Operational Risk Capital Estimation and Planning: Exact Sensitivity Analysis and Business Decision Making Using the Influence Function*

Chapter Submission for Operational Risk: New Frontiers Explored, Davis, E., ed., Risk Books, London.

John ("J.D.") Opdyke§
Principal, Bates White LLC, JD.Opdyke@BatesWhite.com

Alexander Cavallo‡
Vice President, Corporate Risk Analytics and Insurance, Northern Trust, alc7@ntrs.com

1. Introduction

1.1. Executive Summary

Financial institutions have invested tremendous resources to develop operational risk capital models within the framework of the Advanced Measurement Approach (AMA) of the Basel II Accord. Most of this effort has focused on satisfying evolving regulatory requirements in the near term rather than risk-conscious business decision making in the long term. However, a critical objective of the Basel II Accord is to move institutions beyond viewing operational risk capital modeling as a mere regulatory exercise to embedding operational risk awareness into risk-informed decision making throughout the institution. To this end, we illustrate in this chapter the use of the Influence Function as a powerful analytical tool that allows the operational risk practitioner to leverage existing AMA models to generate critical quantitative insights for direct business decision-making by users of operational risk capital estimates.

Borrowed from the robust statistics literature, the Influence Function (IF) is an extremely useful and relevant methodology that provides a theoretical basis for capital planning and business decision making via exact sensitivity analysis. Because it is based on analytic derivations, the IF avoids the need to perform often resource-intensive, arguably subjective, and often inconclusive or inaccurate simulations. We clearly demonstrate how the IF utilizes any given estimator of the severity model (easily the main driver of estimated capital requirements), the values of its

* The views expressed in this paper are solely the views of the authors and do not necessarily reflect the opinions of Bates White LLC or Northern Trust Corporation. All analyses were performed by J.D. Opdyke using SAS®.

§ First author, corresponding author. The first author wishes to thank his colleague at Bates White LLC, Randal Heeb, Ph.D., for his encouragement and support, and expresses his sincere gratitude to Toyo R. Johnson and Nicole A.J. Opdyke for their thoughtful insights.

‡ The second author gratefully acknowledges the assistance and encouragement of colleagues at Northern Trust including Benjamin Rosenthal, Regina Desler, David Humke, Shang Xue, and Devon Brooks, and the patient support of Giovanna, Allison, Natalie, and Nicholas Cavallo.

parameter estimates, and an assumed forward looking frequency to define Exact Sensitivity Curves for Regulatory Capital and Economic Capital. These curves can be used to conduct exact sensitivity analyses on the capital impacts of hypothetical changes to the underlying loss data. Hypothetical loss scenarios of interest to bank management may be current or prospective, such as assessing the potential capital impact of a single hypothetical "tail" event of differing magnitudes. Relevant loss scenarios also may be retrospective, providing "but for" and exact attribution analyses as to why capital changed from one quarter to another. The information generated from these sensitivity analyses can suggest potential enhancements to the estimation of severity model parameters, and more broadly, better inform decision-making based on a more precisely defined risk profile.

1.2. Background

The Basel II Accord represents a major step forward in the regulation and supervision of the international financial and banking system. The risk measurement and risk management principles and guidelines put forth in the Basel II Accord aim to increase the stability and soundness of the banking system through comprehensive capital adequacy regulation. The approach includes the "three pillars" concept, in which stability and soundness is enhanced through minimum capital requirements (Pillar 1), supervisory review (Pillar 2), and market discipline via public disclosure requirements (Pillar 3). Among the major changes in this second Basel Accord are a greater reliance on banks' internal data systems to provide inputs to capital calculations and the extension of capital requirements to a new risk class, operational risk. This chapter focuses on the quantitative use of banks' internal data for assessing operational risk exposure.[1]

Operational risk is the risk of financial loss due to external events or due to inadequate or failed internal processes, people, or systems, including legal risk, but not reputational or strategic risk. Essentially, operational losses are the many different ways that a financial institution may incur a financial loss in the course of business aside from market, credit, or liquidity related exposure.[2] The Basel II Accord describes three potential methods for calculating capital charges for operational risk, and our focus in this chapter is on the most empirically sophisticated of the three – the Advanced Measurement Approach (AMA).[3] National bank regulators typically require internationally active banks and banks with significant operational risk exposure

---

[1] The methods examined and developed herein are readily applicable to the use of external loss data as well, such as that proffered by any of several banking consortiums. But the capital management and business planning that are informed by these methods would, in all likelihood, take place at the level of the individual bank or financial organization, and hence, make internal loss data most relevant.

[2] Additional information about the Basel II Accord and it specific framework for operational risk, including the definition of operational risk and standardized classification schemes for loss events according to business line (Annex 8) and event type (Annex 9), can be found in Basel Committee on Banking Supervision (hereafter, BCBS) (2006).

[3] The other two approaches in the Basel II framework are The Standardized Approach (TSA) and the Basic Indicator Approach (BIA). See BCBS (2006).

(generally, the largest banks) to use the AMA.[4] Some of the advantages of the AMA are that it permits financial institutions to develop a customized quantification system that makes use of historical data (which may include both internal and external loss data), bank specific information on internal controls and other relevant business factors, and the forward-looking assessments of potential risk generated by the bank's business experts via scenario analysis. The flexibility of the AMA arguably is also one of its major limitations – operational risk practitioners, regulators, and academics have engaged in vigorous debates on issues of methodology and best practices, yet many challenges remain unresolved since the framework was finalized in 2004 despite more than eight years of concerted efforts.[5]

In the AMA framework, an institution attempts to quantify operational risk exposure at a very high percentile of the enterprise aggregate annual loss distribution. Using the Value-at-Risk (VaR) risk measure, regulatory capital for operational risk is estimated at the 99.9th percentile (which corresponds to the size of total annual loss that would be exceeded no more frequently than once in 1,000 years). Economic capital is estimated at an even higher percentile of the distribution (usually between the 99.95th to 99.98th percentiles).[6]

Within the AMA framework, the Loss Distribution Approach (LDA) is the most commonly used method for estimating an aggregate annual loss distribution with parametric models. The LDA decomposes operational risk exposure into its frequency and severity components (that is, distributions of the number and magnitude of losses, respectively). Most institutions find a tractable solution to this empirical task by breaking the problem into a number of sequential stages with the ultimate goal being the estimation of the enterprise level aggregate annual loss distribution from its underlying components.

Yet even after carefully splitting the problem into distinct components (reviewed below), the fundamental statistical challenge remains: how can banks reliably estimate such a high percentile of the enterprise level aggregate annual loss distribution with sufficient precision, accuracy, and robustness so that it is actually useful in practice? To explain by way of example, a statistically correct estimate of required capital of $250m, based on LDA, that has a 95% confidence window of $200m on either side does not add much value: a range on the estimate of required capital from $50m to $450m obviously is not precise enough to use for making actual business decisions, but this range actually is more narrow than many in practice. Even if the percentile somehow was estimated with greater precision, if a single, new loss that deviated somewhat from the parametric assumptions of the statistical model (not even necessarily a large

[4] Some institutions benchmark their AMA capital estimates against estimates generated from the simpler and less risk sensitive Basic Indicator Approach or The Standardized Approach.

[5] The Basel II framework for operational risk was first formally proposed by the Basel Committee on Banking Supervision in June 1999, with substantial revisions released in January 2001 and April 2003, and was finalized in June 2004. The regulations implementing the Basel II Accord in the United States were finalized in 2007.

[6] Economic capital is defined as the amount of capital required to support an institution's risk in alignment with the institution's financial strength or creditworthiness. The enterprise level aggregate annual loss distribution is estimated using the institution's capital quantification system. The institution then selects a solvency standard (probability of default due to operational losses) that is acceptable, often referring to external benchmarks of credit risk. For example, over a one year time horizon, firms with a Moody's credit rating of Aa have a historical probability of default of 0.03%. To support a solvency standard equivalent to a Moody's Aa rating, economic capital could be determined with a VaR percentile of 99.97%. See McNeil et al. (2005) for further discussion.

loss) threw off the estimate by doubling it from one quarter to the next, and then dropping it by a factor of 3 in the following quarter, the estimator clearly is not robust enough to be considered reliable for actual business decision making. Yet this, too, is very typical of the quarterly behavior of many banks' capital estimates at the unit of measure level.[7] . A number of the statistical challenges arising from both the LDA framework and its application to the limited amount of extant operational loss data were raised as early as 2001, during the initial consultative period for the Basel II Accord.[8] Research after the Basel II Accord was finalized has begun to provide a stronger theoretical and empirical understanding of the limitations of some of the widely used estimation methodologies when applied to operational risk capital quantification.[9]

The inescapable challenge, however, is that the 99.9th percentile is so far out-of-sample, even when pooling operational loss data across many institutions, that to make any progress at all the practitioner must make very far-reaching, out-of-sample extrapolations using parametric models: that is, he or she must fit an assumed statistical distribution as closely as possible to the existing data, and then use this "best fit" to presume what the losses look like far out into the right tail of the statistical distribution (even though no (or very, very little) observed loss data exists so far out into the right tail). Consequently, the component models that contribute to this estimated percentile receive a high level of scrutiny from internal auditors, model validators, and regulatory supervisors. Of the empirical models and methods used in the AMA framework, the severity models generally pose much greater modeling challenges, have by far the largest impact on the ultimate estimates of economic and regulatory capital, and are an active area of research among industry practitioners, academics, and regulators.[10]

In this paper, we demonstrate how, for a given set of parameter estimates of a severity model, the Influence Function (IF) can be used to perform exact sensitivity analysis on capital requirements for various current or prospective, and even retrospective changes to the underlying loss data. In other words, the IF can be used to inform us of ***exactly*** what the change in capital would be if the bank experienced a new loss of, say, \$1m, or \$10m, or \$500m in the next quarter. The statistical theory behind the Influence Function has been richly developed in the Robust Statistics literature

---

[7] A unit of measure is a grouping of loss event data for which a bank estimates a distinct operational risk exposure.

[8] In a May 2001 report on Basel II, Daníelson et al. (2001) argue that operational risk simply cannot be measured reliably due to the lack of comprehensive operational loss data. At that point in time, few financial institutions were systematically collecting operational loss data on <u>all</u> business lines and <u>all</u> operational risk event types. Because of this, operational risk analysis made extensive use external loss event from vended database products. de Fontnouvelle et al. (2003) develop empirical models to address the substantial biases that can arise when modeling operational risk with such data including: data capture bias (because only losses beyond a specific threshold are recorded) and reporting bias (because only losses above some randomly varying threshold become public knowledge or are claimed against an insurance policy).

[9] Using the Operational Riskdata eXchange database (an extensive database of operational losses occurring at member institutions), Cope et al. (2009) demonstrate that data sufficiency and the regulatory requirements to extrapolate to the 99.9^{th} percentile of the loss distribution are major sources of instability and sensitivity of capital estimates. More recently, Opdyke and Cavallo (2012) demonstrate that the inherent non-robustness of Maximum Likelihood Estimation (MLE) is exacerbated by the use of truncated distributions, and that the extrapolations required for estimating regulatory and especially economic capital systematically and, in many cases, materially, overstate capital requirements due to Jensen's inequality.

[10] Frachot et al. (2004) demonstrate that the vast majority of variation in capital estimates is due to the variation of the estimated severity parameters, as opposed to the variation of the estimated frequency parameter(s).

for nearly half a century; only its application to operational risk is relatively new.[11] In this setting, the IF needs only three inputs to define exact capital sensitivity curves: i) the estimator used in the severity model; ii) the values of its parameter estimates; and iii) an assumed forward looking frequency. With these inputs the IF defines a deterministic, non-stochastic mathematical formula that exactly describes the impact of data changes (in the form of additional or changed losses) on the parameters of the severity distribution.[12] Because these parameters directly define capital requirements, the IF formula directly determines the exact capital changes caused by hypothetical or actual changes in the loss data – hence, the IF provides exact capital sensitivity analyses. Consequently, the IF arms business users of operational risk capital estimates with relevant information about potential capital needs and precisely defined risks, as long as they carefully align the hypothetical data changes to realistic capital planning and business decisions.

A major benefit for business users lies in the fact that the Influence Function is an exact formula, based on analytic derivations: to understand how capital changes under different scenarios, one need only use the formula, thus avoiding the need to perform extensive simulations that are often resource-intensive, subjectively interpreted, and often inconclusive or inaccurate regarding capital outcomes. Simply put, the IF provides the definitive, exact answer to the question, "How will capital requirements change if there is a new loss of $50m in the next quarter? Or just $500k? Or even $500m?"

Fortunately, the IF has a very wide range of application. It can be used with any of the commonly used severity distributions as well as with virtually any estimator of the severity distribution parameters. We illustrate the use of the Influence Function here with the most widely used operational risk severity estimator, the Maximum Likelihood Estimator (MLE). In spite of its known limitations, MLE continues to be most popular amongst practitioners and is almost universally accepted by regulatory authorities. The appeal of MLE for estimating the parameters of the severity distribution is its desirable statistical properties when the MLE modeling assumptions are satisfied, that is, when loss data is independent and identically distributed ("i.i.d.").[13] Under these conditions, MLE estimators are accurate (asymptotically unbiased), asymptotically normal, and maximally efficient (precise).

Under an extensive range of hypothetical changes in the loss data, we apply the IF to the MLE estimators of the parameters of multiple severity distributions to demonstrate, on both relative and absolute bases, the exact impacts of the data changes on the estimated capital requirements. These are the Exact Capital Sensitivity Curves mentioned above. This is extremely valuable

[11] Some recent applications of robust statistics to operational risk severity estimation include Opdyke and Cavallo (2012), Ruckdeschel and Horbenko (2010), and Horbenko, Ruckdeschel, and Bae (2011). Older publications include Chernobai and Rachev (2006) and Dell'Aquila and Embrechts (2006).

[12] Another way of stating this is that, as an exact formula, the IF introduces no additional estimation error beyond what has been estimated already, namely, the severity and frequency parameters.

[13] The i.i.d. assumption describes two important aspects of a data sample. First, an observed sample of data points is independent "when no form of dependence or correlation is identifiable across them" (BCBS 2011, fn. 29). Second, an observed sample of data points is identically distributed (homogeneous) when the data are generated by exactly the same data generating process, such as one that follows a parametric probability density function, or "are of the same or similar nature under the operational risk profile" (BCBS 2011, fn. 29). These textbook conditions are mathematical conveniences that rarely occur with actual, real-world data, let alone "messy" operational risk loss event data.

information that is useful in two major ways: first, for capital planning, as they are, by definition, exact sensitivity analyses whereby the capital effects of different scenarios, based on hypothetical changes to the underlying loss data, can be seen directly. Scenarios of interest to bank management may be prospective, such as assessing the potential capital impact of hypothetical "tail" events of differing magnitudes, or retrospective, allowing for exact attribution or “but for” analysis to provide insight into the reasons why capital changed the way it did from one quarter to another. Secondly, statistically, the IF and the Capital Curves it generates can guide severity estimator choice and development to potentially increase both the robustness and efficiency of the capital distribution (as distinct from the distribution of the severity parameter estimates), while mitigating material bias via previously unidentified but important statistical effects, like Jensen’s inequality (see Opdyke and Cavallo, 2012). Taken together, the IF and its associated Exact Capital Sensitivity Curves not only can suggest major potential enhancements to the severity model, but also better inform decision-making and capital planning based on a more precisely and accurately defined risk profile.

So the IF can be used 1) as an essential tool to inform estimator choice and development when tackling the fundamental statistical problem of obtaining estimates of a very high percentile of the loss distribution that are more precise, less biased, and more robust; and 2) directly in the capital planning process, once an estimator is selected, to generate corresponding Exact Capital Sensitivity Curves. Opdyke and Cavallo (2012) focus on the former, but in this paper we focus on the latter, while noting the many benefits associated with a unified methodological framework that relies on the IF for *both* 1) and 2).

In Section 2, we describe the basic capital estimation problem under LDA, including a discussion of the empirical challenges of estimating a severity distribution and operational risk capital from historical data. This is followed by a discussion of the M-Class estimation framework, as MLE is an M-Class estimator. In Section 3, we discuss the Influence Function and its central role in the robust statistics framework, with a focus on how the IF provides a widely accepted and well established statistical definition of “robustness” (specifically, “B-robustness”). Here we also present the Empirical Influence Function, and analytically derive the Influence Functions of MLE estimators of parameters for some of the most commonly used medium- to heavy-tailed loss severity distributions, both with and without data truncation.[14] In Section 4 we review a series of “case studies” to show how the Exact Capital Sensitivity Curves arise in real-world situations of relevance to business decision makers. Section 5 concludes the chapter with a summary and a discussion of the implications of the results, as well as suggested related topics for future applied research.

2. The Capital Estimation Problem in Operational Risk

The capital estimation problem is generally approached by segmenting risk into suitably homogeneous risk classes, applying the LDA to these risk classes, and then aggregating risk to

[14] Most institutions collect information on operational losses only above a specified data collection threshold. The most common method of accounting for this threshold is to fit truncated severity distributions to the available loss data.

the enterprise level. Typical stages of a modeling process may include the following:

1. Unit of measure definition: Historical loss data (which may be internal to the bank or include external data as well) are partitioned or segmented into non-overlapping and homogeneous "units of measure" that share the same basic statistical properties and patterns.

2. Frequency estimation: Empirical models are developed to describe or estimate the distribution of annual loss frequency in each unit of measure.

3. Severity estimation: Empirical models for the severity of losses within each unit of measure are developed.

4. Estimation of aggregate annual loss distributions: By combining the frequency and severity components within each unit of measure, the distribution of the annual total value of operational losses is computed for each unit of measure.

5. Top-of-house risk aggregation: The aggregate annual loss distributions for each unit of measure are combined to represent the enterprise level aggregate annual loss distribution and related capital requirements. The Basel II framework presumes perfect dependence of risk across units of measure as a default assumption, which can be implemented for the VaR risk measures currently used by simply summing all the VaR estimates of all the units of measure. A reduction in enterprise level capital can be obtained if an institution can successfully demonstrate that the risks of all the units of measure are less than perfectly dependent. This is often accomplished via correlation matrices or copula models.

2.1. The Loss Distribution Approach

The Loss Distribution Approach (LDA) is an actuarial modeling technique widely used to estimate aggregate loss distributions for many types of risk, including operational risk. Total annual loss in unit of measure i is given by $S_i = \sum_{j=1}^{n_i} x_{ij}$, where n_i is the total number of losses in the year for unit of measure i.

To estimate the probability distribution of total annual loss amounts, the LDA decomposes the distribution of S_i into its frequency and severity components. The application of the LDA within unit of measure i requires the following fundamental modeling assumptions:

- Annual loss frequency (n_{it}) is independently and identically distributed within unit of measure i with some probability distribution: $n_{it} \sim H_i(\lambda_i)$
- Each of the n_i individual loss severities (x_{ij}) in unit of measure i are independently and identically distributed with some probability distribution: $x_{ij} \sim F_i(\theta_i)$ with $j = 1, \ldots, n_i$

- Loss frequency is independent of loss severity[15]

Current industry practice is that banks use essentially all available data points for estimating the parameters of the severity distributions in each unit of measure. Some institutions exclude certain individual data points when they are not representative of operational risk exposure on a current or forward looking basis. Typically, banking supervisors require a substantial level of documentation, justification, and internal governance for the exclusion of historical loss events from the estimation samples.[16] The level of scrutiny for excluding external losses is substantially lower than for excluding internal losses.[17]

In contrast, the parameters of the frequency distributions often are estimated on subsets of the available data, for example, the five most recent years of data. Internal data quality considerations may also affect an institution's decision on how many years of loss frequency data to use in estimating frequency parameters.

2.2. The Setting: Empirical Challenges to Operational Risk Severity Modeling

The very nature of operational loss data makes estimating severity parameters challenging for several reasons:

- Limited Historical Data: Observed samples of historical data are quite limited in sample size since systematic collection of operational loss data by banks is a relatively recent development. Sample sizes for the critical low frequency – high severity loss events are even smaller.[18]

[15] However, Ergashev (2008) notes that this assumption is violated for truncated distributions. Whether this violation is material to the estimation of either severity parameters or capital estimates is not explored.

[16] Both industry practitioners and banking supervisors appear to accept the notion that some external loss events may not reflect an institution's risk profile for a variety of reasons. Filtering of external data for relevance according to business lines, geographic areas, and other salient characteristics is widely accepted. Scaling models are sometimes used to make external losses more representative of an institution's risk profile. Banking supervisors also appear to be receptive to arguments that specific individual external loss events may be excluded, but banks must typically acquire very detailed information in order to make acceptable arguments. For example, certain types of events are "industry" events that occur at multiple institutions in the same general time period, such as the wave of legal settlements related to allegations of mutual fund market timing. An institution that itself incurred one or more such losses may be justified in excluding other institutions' losses (if they can be identified in the external data) since that specific industry event is already represented by an internal loss in the bank's loss database.

[17] There is greater industry range of practice across the different national banking jurisdictions with respect to the exclusion of internal loss events from estimation samples. Some jurisdictions permit wholesale exclusion of losses for disposed businesses (i.e. when business units or business lines are no longer part of the institution). Some jurisdictions require more detailed analysis to determine which losses may be excluded (e.g. banking supervisors may require the inclusion of loss events related to employment practices, since these policies are typically established at a corporate level).

[18] Pooling data from multiple financial institutions in the 2002 Loss Data Collection Exercise, Moscadelli (2004) estimates GPD distributions on as few as 42 data points. Chapelle et al. (2008) estimate GPD distributions with sample sizes of only 30 to 50 losses, and other parametric distributions with sample sizes of approximately 200, 700,

- Heterogeneity: Obtaining reasonably large sample sizes for severity modeling necessarily requires making the unit of measure more heterogeneous, either by incorporating external loss data or by pooling internal loss data across multiple product lines, business units, etc.[19] Due to this inherent heterogeneity in the loss data, it is highly improbable that the critical MLE assumptions of i.i.d. data are satisfied.

- Heavy-tailed Severity Distributions: Operational risk practitioners have observed that parameter estimates for heavy-tailed loss distributions typically are quite unstable and can be extremely sensitive to individual data points. The extreme sensitivity of parameter estimates and capital estimates to large losses is well documented, but the comparable sensitivity, and sometimes even *greater* sensitivity, of parameter estimates to small losses has been almost completely missed in the literature.[20]

- Truncated Severity Distributions: Most institutions collect information on operational losses above a specific data collection threshold. The most common method to account for this threshold is to fit truncated severity distributions to the loss data, but truncation creates additional computational complexity and is a major source of parameter (and capital) instability. This is due at least in part to the fact that it creates much more heavy-tailed distributions. Additionally, for some estimators, such as MLE, truncation induces or augments existing covariance between the estimated parameters, which further heightens instability and non-robustness in the parameter estimates (see Opdyke and Cavallo, 2012).

- Changing Data: An often underappreciated fact of real-world operational loss event databases is that the data itself evolves over time. Some institutions include specific provisions or reserves to account for this, because over time the severity of some losses may be adjusted upward or downward to reflect additional information about the event (litigation-related events of notable durations are common examples).[21] Also, due to the inherent application of judgment in interpreting and applying unit-of-measure classification schemes, individual loss events can even be reclassified into other business lines or event types as additional loss event details are discovered or understood. As a

---

and 3,000 losses. The smaller sizes are in stark contrast to important publications in the literature, both seminal (see Embrechts et al., 1997) and directly related to operational risk VaR estimation (see Embrechts et al., 2003) which make a very strong case, via "Hill Horror Plots" and similar analyses, for the need for sample sizes much larger to even begin to approach stability in parameter estimates.

[19] Heterogeneity of operational loss data has been flagged as a major problem by a number of authors. Daníelson et al. (2001) state "the loss intensity process will be very complicated, depending on numerous economic and business related variables" (p. 13). For example, Cope and Labbi (2008) and Cope (2010) make use of country level characteristics and bank gross income to build location-scale models that define more homogeneous units of measure, without which, of course, the units of measure would have been (much more) heterogeneous.

[20] Cope (2011) documents the substantial sensitivity of MLE parameter estimates to large losses using a mixture approach to induce misspecification in the right tail. The analysis does not examine the ultimate impact to operational risk capital, nor does it examine misspecification in the left tail. Opdyke and Cavallo (2012) is the only paper known to these authors that does all three, and it finds, under certain circumstances, potentially massive instability due to left tail misspecification, both in terms of parameter estimation and in terms of capital estimation.

[21] Recent AMA-related guidance states that banks must have a process for updating legal event exposure after it is financially recognized on the general ledger until the final settlement amount is established. See BCBS (2011).

result, business users of operational risk capital estimates have a strong need for a well-defined, theoretically justified, and easily implemented tool to assess potential capital needs for certain hypothetical data changes, because in practice, such changes happen all the time, even without considering new loss events in each new quarter.

2.3. M-Class Estimation

MLE is amongst the class of M-Class estimators, so called because they generalize "M"aximum Likelihood Estimation. M-Class estimators include a wide range of statistical models for which the optimal values of the parameters are determined by computing sums of sample quantities. The general form of the estimator (1) is extremely flexible and can accommodate a wide range of objective functions, including the MLE approach and various robust estimators.

$$\hat{\theta}_M = \underset{\theta \in \Theta}{\arg\min} \sum_{i=1}^{n} \rho(x_i, \theta) \qquad (1)$$

Assuming the regularity conditions commonly assumed when using MLE,[22] all M-Class estimators are asymptotically normal and consistent (asymptotically unbiased),[23] which are very useful properties for statistical inference.

Maximum Likelihood Estimation is considered a "classical" approach to parameter estimation from a frequentist perspective. In the Basel II framework, MLE is the typical choice for estimating the parameters of severity distributions in operational risk.[24] To maintain its desirable statistical properties (described below), MLE requires the following assumptions:

- A1) Independence: Individual loss severities are statistically independent from one another
- A2) Homogeneity: Loss severities are identically distributed within a unit of measure (perfect homogeneity)
- A3) Correct Model: The probability model of the severity distribution is correctly specified (this is distinct from the model's parameters, which must be estimated)

Under these restrictive and idealized textbook assumptions, MLE is known to be not only asymptotically unbiased ("consistent") and asymptotically normal, but also asymptotically efficient.[25] Given an i.i.d. sample of losses $(x_1, x_2, \cdots x_n)$ and knowledge of the "true" family of

[22] A discussion of the regularity conditions required for the application of MLE is included in many statistics and econometrics textbooks. See Greene (2007) for one example.

[23] A summary of the regularity conditions needed for the consistency and asymptotic normality of M-Class estimators generally can be found in many textbooks on robust statistics, such as Huber and Ronchetti (2009).

[24] The recent AMA guidance from the Basel Committee acknowledges the recent application of robust statistics in operational risk, but refers to Maximum Likelihood Estimation and Probability Weighted Moments as "classical" methods. See BCBS (2011) ¶ 205.

[25] The term "efficient" here is used in the absolute sense, indicating an estimator that achieves the Cramér-Rao lower bound – the inverse of the information matrix, or the negative of the expected value of the second-order derivative

the probability density function $f(x|\theta)$ (that is, knowledge of the pdf, but not its parameter values), the MLE parameter estimates are the values of $\hat{\theta}_{MLE}$ that maximize the likelihood function, or equivalently, that minimize the objective function $\rho(x,\theta)$.

The MLE estimator is an M-class estimator with $\rho(x,\theta) = -\ln\left[f(x|\theta)\right]$, so

$$\hat{\theta}_{MLE} = \underset{\theta\in\Theta}{\arg\min}\sum_{i=1}^{n}\rho(x_i,\theta) = \underset{\theta\in\Theta}{\arg\min}\sum_{i=1}^{n}-\ln f(x_i|\theta). \tag{2a}$$

An equivalent expression is obtained by maximizing the usual log-likelihood function,

$$\hat{l}(\theta|x_1,x_2,\cdots x_n) = \ln\left[L(\theta|x)\right] = \sum_{i=1}^{n}\ln\left[f(x_i|\theta)\right],$$

so

$$\hat{\theta}_{MLE} = \underset{\theta\in\Theta}{\arg\max}\left[\hat{l}(\theta|x_1,x_2,\cdots x_n)\right]. \tag{2b}$$

An objective assessment of real-world operational risk data must acknowledge that each one of the key assumptions required for MLE estimators to retain their desirable statistical properties (consistency, asymptotic efficiency, and asymptotic normality) is unlikely to hold in practice.

- Independence (A1): If the severity of operational losses has both deterministic and stochastic components, then operational losses by definition fail to be independent due to common determinants. For example, systematic differences in loss severity may be explained by event characteristics such as geography, legal system, client segment, time period effects, etc.[26] The impact of potential failures of the independence assumption is beyond the scope of this chapter, but is known to be nontrivial (see van Belle, 2002, pp.7-11).

- Identically distributed (A2): Because a unit of measure typically pools internal loss events from multiple business processes that undoubtedly have different data generating processes, achieving perfect homogeneity, as required by MLE, is virtually impossible. The pooling of internal and external loss data for severity modeling further augments heterogeneity. Each institution's operational risk profile is unique and is moderated by its specific characteristics – the specific combination of products and service offerings, technology, policies, internal controls, culture, risk appetite, scale of operation, governance, and other factors. Understanding the behavior of severity estimates and

of the log-likelihood function. This is the minimal variance achievable by an estimator. See Greene (2007) for more details. The term "efficient" also can be used in a relative sense, when one estimator is more efficient – that is, all else equal, it achieves a smaller variance (typically assuming unbiasedness) – than another.

[26] Cope et al. (2011) find systematic variation in loss severity by region, country characteristics, and certain macroeconomic variables.

capital estimates when confronted with both modest and extreme deviations from perfectly homogeneous data is the central focus of this chapter.

- Correctly specified model (A3): MLE has desirable asymptotic statistical properties only when the correct form of the loss distribution is "known" and correctly specified. Under idealized i.i.d. data conditions, MLE remains consistent, asymptotically efficient, and asymptotically normal, but if there is the possibility (probability) that some losses do not come from the statistical distribution assumed by the model, then MLE can perform poorly, losing most or all of these desirable statistical qualities. This has been shown in the literature (see Dupuis, 1999, and Opdyke and Cavallo, 2012) and is consistent with findings presented herein in the following sections.

3. The Influence Function and the Robust Statistics Framework

Robust Statistics is a general approach to estimation that explicitly recognizes and accounts for the fact that all statistical models are by necessity idealized and simplified approximations of complex realities. As a result, a key objective of the robust statistics framework is to bound or limit the *influence* on parameter estimates of a small to moderate number of data points in the sample which happen to deviate from the assumed statistical model. Since actual operational loss data samples generated by real-world processes do not exactly follow mathematically convenient textbook assumptions (e.g. all data points are not perfectly i.i.d., and rarely, if ever, exactly follow parametric distributions), this framework would appear to be well suited for operational risk severity modeling.

The Influence Function is central to the Robust Statistics framework, the theory behind which is very well developed and has been in use for almost half a century. Some of the seminal theoretical results that established the field of robust statistics include Tukey (1960), Huber (1964), and Hampel (1968). Classic textbooks on robust statistics such as Huber (1981) and Hampel et al. (1986) have been widely used for more than 30 years. The dramatic increases in computing power over the last 20 years also have enabled the theoretical development and practical use of computationally intensive methods for computing robust statistics.[27] Robust statistics have been used widely in many different applications, including the analysis of extreme values arising from both natural phenomena and financial outcomes.[28] These applications of robust statistics use many of the medium- to heavy-tailed loss distributions of greatest relevance in operational risk. A unified approach to comparing the relative merits of robust statistics and classical statistics, both in terms of parameter estimation, directly, and capital estimation, ultimately, can be made with the Influence Function. This is the focus of Opdyke and Cavallo (2012), but this chapter focuses on the application of the IF to MLE-based estimates of capital to better inform business decision making.

[27] See Huber and Ronchetti (2009) for a discussion of these more recent advances.

[28] A detailed summary table of applications of robust statistics across many disciplines is available from the authors upon request.

3.1. The Influence Function

The Influence Function is an essential tool from the robust statistics framework that allows researchers to analytically determine and describe the sensitivity of parameter estimates to arbitrary deviations from the assumed statistical model. The Influence Function can be thought of as indicating the impact of a marginal deviation at severity amount *x* on the parameter estimates. Simply put, it answers the question: "How does the parameter estimate change when the data sample changes with a new loss of amount = $*x*?"

One requirement of the IF is that the estimator being assessed is expressed as a statistical functional, that is, as a function of the assumed distribution of the model: $T\left(F\left(y,\theta\right)\right)=T\left(F\right)$, where $F\left(y,\theta\right)$ is the assumed severity distribution. Fortunately, almost all relevant estimators can be expressed as statistical functionals. Common examples are the first and second moments of a known distribution $F\left(y,\theta\right)$ as

$$m_1 = T_1\left(F\left(y,\theta\right)\right) = \int y dF\left(y,\theta\right) = \int y f\left(y,\theta\right) dy$$

and

$$m_2 = T_2\left(F\left(y,\theta\right)\right) = \int y^2 dF\left(y,\theta\right) = \int y^2 f\left(y,\theta\right) dy$$

respectively.

The Influence Function is an analytic formula for assessing this impact of an infinitesimal deviation from the assumed distribution occurring at a severity amount of *x*:

$$IF\left(x;T,F\right) = \lim_{\varepsilon\to 0}\left[\frac{T\left\{\left(1-\varepsilon\right)F+\varepsilon\delta_x\right\}-T\left(F\right)}{\varepsilon}\right] = \lim_{\varepsilon\to 0}\left[\frac{T\left(F_\varepsilon\right)-T\left(F\right)}{\varepsilon}\right] \qquad (3)$$

where

- $F\left(y,\theta\right)=F\left(\ \right)$ is the assumed severity distribution,
- $T\left(F\left(y,\theta\right)\right)=T\left(F\right)$ is the statistical functional for the specified estimator under the assumed distribution,
- x is the location of the deviation,
- ε is the fraction of data that is deviating,
- δ_x is the cumulative distribution function of the Dirac delta function D_x, a probability measure that puts mass 1 at the point *x*:

  $$D_x\left(y\right)=\begin{Bmatrix}1 & \text{if } y=x\\ 0 & \text{otherwise}\end{Bmatrix} \quad \text{and} \quad \delta_x\left(y\right)=\begin{Bmatrix}1 & \text{if } y\geq x\\ 0 & \text{otherwise}\end{Bmatrix},$$

- $T\left\{\left(1-\varepsilon\right)F+\varepsilon\delta_x\right\}=T\left(F_\varepsilon\right)$ is simply the estimator evaluated with contamination.[29]

[29] The term "statistical contamination" does not indicate a problem with data quality per se, but instead reflects the realistic possibility (probability) that most of the data follows the assumed distribution, but some fraction of the data

The Influence Function has a simple and direct interpretation as the difference between the parameter estimates when the data sample is "contaminated," that is, having distribution F_ε which is "contaminated" by a new loss amount *x*, and the parameter estimates under the assumed distribution, F (with no "contamination"); this difference is then normalized by the amount of contamination, ε. This framework allows us to compare the exact asymptotic behavior of a selected estimator across a range of arbitrary data deviations, or even make comparisons across multiple estimators, when faced with less-than-ideal, non-textbook data, regardless of the nature of the deviating data (i.e. regardless of the distribution from which the "contamination" or "arbitrary deviation" came). By systematically varying the location of the arbitrary deviation (*x*) the impact of data contamination from any distribution can be assessed. The Influence Function is an asymptotic result because it is the limiting value as the amount of deviating data approaches zero (or equivalently, as the sample size increases without bound since ε is a function of the number of contaminated data points).[30] The Influence Function is an extremely powerful tool with many uses and serves as the foundation for our analytic approach to capital sensitivity analysis using hypothetical scenarios based on changes in loss data.[31]

3.2. *The Empirical Influence Function*

Often used in conjunction with the IF, the Empirical Influence Function (EIF) is simply the finite sample approximation of the IF, that is, the IF applied to the empirical distribution of the data sample at hand. Arbitrary deviations, *x*, are used in the same way to trace the EIF as a function of *x*, and since the empirical distribution is used, typically $\varepsilon = 1/n$. So to define the EIF, we use equation (3) with the empirical cumulative distribution function, so $F(\ ) = \hat{F}(\ )$:

$$EIF\left(x;T,\hat{F}\right) = \lim_{\varepsilon\to 0}\left[\frac{T\left\{\left(1-\varepsilon\right)\hat{F}+\varepsilon\delta_x\right\}-T\left(\hat{F}\right)}{\varepsilon}\right] = \lim_{\varepsilon\to 0}\left[\frac{T\left(\hat{F}_\varepsilon\right)-T\left(\hat{F}\right)}{\varepsilon}\right] \qquad (4)$$

where all terms agree with (3) except that practical application dictates that $\varepsilon = 1/n$ (so $\varepsilon \to 0$ still as $n \to \infty$)

Because EIF converges to IF quickly, that is, even with relatively small sample sizes, EIF is a

comes from a different distribution (this portion is called "contaminated"). In the remainder of this paper, we use the more neutral term "arbitrary deviation" synonymously with "statistical contamination."

[30] The conditions required for the existence of the Influence Function are detailed in Hampel et al. (1986), and Huber (1977). The Influence Function is a special case of a Gâteaux derivative and it requires even weaker conditions for existence than a Gâteaux derivative (see Hampel et al. 1986, and Huber 1977). The IF can be defined for any of the commonly used operational risk severity distributions.

[31] See Hampel et al. (1986) for an extensive and detailed description of the many uses of the influence function.

good practical tool for validating IF derivations, or for approximating IF for other reasons (we show examples of this in practice later in the paper, and this is one of the reasons it is good practice to implement both EIF and IF simultaneously).

3.3. B-Robustness

The entire point of robust statistics is to obtain reliable estimates of a parametric model even when the assumed model is only approximately correct, specifically, when some of the data may come from a different underlying distribution than the bulk of the data. The definitions of the Influence Function suggest a simple and useful definition of "robustness" to such deviating data points. An estimator is said to be "B-robust" for the distribution F ( ) if the Influence Function is bounded (meaning it does not diverge toward $\pm\infty$) over the domain of F ( ). If the Influence Function is not bounded, then the estimator is *not* "B-Robust" for a particular parameter of distribution F ( ).[32] When the Influence Function is unbounded, on the other hand, then an arbitrary deviant data point can result in meaningless or practically unusable parameter estimates, as when parameter estimates become arbitrarily large or small (i.e. divergent toward $\pm\infty$). This type of extreme sensitivity of parameter estimates is precisely the type of "bias" that can result from heterogeneous data.

A comparison of the influence functions of two estimators of central tendency, the mean and the median, is useful for illustrating the concept of B-Robustness. Assume the data follows a standard normal distribution, $F = \Phi$. The statistical functional for the mean is $T(F) = \int y dF(y) = \int y f(y) dy$, so to derive the IF of the mean, we have:

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

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

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

$$IF(x;T,F) = x \tag{5}$$

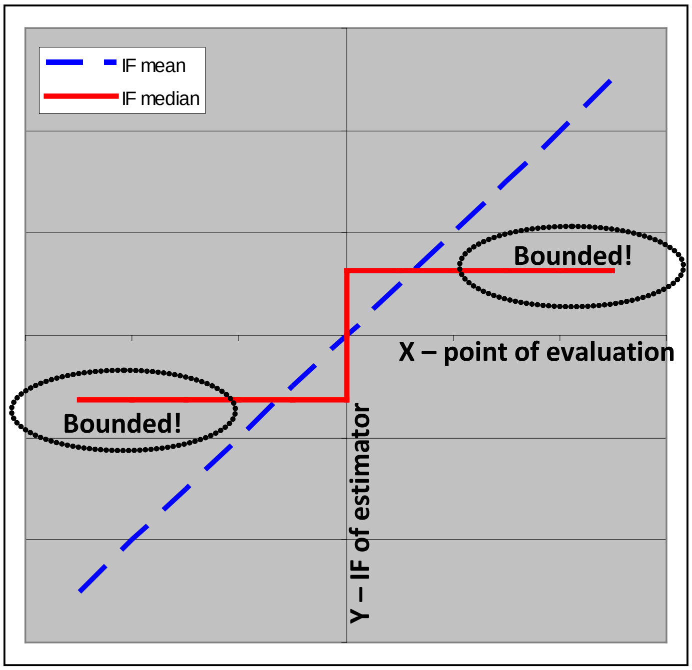
From the mathematical derivation above, it is evident that the Influence Function for the mean of a standard normal random variable is unbounded. As the point of arbitrary deviation (x) increases to $+\infty$, so does the Influence Function, and as a result, the mean becomes arbitrarily

[32] Note that the "B" in "B-robust" signifies limiting the "B"ias of an estimator, because if the estimator itself is bounded, so too must be its bias (if any).

large and meaningless. Similarly, as the point of deviation decreases to –∞, the Influence Function does as well, and the mean becomes arbitrarily small and meaningless.

Figure 1 displays the Influence Functions of the mean and median of a standard normal distribution.[33] Consistent with the mathematical derivation above, the Influence Function for the mean has a positive slope of 1 and increases without bound in both directions. In contrast, the Influence Function for the median is bounded and never tends toward ±∞.

Graph 1: Influence Functions of the Mean and Median



As expected, the non-robustness and B-robustness of the mean and median, respectively, hold even when the $F(\ )$ is not the standard normal distribution, as shown below for the mean.

$$IF(x;T,F) = \lim_{\varepsilon \to 0}\left[\frac{(1-\varepsilon)\int y dF(y) + \varepsilon \int y d\delta_x(y) - \int y dF(y)}{\varepsilon}\right] = \lim_{\varepsilon \to 0}\left[\frac{\varepsilon x - \varepsilon \mu}{\varepsilon}\right] = x - \mu$$

where μ is the mean, so

$$IF(x;T,F) = x - \mu$$

As should now be apparent from the above, estimators that are not B-robust run the very real risk of generating capital estimates that, due to unanticipated, unmeasured, or unmeasurable heterogeneity (even if it is relatively small), are extremely "biased" relative to that which would be generated by the assumed severity distribution of the model, and potentially grossly inflated.

[33] See Hampel et al. (1986) for a derivation of the influence function of the median.

The IF is the analytic tool to use to identify such conditions. In the next section, we present the formula for the Influence Function for M-class estimators generally and MLE specifically. We use this to derive and present the MLE Influence Functions of parameters of specific and widely used severity distributions in operational risk. The capital estimates based on these estimators are then generated and assessed, with one criteria for evaluation being whether they are B-robust, and how this robustness, or lack thereof, impacts capital estimation from the standpoint of practical implementation.

3.4. The Influence Function for M-Class Estimators

M-class estimators generally are defined as any estimator $T_n = T_n\left(X_1,\cdots,X_n\right)$ whose optimized objective function satisfies $\sum_{i=1}^{n}\varphi\left(X_i,T_n\right)=0$, or equivalently, $\sum_{i=1}^{n}\rho\left(X_i,T_n\right)=\underset{T_n}{\min}!$, where $\varphi\left(x,\theta\right)=\frac{\partial\rho\left(x,\theta\right)}{\partial\theta}$ is the derivative of ρ which is defined over $\wp\times\Theta$, the sample space and parameter space, respectively.

The first order conditions for this optimization problem are
$\varphi\left(x,\theta\right)=\frac{\partial\rho\left(x,\theta\right)}{\partial\theta}=0$

The second order conditions are satisfied when the Hessian of the objective function $\varphi'_\theta\left(x,\theta\right)=\frac{\partial\varphi_\theta\left(x,\theta\right)}{\partial\theta}$ is positive-definite.[34]

Hampel et al. (1986) show that, conveniently, the Influence Function for all M-class estimators is

$$IF_\theta\left(x;\theta,T\right)=\frac{\varphi_\theta\left(x,\theta\right)}{-\int_a^b\varphi'_\theta\left(y,\theta\right)dF\left(y\right)} \tag{6}$$

where *a* and *b* are the endpoints of support for the distribution. When multiple parameters are being estimated, as with most operational risk severity distributions, the possibility of (non-zero) parameter covariance must be taken into account with the matrix form of (6) as shown in (7) below:

[34] This is the second partial derivative test for more than one variable (in this case, more than one parameter). The Hessian is positive-definite if all eigenvalues are positive in which case *f* ( ) attains a local minimum at *x*, the point at which it is evaluated.

$$IF_\theta(x;\theta,T) = A(\theta)^{-1}\varphi_\theta = \begin{bmatrix} -\int_a^b \frac{\partial \varphi_{\theta_1}}{\partial \theta_1} dF(y) & -\int_a^b \frac{\partial \varphi_{\theta_1}}{\partial \theta_2} dF(y) \\ -\int_a^b \frac{\partial \varphi_{\theta_2}}{\partial \theta_1} dF(y) & -\int_a^b \frac{\partial \varphi_{\theta_2}}{\partial \theta_2} dF(y) \end{bmatrix}^{-1} \begin{bmatrix} \varphi_{\theta_1} \\ \varphi_{\theta_2} \end{bmatrix} \quad (7)$$

(see Stefanski and Boos, 2000).

3.5. The Influence Function for MLE Estimators

MLE is an M-Class estimator with objective function

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

In this case, the derivative of the objective function with respect to the parameters is simply the negative of the score function

$$\varphi_\theta(x,\theta) = \frac{\partial \rho(x,\theta)}{\partial \theta} = -\frac{\partial f(x,\theta)}{\partial \theta} \Big/ f(x,\theta)$$

and the Hessian is

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

So for the specific case of MLE, the Influence Function shown in (6) and (7) is simply the score function normalized by its variance (the negative of the expected value of the second order derivative, or the Fisher Information). And as noted above, in this setting it is important to note and account for potential covariance of the severity distribution's parameters by evaluating the cross-partial derivative terms for each parameter in (7). This is shown below to sometimes have very large, and even counterintuitive effects on the estimators under common conditions. The IF is the tool that can establish such effects, definitively, as the analytic behavior of the estimators, and not as an uncertain function of simulations that can be misspecified or subjectively interpreted, with inferences resting largely on the specific and narrow ranges of input parameter values. This is one of the tremendous advantages of using IF: it is an analytic derivation describing the exact behavior of the estimator under any degree of arbitrary deviation from the assumed severity distribution. IF is not only more accurate than any simulation could be, it makes behavioral simulations moot because it ***is*** the formulaic answer to the question, "exactly how does the parameter estimate change when loss event, *x*, is added to my sample of loss data?" And regarding the possible "B-robustness" of the MLE estimator(s) for a specific distribution, this can be determined, based on the above, simply by determining whether the score function is bounded, as long as it is monotonic over the relevant domain (see Huber, 1981).

3.6. The Influence Function for MLE Estimators of Truncated Severity Distributions

Most banks record losses only above a certain threshold *H* (typically *H* = $5,000, $10,000, or €20,000 for the case of some external consortium data), so data on smaller losses generally are not available. The reason for this is that many business processes at a financial institution generate large numbers of operational loss events with de minimis impact that do not threaten bank solvency. It is much more efficient for banks to gather operational loss data on the smaller set of operational loss events that have material impact to earnings, may threaten bank solvency, may generate reputational risk, and/or may be preventable with appropriate changes in bank policies and procedures.

When data is collected subject to a data collection threshold, the most widely accepted and utilized method to account for incomplete observation of the data sample is to assume that losses below the threshold follow the same parametric loss distribution, *f* ( ), as those above it, whereby the severity distribution becomes *g*( ), a (left) truncated distribution, with pdf and cdf below.

$$g(x,\theta,H)=\frac{f(x,\theta)}{1-F(H,\theta)} \quad \text{and} \quad G(x,\theta,H)=1-\frac{1-F(x,\theta)}{1-F(H,\theta)}$$

Under truncation, the terms of the Influence Function for the MLE estimator now become

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

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

and

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

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

When the severity distribution has only one parameter, the general form of the Influence Function is:

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

$$=\frac{-\frac{\frac{\partial f\left(x,\theta\right)}{\partial\theta}}{f\left(x,\theta\right)}-\frac{\frac{\partial F\left(H,\theta\right)}{\partial\theta}}{1-F\left(H,\theta\right)}}{-\frac{1}{1-F\left(H,\theta\right)}\int_a^b\frac{\left[\frac{\partial f\left(y,\theta\right)}{\partial\theta}\right]^2-\frac{\partial^2 f\left(y,\theta\right)}{\partial\theta^2}\cdot f\left(y,\theta\right)}{f\left(y,\theta\right)}dy+\frac{\left[\frac{\partial F\left(H,\theta\right)}{\partial\theta}\right]^2+\frac{\partial^2 F\left(H,\theta\right)}{\partial\theta^2}\cdot\left[1-F\left(H,\theta\right)\right]}{\left[1-F\left(H,\theta\right)\right]^2}} \quad (8)$$

where *a* and *b* define the endpoints of support, which are now *H* and (for all relevant severity distributions) +∞, respectively. When the severity distribution has more than one parameter, the Influence Function has the general form:

$$IF_{\theta}\left(x;\theta,T\right)=A\left(\theta\right)^{-1}\varphi_{\theta}=\begin{bmatrix}-\int_a^b\frac{\partial\varphi_{\theta_1}}{\partial\theta_1}dG\left(y\right) & -\int_a^b\frac{\partial\varphi_{\theta_1}}{\partial\theta_2}dG\left(y\right)\\ -\int_a^b\frac{\partial\varphi_{\theta_2}}{\partial\theta_1}dG\left(y\right) & -\int_a^b\frac{\partial\varphi_{\theta_2}}{\partial\theta_2}dG\left(y\right)\end{bmatrix}^{-1}\begin{bmatrix}\varphi_{\theta_1}\\ \varphi_{\theta_2}\end{bmatrix}. \quad (9)$$

The structure of the multi-parameter (typically two-parameter) version of the IF does not change from equation (7) except that the differential, of course, corresponds with the cdf of the truncated severity distribution, *G*( ).

Comparing (6) and (8) we can see that the numerator of the Influence Function of a truncated distribution simply is a shift of the score function for the non-truncated distribution, and the magnitude of the shift depends only on the threshold *H* and the parameter values *θ*, but **not** on the location of the arbitrary deviation *x*. The denominator of the Influence Function for a truncated distribution differs substantially from that of its non-truncated distribution. The expected value of the Hessian is computed over the truncated domain (*H*, ∞), multiplied by a truncation constant, and added to an additional constant term in each second derivative. As is the case for the *φ* function, the constant terms depend only on the threshold *H* and the parameter values *θ*, but **not** on the location of the arbitrary deviation *x*. These changes in the Fisher information matrix (relative to the non-truncated case) fundamentally alter the correlation structure of the parameters of the distribution, introducing dependence, or magnifying it if already present before truncation.

Even when analyzing data collected with a data collection threshold, the MLE Influence Function is an analytically determined function given an assumed distribution and parameter values. With it, no simulation is required to assess the behavior of MLE parameter estimators because this remains a definitive, analytic result. All that is required to perform the analysis is the calculation of the derivatives and integration in (6) and (7) or (8) and (9), and these derivatives, of course, differ for each different severity distribution. So for each severity distribution, one must calculate

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

and under truncation, also calculate

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

(the derivatives of the cumulative distribution function with respect to the parameters can be computed using Leibniz's Rule). Once this is done, (7) and (9) apply to all non-truncated and truncated severity distributions, respectively, which is very convenient as it makes calculations and testing for multiple severity distributions considerably easier.

3.7. The Influence Function for LogNormal, LogGamma, and GPD Distributions, With and Without Truncation

Using the analytic formulas for the Influence Function under MLE estimation in (7) and (9), we summarize below the key mathematical results to obtain the IF for the parameters of the LogNormal, LogGamma, and GPD distributions when there is no truncation. The same results under truncation can be found in the Appendix. Complete derivations for all IFs can be found in Opdyke and Cavallo (2012). Below we also present graphs, for all six truncated and non-truncated cases, of all the IFs and the EIFs so the behavior of the former, and the convergence of the latter, is clear.

<u>LogNormal</u>:

The pdf and cdf of the LogNormal distribution are defined as:

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

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

Inserting the derivatives of $\frac{\partial f(y;\theta)}{\partial\theta_1},\frac{\partial f(y;\theta)}{\partial\theta_2},\frac{\partial^2 f(y;\theta)}{\partial\theta_1\partial\theta_2},\frac{\partial^2 f(y;\theta)}{\partial\theta_1^2}$, and $\frac{\partial^2 f(y;\theta)}{\partial\theta_2^2}$ into the Fisher Information yields

$$A(\theta)=\begin{bmatrix}-1/\sigma^2 & 0\\ 0 & -2/\sigma^2\end{bmatrix}\text{ and into the psi function yields }\varphi_\theta=\begin{bmatrix}\frac{\mu-\ln(x)}{\sigma^2}\\ \frac{1}{\sigma}-\frac{(\ln(x)-\mu)^2}{\sigma^3}\end{bmatrix}$$

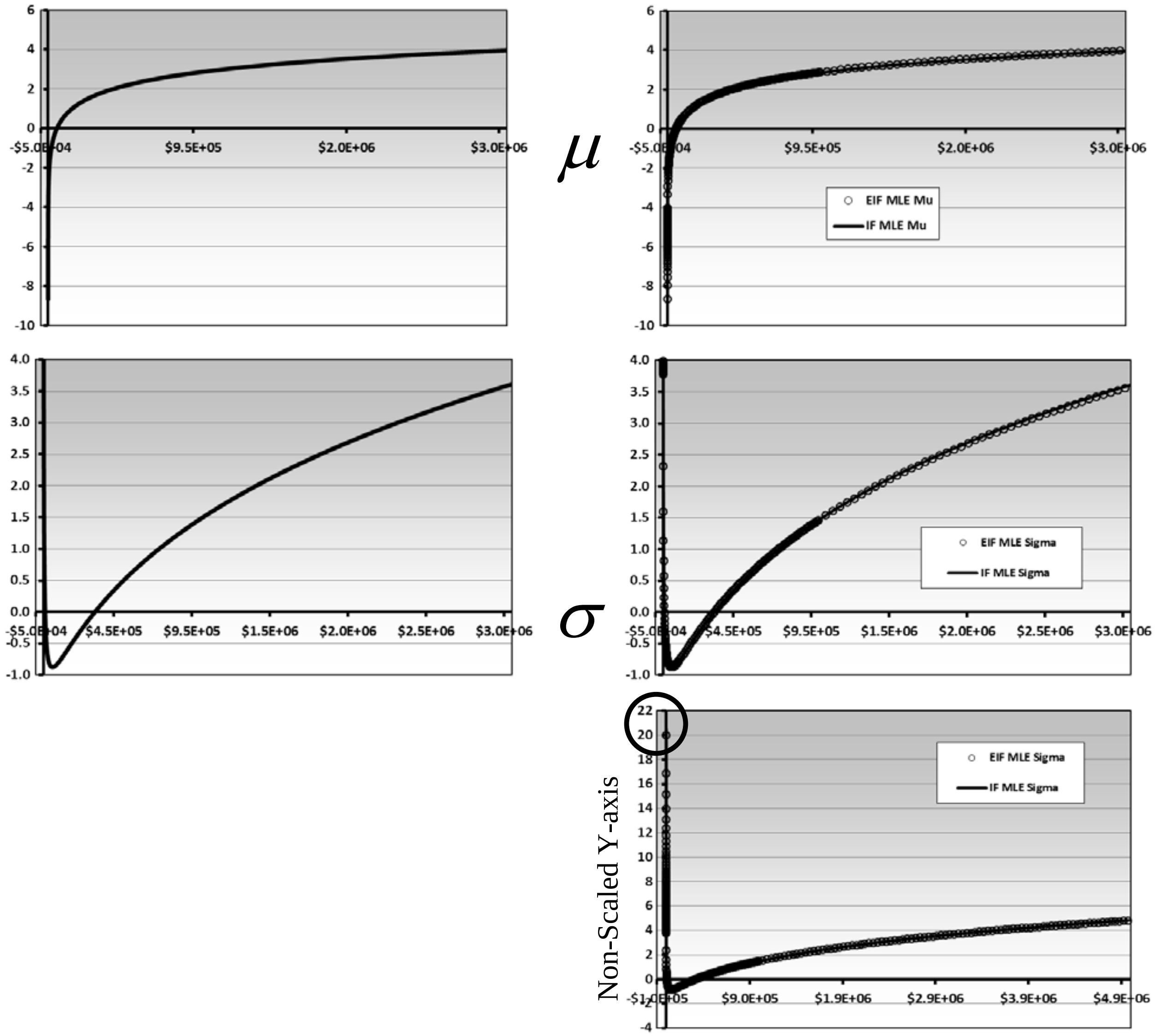
So via (7) the Influence Function of the MLE parameters of the LogNormal severity is

$$IF_{\theta}\left(x;\theta,T\right)=A\left(\theta\right)^{-1}\varphi_{\theta}=\begin{bmatrix}-\sigma^{2} & 0\\ 0 & -\sigma^{2}/2\end{bmatrix}\begin{bmatrix}\dfrac{\mu-\ln\left(x\right)}{\sigma^{2}}\\ \dfrac{1}{\sigma}-\dfrac{\left(\ln\left(x\right)-\mu\right)^{2}}{\sigma^{3}}\end{bmatrix}=\begin{bmatrix}\ln\left(x\right)-\mu\\ \dfrac{\left(\ln\left(x\right)-\mu\right)^{2}-\sigma^{2}}{2\sigma}\end{bmatrix} \tag{10}$$

Importantly, note that the zero cross derivatives in $A(\theta)$ indicate parameter independence in *x*, which is discussed further below.

This result (10) for the LogNormal is a well known. Graphs of the IF for the MLE parameters of the LogNormal severity (with μ = 10.95 and σ = 1.75), compared with their EIF counterparts, are shown below.

Graphs 2a-2e: IF v. EIF of LogNormal (μ = 10.95 and σ = 1.75)
for MLE Parameter Estimates by Arbitrary Deviation, *x*



First note, as mentioned above, the quick convergence of EIF to IF even for *n* not very large (here, *n* = 250). Secondly, note the asymptotic behavior of both μ and σ as $x \to 0^+$: according to (10), $\mu \to -\infty$ and $\sigma \to +\infty$. But note that $\sigma \to +\infty$ at a much faster rate because of the squared term in the numerator of its IF, so this would indicate a capital estimate $\to +\infty$ as $x \to 0^+$, that is, a larger and larger capital estimate caused by smaller and smaller arbitrary deviations in the left tail. This is an important, counter-intuitive result shown previously only in Opdyke and Cavallo (2012), and which we shall explore more in the next section when we present the exact capital sensitivity curves.

LogGamma:

We present the same derivations below for the LogGamma:

$$f\left(x;a,b\right)=\frac{b^{a}\left(\log\left(x\right)\right)^{\left(a-1\right)}}{\Gamma\left(a\right)x^{b+1}} \text{ and } F\left(x;a,b\right)=\frac{b^{a}}{\Gamma\left(a\right)}\int_{\ln\left(0^{+}\right)}^{\ln\left(x\right)} y^{\left(a-1\right)}\exp\left(-yb\right)dy$$

for $0<x<\infty;\ 0<a;\ 0<b$.[35]

Inserting the derivatives of $\frac{\partial f\left(y;\theta\right)}{\partial\theta_1},\frac{\partial f\left(y;\theta\right)}{\partial\theta_2},\frac{\partial^2 f\left(y;\theta\right)}{\partial\theta_1\partial\theta_2},\frac{\partial^2 f\left(y;\theta\right)}{\partial\theta_1^2}$, and $\frac{\partial^2 f\left(y;\theta\right)}{\partial\theta_2^2}$ into the Fisher Information yields

$A\left(\theta\right)=\begin{bmatrix}-trigamma\left(a\right) & 1/b\\ 1/b & -a/b^2\end{bmatrix}$ and into the psi function yields $\varphi_\theta=\begin{bmatrix}-\ln\left(b\right)-\ln\left(\ln\left(x\right)\right)+digamma\left(a\right)\\ -\frac{a}{b}+\ln\left(x\right)\end{bmatrix}$

So via (7) the Influence Function of the MLE parameters of the LogGamma severity is

$$IF_\theta\left(x;\theta,T\right)=A\left(\theta\right)^{-1}\varphi_\theta=\frac{1}{\left(-a/b^2\right)\cdot trigamma\left(a\right)-1/b^2}\begin{bmatrix}-a/b^2 & -1/b\\ -1/b & -trigamma\left(a\right)\end{bmatrix}\begin{bmatrix}-\ln\left(b\right)-\ln\left(\ln\left(x\right)\right)+digamma\left(a\right)\\ -\frac{a}{b}+\ln\left(x\right)\end{bmatrix}=$$

$$=\begin{bmatrix}\dfrac{\frac{a}{b^2}\left[\ln\left(b\right)+\ln\left(\ln\left(x\right)\right)-digamma\left(a\right)\right]-\frac{1}{b}\left[\ln\left(x\right)-\frac{a}{b}\right]}{trigamma\left(a\right)\left(\frac{a}{b^2}\right)-\frac{1}{b^2}}\\ \dfrac{\frac{1}{b}\left[\ln\left(b\right)+\ln\left(\ln\left(x\right)\right)-digamma\left(a\right)\right]-trigamma\left(a\right)\left[\ln\left(x\right)-\frac{a}{b}\right]}{trigamma\left(a\right)\left(\frac{a}{b^2}\right)-\frac{1}{b^2}}\end{bmatrix} \tag{11}$$

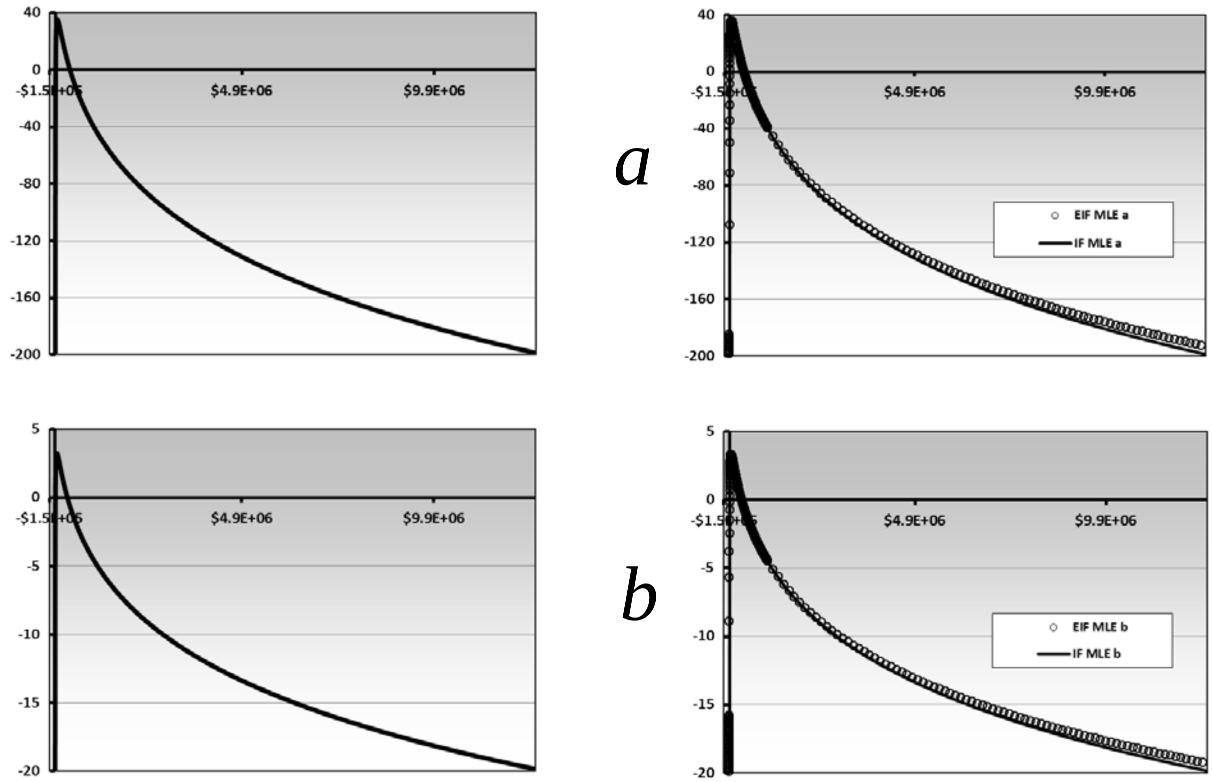
Importantly, note that the non-zero cross derivatives in $A\left(\theta\right)$ indicate parameter dependence in *x*, the effects of which are discussed below.

Graphs of the IF for the MLE parameters of the LogGamma severity (with *a* = 35.5 and *b* = 3.25), compared with their EIF counterparts, are shown below.

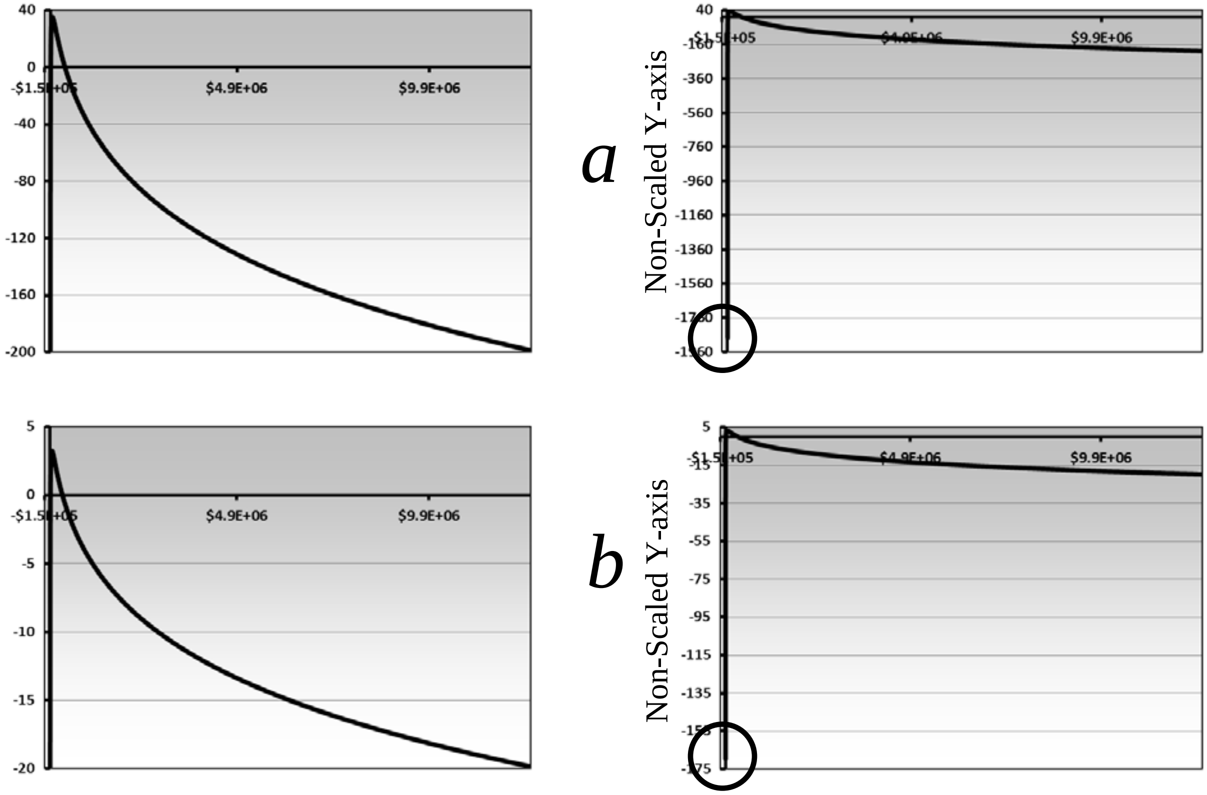
---

[35] A common definition of the LogGamma distributed random variable, *z*, is $z=\exp\left(q\right)$, where *q* is a random variate that follows the Gamma distribution with endpoints of support $0<q<\infty$. This inconveniently makes the endpoints of support for the LogGamma $1<z<\infty$. Conventional practice, when using this definition, is to subtract the value 1 (one) ex post so that the endpoints of support for the LogGamma become $0<z<\infty$.

**Graphs 3a-3d: IF v. EIF of LogGamma (*a* = 35.5 and *b* = 3.25)
for MLE Parameter Estimates by Arbitrary Deviation, *x***



Graphs 3e-3h: IF Y-Scaled v. IF of LogGamma (*a* = 35.5 and *b* = 3.25) for MLE Parameter Estimates by Arbitrary Deviation, *x*



Note again the quick, if imperfect, convergence of EIF to IF even for *n* not very large (*n* = 250). Secondly, note the asymptotic behavior of both *a* and *b* as $x \to 1^{+}$: $a \to -\infty$ and $b \to -\infty$, which we can see in the graphs above where the y-axis is not scaled. But note that while smaller *a* indicates smaller quantiles for the LogGamma, smaller *b* indicates *larger* quantiles for the LogGamma. What does this mean for capital estimation (which is essentially a high quantile estimate of the severity distribution)? The effect of the *b* term in (11) ends up dominating that of the *a* term because of the relative size of the constants in the numerators of both terms, and estimated capital $\to +\infty$ as $x \to 1^{+}$; that is, like the LogNormal, a larger and larger capital estimate results from a smaller and smaller arbitrary deviation in the left tail. Again, a counterintuitive result, and as we shall see in the next section, a very large one.

<u>Generalized Pareto Distribution</u>:

For the GPD severity, we have:

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

for $0 \le x < \infty;\ 0 < \beta;$ assuming $\varepsilon > 0$ (which is appropriate in this setting)

Inserting the derivatives of $\frac{\partial f(y;\theta)}{\partial\theta_1},\frac{\partial f(y;\theta)}{\partial\theta_2},\frac{\partial^2 f(y;\theta)}{\partial\theta_1\partial\theta_2},\frac{\partial^2 f(y;\theta)}{\partial\theta_1^2}$, and $\frac{\partial^2 f(y;\theta)}{\partial\theta_2^2}$ into the cells of the Fisher Information $A(\theta)=\begin{bmatrix} -\int_a^b \frac{\partial\varphi_{\theta_1}}{\partial\theta_1}dF(y) & -\int_a^b \frac{\partial\varphi_{\theta_1}}{\partial\theta_2}dF(y) \\ -\int_a^b \frac{\partial\varphi_{\theta_2}}{\partial\theta_1}dF(y) & -\int_a^b \frac{\partial\varphi_{\theta_2}}{\partial\theta_2}dF(y) \end{bmatrix}$ yields

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

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

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

and into the psi function yields

$$\varphi_\theta = \begin{bmatrix} \frac{1}{\beta}\left[\frac{\beta-x}{\beta+\varepsilon x}\right] \\ -\left[\left(\frac{-x(1+\varepsilon)}{\beta\varepsilon+\varepsilon^2 x}\right) + \frac{\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right] \end{bmatrix} \quad (12)$$

Via (7), the Influence Function of the MLE parameters of the GPD severity –

$IF_\theta(x;\theta,T)=A(\theta)^{-1}\varphi_\theta$ – is solved numerically. However, note that Smith (1987),[36] for the GPD specifically, was able to conveniently simplify the Fisher Information to yield

$$A(\theta)^{-1} = (1+\xi)\begin{bmatrix} 1+\xi & -\beta \\ -\beta & 2\beta^2 \end{bmatrix} \quad (13)$$

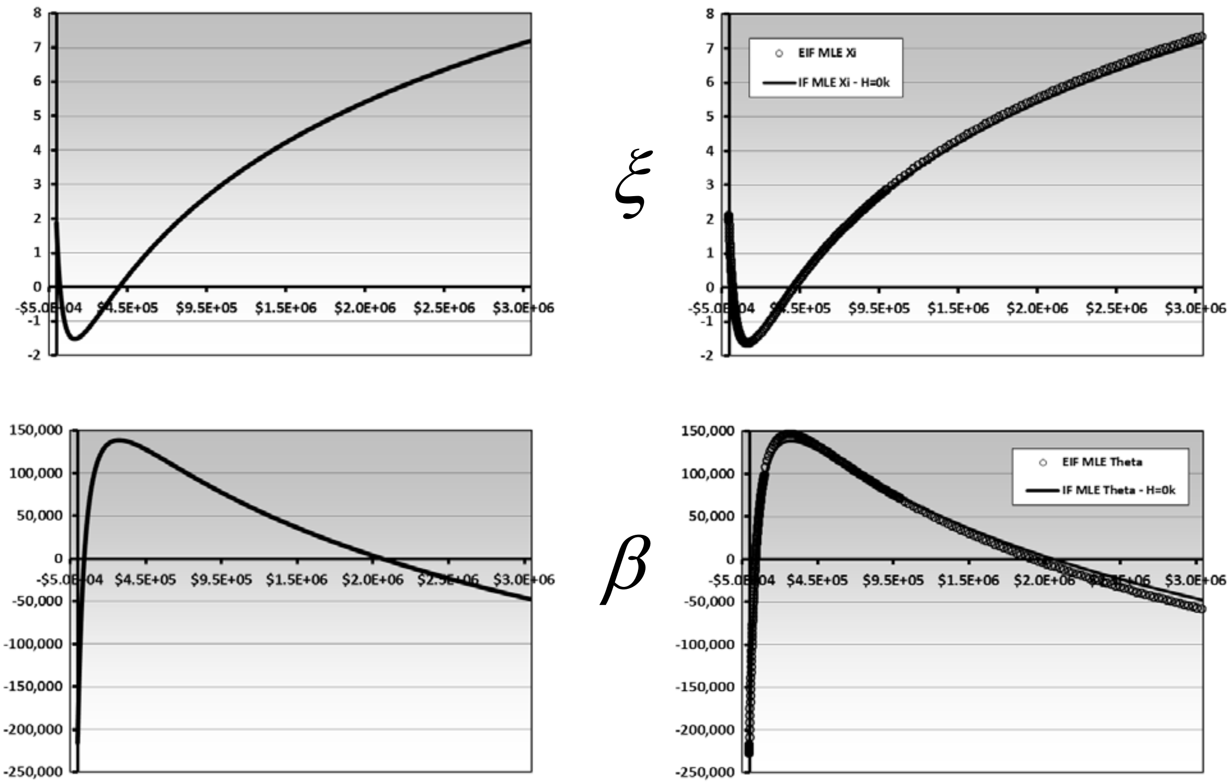
(Ruckdeschel and Horbenko (2010) later re-present this result in the operational risk setting). This gives the exact same result, as shown in the graphs below, as the numerical implementation of (12) above, and provides further independent validation of the more general framework presented herein (which, of course, can be used with *all* commonly used severity distributions).

[36] Smith (1987) was the earliest example of this result that we were able to find in the literature.

Importantly, note that the non-zero cross derivatives in (12), as well as in (13), indicate parameter dependence in *x*, the effects of which are discussed below.

Graphs of the IF for the MLE parameters of the GPD severity (with ξ = 0.875 and β = 57,500), compared with their EIF counterparts, are shown below.

Graphs 4a-4d: IF v. EIF of GPD (ξ = 0.875 and β = 57,500) for MLE Parameter Estimates by Arbitrary Deviation, *x*



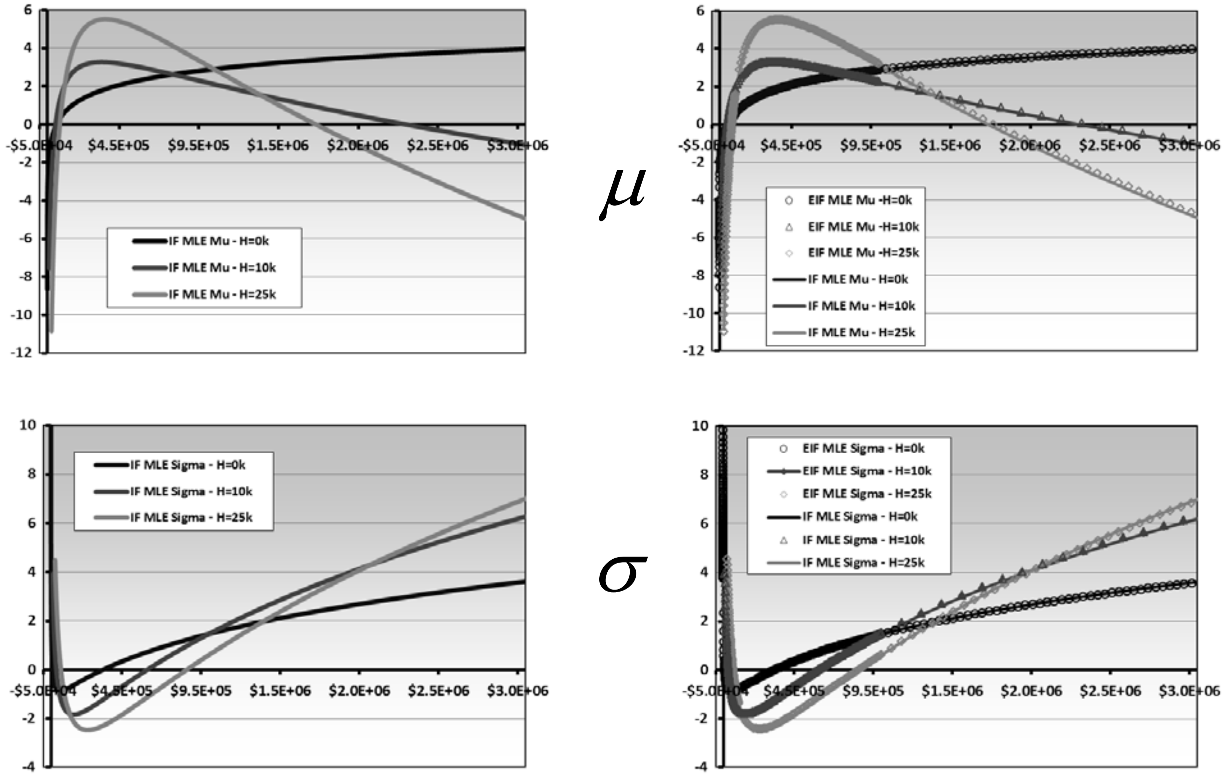
Again note the quick, if imperfect, convergence of EIF to IF even for *n* not very large (*n* = 250). Secondly, note that as $x \to +\infty$ apparently $\xi \to +\infty$ and $\beta \to -\infty$, which will undoubtedly be reflected in the exact capital sensitivity curves in the following sections.

The mathematical results for the IFs of the LogNormal, LogGamma, and GPD severities under truncation all are presented in the Appendix. Their corresponding graphs, for thresholds (*H*) of \$0, \$10k and \$25k, are presented below, with IFs presented side-by-side with EIFs.

Truncated LogNormal:

For the Truncated LogNormal, we have:

Graphs 5a-5d: IF v. EIF of Truncated LogNormal (μ = 10.95 and σ = 1.75) for MLE Parameter Estimates by Arbitrary Deviation, *x*



Again, EIF converges to IF even for n not very large (n = 250). Secondly, note that 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 $\mu(x)$ 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***.[37] 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 definitively 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 probably the source, at least in part, of the extreme sensitivity reported in the literature of MLE parameter estimates of the truncated LogNormal.

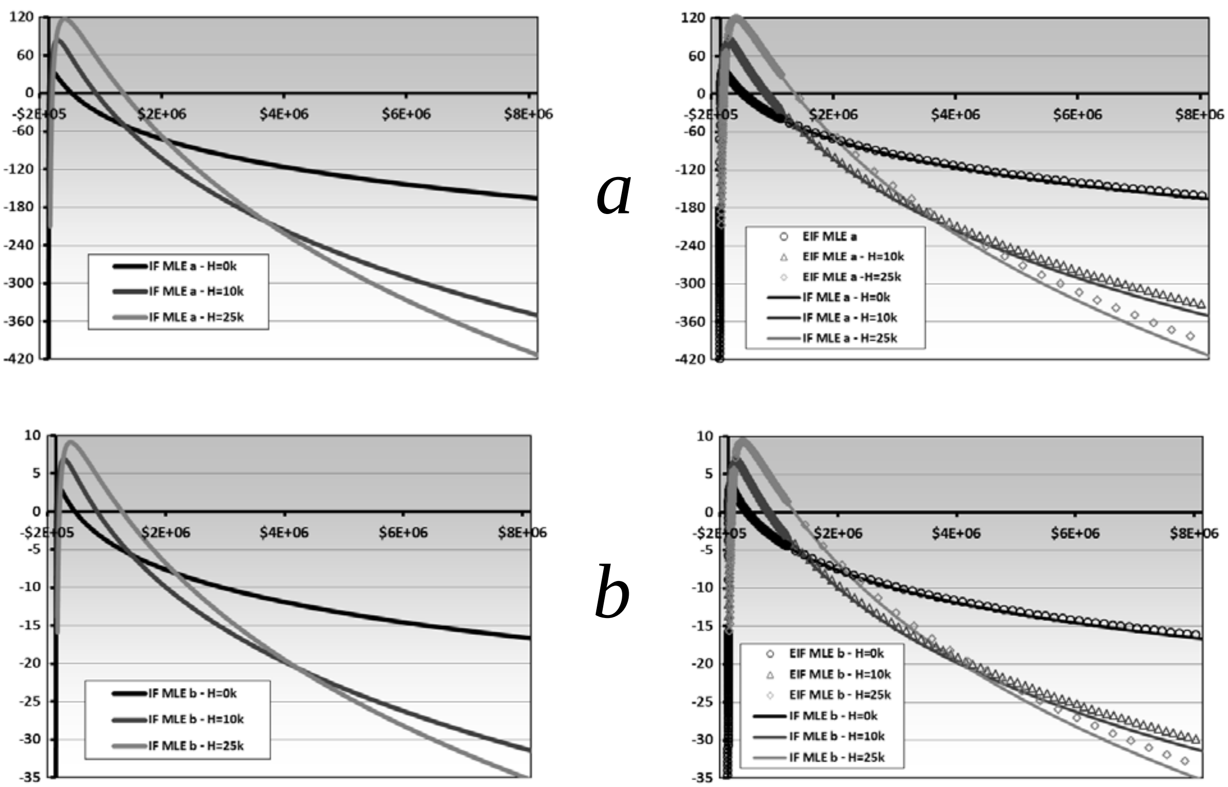
[37] In fact, exp($\boldsymbol{\mu}$) is the scale parameter of the LogNormal.

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 inconclusive musing.

Truncated LogGamma:

For the Truncated LogGamma, we have:

Graphs 6a-6d: IF v. EIF of Truncated LogGamma (*a* = 35.5 and *b* = 3.25)
for MLE Parameter Estimates by Arbitrary Deviation, *x*

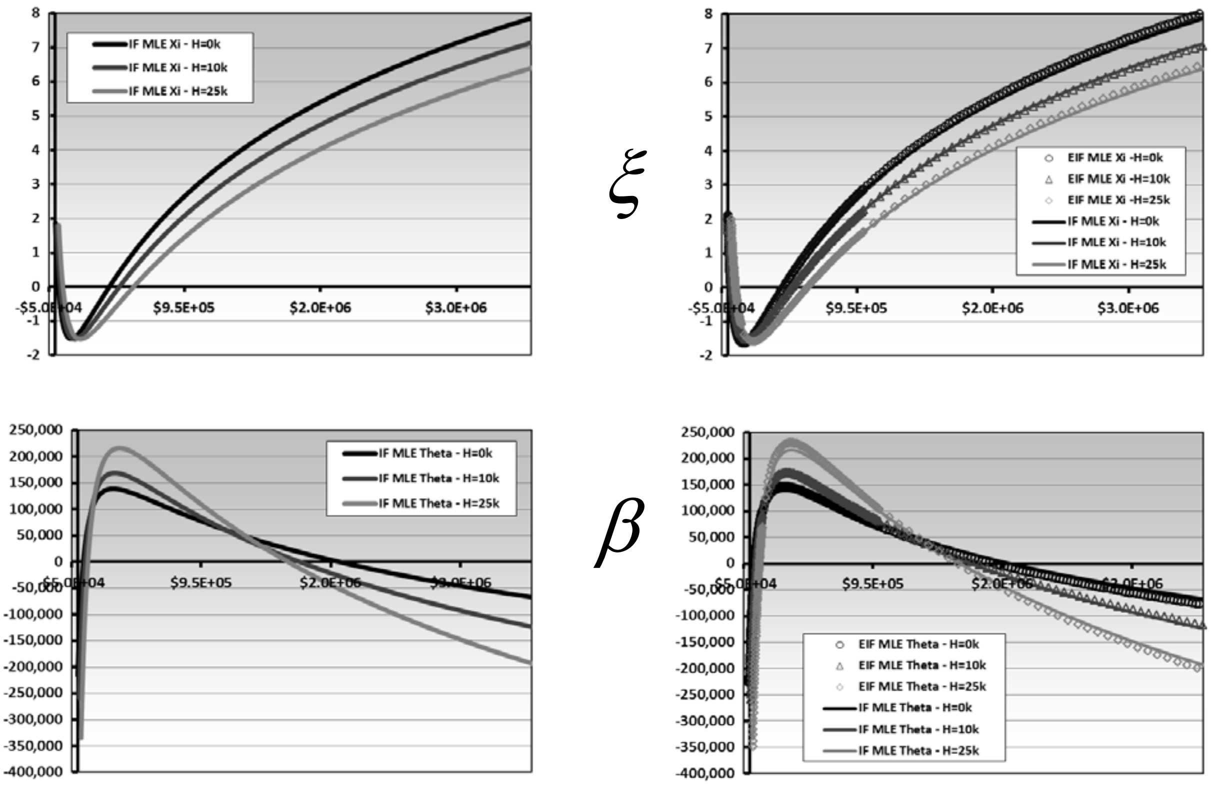


Note again the quick, if imperfect convergence of EIF to IF even for *n* not very large (*n* = 250). Also note that the extreme asymptotic behavior of both parameters as $x \to 1^+$ is mitigated somewhat by truncation, just as with the LogNormal. However, both parameters' diverge much more quickly to negative infinity as $x \to +\infty$, and this more rapid divergence is also like the LogNormal (but in the opposite direction for μ). So while in the case of the LogNormal truncation *caused* parameter dependence, in the case of the LogGamma it *augmented* dependence that was already there, as shown in the non-zero cross derivative terms of $A(\theta)$ in (11) above.

Truncated GPD:

For the GPD, we have:

Graphs 7a-7d: IF v. EIF of Truncated GPD (ξ = 0.875 and β = 57,500)
for MLE Parameter Estimates by Arbitrary Deviation, *x*



EIF again converges to IF fairly quickly, if imperfectly, even for *n* not very large (*n* = 250). Unlike the LogNormal and the LogGamma, truncation does not mitigate parameter variance as $x \to 0^+$ (perhaps because there was somewhat less to begin with), but like the other two severity distributions it does cause much more rapid divergence to negative infinity for β as $x \to +\infty$, while ξ mostly just shifts to the right, which is consistent with its role as the tail index. So the negative parameter covariance in *x* that already was present in the non-truncated case, as seen in (12) and (13), remains in the truncated case, as seen in (A.3) in the Appendix.

3.8. Capital Estimation

The entire point of the statistical exercise of estimating severity distribution parameters is to estimate a capital requirement. As the convolution of the frequency and severity distributions, the aggregate loss distribution, for which we must obtain a VaR, has no general closed form solution, so large scale Monte Carlo simulations are the gold standard for obtaining the "true" capital requirement for a given set of frequency and severity distribution parameters. However, a number of less computationally intensive methods exist, the most convenient of which is the mean adjusted Single Loss Approximation (SLA) of Degen (2010).[38] Given a desired level of statistical confidence (α), an estimate of forward-looking annual loss frequency (λ), an assumed severity distribution ($F(\,)$), and values for the parameters of the severity distribution (θ), capital requirements are approximately given by

$$C_\alpha \approx F^{-1}\left(1-\frac{1-\alpha}{\lambda}\right)+\lambda\mu \quad \text{where} \tag{14}$$

α = single year loss quantile (0.999 for regulatory capital; 0.9997 for economic capital)
λ is the average number of losses occurring within one year (the frequency estimate)
μ is the mean of the estimated severity distribution

This provides us with a very accurate approximation of the VaR of the aggregate loss distribution without having to simulate it.[39] Note that from (14) we can see that the VaR of the aggregate loss distribution is essentially just a high quantile of the severity distribution on a single loss (the first term) with a mean adjustment that typically is small (the second term) relative to the first term (in the remainder of the paper, our use of "SLA" refers to the mean adjusted SLA of (14)). And for the case of severity distributions with infinite mean, say, a GPD severity with $\xi \geq 1$, Degen derives an SLA approximation that is not dependent upon the mean of the distribution:

$$C_\alpha \approx F^{-1}\left(1-\frac{1-\alpha}{\lambda}\right)-(1-\alpha)F^{-1}\left(1-\frac{1-\alpha}{\lambda}\right)\cdot\left(\frac{c_\xi}{1-1/\xi}\right) \tag{15}$$

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 $\xi = 1$

So the capital estimates, based on (14) and (15), are functions of the severity distribution parameter estimates θ (whether via MLE or some other estimator) which define $F^{-1}(x,\theta)$. Since the IFs define the exact behavior of the parameter estimator, and the parameter estimator defines the exact behavior of capital estimates, all as a function of new losses, the Exact Capital Sensitivity Curve can thus be drawn as a function of new losses, based directly on the IF. We now have a way to perform exact sensitivity analyses (no simulations required) based on hypothetical new losses: simply evaluate the IF at the value of the new loss, then multiply IF by ε (typically $1/n$) and subtract the parameter estimate based on the original sample to get the value of the new parameter estimate. Then use the new parameter estimate to obtain the new capital

[38] Degen's (2010) fomula is supported by analytic derivations, whereas that of Böcker and Sprittulla (2006), which is commonly used, is based on empirical observation (although both are very similar).

[39] Capital estimates based on fully simulated aggregate loss distributions, with both frequency and severity parameters simulated (which is the gold standard here) are compared to SLA approximations in Opdyke and Cavallo (2012) and the later are shown to be very accurate for practical purposes.

requirement (this is shown below in (16)). This provides a capital estimate with no additional estimation error (beyond that of the original frequency and severity parameter estimation), and is described in more detail below.

4. Using the Influence Function to Inform Business Decisions

In this section we demonstrate how the Influence Function is directly used to define, under a wide range of scenarios, exact capital requirements for business decision makers. The basic framework begins with a set of baseline parameters for the loss severity distribution $F\left(x,\theta_0\right)$ using a particular estimator on a sample of size *n*. The mean-adjusted single loss approximation can be used to generate a baseline estimate of capital given the relevant forward looking annual loss frequency (λ) and the required percentile of the aggregate annual loss distribution (e.g. α = 99.9% for regulatory capital and, typically, α = 99.95% to 99.98% for economic capital).

The exact asymptotic behavior of the MLE estimator when faced with potential deviating data, i.e. a new loss event, *x*, is given by the analytic formulae presented above in Section 3.7 and in the Appendix. The exact capital impact of the new loss event can be assessed by using these analytic formulae in combination with the specific information for the hypothetical data change scenario. For example, for a data change scenario in which an additional loss is to be included in the sample, the Influence Function describes the impact of adding an infinitesimal fraction of deviating data at severity amount *x*, so we obtain the new parameter estimates via:

$$T\left(F_\varepsilon\right) \approx \varepsilon \cdot IF\left(x;T,F\right) + T\left(F\right) \quad \text{where } \varepsilon = 1/n \text{ is used in practice} \qquad (16)$$

This new parameter estimate is then used in the SLA formula to generate a new estimate of capital, and the difference between this new capital and the baseline capital is the change in capital resulting from the new (or dropped) loss event.

The examples that follow represent some of the real-world business situations whose changing capital requirements can be informed, directly and exactly, by the Influence Function via the above. The specific values for the parameters of the severity distributions and the hypothetical loss events in each scenario have been modified to protect confidential and proprietary information. The examples apply the Influence Function approach to samples of 250 loss events from LogNormal, LogGamma, and GPD distributions, with data collection thresholds of $0, $10,000, and $25,000. By examining the resulting Exact Capital Sensitivity Curves we can see how deviations from the assumed distributions differentially affect the capital estimates based on different severity distributions, and (very) differentially affect capital estimates over different ranges of deviating loss values. A summary of the baseline parameters from the samples and the resulting capital estimates based on the SLA (12), assuming an annual loss frequency of λ = 25, are presented in Table 1.

Table 1: Baseline Parameter and Capital Estimates

Distribution	Parameter Names	Data Collect. Threshold (H)	Total "Historical" Loss [n = 250] ($ mill)	MLE Baseline Results: Parameter 1	Parameter 2	Capital ($mill) Regulatory ($\alpha$=0.999)	Economic (α=0.9997)
LogNormal	μ, σ	0	$61.2	10.953	1.749	$63.3	$99.0
LogNormal	μ, σ	10,000	$77.6	10.954	1.750	$69.1	$107.4
LogNormal	μ, σ	25,000	$77.3	10.917	1.749	$73.5	$113.0
LogGamma	α, β	0	$71.0	35.484	3.252	$359.0	$755.4
LogGamma	α, β	10,000	$94.0	35.513	3.263	$387.3	$809.3
LogGamma	α, β	25,000	$146.2	35.410	3.252	$464.8	$960.7
GPD	ξ, β	0	$48.9	0.8713	57,584	$459.8	$1,291.8
GPD	ξ, β	10,000	$64.8	0.8825	57,484	$583.1	$1,670.3
GPD	ξ, β	25,000	$77.1	0.8798	57,340	$680.8	$1,939.6

A very important finding to remember when considering the following capital results is MLE's apparent lack of B-robustness. Although we have left mathematical proofs for the less obvious cases for another paper, based on the derivations and results shown in Section 3.7 and the Appendix, all evidence points to non-robustness of MLE for all of the parameters of all of the severity distributions examined. More importantly, however, is that this lack of robustness is reflected in the behavior of the capital estimates shown in the next section. Across the entire domain of relevant loss events, this MLE non-robustness directly impacts capital estimates in very material, sometimes unexpected, and even completely counter-intuitive ways.

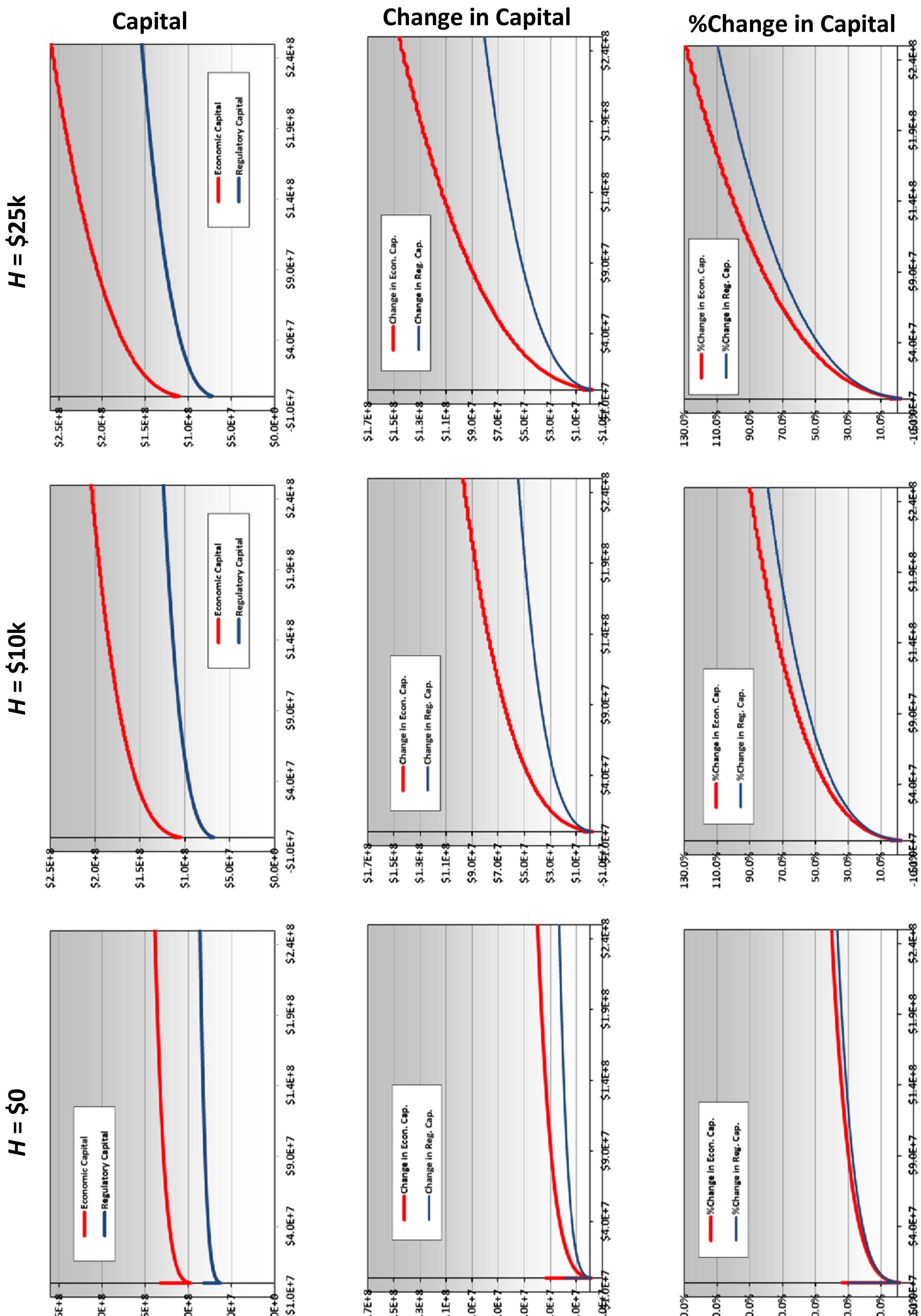
4.1. Case Study 1: New right tail loss of different possible severity amounts

Operational losses associated with litigation are a common occurrence in the banking industry. The existence of potential litigation brings an element of uncertainty into the capital planning process on the part of management. In some cases, management may request information on the potential capital requirements assuming alternative outcomes for the litigation.

Suppose that the institution faces a legal claim for an alleged operational loss related to the Advisory Services event subtype of CPBP (Clients, Products, and Business Practices), and that on the advice of counsel, it is determined that a loss reserve of $100 million be established in accordance with U.S. GAAP accounting rules (so based on available information, a loss of $100 million is probable and reasonably estimable). Suppose that this loss is recognized after the regular quarterly cycle of capital modeling and reporting has completed. Although the best estimate of the potential loss is $100 million at the time the loss is financially recognized, suppose that it is determined that the loss could be as low as $15 million if the litigation were to resolve favorably and could be as high as $200 million in the case of very adverse discovery or motion rulings. Management could very reasonably request an assessment of the potential capital implications of the three alternative loss scenarios.

The loss scenarios are the addition of a single individual loss with severity of $15, $100, or $200 million. This can be evaluated within the Influence Function framework via (16), or using the EIF by augmenting the baseline data sample with the additional hypothetical loss and re-estimating the severity parameters. An updated set of capital estimates is then calculated making use of the revised severity parameter estimates. If this process is repeated over a range of relevant loss severities, then the capital curves as in Figures 8a-i, 9a-i, and 10a-i can be plotted. Table 2 summarizes overall dollar impact of the hypothetical loss scenarios for Case 1 (addition of a right tail loss).

LogNormal Exact Sensitivity Capital Curves (n = 250, λ = 25)



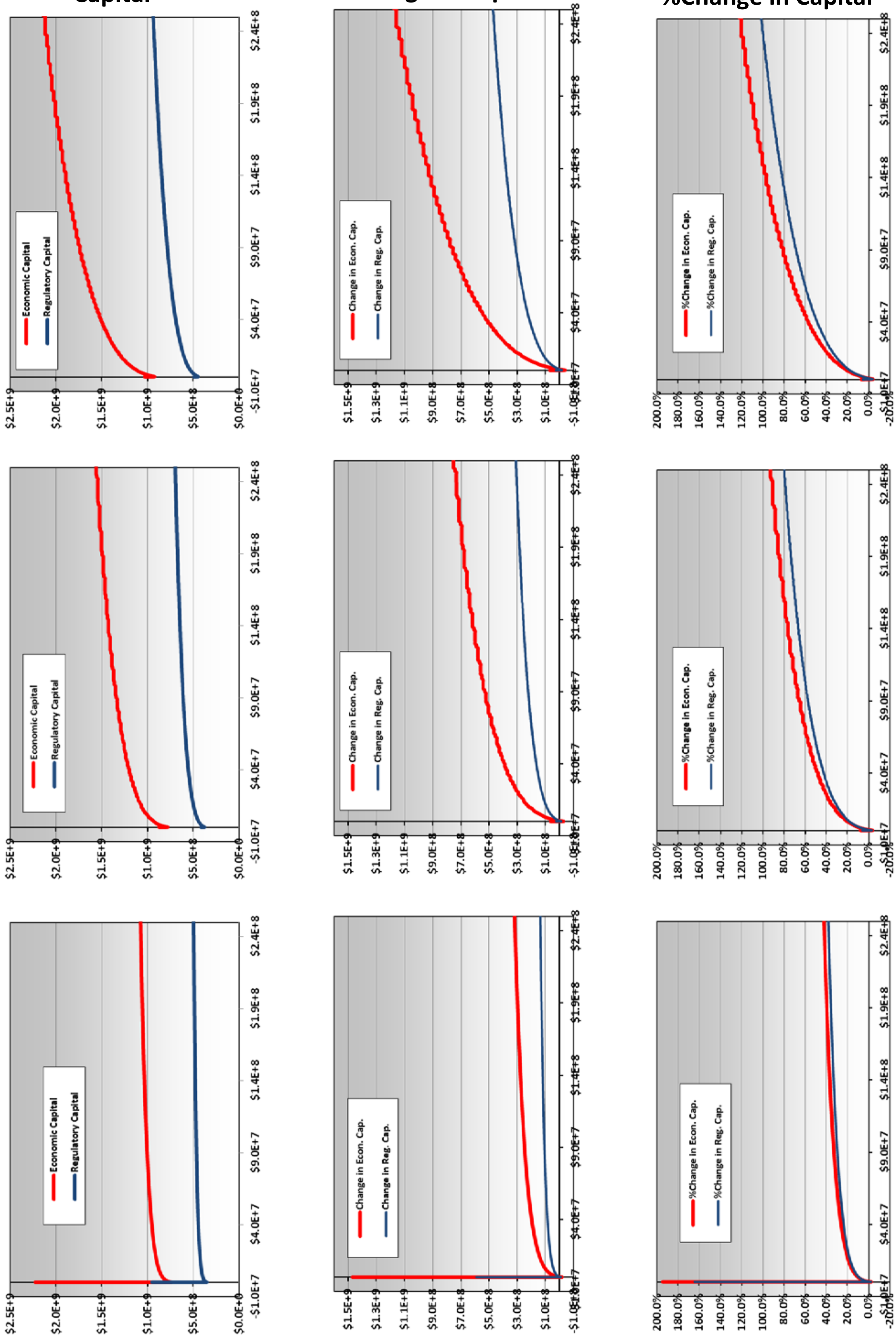
LogGamma Exact Sensitivity Capital Curves (n = 250, λ = 25)

Capital | **Change in Capital** | **%Change in Capital**

H = $25k

H = $10k

H = $0



GPD Exact Sensitivity Capital Curves (n = 250, λ = 25)

Capital | Change in Capital | %Change in Capital

H = \$25k

H = \$10k

H = \$0

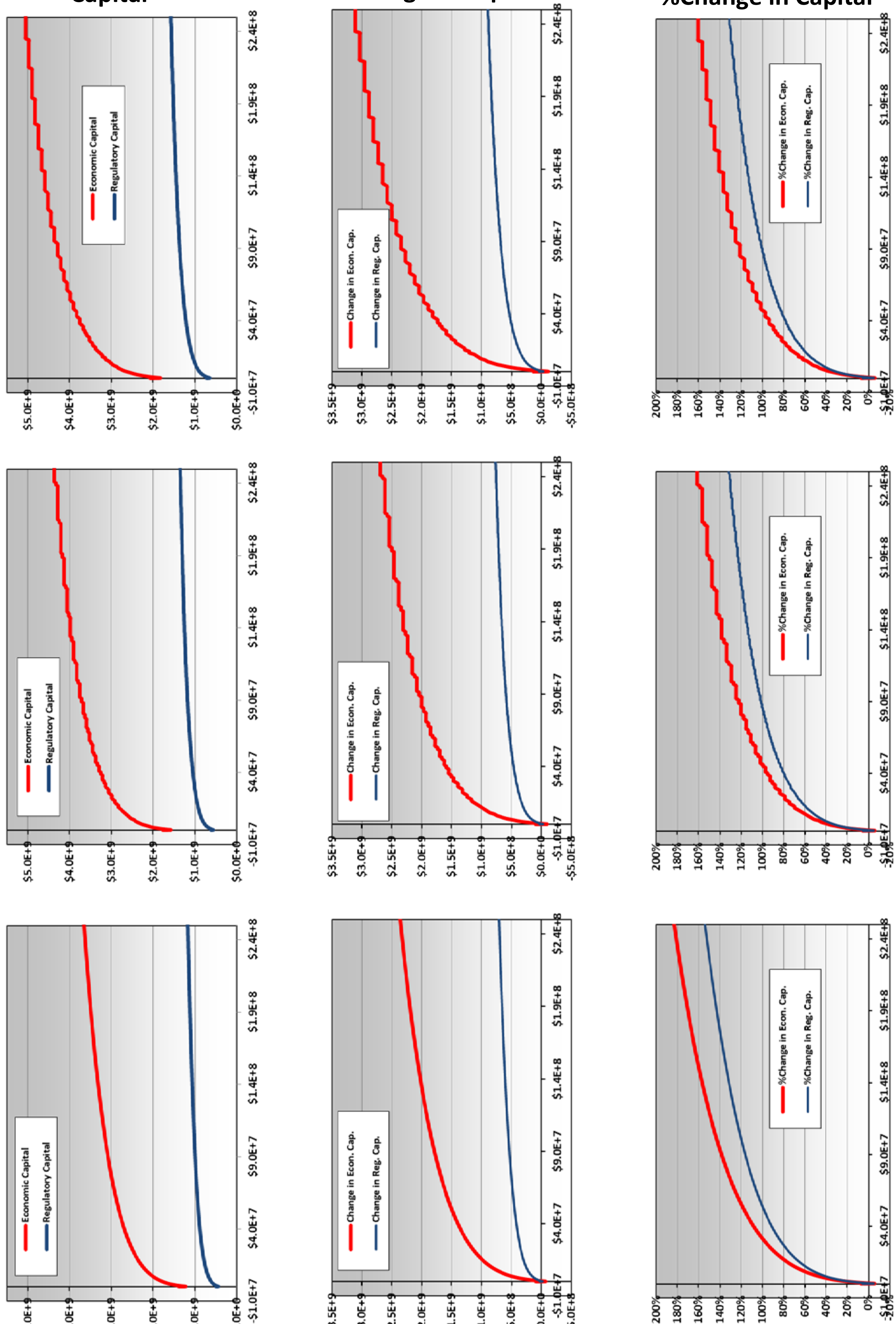


Table 2: Dollar Impact of an Additional Hypothetical Loss in the Right Tail

		Data Collection		Change from MLE Baseline due to Additional Loss					
		Threshold	Total	$15 million		$100 million		$200 million	
Estimator	Distribution	(H)	Loss ($m)	RC	EC	RC	EC	RC	EC
MLE	LogNormal	0	$61.2	$9.5	$16.0	$17.9	$30.6	$21.8	$37.4
MLE	LogNormal	10,000	$77.6	$17.2	$30.0	$38.9	$67.8	$50.5	$89.7
MLE	LogNormal	25,000	$77.3	$21.5	$37.9	$54.4	$97.4	$73.4	$132.6
MLE	LogGamma	0	$71.0	$62.5	$144.6	$109.3	$254.9	$129.5	$302.8
MLE	LogGamma	10,000	$94.0	$110.0	$254.3	$229.4	$551.3	$289.6	$695.0
MLE	LogGamma	25,000	$146.2	$140.2	$335.7	$332.9	$814.5	$433.9	$1,082.4
MLE	GPD	0	$48.9	$298.5	$987.1	$550.8	$1,844.7	$666.7	$2,229.2
MLE	GPD	10,000	$64.8	$331.5	$1,127.8	$609.7	$2,083.2	$724.8	$2,541.6
MLE	GPD	25,000	$77.1	$374.7	$1,280.7	$700.9	$2,427.0	$844.7	$2,961.9

		Data Collection		%Change from MLE Baseline due to Additional Loss					
		Threshold	Total	$15 million		$100 million		$200 million	
Estimator	Distribution	(H)	Loss ($m)	RC	EC	RC	EC	RC	EC
MLE	LogNormal	0	$61.2	15%	16%	28%	31%	34%	38%
MLE	LogNormal	10,000	$77.6	25%	28%	56%	63%	73%	83%
MLE	LogNormal	25,000	$77.3	29%	34%	74%	86%	100%	117%
MLE	LogGamma	0	$71.0	17%	19%	30%	34%	36%	40%
MLE	LogGamma	10,000	$94.0	28%	31%	59%	68%	75%	86%
MLE	LogGamma	25,000	$146.2	30%	35%	72%	85%	93%	113%
MLE	GPD	0	$48.9	65%	76%	120%	143%	145%	173%
MLE	GPD	10,000	$64.8	57%	68%	105%	125%	124%	152%
MLE	GPD	25,000	$77.1	55%	66%	103%	125%	124%	153%

		Data Collection		% of Total Loss					
		Threshold	Total	$15 million		$100 million		$200 million	
Estimator	Distribution	(H)	Loss ($m)	RC	EC	RC	EC	RC	EC
MLE	LogNormal	0	$61.2	15%	26%	29%	50%	36%	61%
MLE	LogNormal	10,000	$77.6	22%	39%	50%	87%	65%	116%
MLE	LogNormal	25,000	$77.3	28%	49%	70%	126%	95%	172%
			$0.0						
MLE	LogGamma	0	$71.0	88%	204%	154%	359%	183%	427%
MLE	LogGamma	10,000	$94.0	117%	270%	244%	586%	308%	739%
MLE	LogGamma	25,000	$146.2	96%	230%	228%	557%	297%	740%
			$0.0						
MLE	GPD	0	$48.9	610%	2018%	1126%	3771%	1363%	4557%
MLE	GPD	10,000	$64.8	512%	1742%	942%	3217%	1119%	3925%
MLE	GPD	25,000	$77.1	486%	1661%	909%	3148%	1096%	3842%

The results above make clear that the sensitivity of capital to a new large loss is greatly impacted by the assumed distribution of losses. The more heavy-tailed the loss distribution (i.e. having a larger data collection threshold within a distributional family or for LogGamma and GPD compared to the LogNormal), the greater the impact of an additional loss in the right tail. Moreover, the impact relative to the total loss in the loss sample can be extremely large, often 10, 20, and even more than 30 times the size of all previous losses put together! This brings into question the plausibility of the LDA framework, or at the very least, the use of MLE as a severity estimator, as the severity distributions above are widely used in this setting.

4.2. Case Study 2: New left tail loss

As financial institutions increase the business use of operational risk quantification models, operational risk practitioners are increasingly asked to explain to management why capital has changed from the prior period estimate. Many business users of operational risk capital estimates reasonably believe that capital estimates should be stable from quarter to quarter when the institution's risk profile is stable; should increase when the institution's risk profile increases (e.g. due to an increase in the scale of operations through acquisition, or due to the realization of losses in the right tail of the distribution); and should decrease when the institution's risk profile decreases (e.g. due to a decrease in loss frequency). Operational risk practitioners note that satisfactory attribution analysis of capital changes for the business audience can be quite elusive, because even though mathematical or statistical explanations can be provided, the resulting capital impacts can be quite unexpected to all constituents, both business users and operational risk quantitative experts. This case study demonstrates how the influence function can be used as a statistical tool to explain changes in capital, especially when such changes are counter-intuitive, and simulation approaches, unlike the IF, provide no definitive answers as to why.

Suppose that an institution has observed a total of 250 loss events within the unit of measure and that only one additional loss is expected to enter the loss database in the next period. Suppose that this individual event happened to be very close to the data collection threshold, specifically, the data collection threshold plus $1,000 (which is generally below the 5^{th} percentile of each of the loss distributions, but certainly not extremely unlikely).

The loss scenarios are the addition of a single individual loss with severity of $1,000, $11,000, or $26,000 (for *H* = $0, $10,000, and $25,000, respectively). As in Case 1, this can be studied within the Influence Function framework by using (16) or augmenting the baseline data sample with the additional hypothetical loss and re-estimating the severity parameters. An updated set of capital estimates is then calculated making use of the revised severity parameter estimates.

Table 3: Dollar Impact of an Additional Hypothetical Loss in the Left Tail

		Data Collection		Change from MLE Baseline due to Additional Loss					
		Threshold	Total	$1,000		$11,000		$26,000	
Estimator	Distribution	(H)	Loss ($m)	RC	EC	RC	EC	RC	EC
MLE	LogNormal	0	$61.2	$2.6	$4.5	-$0.5	-$0.8	-$0.8	-$1.4
MLE	LogNormal	10,000	$77.6			$2.3	$4.2	$0.0	$0.0
MLE	LogNormal	25,000	$77.3					$2.4	$4.2
MLE	LogGamma	0	$71.0	$29.4	$70.4	-$5.2	-$11.5	-$7.8	-$17.9
MLE	LogGamma	10,000	$94.0			$21.6	$52.6	-$0.1	-$0.1
MLE	LogGamma	25,000	$146.2					$24.0	$66.9
MLE	GPD	0	$48.9	$25.9	$85.7	$10.1	$34.1	-$4.2	-$12.8
MLE	GPD	10,000	$64.8			$28.8	$95.5	$4.8	$19.1
MLE	GPD	25,000	$77.1					$38.4	$133.7

		Data Collection		%Change from MLE Baseline due to Additional Loss					
		Threshold	Total	$1,000		$11,000		$26,000	
Estimator	Distribution	(H)	Loss ($m)	RC	EC	RC	EC	RC	EC
MLE	LogNormal	0	$61.2	4%	5%	-1%	-1%	-1%	-1%
MLE	LogNormal	10,000	$77.6			3%	4%	0%	0%
MLE	LogNormal	25,000	$77.3					3%	4%
MLE	LogGamma	0	$71.0	8%	9%	-1%	-2%	-2%	-2%
MLE	LogGamma	10,000	$94.0			6%	7%	0%	0%
MLE	LogGamma	25,000	$146.2					5%	7%
MLE	GPD	0	$48.9	6%	7%	2%	3%	-1%	-1%
MLE	GPD	10,000	$64.8			5%	6%	1%	1%
MLE	GPD	25,000	$77.1					6%	7%

		Data Collection		% of Total Loss					
		Threshold	Total	$1,000		$11,000		$26,000	
Estimator	Distribution	(H)	Loss ($m)	RC	EC	RC	EC	RC	EC
MLE	LogNormal	0	$61.2	4%	7%	-1%	-1%	-1%	-2%
MLE	LogNormal	10,000	$77.6			3%	5%	0%	0%
MLE	LogNormal	25,000	$77.3					3%	5%
MLE	LogGamma	0	$71.0	41%	99%	-7%	-16%	-11%	-25%
MLE	LogGamma	10,000	$94.0			23%	56%	0%	0%
MLE	LogGamma	25,000	$146.2					16%	46%
MLE	GPD	0	$48.9	53%	175%	21%	70%	-9%	-26%
MLE	GPD	10,000	$64.8			44%	148%	7%	29%
MLE	GPD	25,000	$77.1					50%	173%

As Table 3 shows, in Case 2 the story is less the change in capital relative to the baseline and more the change relative to the size of the new loss. These results are quite dramatic: a $4.5m increase in economic capital results from a $1,000 loss under a LogNormal model; a $21m increase in regulatory capital results from a $11,000 loss under a Truncated LogGamma

(*H*=$10k) model; and a $133.7m increase in economic capital results from a $26,000 loss under a Truncated GPD (*H*=$25k) model. These capital increases appear extremely disproportionate with the new loss amount, and yet they are completely consistent with the IFs derived above in Section 3.7 and in the Appendix. For example, recall the IF of the MLE LogNormal parameters in equation (10):

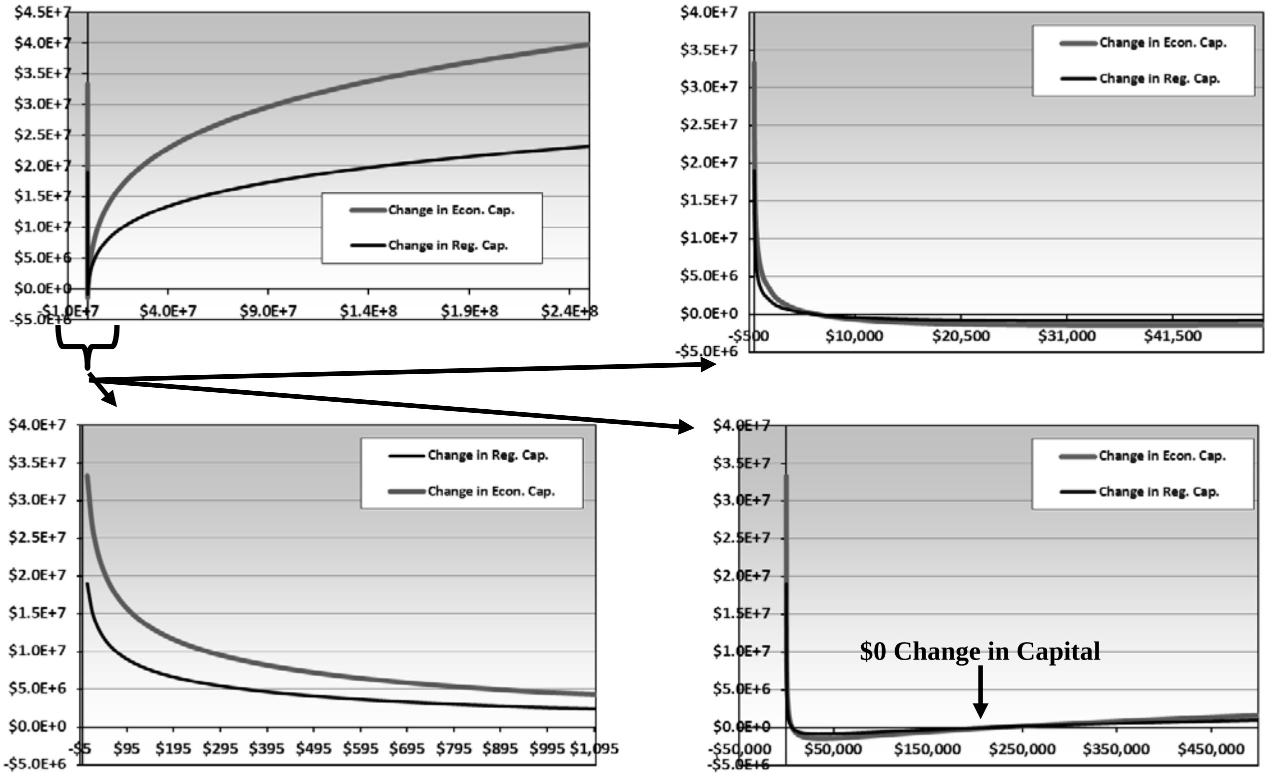
$$IF_{\theta}\left(x;\theta,T\right)=\begin{bmatrix}\ln\left(x\right)-\mu \\ \dfrac{\left(\ln\left(x\right)-\mu\right)^{2}-\sigma^{2}}{2\sigma}\end{bmatrix} \tag{10}$$

As seen in (10), and as Graphs 2a-e showed, when $x\rightarrow 0^{+}$, $\sigma\rightarrow+\infty$ much faster than $\mu\rightarrow-\infty$, because $\ln\left(x\right)$, which becomes a very large negative number as $x\rightarrow 0^{+}$, is squared in IF_{σ}, but not IF_{μ}. So σ will increase without bound, causing the entire LogNormal severity – all of its percentiles – to increase without bound. This causes the capital estimate based on (14) to increase without bound, because (14) is a direct function of the specified percentile of the LogNormal severity. So if a new loss was even smaller than $1,000, say $10, capital would increase even more – increases of $19,019,123 and $33,292,687, in fact, for regulatory and economic capital, respectively. The same increases under the LogGamma, which is characterized by an even more extreme asymptotic behavior as $x\rightarrow 1^{+}$, would be $590,889,232 and $1,469,816,763, respectively. Of course these numbers are absurd, but they are inescapable mathematical consequences of using MLE estimators. And while few, if any banks would include $10 losses in their severity models, every single one, by definition, will be conducting severity modeling on loss event datasets with losses within a few thousand dollars of their respective data collection thresholds. Truncation does mitigate to some degree the extreme asymptotic behavior of the MLE estimators and the capital estimates based on them, but as shown in Table 3 above, it certainly does not eliminate it altogether.

These extreme results are actually shown in the full Exact Capital Sensitivity Curves in Graphs 8a-i, 9a-i, and 10a-i, but they are difficult to see because of the large scales of the axes. Below are the same graphs, for changes in capital, with expanded scales for the LogNormal as an example, to highlight the counterintuitive and dramatic affect that small left-tail losses have on capital estimation.

Graphs 11a-11d: Exact Capital Sensitivity Curves (MLE), Scaled and Not Scaled, based on LogNormal (μ = 10.95 and σ = 1.75), by Arbitrary Deviation, *x*

Complete LogNormal ECSC **Scaled Axes LogNormal ECSC**



Given Table 3 and Graphs 11a-11d, it is no wonder that most banks relying on MLE-based LDA experience tremendous quarter-to-quarter instability in their capital requirements. All it takes is a few losses near, say, the $10,000 threshold to add many tens of millions of dollars to estimates of required capital – and this is a "correct" result, based on MLE estimators!

4.3. Case Study 3: Removal of a current loss

In practice, operational loss event databases evolve over time and financial institutions estimate capital requirements using the state of the database at a particular point in time. Depending on an institution's data recording policies and internal governance, individual loss events may appear in the loss event database only for a period of time. Many institutions have a thorough data review and approval process for entries to the loss event database but include events in "draft" status in the modeling dataset while the review and approval process is underway. A loss event may be removed from the loss event database for any of number of potential reasons, such as:

- Change to loss severity – the severity of an operational loss may exceed the data collection threshold at the time the record was first entered to the loss event database, but may be found to fall below the threshold during the review process.

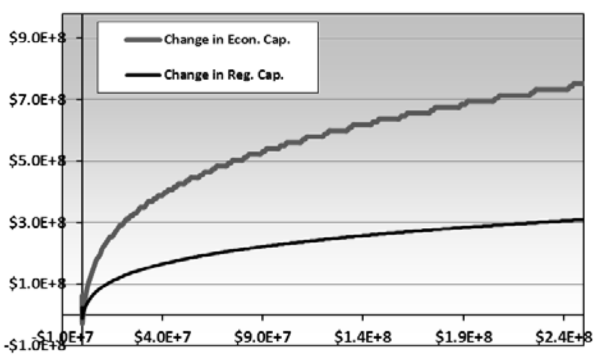
- <u>Reclassification of loss event</u> – the precise determination of whether a particular event is an operational risk, or represents some other type of risk such as credit risk, business risk, or strategic risk can take some time to complete. Modeling datasets include only losses due to operational risk, and such events are removed when determined to represent a different risk class.

This type of data change can affect individual loss events across the entire range of the loss distribution.

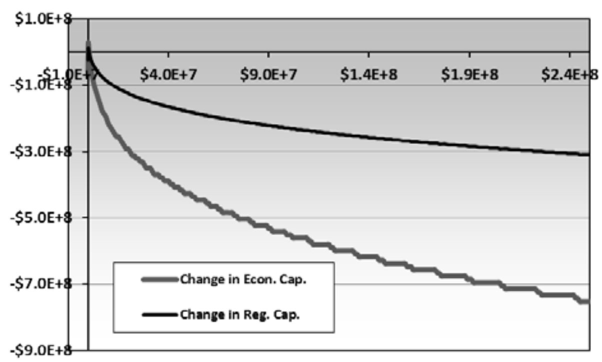
Suppose that an institution has observed a total of 250 loss events within the unit of measure and that one event in draft status will be removed from the loss event database in the subsequent quarter. Suppose that this individual event is either a left tail event (very close to the data collection threshold) or a right tail event. Since these hypothetical data change scenarios reflect the removal of a loss, we simply change the sign of the usual Influence Function to capture the impact of removing data with arbitrary loss amount *x*. The resulting capital curves are the same as Graphs 8a-i, 9a-i, and 10a-i, but the change in capital is multiplied by negative one. Only one illustrative example – that of the IF of the MLE estimators of the Truncated LogGamma – is shown below for purposes of brevity.

Graphs 12a-12b: Exact Capital Sensitivity Curves (MLE), Loss Added / Dropped, based on Truncated LogGamma (*a* = 35.5, *b* = 3.25, *H*=$10k), by Arbitrary Deviation, *x*

Loss Added



Loss Dropped



4.4. Case Study 4: Reclassification of an individual loss

Since operational loss event databases evolve over time, any number of important data fields may change as more information is learned about an individual loss event. Any number of issues may result in the reclassification of a loss in such a way that it no longer belongs in a particular unit of measure. Some examples include:

- <u>Reclassification of loss event type</u> – when a particular operational loss is first financially recognized, the details of the loss pathway may be unclear. For example, a business unit may recognize an operational loss for an event believing it to result from a type of improper business practice (a subclass of Basel loss event type CPBP) when subsequent review identifies the loss as a transaction error event (a subclass of EDPM (Execution Delivery and Process Management)).

- Reclassification of business line – individual business units or subgroups may dispute which business is responsible for a particular operational loss. Subsequent internal discussions may result in a shared allocation of a loss event or even complete reassignment from one business to another.

- Treatment of corporate events – in some cases, an operational loss initially may be recognized by a particular business unit, but upon further internal review, may be reassigned to the corporation as a whole.

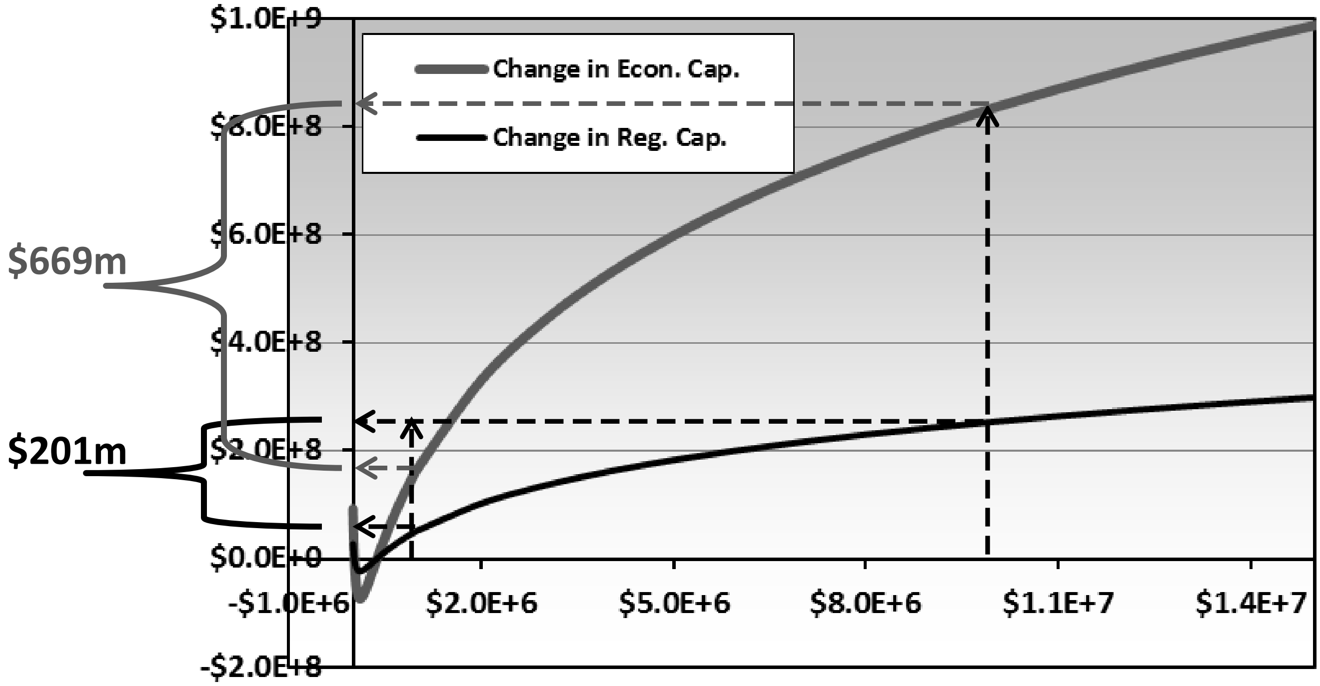
Such reclassifications can affect individual loss events across the entire range of the loss distribution. For the unit of measure that receives the reclassified loss event, the Influence Function analysis follows the form of Case 1 or Case 2. For the unit of measure from which the loss event is removed, the Influence Function analysis follows the form of Case 3. The aggregate effect is calculated when the VaR's are combined across all units of measure.

4.5. Case Study 5: Revision of a current loss

As previously explained, there are a variety of reasons why the characteristics of an operational loss event may change over time in the database. The examples above have focused on the capital impacts of adding or removing data points from the estimation sample, but it is quite common for the severity of individual loss events to change during the current period. This can occur for events that are in "draft" status, or due to updates of certain components of the loss amount (such as transaction fees, taxes, penalties, attorney fees, etc.). Suppose the original loss severity is an amount x_0, and the revised loss amount is x_1. To accommodate this type of data change in the Influence Function framework, (16) is simply applied using x_0 and a new capital requirement is calculated; and then (16) is applied a second time using x_1 and another capital requirement is calculated. The difference between these two capital requirements is the difference in the *expected* capital effect.

Suppose that an institution has observed a total of 250 loss events within the unit of measure and that, under a GPD (ξ = 0.875 and β = 57,500) severity model, a $1 million loss event has its severity revised to $10 million. For such data changes, the resulting change in *expected* change in capital is just the difference between the two points on the capital curve; that is, the change in capital associated with the $1m loss subtracted from the change in capital associated with the $10m loss, as described above. This is shown on Graph 13 below. Regulatory capital would have increased by $52.1m due to the $1m loss, but it changed by $253.5m because the loss was really $10m, for a difference of about $201m. Economic capital would have increased by $166.3m but it actually changed by $834.9m for a difference of about $668.5m.

Graph 13: Exact Capital Sensitivity Curves (MLE), Current Loss Changed, based on GPD (ξ = 0.875 and β = 57,500), by Arbitrary Deviation, *x*



4.6. Case Study 6: A Retrospective Exact Attribution/ "But For" Analyses

Due in part to the quarter to quarter instability in capital estimation, bank management may request an attribution analysis in an attempt to understand why capital requirements changed from the previous period. This is especially true when no major new losses were recorded and the bank's risk profile did not change in any notable way, but nonetheless, a dramatic movement in estimated required capital is observed (Case 2 is one way this can happen).

Suppose that an institution has observed a total of 250 loss events within the unit of measure and that three losses were recorded in the previous quarter. Management could reasonably ask what capital requirements would have been "but for" loss #2. This is simply Case 3 above, where the loss dataset used is the one that existed at the time of the quarter in question, and the dropped loss is loss #2. Or management could ask, "What would capital requirements have been if only loss #1, or only loss #2, or only loss #3 occurred?" This is simply Case 1 or 2, applied to each loss separately, and the loss dataset used is the one that existed at the time of the quarter in question excluding the other two losses. The three resulting capital estimates can then be compared to provide some measure of the contribution of each to the overall change in capital. The effects of each could be offsetting, or in the same direction, all augmenting the overall change in capital, but either way such an analysis would identify whether, for example, one small left tail loss event was driving 99% of the change in capital; or whether a loss of, say, about $220,000 actually had no effect on capital whatsoever, as shown in Graph 11d above.

This type of "but for" exact sensitivity analysis is particularly helpful in explaining the surprising capital response to a cluster losses of similar severity that are very near the data collection threshold, which is not an uncommon occurrence. When a bank's internal constituents (such as senior management or other business users of operational risk capital estimates) or external constituents (regulatory supervisors) find quarter to quarter capital changes to be "out of proportion" to the underlying data changes (e.g. Case 2 above), there may be calls for independent validation of the results by the bank's internal audit group or the bank's model validation group The Influence Function and its associated Exact Capital Sensitivity Curves, however, can make immediately apparent the sources of the "out of proportion" effects, and demonstrate definitively and absolutely how capital estimates using MLE-based severity parameter estimates can very easily display unexpected, material, and even "counter-intuitive" capital requirements.

5. Conclusions

Given the well documented and extensive empirical challenges of operational risk loss data and the methodological challenges inherent in the AMA framework (specifically, estimating very high percentiles of the aggregate loss distribution), reliable estimation of both economic and regulatory capital without bias, with acceptable precision, and with reasonable robustness remains a very formidable exercise. In this chapter, we demonstrated how the Influence Function can be used in this effort as a definitive, analytic tool for two essential purposes: a) to inform the development and choice of severity estimators, which unarguably remain the main drivers of capital estimation in the Loss Distribution Approach framework; and b) to perform direct capital planning, once an estimator has been selected, that permits the *exact* determination of capital needs under alternative hypothetical changes to the loss data used in severity modeling.

For the former objective, we demonstrated how the Influence Function can very effectively highlight the failure of the most widely used estimator (namely, MLE) to provide robust, reliable, and stable capital estimates under a wide range of commonly encountered conditions. But our main focus in this chapter has been on the latter objective. The main advantage of the IF for capital planning lies in the fact that it is an analytically derived, deterministic formula. As such, it is the superior alternative, when assessing the behavior of capital estimates under varying conditions, to simulation-based approaches that are often resource-intensive, arguably subjective, and often inconclusive as they are unable to definitively confirm or invalidate counter-intuitive results. The IF literally provides the definitive answer to the question, "If my bank is subjected to a new $10m loss, or a $50m loss, or even a $200m loss, what will be the exact effect on my capital requirements?" As a relatively straightforward formula, the IF provides this exact answer with no additional estimation error beyond that of the already estimated severity and frequency parameters.

The Influence Function is readily programmed in most software systems for statistical or mathematical modeling, and in fact this was done to provide the results presented herein. We provide the blueprint for doing this by presenting the derivations for the Influence Functions of the Maximum Likelihood Estimators for the parameters of some of the most commonly used loss distributions, namely, the LogNormal, LogGamma, and GPD distributions. We also provide the derivations for each of these distributions when they are truncated on the left, which is the most common and accepted method for dealing with loss data recorded subject to a data collection threshold. In addition, we describe how the Empirical Influence Function (EIF) can be used as a very simple yet accurate approximation of the asymptotic Influence Function. It is good practice to use the IF and the EIF simultaneously, to both know definitively the behavior of the estimator over the entire domain of possible loss events, as well as to have a useful and easily implemented verification of the more involved calculations required for some IFs.

Finally, we illustrated the practical use of the Influence Function for capital planning by using its results to generate Exact Capital Sensitivity Curves. These demonstrate the definitive, exact capital impacts under six realistic data-change scenarios that might arise within a financial institution. Through these scenarios, we demonstrate the inherent instability of capital estimates using the LDA (Loss Distribution Approach) with MLE-estimated severity parameters. Unfortunately, MLE's non-robustness directly translates into non-robustness in its capital estimates. The instability of MLE-based capital estimates is sometimes very dramatic and worse, occurs under unexpected conditions (e.g. new, small left-tail losses) and in counter-intuitive ways, increasing dramatically and without bound as the severity of a new loss actually *decreases* dramatically. This behavior is exactly the opposite of the business requirements for operational risk capital estimates: stability, reliability, robustness, and precision. Given the potentially material changes in capital that can result from changes to the underlying data using the LDA/MLE method, it may be prudent for operational risk practitioners to inform management and business users of operational risk capital estimates about the range of potential capital outcomes for different data change scenarios. Management and other business users need a better understanding of how capital requirements under LDA/MLE may be impacted by data changes, as capital instability may negatively impact medium to long term strategic plans. And we believe there is no more effective tool to communicate this than the Influence Function.

Regarding next steps, as described above the IF can be used to assess the behavior of virtually any estimator, applied to any of the commonly used severity distributions in operational risk modeling. Alternatives to MLE should be sought out and/or developed. In fact, Opdyke and Cavallo (2012) present initial results of similar tests on a widely used B-robust estimator, the OBRE (Optimally Bias-Robust Estimator).[40] Preliminary results of OBRE-based capital estimates show a respectable mitigation of MLE's extreme sensitivity vis-à-vis new, small, left-tail loss events. However, under some conditions, OBRE-based capital estimates can exhibit what is arguably too much robustness on the other extreme, with relatively flat capital requirements over large ranges of very large new losses. OBRE's robustness tuning parameter may provide an effective method for getting around this possible limitation, and this is currently

[40] OBRE is a B-robust estimator that is essentially a constrained MLE. As such it preserves efficiency under data-change conditions consistent with the presumed severity distribution, but is resistant to extreme data-change conditions inconsistent with the presumed severity distribution.

being researched. Regardless of the estimators ultimately used, given the depth of the challenges in estimating operational risk capital (both empirical and methodological), and the inherent limitations of the LDA framework, it is unlikely that any estimator will serve as a panacea under all possible severity distributions and all possible data conditions and all possible new or revised loss scenarios, providing universally reasonable capital estimates and a capital distribution that is robust, unbiased, and reasonably precise. However, we have demonstrated that an absolutely essential tool in this effort that will aid in estimator development and choice, and especially in direct capital planning, under the complete domain of data-change conditions, any estimator, and any severity distribution, is the Influence Function.

APPENDIX: Influence Functions of LogNormal, LogGamma, and GPD Parameters Under Truncation

Truncated LogNormal:

The Truncated LogNormal distribution is defined as:

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

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

where f () and F() are the pdf and cdf of the LogNormal (see section 3.7 above).

Inserting the derivatives 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 $A(\theta)=\begin{bmatrix} -\int_a^b \frac{\partial\varphi_{\theta_1}}{\partial\theta_1}dK(y) & -\int_a^b \frac{\partial\varphi_{\theta_1}}{\partial\theta_2}dK(y) \\ -\int_a^b \frac{\partial\varphi_{\theta_2}}{\partial\theta_1}dK(y) & -\int_a^b \frac{\partial\varphi_{\theta_2}}{\partial\theta_2}dK(y)\end{bmatrix}$ yields

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

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

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

and into the psi function yields $$\varphi_\theta = \begin{bmatrix} -\left[\frac{\ln(x)-\mu}{\sigma^2}\right] - \frac{\int_0^H \left[\frac{\ln(y)-\mu}{\sigma^2}\right] f(y;\mu,\sigma)dy}{1-F(H;\mu,\sigma)} \\ -\left[\frac{\left(\ln(x)-\mu\right)^2}{\sigma^3} - \frac{1}{\sigma}\right] - \frac{\int_0^H \left[\frac{\left(\ln(y)-\mu\right)^2}{\sigma^3} - \frac{1}{\sigma}\right] f(y;\mu,\sigma)dy}{1-F(H;\mu,\sigma)} \end{bmatrix} \quad \text{(A.1)}$$

So via (9), the Influence Function of the MLE parameters of the Truncated LogNormal severity – $IF_\theta(x;\theta,T) = A(\theta)^{-1}\varphi_\theta$ – is computed numerically.

As seen in the graphs in section 3.7, note that the non-zero cross derivatives in $A(\theta)$ above introduces parameter dependence in *x*, which dramatically changes the behavior of the parameters and the resulting capital estimates as a function of *x*.

<u>Truncated LogGamma</u>:

For the Truncated LogGamma, we have:

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

for $0<H \le x<\infty;\ 0<a;\ 0<b$

where *f* () and *F*() are the pdf and cdf of the LogGamma (see section 3.7 above).

Inserting the derivatives 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 $A(\theta)=\begin{bmatrix} -\int_a^b \frac{\partial\varphi_{\theta_1}}{\partial\theta_1}dK\left(y\right) & -\int_a^b \frac{\partial\varphi_{\theta_1}}{\partial\theta_2}dK\left(y\right) \\ -\int_a^b \frac{\partial\varphi_{\theta_2}}{\partial\theta_1}dK\left(y\right) & -\int_a^b \frac{\partial\varphi_{\theta_2}}{\partial\theta_2}dK\left(y\right) \end{bmatrix}$ yields

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

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

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

and into the psi function yields

$$\varphi_\theta = \begin{bmatrix} -\left[\ln(b)+\ln(\ln(y))-digam(a)\right]-\dfrac{\int_1^H\left[\ln(b)+\ln(\ln(y))-digam(a)\right]f(y;a,b)dy}{1-F(H;\mu,\sigma)} \\ -\left[\dfrac{a}{b}-\ln(y)\right]-\dfrac{\int_1^H\left[\dfrac{a}{b}-\ln(y)\right]f(y;a,b)dy}{1-F(H;\mu,\sigma)} \end{bmatrix} \quad \text{(A.2)}$$

So via (9), the Influence Function of the MLE parameters of the Truncated LogGamma severity – $IF_\theta(x;\theta,T)=A(\theta)^{-1}\varphi_\theta$ – is computed numerically.

As seen in the graphs in section 3.7, note that the non-zero cross derivatives in $A(\theta)$ above augments parameter dependence in *x*, which changes the behavior of the parameters and the resulting capital estimates as a function of *x*.

<u>Truncated GPD</u>:

For the Truncated GPD, we have:

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

for $0<H\le x<\infty;\ 0<\beta;$ assuming $\varepsilon>0$ (which is appropriate in this setting) where *f* () and *F*() are the pdf and cdf of the GPD (see section 3.7 above).

Inserting the derivatives 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 $A(\theta)=\begin{bmatrix} -\int_a^b \frac{\partial\varphi_{\theta_1}}{\partial\theta_1}dK(y) & -\int_a^b \frac{\partial\varphi_{\theta_1}}{\partial\theta_2}dK(y) \\ -\int_a^b \frac{\partial\varphi_{\theta_2}}{\partial\theta_1}dK(y) & -\int_a^b \frac{\partial\varphi_{\theta_2}}{\partial\theta_2}dK(y) \end{bmatrix}$ yields

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

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

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

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

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

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

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

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

and into the psi function yields

$$\varphi_\theta = \left[\begin{array}{c} -\left[-\frac{1}{\beta}\left[\frac{\beta - x}{\beta + \varepsilon x}\right]\right] - \frac{\int_0^H -\frac{1}{\beta}\left[\frac{\beta - x}{\beta+\varepsilon x}\right] f(x;\beta,\varepsilon)dx}{1 - F(H;\mu,\sigma)} \\ -\left[\left(\frac{-x(1+\varepsilon)}{\beta\varepsilon + \varepsilon^2 x}\right) + \frac{\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right] - \frac{\int_0^H \left[\left(\frac{-x(1+\varepsilon)}{\beta\varepsilon+\varepsilon^2 x}\right) + \frac{\ln\left(1+\frac{\varepsilon x}{\beta}\right)}{\varepsilon^2}\right] f(x;\beta,\varepsilon)dx}{1 - F(H;\mu,\sigma)} \end{array} \right] \quad \text{(A.3)}$$

So via (9), the Influence Function of the MLE parameters of the Truncated GPD severity – $IF_\theta(x;\theta,T) = A(\theta)^{-1}\varphi_\theta$ – is computed numerically. As seen in the graphs in section 3.7, note that the non-zero cross derivatives in $A(\theta)$ above indicates parameter independence in *x*, as existed in the non-truncated case.

References

Alaiz, P., and Victoria-Feser, M. (1996), "Modelling Income Distribution in Spain: A Robust Parametric Approach," DARP Discussion Paper, No. 20 (PDF), STICERD, LSE. Available online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1094765.

Basel Committee on Banking Supervision. (2006), "International convergence of capital measurement and capital standards," Bank of International Settlements. Available online at http://www.bis.org/publ/bcbs128.htm.

Basel Committee on Banking Supervision. (2011), "Operational Risk – Supervisory Guidelines for the Advanced Measurement Approaches," Bank of International Settlements. Available online at http://www.bis.org/publ/bcbs196.htm.

Böcker, K., and Sprittulla, J. (2006), "Operational VaR: meaningful means." *RISK Magazine*, 19 (12), 96–98.

Chapelle A., Crama Y., Hubner G., and Peters J.P., (2008), "Practical Methods for Measuring and Managing Operational Risk in the Financial Sector: A Clinical Study," *Journal of Banking & Finance*, 32, 1049-1061.

Chernobai, A., and Rachev, S. (2006), "Applying robust methods to operational risk modelling." *The Journal of Operational Risk*, 1 (1), 27–41.

Cope, E. (2010), "Modeling operational loss severity distributions from consortium data," *The Journal of Operational Risk*, 5 (4), 35-64.

Cope, E. (2011), "Penalized likelihood estimators for truncated data," *Journal of Statistical Planning and Inference*, 141 (1), 345-358.

Cope, E., and Labbi, A. (2008), "Operational risk scaling by exposure indicators: evidence from the ORX database," *The Journal of Operational Risk*, 3 (4), 25–46.

Cope, E., Mignola, G., Antonini, G., Ugoccioni, R. (2009), "Challenges and pitfalls in measuring operational risk from loss data," *The Journal of Operational Risk*, 4 (4), 3–27.

Cope, E., Piche, M., and Walter, J. (2011), "Macroenvironmental determinants of operational loss severity," *Journal of Banking and Finance*, 36(5), 1362-1380.

Daníelson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O. and Shin, H.S. (2001), "An Academic Response to Basel II," LSE Financial Markets Group, Special Paper No 130. Available online at http://www.bis.org/bcbs/ca/fmg.pdf.

de Fontnouvelle, P., DeJesus-Rueff, V., Jordan, J., and Rosengren, E. (2003), "Capital and Risk: New Evidence on Implications of Large Operational Losses," Working Paper 03-5, Federal Reserve Bank of Boston. Available online at http://www.bos.frb.org/economic/wp/wp2003/wp035.htm.

de Fontnouvelle, P., Jordan, J., and Rosengren, E. (2006), "Implications of Alternative Operational Risk Modelling Techniques," in M. Carey and R. Stulz, eds., The Risks of Financial Institutions, NBER/University of Chicago Press.

Degen, M. (2010), "The Calculation of Minimum Regulatory Capital Using Single-loss Approximations," *The Journal of Operational Risk*, 5 (4), 3-17.

Dell'Aguila, R., and Embrechts, P., (2006), "Extremes and Robustness: A Contradiction?" *Financial Markets and Portfolio Management*, Vol. 20, 103-118.

Dupuis, D.J. (1999), "Exceedances Over High Thresholds: A Guide to Threshold Selection," *Extremes*, 1 (3), 251-261.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

Embrechts, P., Furrer, H. and Kaufmann, R. (2003), "Quantifying Regulatory Caiptal for Operational Risk," working paper, research supported by Credit Suisse Group, Swiss Re and UBS AG through RiskLab, Switzerland. http://www.bis.org/bcbs/cp3/embfurkau.pdf

Ergashev, B. (2008), "Should risk managers rely on the maximum likelihood estimation method while quantifying operational risk?" *The Journal of Operational Risk*, 3 (2), 63-86.

Frachot, A., Moudoulaud, O., Roncalli, T. (2004), "Loss Distribution Approach in practice," In Ong, M. (Ed.), The Basel Handbook: A Guide for Financial Practitioners, Risk Books, London.

Greene, W. (2007), Econometric Analysis, 6th Edition. Prentice Hall, Upper Saddle River, NJ.

Hampel, F.R. (1968), "Contributions to the Theory of Robust Estimation," Ph.D. thesis, University of California, Berkeley.

Hampel, F.R., E. Ronchetti, P. Rousseeuw, and W. Stahel. (1986), Robust Statistics: The Approach Based on Influence Functions, John Wiley and Sons, New York.

Horbenko, N., Ruckdeschel, P. and Bae, T. (2011), "Robust Estimation of Operational Risk," *The Journal of Operational Risk*, 6 (2), 3-30.

Huber, P.J. (1964), "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35, 73-101.

Huber, P.J. (1981), Robust Statistics, John Wiley & Sons, New York.

Huber, P.J., and Ronchetti, E. (2009), Robust Statistics, 2nd Edition, John Wiley & 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.

McNeil, A.J., Frey, R., and Embrechts, P. (2005). Quantitative Risk Management: Concepts, Techniques and Tools, Princeton University Press, Princeton, NJ.

Moscadelli, M. (2004). "The modeling of operational risk: The experience from the analysis of the data collected by the Risk Management Group of the Basel Committee," Working Paper No. 517, Bank of Italy.

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," *The Journal of Operational Risk*, forthcoming.

Ruckdeschel, P., and Horbenko, N. (2010), "Robustness properties of estimators in generalized Pareto models," Technical Report ITWM 182. Available online at http://www.itwm.fraunhofer.de/presse/berichte-des-itwm.html.

Smith, J. (1987), "Estimating the Upper Tail of Flood Frequency Distributions," *Water Resources Research*, Vol.23, No.8, 1657-1666.

Stefanski, L., and Boos, D. (2002), "The Calculus of M-Estimation," *The American Statistician*, 56 (1), 29-38.

Tukey, J.W. (1960), "A Survey of Sampling from Contaminated Distributions," in I. Olkin, ed., Contributions to Probability and Statistics, Stanford University Press, Stanford, CA., 448-485.

van Belle, G. (2002), Statistical Rules of Thumb, John Wiley & Sons, Inc., New York, New York.