), only Taleb & Douady (2014 & 2015) explicitly identify convexity, i.e. Jensen's inequality, as the source of the bias in parametrically-based estimates of VaR (Guégan, et al., 2017 shows that the distribution of VaR-SLD (not VaR-CLD) is skewed, but does not connect this to Jensen's inequality and the resulting bias in its estimation).**
- **One cannot develop a defensible solution to a problem without first spelling out WHY it was missed in previous research… We believe some of the reasons this was missed here include:**
  - **A narrow focus on the statistical properties of parameter estimators, rather than exploring/recognizing the possibility that the function that LINKS the parameter estimates to the VaR is nonlinear, and thus, skews the VaR-CLD distribution and biases its first moment.**
  - **Statisticians like asymptotic results, and asymptotically, bias goes away, since variance in the parameter estimate, which allows convexity-induced VaR bias, goes away; so a focus on asymptotic results overlooks this real-world Jensen Inequality-induced VaR bias.**
  - **The empirical VaR estimator is normally distributed and well-behaved, which might have thrown some off the scent of the VaR-CLD estimator being convex.**
  - **Convexity-induced VaR-CLD bias does not manifest when severity distributions are not heavy-tailed, and many researchers have a tendency to default to (mathematically convenient) Gaussian assumptions, which often, if not typically, do not reflect reality when it comes to loss/returns distributions.**
  - **This bias only manifests materially when alpha is large (at least ≥ 0.995) (Appendix 1 addresses the issue of whether large alpha are appropriate for financial institutions, as opposed to estimating, say, 1,000-year floods).**
  - **This bias often only manifests materially when the COMBINATION of most or all of these conditions exist concurrently: alpha is large (≥ 0.995), severity tails are heavy, AND sample sizes are small to moderately sized.**
  - **Finally, Opdyke (2014 and 2017) and Abdymomunov et al. (2019) warn of the importance of separately treating the 3 sources of error in this setting – Approximation, Estimation, and Model error. Failure to do this makes it very difficult, if not impossible based on empirical results alone, to correctly identify the MECHANISM causing the VaR bias, even when identifying it as a factual/empirical matter. In the operational risk setting, this resulted in extensive (and misguided) efforts to 'fix' biased VaR-CLD estimates via different types of parameter estimation. Appropriately separating distinct sources of error requires analytic discipline, but can avoid such wrong paths altogether, not to mention related papers that can cross the line from the misguided to the obfuscatory (see Larsen, unpublished, 2015; see APPENDIX 2 for more detail) when promulgating irrelevant schemes (e.g. so-called 'median bias'; see APPENDIX 2).**

IV. How can Jensen's Inequality (1906!) Have Been Missed?

- **Again, to reemphasize: the particular PARAMETER estimator used here generally matters little under these conditions, since (almost) all are unbiased and most are very similar in terms of efficiency (e.g. Maximum Likelihood Estimation (MLE), Penalized Likelihood Estimation (PLE), Method of Moments, Generalized Method of Moments, Probability Weighted Moments, etc. generally, all M-Class Estimators work well here; this includes many Robust Estimators (OBRE & CvM (see Opdyke and Cavallo, 2012b), Quantile Distance (see Ergashev, 2008), etc.), which are the only exception to the above as these do have SOME indirect, although not much, mitigating effect on Jensen's-induced VaR bias).**
- **While Taleb & Douady (2015) identify VaR-CLD bias due to Jensen's inequality, and base their fragility heuristic on it to serve as an effective 'red flag' indicating its presence (see Taleb & Douady, 2014), they do not develop an ESTIMATOR to mitigate its effects.**
- **The only other estimator in the extant literature to directly address VaR-CLD bias due to Jensen's inequality – the RCE estimator from Opdyke (2014) – yields similar empirical results to those developed herein. But RCE requires the specification of a tuning parameter value, and relies on asymptotic Fisher information matrices as opposed to the empirical versions used herein. Both issues arguably can be viewed as drawbacks compared to the JAEQE developed below.**
- **The point of the present research is to develop a very general ESTIMATOR of VaR-CLD that, under the widest possible range of distributions, directly estimates its bias-inducing convexity to eliminate it; this dramatically improves the accuracy, precision, RMSE, and robustness of the estimator compared to competitors under the conditions cited above (small-to-medium sample size, large alpha far into the tail, and heavy-tails). This is consistent with Harrington's dictate!**
- **The Jensen-Adjusted Extreme Quantile Estimate (JAEQE) is described below.**

V. Fixing Approximation Error with MISLA

1. **Approximation Error is irrelevant for VaR-SLD because this is just the quantile of a single (severity) distribution.* However, there very rarely are closed-form Compound Loss Distributions, so VaR-CLD must be approximated. Approximation error is effectively eliminated via both MISLA of Opdyke (2017) and PE2 of Hernandez et al. (2014), and while the latter is (very) slightly more accurate, the former is much faster to calculate under most circumstances, that is, across the entire range of relevant (severity) parameter values. So MISLA is the preferred approximation and is used throughout this presentation.**

***NOTE: Again, it is very notable that the results shown herein regarding Jensen's Inequality-induced bias in VaR-CLD also apply to VaR-SLD based on a single (severity) loss distribution like, say, a GPD; the only difference being that the variance and bias of VaR-SLD are less than those of VaR-CLD, all else equal.**

VI. Fixing Estimation Bias with JAEQE

1. The Source of Estimation Bias under the defined conditions (small-to-medium sample size, large alpha far into the tail, and heavy-tails) is Jensen' Inequality: convexity in VaR-CLD as a function of the severity parameter(s) estimates. This has been missed by almost all the extant literature.
2. Misguided attempts to improve parameter estimation will never address this convexity-induced bias in VaR-CLD.
3. So how do we fix it? … We have to broaden our estimation process beyond parameter estimates and explicitly build the convexity of VaR as a function of the severity parameter estimates into the VaR estimation process, and this is done using JAEQE.
4. The point of the present research is to develop a very general ESTIMATOR of VaR-CLD that, under the widest possible range of distributions, directly estimates its bias-inducing convexity to eliminate it; this dramatically improves the accuracy, precision, RMSE, and robustness of the estimator compared to competitors under the conditions cited above (small-to-medium sample size, large alpha, and heavy-tails). The Jensen-Adjusted Extreme Quantile Estimate (JAEQE) described below accomplishes these objectives.

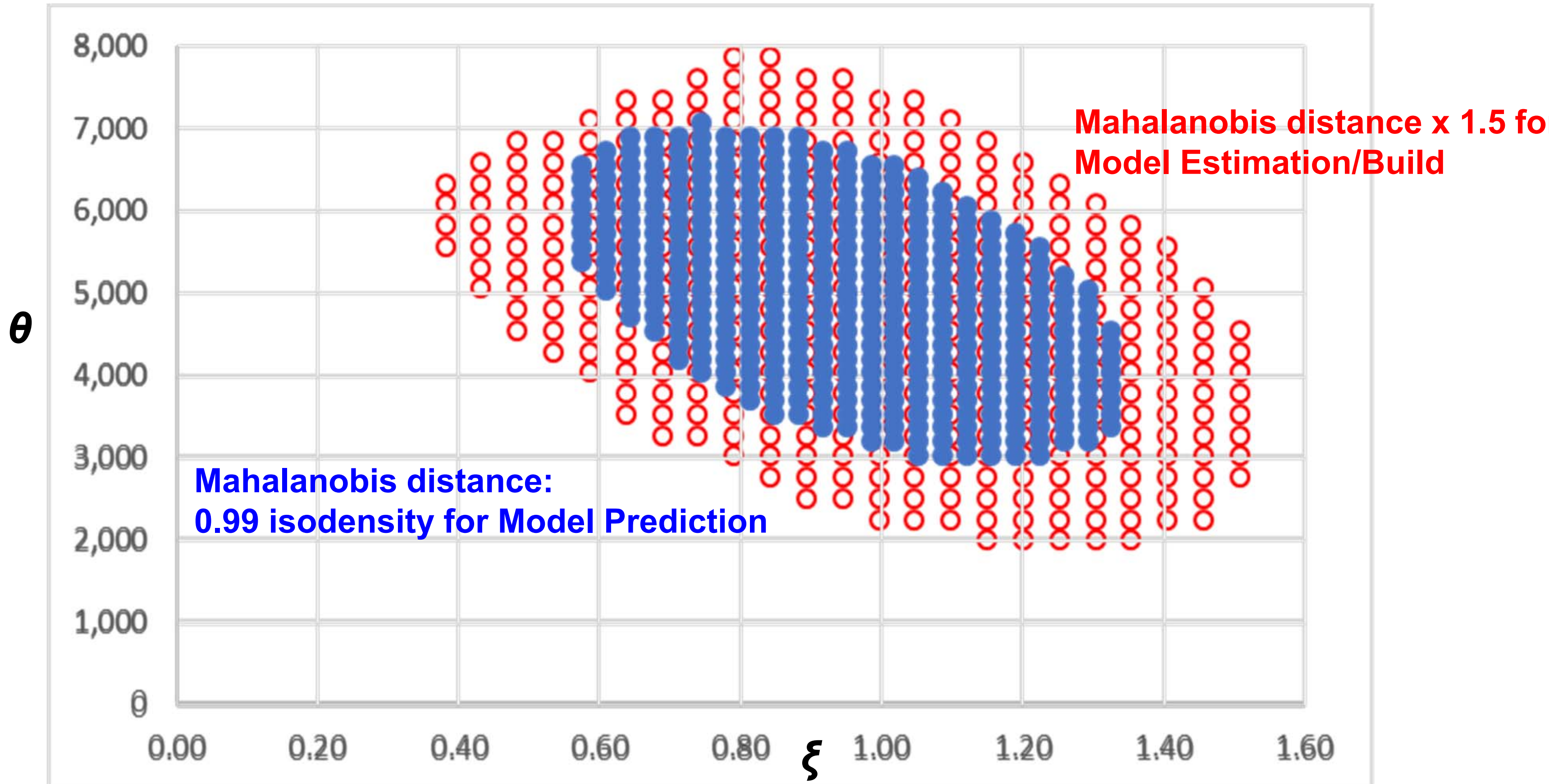
VI. Fixing Estimation Bias with JAEQE

The Jensen-Adjusted Extreme Quantile Estimate (JAEQE) is defined below:

1. **Estimate frequency and severity parameters from data sample of losses/returns** $\left(\hat{\lambda},\ \hat{\theta}_i \text{ where } i = 1,2\right)$
2. **Simulate 100k samples where all (3) parameters are random variables (severity dist. is selected!) and estimate the parameters (using MLE) for every sample.**
3. **Define the empirical 0.99 isodensity of** θ_i **using the Mahalanobis distance from the original 1.** $\hat{\theta}_i$ **estimates and define a grid within it of 400-500 of pairs of evenly spaced values of** θ_i **with** $\hat{\theta}_i$ **at the center.**
4. **Define a second grid like 3. but based on the Mahalanobis distance x d=1.5 (verifying that d=1.5 does not violate the parameter domain; if it does, adjust d accordingly).**
5. **Using the lattice from 4., generate 5k samples for each lattice point using the corresponding values for** θ_i**, calculate the corresponding values of VaR-CLD, calculate the mean of these 5k VaR-CLD's, and then the difference between this mean and the 'true' VaR-CLD using the values of** θ_i **for each lattice point. This difference is the estimated bias due to Jensen's Inequality.**
6. **Define Y = ['true' VaR-CLD / avg VaR-CLD] (ranging from 0 < Y ≤ 1) for each lattice point, and regress X's on Y using the Adaptive LASSO of Zou (2006) with** $\hat{Y}$ **= 1.0 where X's are polynomials of** θ_i **into the high teens, and all their interactions (when dependence exists between the parameters; see Table 1).**
7. **Take the lattice defined in 3. (i.e. the 0.99 isodensity) and generate 5k samples for each lattice point and 5k corresponding VaR-CLD values, apply the model built in 6. to deflate each of the 5k VaR-CLD values by the percentage that is estimated to be bias, and take the mean of the adjusted 5k VaRs.**
8. **Finally, define %diff=(avg adj VaRs – 'true' VaRs)/ 'true' VaRs, take avg %diff across all 400-500 values, then final VaRs = avg adj VaRs – avg%diff*('true' VaRs): this shift yields an unbiased VaR on a RELATIVE rather than ABSOLUTE basis (to avoid overfitting on the border of the 0.99 isodensity). This adjustment applied to VaR-CLD based on the original parameter estimates** $\hat{\theta}_i$ **yields the final JAEQE VaR-CLD estimate. For any new quarters of data, use the model from 6. & repeat 7 (only re-estimate model if beyond 0.99 isodensity).**

VI. Fixing Estimation Bias with JAEQE

<u>Figure 5: JAEQE Finite Sample Parameter Lattices</u> – GPD Severity (ξ = 0.95 θ = 5,000)



Note Upper Left Lattice for Model Build: Fewer than Half of the 5k samples for these Grid points had convergent parameter estimates, and so were discarded.

$Mdist = \sqrt{(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})}$ where $\vec{x}$ is the vector of data points, $\vec{\mu}$ is the vector of corresponding means, and Σ is the covariance matrix.

VI. Fixing Estimation Bias with JAEQE

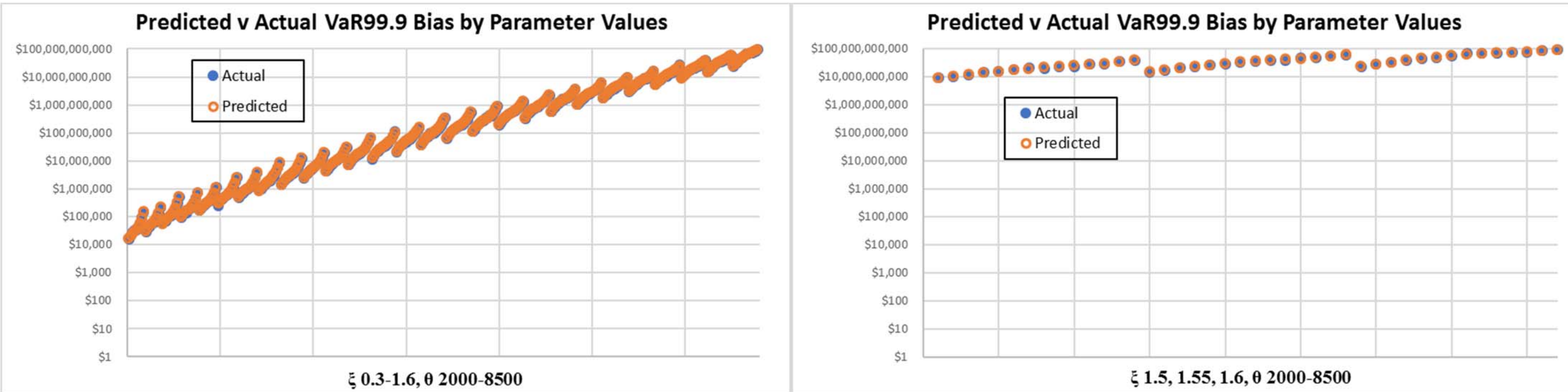
Table 2: Example of A-LASSO Model Statistics/Fit

VaR99.9														
ALasso Parms	**ALasso Est**		**Model Fit Stats**				**DF**	**Sum Sqrs**	**Mean Sqr**	**F Value**	**Prob>F**	**Model Sel Descrip**	**Value**	**ValueN**
int	0.930325		Root MSE	0.01456		Model	8	2.67859	0.33482	1579.6	<.0001	Selection Method	Adaptive LASSO	
parm1^4*parm2	-0.000011759		R-Square	0.9781		Error	283	0.05999	0.000212	_	_	Stop Criterion	SBC	
parm1^2	-0.5177		Adj R-Sq	0.97748		Corrected Total	291	2.73858	_	_	_	Effect Hierarchy	None	
parm1^4	0.190552		AIC	-2167.20022								Stop Horizon	3	3
parm1^8	-0.00518		AICC	-2166.4173										
parm1*parm2^4	1.72E-16		SBC	-2428.10943										
parm2^10	6.66E-41		ASE	0.00020543										
parm2^3	3.50E-13													
parm2^5	-3.02E-20													
VaR99.95														
int	0.908909		Root MSE	0.01629		Model	7	2.86971	0.40996	1545.8	<.0001	Selection Method	Adaptive LASSO	
parm1^2*parm2^2	1.37E-08		R-Square	0.97442		Error	284	0.07532	0.0002652	_	_	Stop Criterion	SBC	
parm1^3*parm2	-0.000073157		Adj R-Sq	0.97379		Corrected Total	291	2.94503	_	_	_	Effect Hierarchy	None	
parm1^2	-0.557937		AIC	-2102.72879								Stop Horizon	3	3
parm1^4	0.281613		AICC	-2102.09049										
parm1^8	-0.008142		SBC	-2367.31476										
parm2^5	-1.92E-20		ASE	0.00025794										
parm2^8	7.82E-33													

VI. Fixing Estimation Bias with JAEQE

Figure 6a: A-LASSO Model–VaR99.9 Bias Predictions v. Bias Actuals, **λ=25, GPD Severity (ξ=0.925 θ=6000) (center)**

Log-Scale



NOT Log-Scale (**NOTE THE SPANNING OF MANY ORDERS-of-MAGNITUDE in VaR to fully cover ellipses!**)

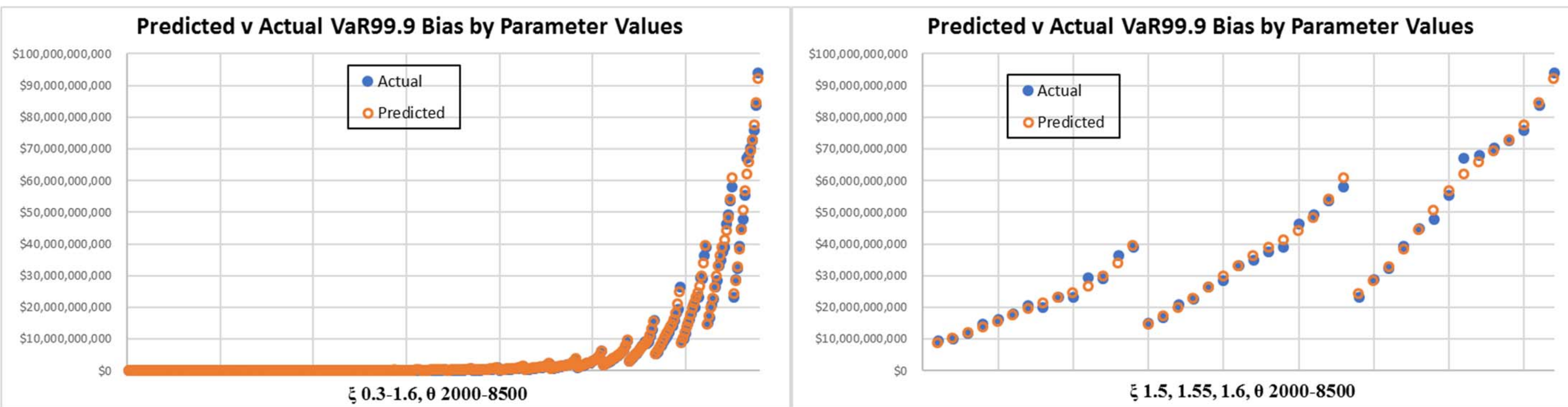
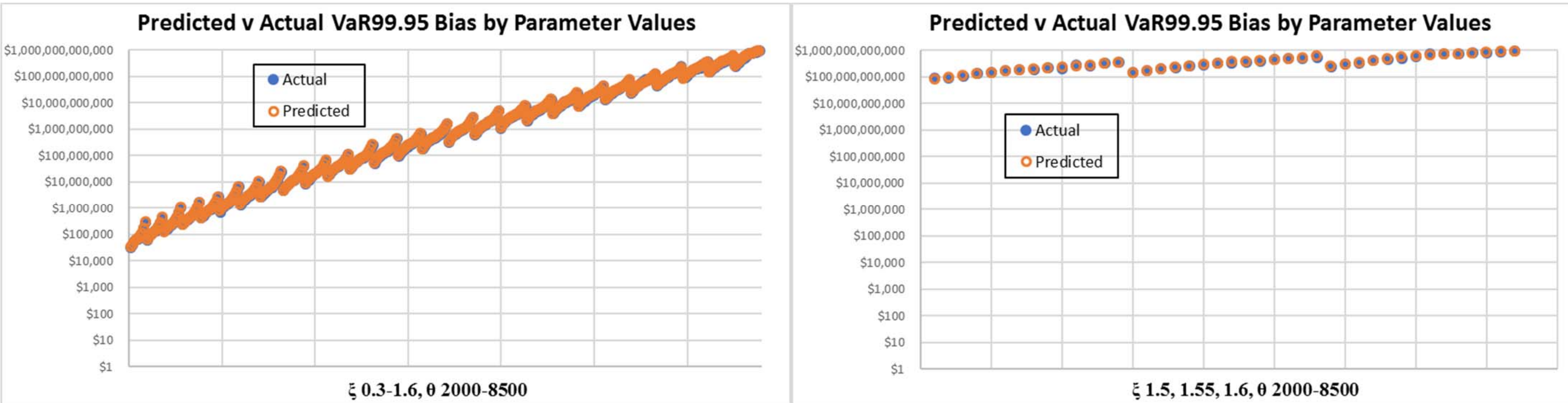
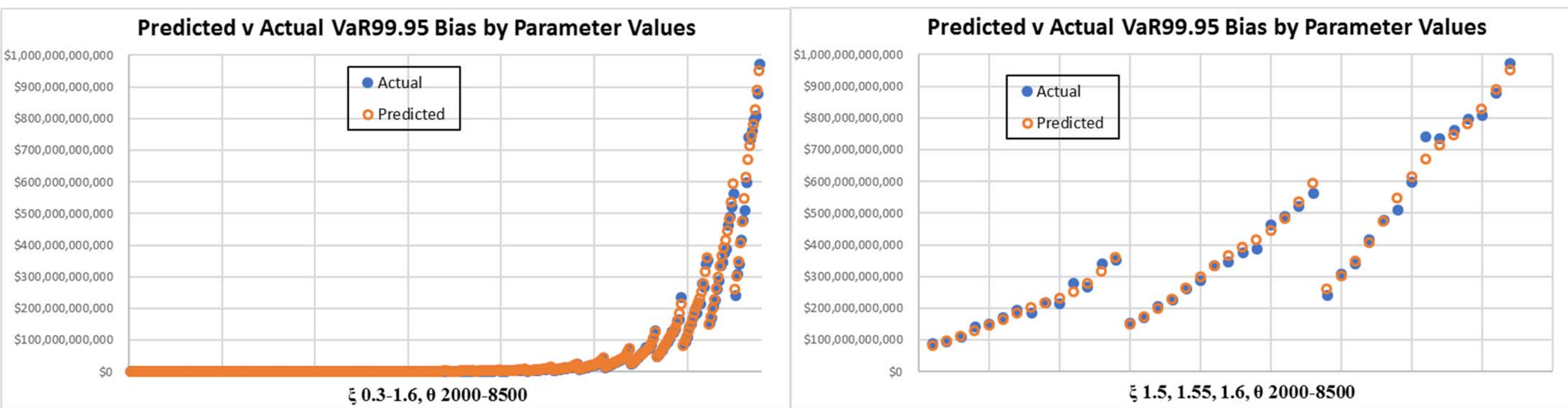


Figure 6b: A-LASSO Model, VaR99.95 Bias Predictions v Bias Actuals λ=25, GPD Severity (ξ=0.925 θ=6000) (center)

Log-Scale

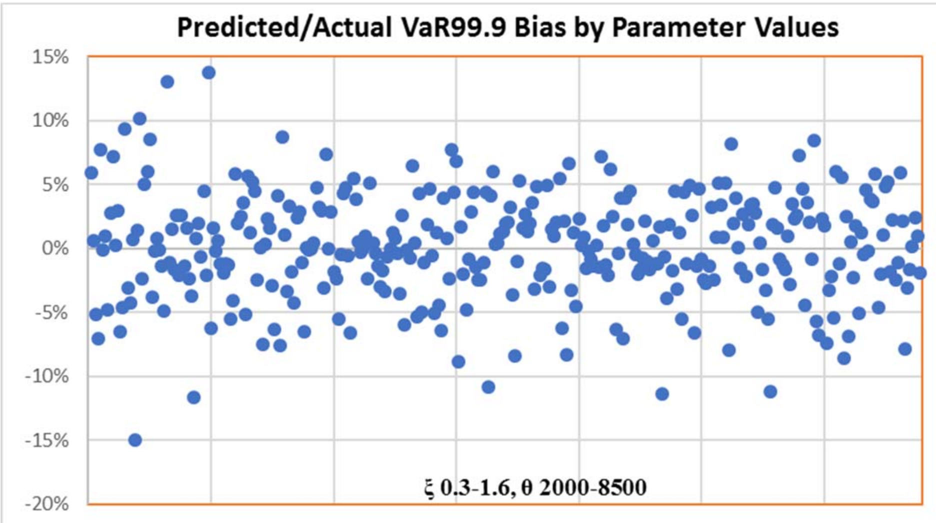


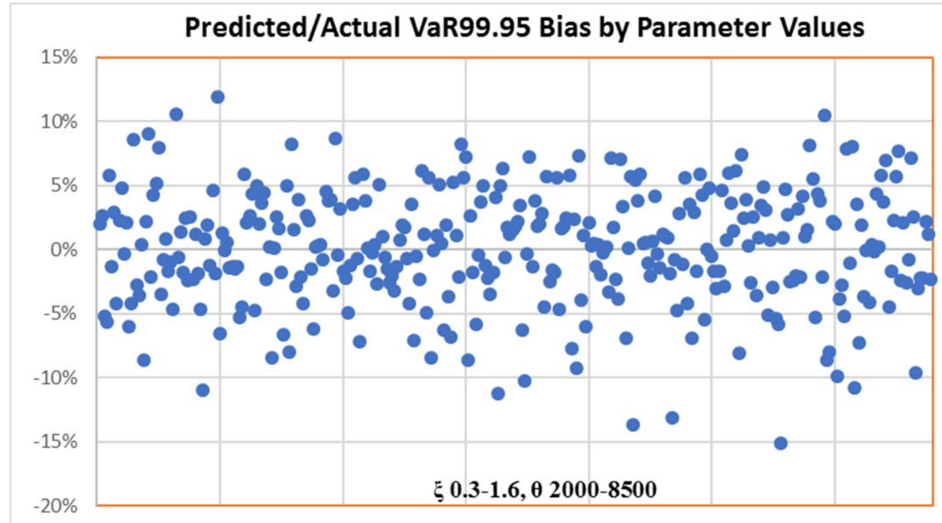
NOT Log-Scale (**NOTE THE ORDER-of-MAGNITUDE DIFFERENCE BETWEEN VaR99.9 (p.27) & VaR99.95!**)

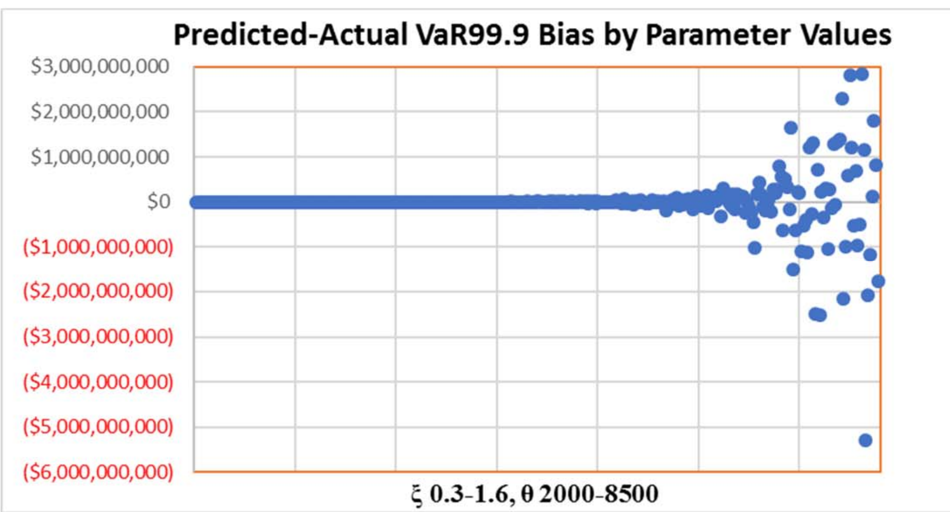


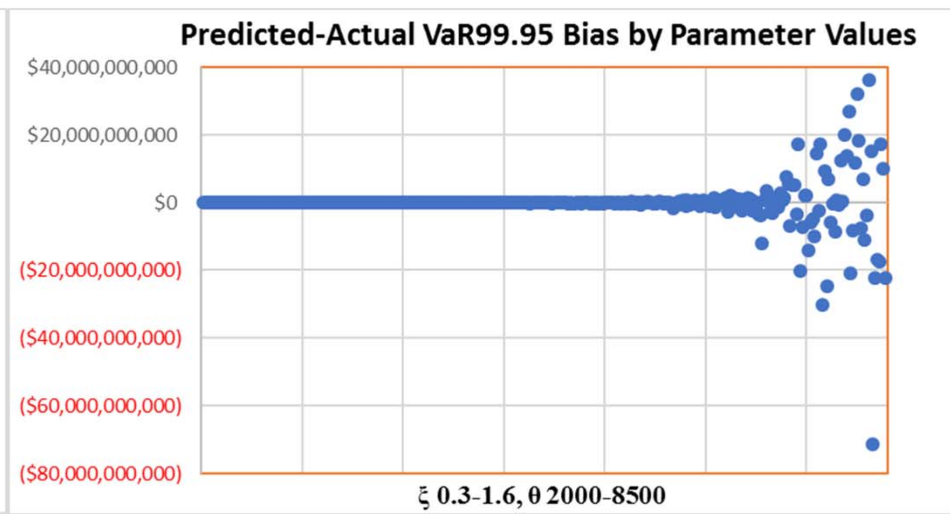
VI. Fixing Estimation Bias with JAEQE

<u>Figure 6c: A-LASSO Model – Bias Predictions vs. Bias Actuals</u> λ=25, GPD Severity (ξ=0.925 θ=6000) (center)





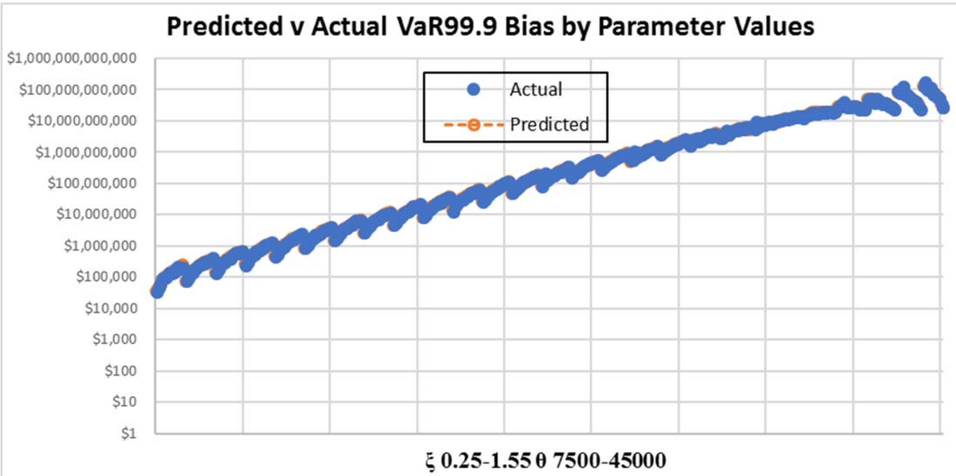


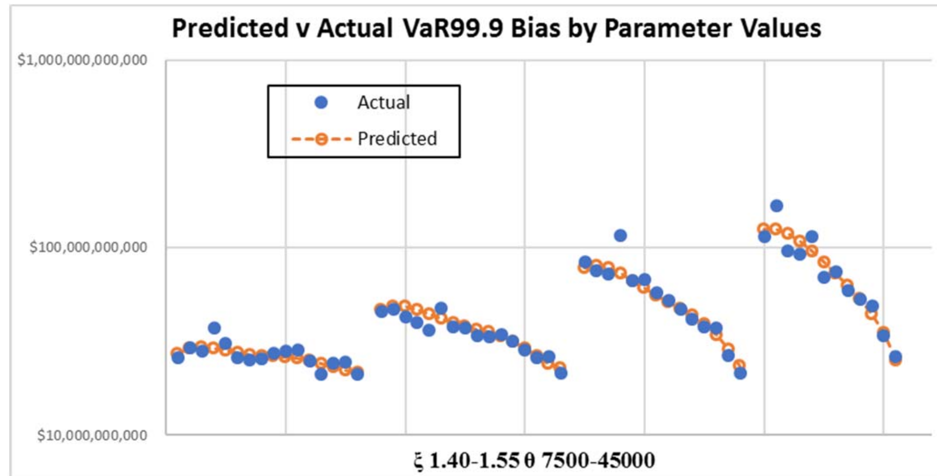


VI. Fixing Estimation Bias with JAEQE

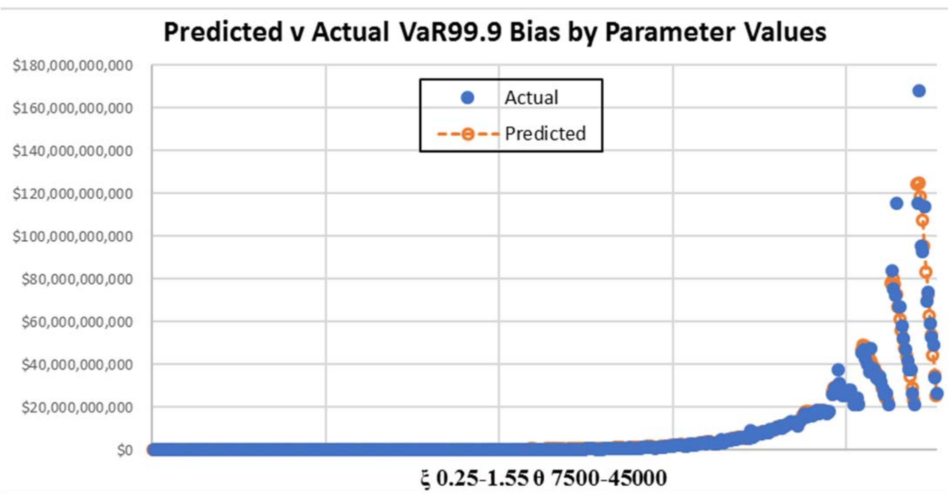
Figure 6d: A-LASSO Model–VaR99.9 Bias Predictions v Bias Actuals, λ=25, TGPD Severity (ξ=0.9 θ=25,000) (center)

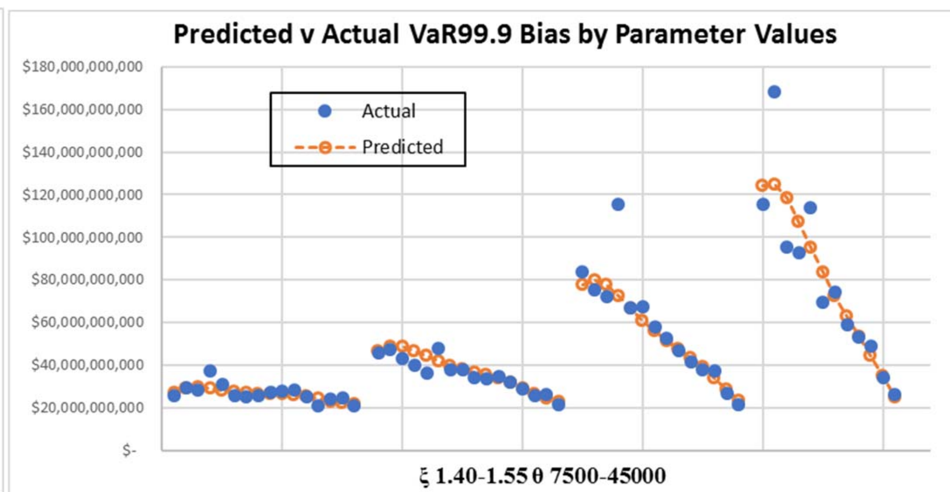
Log-Scale





NOT Log-Scale **CAUTION: TRUNCATION CAN CHANGE DIRECTION OF FUNCTIONAL RELATIONSHIPS!**

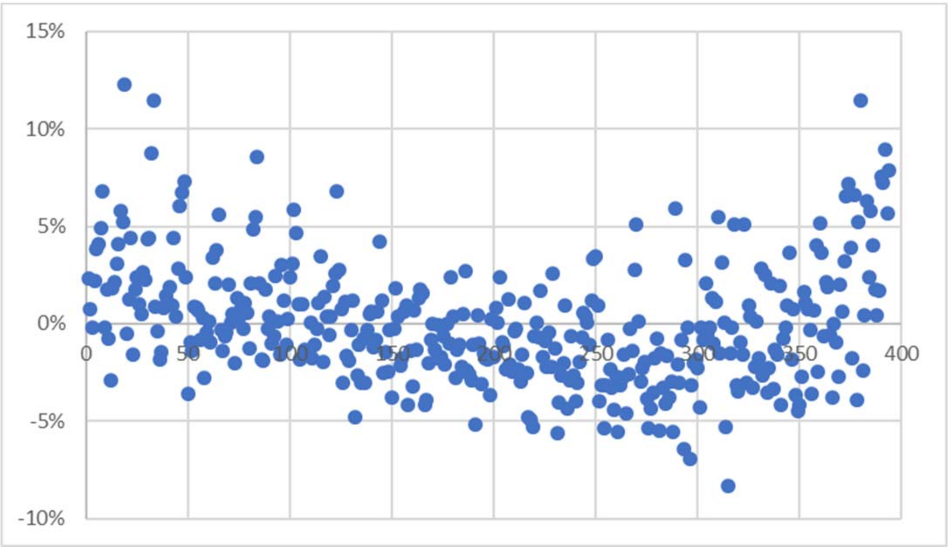




VI. Fixing Estimation Bias with JAEQE

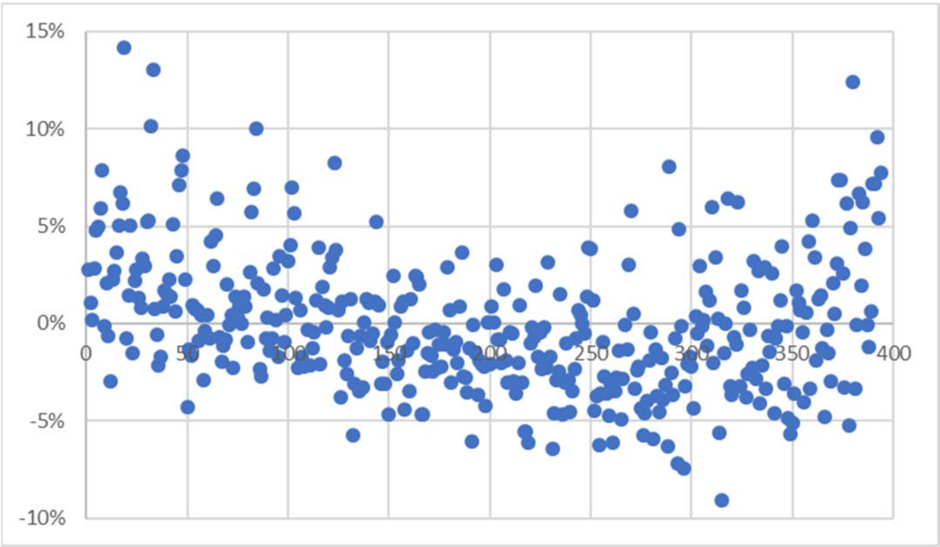
Figure 7a: JAEQE %Deviation of VaR-CLD from True Value by Tail Index Severity Parameter Value
λ=25, GPD Severity (ξ=0.95 θ=5000)
(center)

JAEQE – VaR99.9



ξ : 0.575 – 1.326

JAEQE – VaR99.95



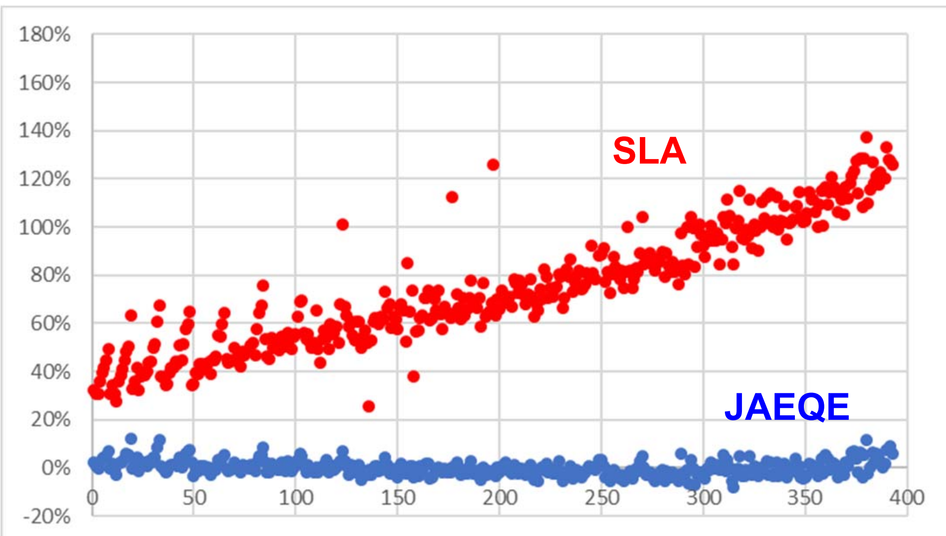
ξ : 0.575 – 1.326

Not perfectly uniform residuals, some curvature in ξ, BUT…

VI. Fixing Estimation Bias with JAEQE

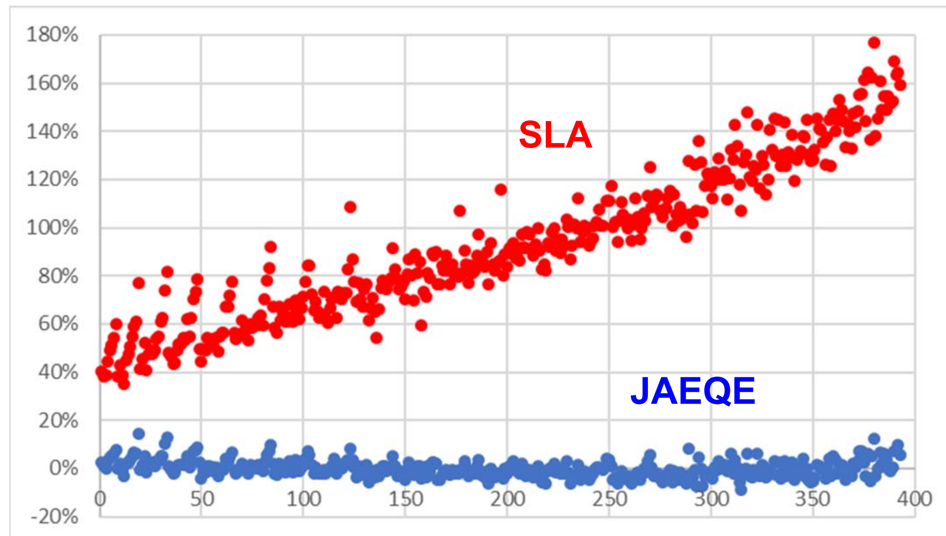
Figure 7b: JAEQE v. SLA*: %Deviation of VaR-CLD from True Value by Tail Index Severity Parameter Value
λ=25, GPD Severity (ξ=0.95 θ=5000)
(SLA Outliers Removed) **(center)**

VaR99.9



ξ : 0.575 – 1.326

VaR99.95



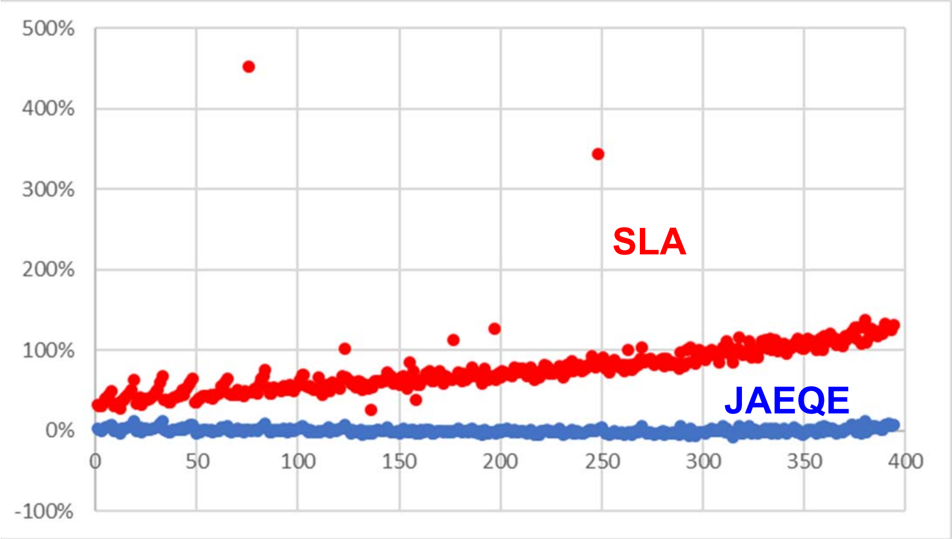
ξ : 0.575 – 1.326

Note that SLA-based VaR bias increases primarily as a function of the tail index parameter, ξ.

VI. Fixing Estimation Bias with JAEQE

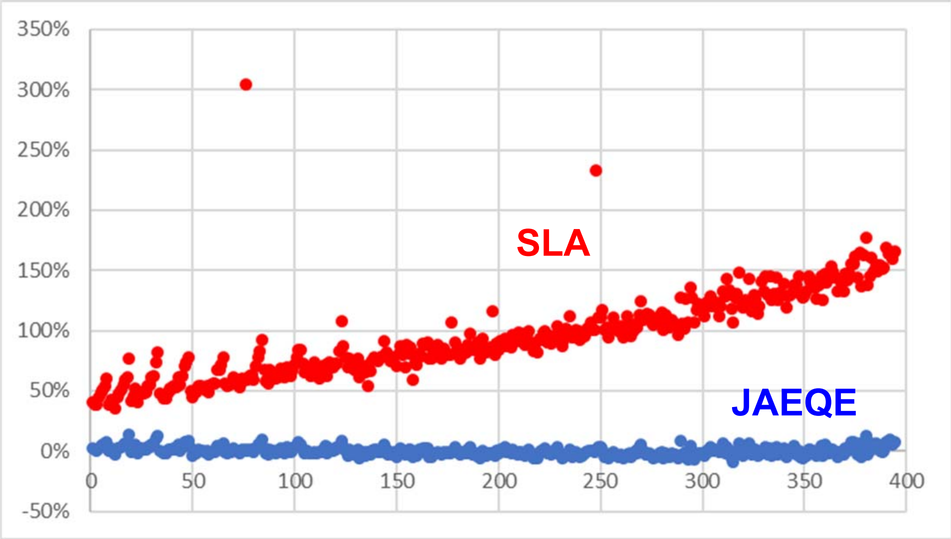
Figure 7c: JAEQE v. SLA: %Deviation of VaR-CLD from True Value by Tail Index Severity Parameter Value
λ=25, GPD Severity (ξ=0.95 θ=5000) (center)
(SLA Outliers NOT Removed – extensive additional evidence exists of SLA's non-robustness)

VaR99.9



ξ : 0.575 – 1.326

VaR99.95



ξ : 0.575 – 1.326

Note that SLA-based VaR bias increases primarily as a function of the tail index parameter, ξ.

VI. Fixing Estimation Bias with JAEQE

JAEQE one sentence summary: Calculate VaR-CLD averages over many samples, for each grid-point of a lattice across a reasonably large range of parameter values (e.g. a 0.99 isodensity), and run a penalized regression of the severity parameter values (polynomials and interactions) on these averages to estimate convexity-induced VaR-CLD bias; then use this regression to predict (and eliminate) the bias due to Jensen's Inequality and obtain a nearly unbiased estimate of VaR-CLD. Why this approach?

- i. **Since VaR-CLD is a convex function of the (severity) parameter(s), it is best to estimate the convexity (so we can discard it) based directly on the parameter values.**
- ii. **The RANGE of parameter values used in estimation should be efficiently defined: this requires knowledge of their joint (finite!) distribution to efficiently define the 0.99 isodensity. The asymptotic joint distributions of the relevant severities (defined in Opdyke (2014) and in APPENDIX 4) can differ materially from their finite counterparts, which also appropriately account for dependence between severity parameters (see APPENDIX 4 and Table 1).**
- iii. **Estimating convexity based on a larger range that that used for prediction helps to guard against overfitting extreme values as well as poor predictions at the edges of the 0.99 isodensity.**
- iv. **Using penalized regression also helps to guard against overfitting, which is a very real trap here because the range of VaR-CLD over the range of relevant severity parameter values changes by ORDERS OF MAGNITUDE! It is not a hard regression model to build and fit due the SHAPE of the response function, which is actually smooth and well behaved, but rather because it spans orders of magnitude in value AND changes very quickly: the rate of change of the gradient continues to increase (i.e. 'accelerate rapidly', vs. maintain 'velocity') towards the edges of the 0.99 isodensity.**
- v. **Given iv., iii. is especially true: it is very easy to fit badly at the edges. But iii. and iv., combined with vi., mitigate it well.**
- vi. **The %diff adjustment of 8. also mitigates difficult predictions on the edge of the 0.99 isodensity.**

VI. Fixing Estimation Bias with JAEQE

vii. Given iv., it is best to define the dependent variable of the regression not on absolute value, but on relative value (i.e. ranging from 0 to 1 where 1=no bias … a semi-log model yields similar results).

viii. Four regularization regression models were tested: LASSO [Tibshirani, 1996], Adaptive LASSO (A-LASSO) [Zou, 2006], Elastic Net [Zou & Hastie, 2005], Scaled Elastic Net [Zou & Hastie, 2005]. Across twelve severity distributions tested [LogNormal, LogGamma, GPD, TruncatedLogNormal, TruncatedLogGamma, TruncatedGPD, LogNormal-Poisson, LogGamma-Poisson, GPD-Poisson, TruncatedLogNormal-Poisson, TruncatedLogGamma-Poisson, TruncatedGPD-Poisson], the ranking in terms of model parsimony, from greatest to least, was LASSO, A-LASSO, Elastic Net, and Scaled Elastic Net. A-LASSO and Elastic Net were very similar, with the former providing a slightly better tradeoff between avoiding overfitting, and thus, oversensitivity to large predictions "on the edges," versus not being able to sufficiently adapt to the order-of-magnitude changes in VaR over the entire range of the 0.99 isodensity.

ix. The approximation methods MISLA (Opdyke, 2014) and PE2 (Hernandez, et al., 2014) also were tested and compared, with no measurable difference in results (although MISLA was orders of magnitude faster, on average).

x. Although by definition a 0.99 isodensity does not cover the entire range of possible parameter values, if the analyst encounters parameter values outside or near the edge of the isodensity, he/she can simply re-center the lattice on the new parameter values and re-estimate the JAEQE VaRs.

xi. JAEQE arguably is computationally demanding (several hours on a modest PC), but not when considering the $ and resources at stake. Also, once the regularization model is estimated, it can be reused in only a few minutes, unless input data changes notably.

VI. Fixing Estimation Bias with JAEQE

Real-World Simulation Study – JAEQE vs. SLA:
40 Cumulative Quarters, Annually Re-estimated, with Annual λ = 25, GPD Severity (ξ = 0.925; θ = 6,000)

1. **Start with 10 Years of Losses**
2. **Add Losses Quarterly, Use Data Cumulatively: Re-estimate A-LASSO Model Every 4 quarters**

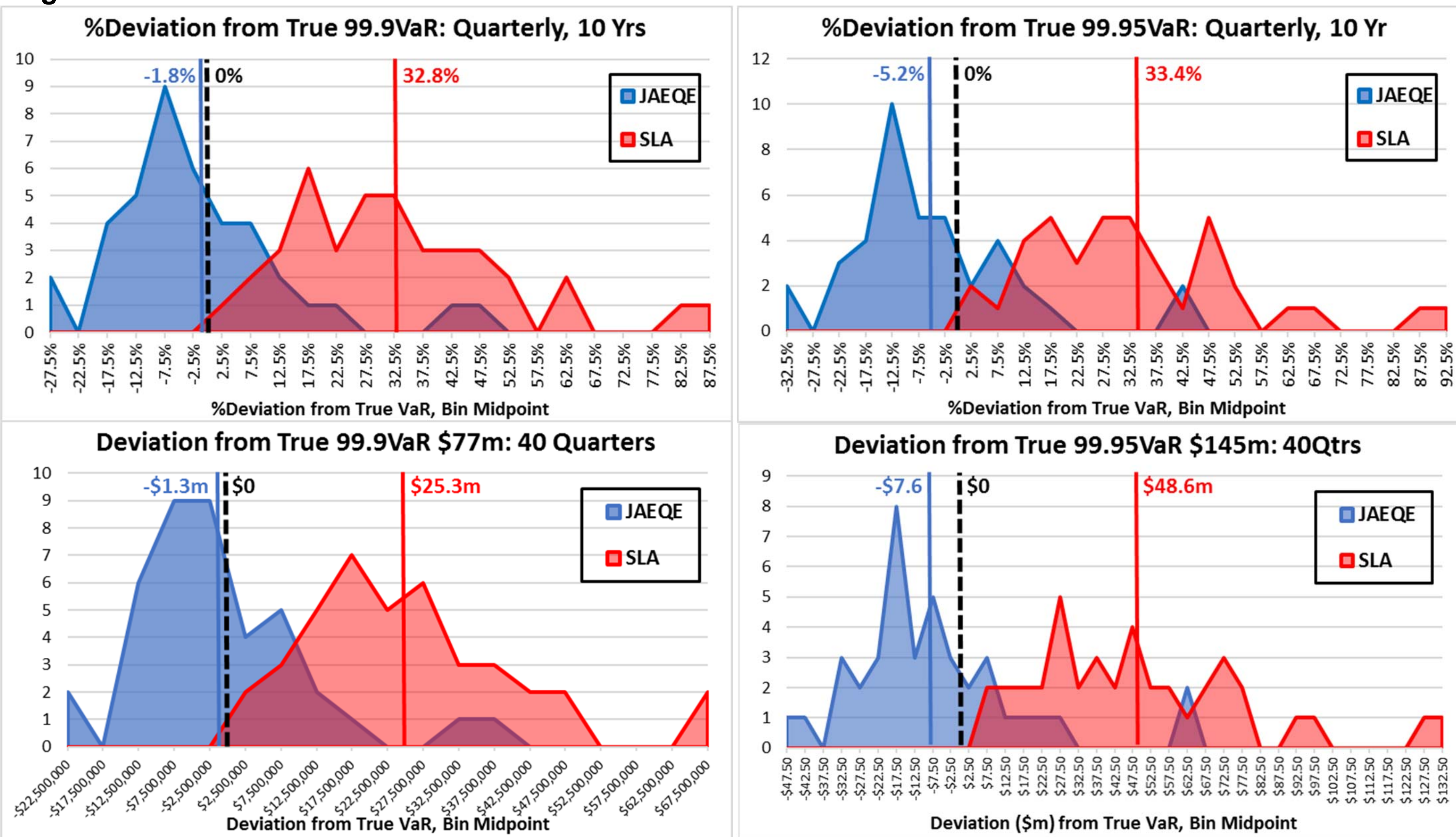
TABLE 3

	+-------- VaR99.9 ($77m) --------+			+-------- VaR99.95 ($145m) --------+		
Absolute Deviation	**SLA**	**JAEQE**	**JAEQE/SLA**	**SLA**	**JAEQE**	**JAEQE/SLA**
Mean	$25,249,348	-$1,387,660	18.20	$48,559,168	-$7,571,148	6.41
Median	$22,358,047	-$3,287,492	6.80	$43,730,374	-$11,914,097	3.67
Q1	$14,983,043	-$7,875,857		$26,998,933	-$20,019,971	
Q3	$33,454,020	$3,567,597		$65,810,239	$1,361,122	
IQR	$18,470,976	$11,443,454	62.0%	$38,811,306	$21,381,093	55.1%
STDEV	$14,903,944	$11,851,928	79.5%	$29,746,828	$23,205,780	78.0%
RMSE	$29,225,056	$11,784,825	**40.3%**	$56,751,606	$24,132,298	**42.5%**
Relative Deviation						
Mean	32.8%	-1.8%		33.4%	-5.2%	
Q1	19.4%	-10.2%		18.6%	-13.8%	
Q3	43.4%	4.6%		45.3%	0.9%	
IQR	24.0%	14.8%		26.7%	14.7%	
STDEV	19.3%	15.4%		20.5%	16.0%	
RMSE	37.9%	15.3%		39.0%	16.6%	

VI. Fixing Estimation Bias with JAEQE

40 Cumulative Quarters, Annually Re-estimated, with Annual λ = 25, GPD Severity (ξ = 0.925; θ = 6,000)

Figure 8



VI. Fixing Estimation Bias with JAEQE

Real-World Simulation Study – JAEQE vs. SLA:
40 Cumulative Quarters, Annually Re-estimated, with Annual λ = 25, GPD Severity (ξ = 0.925; θ = 6,000)

<u>COMMENTS</u>:

1. **These results are compelling: most notable is that RMSE of JAQEQ is well under HALF that of SLA.**
2. **JAEQE's accuracy is also compelling: SLA's bias due to Jensen's Inequality hovers around +33%, which his very material, while JAEQE essentially eliminates this bias altogether.**
3. **Since the above simulation was run, I have had time to re-run this with other parameter values and other severity distributions.**
4. **While the Means for JAEQE above are not atypical, they CAN be much bigger, e.g. well over 20% deviation from true VaR.**
5. **This is due in part because this is not a full simulation study, but rather, only 10-20 years worth of loss data, by design, to replicate the reality facing many financial institutions. Larger numbers of simulations would improve the mean performance of JAEQE and widen the gap even further vs. SLA based on not only accuracy, but also precision, RMSE, and robustness.**
6. **However, the point is less that JAEQE can exhibit notable variance for a specific, small-scale simulation study, than that it will ALWAYS notably outperform, on a relative basis, the alternative of failing to directly address Jensen's Inequality-induced VaR bias.**

VII. …But What About Model Error?! "Don't be sad…"

1. To date, even the best Statistical Goodness-of-Fit (GoF's) Tests have notoriously low statistical power for moderately sized samples, especially for heavy-tailed distributions; in other words, the tests, when used to select the True distribution that generated the data, get it wrong very often. We need better GoF's to address Model Error, which in this case, is the selection of the wrong severity distribution when samples are of moderate size.
2. Until then, we must take solace in the timeless words of the immortal sage Meat Loaf: "Don't be sad, 'cuz 2 out of 3 ain't bad…" For now, we must 'know' the right severity distribution…
3. And in the interim, use JAEQE, which as we have seen above in Figure 8, is transformational:

BEFORE Implementing JAEQE…

AFTER Implementing JAEQE

VIII. Summary and Conclusions

- **For both Compound Loss Distributions (frequency+severity) AND Single Loss Distributions (only severity, constant sample size), VaR is a convex function of the severity parameter(s) under the following conditions: when samples are small-to-moderately sized, severity distributions are heavy-tailed, and extreme quantiles (VaR99.5+) must be estimated.**
- **This convexity means that when VaR is estimated BASED on the parameters, as it must be under these conditions, its estimate will be biased upwards (often dramatically) due to Jensen's Inequality.**
- **This is true regardless of the parameter estimators used (only Robust estimators provide very partial mitigation of this convexity-induced VaR-bias (see Opdyke, 2012)).**
- **Failure to separately treat the 3 distinct sources of error here – Approximation Error, Estimation Error, and Model Error – makes solving this convexity-induced VaR bias difficult, if not impossible.**
- **But Approximation Error is solved by Opdyke's (2017) MISLA.**
- **And Estimation Error, the greatest source of which in this setting is convexity-induced VaR bias, is solved herein with the Jensen-Adjusted Extreme Quantile Estimator. JAEQE directly estimates VaR's convexity using regularization regression, and then shrinks it away.**
- **JAEQE is a very general solution that works on virtually any severity distribution. No other bias-reduction methods in the extant literature (shifted and otherwise manipulated bootstraps, various 'expansions', etc.) work effectively under the above-defined conditions (which are the dominant reality for many firms in many industries).**
- **The improved accuracy, precision, RMSE, and robustness of JAEQE over SLA, the most widely used alternate method, are notable (RMSE is cut by more than half).**
- **Further mitigating Model Error remains as an important area of continued research under these conditions.**

IX. Next Steps

1. **Develop better GoF's** to more effectively address model error for heavy-tailed distributions under finite sample conditions. This is hard and currently an active topic of applied research.
2. **JAEQE is a very general approach that works with virtually all distributions.** For the simpler distributions (e.g. LogNormal and GPD) I HAVE used analytical solutions to estimate VaR accounting for convexity but they involve recursive traces of Fisher information matrices that very quickly become both analytically and empirically intractable for more complicated distributions (especially truncated distributions), even when using symbolic programming platforms like Mathematica, Maple, and mathStatica. **Analytical solutions to this problem that are useable across a wide range of severity distributions would be a significant advance** and strong contribution to the literature.
3. **Using covariates to estimate (severity) parameters in GAMLSS regression** will kill 2 birds with one stone: most importantly, this is the only way to scientifically and defensibly provide "KRI" levers for active, direct risk mitigation and management; secondly, this will reduce variance in parameter (and thus VaR) estimation, all else equal, to achieve more precise risk measurement and consequently, more effective risk management. This can be done within the JAEQE approach (see APPENDIX 3).
4. **Under the CLD model, we must properly account for dependence between severity and frequency distributions**, which undoubtedly exists in most real-world settings. Stahl (2017) is a great start on this front, and there appears to be no conflict with applying this method and JAEQE simultaneously.
5. **We must right-size alpha (the VaR level) in financial settings based on statistics and science as opposed to political considerations.** Protecting against a 1-in-1,000 year FINANCIAL loss is absurd on its face (naturally occurring phenomena are another matter). How many companies, let alone sovereign nations, have existed continuously for 200 years (VaR99.5)? 1,000 years (VaR99.9)? Back to around the time that Ghengis Khan roamed the Eurasian Steppe?! And Economic Capital typically is much LARGER than Regulatory Capital, starting at least at VaR99.95! This brings us back to the Roman Empire and Biblical times!!! Does that really make any sense?! (see APPENDIX 1).

IX. Next Steps

Contact Information:

J.D. Opdyke
VP – Financial Risk and Measurement, ERRM Division
John.Opdyke@Allstate.com
617-943-6463

X. References

- Abdymomunov, A., Curti, F., and Kane, H., (2019), "Calculate Tail Quantiles of Compound Distributions," *The Journal of Operational Risk*, Vol. 22, No. 5, 41-70.
- Bardou, O., Frikha, N., and Pages, G. (2010), "CVaR Hedging Using Quantization Based Stochastic Approximation," working paper.
- Ben-Tal, A. (2007) "An Old-New Concept of Convex Risk Measures: The Optimized Certainty Equivalent," *Mathematical Finance*, 17(3), 449-476.
- Brown, D., (2007) "Large Deviations Bounds for Estimating Conditional Value-at-Risk," *Operations Research Letters*, 35, 722-730.
- Degen, M. (2010), "The Calculation of Minimum Regulatory Capital Using Single-Loss Approximations," *The Journal of Operational Risk*, Vol. 5, No. 4, 3-17.
- Degen, M., Embrechts, P., Lambrigger, D.D. (2007), "The Quantitative Modeling of Operational Risk: Between g-and-h and EVT," *Astin Bulletin*, 37(2), 265-291.
- Embrechs, P., & Frei, M., (2009), "Panjer Recursion versus FFT for Compound Distributions," *Mathematical Methods of Operations Research*, 69(3), 497-508.
- Ergashev B., (2008), "Should Risk Managers Rely on the Maximum Likelihood Estimation Method While Quantifying Operational Risk," *The Journal of Operational Risk*, Vol. 3, No. 2, 63-86.
- Guégan, D., Hassani, B., and Li, K., (2017) "Measuring risks in the extreme tail: The extreme VaR and its confidence interval," *Risk and Decision Analysis*, Vol. 6, No. 3, pp. 213-224.
- Hernandez, L., Tejero, J., Suarez, A., and Carrillo-Menedez, S., (2014), "Closed-form Approximations for Operational Value-at-Risk," *Journal of Operational Risk*, 8(4), 39-54.
- 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.
- Kim, J., & Hardy, M., (2007), "Quantifying and Correcting the Bias in Estimated Risk Measures," *Astin Bulletin*, Vol. 37(2), 365-386.
- Larsen, P. (2015), unpublished paper, "Operational Risk Models and Maximum Likelihood Estimation Error for Small Sample-Sizes"
- Opdyke, J.D., (2012), "Better Capital Planning via Exact Sensitivity Analysis Using the Influence Function," American Bankers Association Operational Risk Modeling Forum, Washington, DC, July 18-20.
- Opdyke, J.D., & Cavallo, A. (2012a), "Estimating Operational Risk Capital: The Challenges of Truncation, The Hazards of Maximum Likelihood Estimation, and the Promise of Robust Statistics," *The Journal of Operational Risk*, Vol. 7, No. 3, 3-90.
- Opdyke, J.D. & Cavallo, A. (2012b), "Operational Risk Capital Estimation and Planning: Exact Sensitivity Analysis and Business Decision Making Using the Influence Function," in Operational Risk: New Frontiers Explored, ed. E. Davis, Risk Books, Incisive Media, London.
- Opdyke, J.D. (2014), "Estimating Operational Risk Capital with Greater Accuracy, Precision, and Robustness," *The Journal of Operational Risk*, Vol. 9, No. 4, 3-78.
- Opdyke, J.D. (2015), "," invited presentation, Quant Summit USA, , July 18-20, 2012.
- Opdyke, J.D., (2017), "Fast, Accurate, Straightforward Extreme Quantiles of Compound Loss Distributions," *The Journal of Operational Risk*, Vol. 12, No. 4, 1-30.
- Rigby RA, Stasinopoulos DM (2001). "The GAMLSS project: a Flexible Approach to Statistical Modelling." In B Klein, L Korsholm (eds.), New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling, pp. 249-256. Odense, Denmark.
- Roehr, A., (2002), "Modelling Operational Losses", *Algo Research Quarterly*, Vol.5, No.2, Summer 2002.
- Schaefer, M., (1976), "Note on the k-dimensional Jensen Inequality," *The Annals of Probability*, 4(3), 502-504.
- Shevchenko, P., (2011), Modelling Operational Risk Using Bayesian Inference, Springer, New York, New York.
- Smith, J. (1987), "Estimating the Upper Tail of Flood Frequency Distributions," *Water Resources Research*, Vol.23, No.8, 1657-1666.
- Stahl, D., 2016, "Operational Loss with Correlated Frequency and Severity: an Analytical Approach," *The Journal of Operational Risk*, Vol. 11, No. 2, June.
- Taleb, N. & Douady, R., (2014),"Mathematical Definition, Mapping, and Detection of (Anti)Fragility," 2014. ffhal-01151340.
- Taleb, N. & Douady, R., (2015), "On the Super-Additivity and Estimation Biases of Quantile Contributions," Physica A: Statistical Mechanics and its Applications 429, 252-260.
- Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, Series B*, 58:267–288.
- Zou, H. (2006), "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101:1418–1429.
- Zou, H., & Hastie, T. (2005), "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society, Series B*, 67:301–320.

Appendix 1 – Right-sizing Alpha in Finance

We must note that the statistics of extreme quantile estimation correspond remarkably well with common sense here. So we must again ask the questions so often asked over the past dozen+ years: "Is it really meaningful to be guarding against a 1-in-1000 year loss? A loss that would only occur, on average, slightly less than once since the time of Gengis Khan? A time before most (all?) of today's sovereign nations, let alone companies, even existed?? Or worse still, for Economic Capital rather than Regulatory Capital, a 1-in-2000 year loss corresponding to VaR99.95?! Back to the Roman Empire and biblical times??!!!"**

"Might a 1-in-200 year loss (99.5%tile, per Solvency II), or even a 1-in-100 year loss (99%tile), be a more realistic, meaningful, and appropriate threshold?"

"Can all the requests for this more meaningful level of capital over the past dozen years really be considered intentional 'gaming' of regulations, motivated solely by unadulterated corporate self-interest?"

It is now 15 years since Basel II was published (2004), and a dozen years since the (US) Joint Final Rule: Risk-Based Capital Standards: Advanced Capital Adequacy Framework–Basel II, November 2, 2007 – Docket Number R-1261, was published, and explicitly identified an 'alpha' ('significance level') of 99.9%tile (with a probability densify function diagram, p. 69291).

**Only now does it appear that the above apoplexy is finally being acknowledged as appropriate (see Risk.net, 9/1/19: "Measuring 1-in-1000 Year Loss Events 'Unrealistic', Researchers Say" A. Campbell). The difference between 99%tile and 99.9%tile is NOT 0.9% – it is 10x larger! An order of magnitude!! This non-linearity was somehow missed in the abovementioned regulatory promulgation, causing untold misdirected resources and misguided risk measurement and risk management efforts.

Appendix 1 – Right-sizing Alpha in Finance

Back to ME??! VaR99.9?
REALLY?!!
Ok, get the presses rolling
(thank the hordes we invented
paper money… I wonder what the
Europeans are gonna do… Is
Basel even part of Europe??)

Puh-leeze! The man
said ECONOMIC
Capital, Not Regulatory
Capital … We're talkin'
back to ME!
VaR99.95+!!! 1-in-2000
Years! (…on average)

Appendix 2 – Confusion Re: Jensen (Larsen, 2015)

- **Unfortunately, there are misstatements and confusion regarding Jensen's Inequality in this setting in an unpublished paper (see Larsen, 2015):**
- **Fn[3] "This mean bias is a central object of study in Opdyke and Cavallo (2012), where they claim that MLE results in capital overestimation. The meaning of this statistic for modeling decisions, however, is not completely clear. … Opdyke and Cavallo (2012) write that the mean OpVaR bias is a consequence of Jensen's inequality, but no further details are given. This would follow if the CDF $F(x|\theta)$ for a heavy-tailed distribution were a convex function. There is no mention whether convexity is with respect to the loss variable x or with respect to the parameters *θ*. For the Jensen's inequality argument of Opdyke and Cavallo (2012) to be valid, convexity must be shown with respect to the parameters *θ*, not the loss amount x.[fn3] Specifically, we would have to show that, for all loss amounts x in a neighborhood of the true OpVaR, the Hessian of $F(x|\theta)$ with respect to *θ* is negative definite (and hence the Hessian of the quantile function of $F(x|\theta)$ would be positive definite). This property is trivial to verify for the Pareto distribution considered here as depending only on one variable, but is less than straightforward for more complicated distributions. That there is still something to prove before invoking Jensen's inequality is mentioned in a subsequent paper (Opdyke, 2014)."**
- **In fact, on page 68, Opdyke and Cavallo (2012) do explicitly state that the convexity of the quantile function is with respect to the estimated severity parameters – no mention is made regarding 'x,' the data values themselves: "This is illustrated in Figure 20 (from Kennedy (1992, p. 37)). This applies to quantile estimation of all commonly used severity distributions: if β is a random variable (here, our severity distribution parameter estimates) and $g(\cdot)$is a (strictly) convex function (here, the inverse of our severity distribution CDF), then $g\left(E\left[\hat{\beta}\right]\right) < E\left[g\left(\hat{\beta}\right)\right]$, and our quantile estimate (capital estimate) is biased upward." In the later paper Larsen (2015) references, Opdyke (2014) makes essentially the same statement on page 12, again with no mention of VaR convexity with respect to 'x,' the loss data itself: "…under these conditions, VaR appears to always be a convex function, like $g(\cdot)$, of the parameters of the severity distribution, which here is the vector β (we can visualize β as a single parameter without loss of generality as the multivariate case for Jensen's inequality is well established (see Schaefer 1976)). Consequently, the capital estimation, $\hat{V} = g\left(\hat{\beta}\right)$ will be biased upward."**
- **Larsen's (2015) comments are not only incorrect as a factual matter, but also misguided. In footnote 3 he examines potential convexity of VaR with respect to "x," the variable representing the size of the loss events. But these are not being ESTIMATED – they are the data points themselves! Jensen's inequality is fundamentally about ESTIMATION, not data per se, so the point of the footnote is incomprehensible. We encourage (re)reading Opdyke and Cavallo (2012a) and Opdyke (2014) above to avoid any confusion regarding the relevance Jensen's inequality in this setting.**

Appendix 2 – Confusion Re: Bias (Larsen, 2015)

a. The 2nd confusion in Larsen (2015), this time regarding bias, is addressed below.

b. It is critical to note here that even though extreme quantile (capital) estimates will be, on average, high 50% of the time and low 50% of the time even under Jensen's inequality (see Figure 3), the AMOUNTS that they are high vs. low are very different: when high, they are often MUCH HIGHER than the true quantile (capital) because the distribution is positively skewed, but when low, they often are NOT MUCH LOWER than true capital. Would you/your bank bet on a nickel gain vs. a dollar loss with equal probability?! If you were to use the median rather than the mean here vis-à-vis Larsen's so-called 'median-bias,' that is what you would be doing.

c. When comparing quantile (capital) estimates to true quantile (capital) values, probability alone is not sufficient here – the absolute DISTANCE from true values matters too, if not predominantly. But quantiles, like the median, ignore 'distance from truth' by design, while the mean (expected value) does not. This is why the mean is more appropriate in this setting of highly skewed VaR distributions.

d. This also is why 'bias' has been defined for well over half a century with respect to the mean and not other measures of central tendency (like the median). Larsen's proposed "median bias" metric does not work in this setting based on b.: it is essentially an irrelevant and obfuscatory artifice to the extent that it distracts from the central and primary role that Jensen's Inequality plays in VaR estimation in this setting.

Appendix 3 – Parameter Est. using KRIs/Covariates

Key Risk Indicator (KRI) Data

Econometric Methods for Establishing Direct, Material, Statistically Causal Relationships between KRIs and i. Capital, ii. Loss Frequency, and iii. Loss Severity

KRI Data can (and should!) be used directly in capital modeling. **Establishing material, statistically causal relationships between KRIs and capital is the only way that operational risk management and mitigation efforts can have direct and desired effects on capital requirements.**

For example, this gives the operational risk capital analyst the means by which to make statements to, say, the head of the trading shop such as, "If you can decrease your system downtime by a standard deviation, or X%, I can take $40m in capital off the table for you, all else equal."

This is accomplished using multivariate econometric (regression) techniques to estimate frequency and severity parameters based directly on the KRI Data. This is directly analogous to knowing the drivers of, say, a PD model when estimating capital for credit risk.

Appendix 3 – Parameter Est. using KRIs/Covariates

Econometric Methods for Establishing Direct, Material, Statistically Causal Relationships between KRIs and i. Capital, ii. Loss Frequency, and iii. Loss Severity

WHY Multivariate Regression?

Multivariate regression is needed to control for covariance betwixt the KRI covariates. **Multivariate regression is the only way to estimate the effect of an independent variable (a particular KRI) on a dependent variable (capital) holding all else constant**, that is, without capturing the effects of other KRIs that to some degree move in tandem with the one in question.

Without a regression to "hold all else constant" and eliminate the confounding effect of, say trading volume, when estimating the effect of system downtime on operational risk capital, the estimate of the effect of system downtime will be biased, and inference based on it will be misinformed, and the mitigation efforts based on it will be misguided and likely ineffective.

Appendix 3 – Parameter Est. using KRIs/Covariates

Econometric Methods for Establishing Direct, Material, Statistically Causal Relationships between KRIs and i. Capital, ii. Loss Frequency, and iii. Loss Severity

WHY Multivariate Regression?

This, of course, presumes that relationships (covariance) exist betwixt relevant KRIs, as it does in the real world (if it did not, there would be no need for multivariate regression here).

Multivariate regression also increases the precision with which we are able to estimate the frequency and severity parameters. We are using additional data in the estimation, which **will increase statistical power** (even though we are not increasing sample size in the form of additional loss events).

Appendix 3 – Parameter Est. using KRIs/Covariates

Econometric Methods for Establishing Direct, Material, Statistically Causal Relationships between KRIs and i. Capital, ii. Loss Frequency, and iii. Loss Severity

WHY Multivariate Regression? ONLY this approach provides

1. **Statistically Causal Relationships between KRIs and Capital, AND KRIs and LOSS FREQUENCY AND SEVERITY (… NOT JUST CAPITAL!!!)**
2. **Magnitude of Effect of Each KRI** on i. Capital AND ii. LOSS FREQUENCY AND SEVERITY Independent of other KRIs
3. **RELATIVE IMPORTANCE of Each KRI's Effect** on i. Capital AND ii. LOSS FREQUENCY AND SEVERITY Independent of other KRIs (key for $allocation for mitigation efforts)
4. **Direction of Effect of Each KRI** on i. Capital AND ii. LOSS FREQUENCY AND SEVERITY Independent of other KRIs
5. **Whether Effect of Each KRI is Statistically Significant** vis-à-vis i. Capital AND ii. LOSS FREQUENCY AND SEVERITY Independent of other KRIs
6. **Whether Effect of Each KRI is Material** vis-à-vis i. Capital AND ii. LOSS FREQUENCY AND SEVERITY Independent of other KRIs
7. **Increase in the Precision of the Estimate of Capital** (AND Frequency and Severity), all else equal

Appendix 3 – Parameter Est. using KRIs/Covariates

Econometric Methods for Establishing Direct, Material, Statistically Causal Relationships between KRIs and i. Capital, ii. Loss Frequency, and iii. Loss Severity

REQUIREMENTS

KRI Data for modeling purposes must be disaggregated at the level of the loss event. In other words, it must be "**granular**," with data points for each KRI collected associated with each individual loss (or timing that concurs with the loss).

This is distinct from what many (non-modelers) refer to as "KRIs," which are typically highly aggregated, descriptive statistics that are tracked over time and used to guide operational risk management and mitigation efforts directly, rather than via an estimation process that links them to capital (or some other outcome measure). **Aggregated KRIs typically are used non-inferentially, to identify "Red Lights," "Amber Lights," and "Green Lights."**

Appendix 3 – Parameter Est. using KRIs/Covariates

Econometric Methods for Establishing Direct, Material, Statistically Causal Relationships between KRIs and i. Capital, ii. Loss Frequency, and iii. Loss Severity

METHODS

Frequency: Poisson and Negative Binomial Regression
- Time tested, decades old methods applied in many fields.
- However, doesn't move the capital needle nearly as much as severity.

Severity: Scale regression
- More recent, main difference is just the link function.
- DOES move the capital needle, sometimes dramatically.
- This is a Scale Regression, and so the Severity requires a scale parameter.

GAMLSS (Generalized Additive Models of Location, Scale, and Shape) Regression:
- Most general, covariates apply to location, scale, and shape parameters.
- In literature and applied use at least as long as Operational Risk has been a discipline (see Rigby and Stasinopoulos, 2001).

Appendix 3 – Parameter Est. using KRIs/Covariates

Econometric Methods for Establishing Direct, Material, Statistically Causal Relationships between KRIs and i. Capital, ii. Loss Frequency, and iii. Loss Severity

METHODS

Frequency: Poisson and Negative Binomial Regression

$$\ln\left(E\left[Y \mid x\right]\right) = \boldsymbol{\beta}'\mathbf{x} \quad \text{where } \mathbf{x} \text{ is a vector of regressor variables.}$$

$$E\left[Y \mid x\right] = \exp\left(\boldsymbol{\beta}'\mathbf{x}\right) = \lambda \qquad f_X\left(x \mid \lambda\right) = \frac{\lambda^x \exp\left(-\lambda\right)}{x!} \qquad f_Y\left(y \mid \mathbf{x};\boldsymbol{\beta}\right) = \frac{\left(\boldsymbol{\beta}'\mathbf{x}\right)^y \exp\left(-\boldsymbol{\beta}'\mathbf{x}\right)}{y!}$$

$$L\left(\boldsymbol{\beta} \mid X,Y\right) = \prod_{i=1}^{n} \frac{\left(\beta' x_i\right)^{y_i} \exp\left(-\beta' x_i\right)}{y_i!}$$

$$l\left(\boldsymbol{\beta} \mid X,Y\right) = \ln\left[L\left(\boldsymbol{\beta} \mid X,Y\right)\right] = \sum_{i=1}^{n}\left(y_i \ln\left(\beta' x_i\right) - \beta' x_i - \ln\left(y_i!\right)\right)$$

$$l\left(\boldsymbol{\beta} \mid X,Y\right) = \sum_{i=1}^{n}\left(y_i \ln\left(\beta' x_i\right) - \beta' x_i\right) \qquad \frac{\partial l\left(\boldsymbol{\beta} \mid X,Y\right)}{\partial \boldsymbol{\beta}} = 0, \text{ no closed form solution,}$$

QuantMinds Americas Co-located with RiskMinds Americas

RiskMinds Americas Co-located with QuantMinds Americas

Appendix 3 – Parameter Est. using KRIs/Covariates

Econometric Methods for Establishing Direct, Material, Statistically Causal Relationships between KRIs and i. Capital, ii. Loss Frequency, and iii. Loss Severity

METHODS

Severity: Scale Regression

$Y \sim \Im(\theta, \Omega)$ such that θ is affected by the regressors as

$$\theta = \theta_0 \cdot \exp\left(\sum_{i=1}^{k} \beta_i x_i \right)$$

where θ_0 is the base value of the scale parameter,

$\Im$ is the distribution of Y with nonscale parameters Ω and scale parameter θ

and x_i are k regressors and β_i are the corresponding parameters.

θ is a scale parameter iff $f\left(x;\theta,\beta\right) = \frac{1}{\theta} f\left(\frac{x}{\theta};1,\beta\right)$ and $F\left(x;\theta,\beta\right) = F\left(\frac{x}{\theta};1,\beta\right)$.

Appendix 3 – Parameter Est. using KRIs/Covariates

Econometric Methods for Establishing Direct, Material, Statistically Causal Relationships between KRIs and i. Capital, ii. Loss Frequency, and iii. Loss Severity

METHODS

GAMLSS Regression

if $Y_i \sim f\left(y_i;\mu_i,\sigma_i,\tau_i\right)$; $i=1,\ldots,N$; and X_{ikj_k} are j_k covariates; $k=1,\ldots,p$ parameters;

$g_k\left(\theta_k\right)=\eta_k=h_k\left(X_k,\beta_k\right)$ and μ_i,σ_i, and τ_i are location, scale, and shape parameters, θ_k

$$g_1\left(\mu\right)=\eta_1=h_1\left(X_1,\beta_1\right)=\beta_{11}+\beta_{12}X_{i12}+\cdots+\beta_{1j_1}X_{i1j_1}$$

$$g_2\left(\sigma\right)=\eta_2=h_2\left(X_2,\beta_2\right)=\beta_{21}+\beta_{22}X_{i22}+\cdots+\beta_{2j_2}X_{i2j_2}$$

$$g_3\left(\tau\right)=\eta_3=h_3\left(X_3,\beta_3\right)=\beta_{31}+\beta_{32}X_{i32}+\cdots+\beta_{3j_3}X_{i3j_3}$$

$$\hat{\theta}_k=\underset{\theta_k\in\Theta}{\arg\max}\left(\sum_{i=1}^{N}\log\left[f\left(y_i\mid\mu_i,\sigma_i,\tau_i\right)\right]\right)$$

for the parametric version, and a penalized log likelihood for the semi-parametric version.

GAMLSS can include both linear & non-linear effects.

Appendix 4 – Distributional Characteristics

- **PDF and CDF of LogNormal:**

$$f\left(x;\mu,\sigma\right)=\frac{1}{\sqrt{2\pi}\sigma x}e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} \quad F\left(x;\mu,\sigma\right)=\frac{1}{2}\left[1+erf\left(\frac{\ln(x)-\mu}{\sqrt{2\sigma^2}}\right)\right] \quad 0<x<\infty,\ 0<\sigma<\infty$$

- **Mean of LogNormal:** $E\left(X\right)=e^{\left(\mu+\sigma^2/2\right)}$

- **Inverse Fisher information of LogNormal:**

$$A(\theta)^{-1}=\begin{bmatrix}\sigma^2 & 0\\ 0 & \sigma^2/2\end{bmatrix}$$

Appendix 4 – Distributional Characteristics

- **PDF and CDF of Truncated LogNormal:**

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

$H < x < \infty,\ 0 < \sigma < \infty$

$f(\ )$ is LogNormal PDF and $F(\ )$ is LogNormal CDF

- **Mean of Truncated LogNormal:**

$$E(X)=e^{\mu+\sigma^2/2}\cdot\Phi\left(\frac{\mu+\sigma^2-\ln(H)}{\sigma}\right)\cdot\frac{1}{\left[1-F(H)\right]}$$

where $\Phi(\ )$ is the standard normal CDF.

- **Inverse Fisher information of Truncated LogNormal:**

Let $u=\frac{\ln(H)-\mu}{\sigma}$, $j=\frac{-u^2/2}{\sqrt{2\pi}}$, $J=\frac{j}{1-\Phi(u)}$, where Φ = CDF of Standard Normal, and $INV=\frac{\sigma^2}{\left[2+J\cdot(J-u)\cdot(u\cdot(J-u)-3)\right]}$

Then $$A(\theta)^{-1}=INV\cdot\begin{bmatrix} 2+J\cdot u\cdot(1-u\cdot(J-u)) & J\cdot(u\cdot(J-u)-1) \\ J\cdot(u\cdot(J-u)-1) & 1-(J\cdot(J-u)) \end{bmatrix}$$

From Roehr (2002). Note that the first cell of this matrix as presented in Roehr, 2002, contains a typo: this is corrected in the presentation above.

Appendix 4 – Distributional Characteristics

- **PDF and CDF of Generalized Pareto Distribution (GPD):**

$$f(x;\xi,\theta)=\frac{1}{\theta}\left[1+\xi\frac{x}{\theta}\right]^{\left[-\frac{1}{\xi}-1\right]} \qquad F(x;\xi,\theta)=1-\left[1+\xi\frac{x}{\theta}\right]^{\left[-\frac{1}{\xi}\right]} \qquad \text{assuming } \xi\geq 0, \text{ for } 0\leq x<\infty;\ 0<\theta<\infty$$

- **Mean of GPD:** $E(X)=\dfrac{\theta}{1-\xi}$ for $\xi<1$ $\left(=\infty \text{ for } \xi\geq 1\right)$

- **Inverse Fisher information of GPD:**

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

from Smith (1987)

- **Tail Index =** ξ

Appendix 4 – Distributional Characteristics

- **PDF and CDF of Truncated GPD:**

$$g(x;\xi,\theta)=\frac{f(x;\xi,\theta)}{1-F(H;\xi,\theta)} \qquad G(x;\xi,\theta)=1-\frac{1-F(x;\xi,\theta)}{1-F(H;\xi,\theta)}$$

assuming $\xi \geq 0$, for $H \leq x < \infty;\ 0<\theta<\infty$

$f(\ )$ is GPD PDF and $F(\ )$ is GPD CDF

- **Mean of Truncated GPD:** $E(X)=\frac{\theta}{\xi}\cdot\left(\frac{\left[1-F(H)\right]^{-\xi}}{1-\xi}-1\right)$ for $\xi<1$ $(=\infty$ for $\xi\geq 1)$

- **Inverse Fisher information of Truncated GPD:**

$$A(\theta)^{-1}=(1+\xi)\cdot\begin{bmatrix}(1+\xi) & -\theta\left(1+(1+2\xi)\left(\frac{H}{\theta}\right)\right)\\ -\theta\left(1+(1+2\xi)\left(\frac{H}{\theta}\right)\right) & \theta^2\left(2+2(1+2\xi)\left(\frac{H}{\theta}\right)+(1+\xi)(1+2\xi)\left(\frac{H}{\theta}\right)^2\right)\end{bmatrix}$$

from Roehr (2002)

- **Tail Index =** ξ

Appendix 4 – Distributional Characteristics

- **PDF and CDF of LogGamma*:**

$$f(x;a,b)=\frac{b^a\left(\log(x)\right)^{(a-1)}}{\Gamma(a)x^{b+1}} \qquad F(x;a,b)=\int_1^x \frac{b^a\left(\log(y)\right)^{(a-1)}}{\Gamma(a)y^{b+1}}dy$$

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

where $\Gamma(a)$ is the complete gamma function

- **Mean of LogGamma:** $E(X)=\left(\frac{b}{b-1}\right)^a$ for $b>1$; otherwise $E(X)=\infty$

- **Inverse Fisher information of LogGamma:**

$$A(\theta)^{-1}=\frac{1}{\left(a/b^2\right)\cdot trigamma(a)-1/b^2}\begin{bmatrix} a/b^2 & 1/b \\ 1/b & trigamma(a) \end{bmatrix}$$

from Opdyke and Cavallo (2012a)

- **Tail Index =** $\frac{1}{b}$

***NOTE that a location parameter can be added to change the lower end of the domain to zero, but this is unnecessary in this setting. Also note that this is the "rate" or "inverse scale" parameterization of the LogGamma, which can also be defined with a "scale" parameterization wherein b = 1/b.**

Appendix 4 – Distributional Characteristics

- **PDF and CDF of Truncated LogGamma*:**

$$g(x;a,b) = \frac{f(x;a,b)}{1-F(H;a,b)} \qquad G(x;a,b) = 1 - \frac{1-F(x;a,b)}{1-F(H;a,b)}$$

$H \leq x < \infty;\ 0 < a;\ 0 < b$

$f(\ )$ is GPD PDF and $F(\ )$ is GPD CDF

- **Mean of Truncated LogGamma:**

$$E(X) = \left(\frac{b}{b-1}\right)^a \cdot \frac{1-J\big(\log(H)(b-1);a,1\big)}{\left[1-F(H)\right]} \text{ for } b>1, \text{ otherwise } E(X)=\infty$$

where $J(\ )$ is the CDF of the Gamma distribution. Alternately,

$$E(X) = \left(\frac{b}{b-1}\right)^a \frac{1-F(H;a,b-1)}{\left[1-F(H;a,b)\right]},\ b>1$$

from Opdyke (2017)

- **Tail Index =** $\frac{1}{b}$

- **Inverse Fisher information of Truncated LogGamma:**

QuantMinds Americas Co-located with RiskMinds Americas

RiskMinds Americas Co-located with QuantMinds Americas

Appendix 4 – Distributional Characteristics

- **Inverse Fisher info. of Truncated LogGamma*:** $A(\theta)^{-1} = \begin{bmatrix} A & B \\ B & D \end{bmatrix}^{-1}$ where

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

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

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

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

from Opdyke and Cavallo (2012b)

***The digamma and trigamma functions are the first and second order logarithmic derivatives of the complete gamma function:**
$\text{digamma}(z) = \partial/\partial z \ln\left[\Gamma(z)\right]$ and $\text{trigamma}(z) = \partial^2/\partial z^2 \ln\left[\Gamma(z)\right]$.

Appendix 4 – Distributional Characteristics

- **Inverse Fisher information of Truncated LogGamma:**

To avoid computationally expensive numeric integration, Opdyke (2014) derives the analytic approximation below:

$$A(\theta)^{-1} = \begin{bmatrix} A & B \\ B & D \end{bmatrix}^{-1} \quad \text{where}$$

$$A = \frac{1}{a^4 UIG^2} \times \left\{ \left[-(GHG2)^2 \right] \cdot (-z)^{2a} + 2a(-z)^a \cdot \left[-UIG \cdot GHG3 + a\Gamma(a) \cdot GHG2 \cdot \left(Log(-z) - digamma(a) \right) \right] \right.$$
$$\left. + a^4 \Gamma(a) \left[-\left(\Gamma(a) - UIG \right) \cdot \left(Log(-z) - digamma(a) \right)^2 + UIG \cdot trigamma(a) \right] \right\}$$

$$B = \frac{1}{a^2 b UIG^2} \times \left\{ t^{-b} \cdot GHG2 \cdot (-z)^{2a} - a^2 \left(t^b UIG^2 + \Gamma(a)(-z)^a \left(Log(-z) - digamma(a) \right) \right) \right\}$$

$$D = \frac{a}{b^2} + \frac{t^{-b}(-z)^a (1 - a - z)}{b^2 UIG} - \frac{t^{-2b}(-z)^{2a}}{b^2 UIG^2}$$

where...

Appendix 4 – Distributional Characteristics

- **Inverse Fisher information of Truncated LogGamma:**

where...

$t =$ data collection (truncation) threshold

$\eta = 0.001$

$adown = a - \eta$

$aup = a + \eta$

$z = -bLog[t]$

divide $a = diva = \dfrac{\Gamma(a+1)}{(-z)^{a}}$

divide $adown = divad = \dfrac{\Gamma(adown+1)}{(-z)^{adown}}$

divide $aup = divau = \dfrac{\Gamma(aup+1)}{(-z)^{aup}}$

$$GHG2 = divad \cdot J(-z;adown,1)\frac{aup}{aup-adown} + divau \cdot J(-z;aup,1)\frac{adown}{adown-aup}$$

$$GHG3 = divad \cdot J(-z;adown,1)\left(\frac{aup}{aup-adown}\right)\left(\frac{a}{a-adown}\right) + diva \cdot J(-z;a,1)\left(\frac{adown}{adown-a}\right)\left(\frac{aup}{aup-a}\right)$$
$$+divau \cdot J(-z;aup,1)\left(\frac{adown}{adown-aup}\right)\left(\frac{a}{a-adown}\right)$$

where $J(\ )$ is the CDF of the Gamma distribution.

$UIG =$ upper incomplete gamma function $= \Gamma(a,-z) = \Gamma(a)\left(1 - J(-z;a,b=1)\right)$

QuantMinds Americas Co-located with RiskMinds Americas

RiskMinds Americas Co-located with QuantMinds Americas