Gradients of Risk Measures: Theory and Application to Catastrophe Risk Management and Reinsurance Pricing

John A. Major, ASA, MAAA

Gradients of Risk Measures: Theory and Application to Catastrophe Risk Management and Reinsurance Pricing

By John A. Major, ASA[1]

As part of pricing, many reinsurers would like to know the incremental impact that adding a new contract or canceling an existing contract might have on the capital needed to support the entire portfolio of business. Typically, catastrophe models take hours or days to run, ruling out a straightforward approach. A method of assessing incremental impact which did not require repeatedly simulating losses to the entire portfolio would therefore be quite useful. Efficient procedures for calculating the first derivatives of widely-used risk measures (such as Value at Risk and Tail Value at Risk) with respect to portfolio parameters would support the development of such a method. This paper presents general formulas for gradients of risk measures including VaR and TVaR. While the derivative of VaR in the case of linear risk models is widely known, this paper presents the general solution applicable not only to linear portfolio weights, but also to nonlinear parameters, such as retentions and limits. Implementation of the theoretical formulas within existing catastrophe simulation models is elaborated. A normal mixture approximation leads to a closed form solution for the incremental impact on VaR or TVaR of adding or removing a contract from a portfolio of excess-of-loss contracts.

1. Introduction

Reinsurance pricing is, in part, a portfolio problem: a matter of determining a sufficiently high expected return (margin in the premium) to compensate the firm for the risk it is taking on. In a survey of the practice of catastrophe excess-of-loss (XOL) pricing among reinsurers, Major and Kreps [2002] find the more sophisticated markets "attempt to assess the incremental impact a new contract would have on the capital needed to support the overall portfolio...." (Models of such an approach can be seen in Kreps [1990; 1999]. Tasche [1999] centers a theory of capital allocation and portfolio component returns on the first derivative of the risk measure.) They also state that most markets "lacked the analytical skills or the brute force computing power needed to accomplish it." Typically, catastrophe models running on the entire portfolio take hours or days so runs are done only a few times per year, ruling out incremental analysis of new contract proposals. Therefore a method of assessing incremental impact which did not require simulating losses to the entire portfolio would be quite useful.

Key statistics for measuring risk in the context of insurance capital requirements include Probable Maximum Loss, also known as Value at Risk or VaR, and Expected Shortfall, which is closely related to Tail Value at Risk or TVaR. Efficient procedures for calculating the first derivatives of these risk measures with respect to portfolio parameters would support the development of desired pricing algorithms. These risk measures also extend far beyond the domain of catastrophe insurance. Because of the Basel Committee [1996], VaR is used heavily by nearly all banks and financial institutions. TVaR is also becoming of considerable interest to financial risk managers (Andersson et. al. [2000], Gaivoronski and Pflug [2000], Yamai and Yoshiba [2002]). Efficient computation of gradients would then support efforts in risk management throughout the financial services industry.

This paper presents general formulas for the gradients of risk measures including VaR and TVaR with respect to parameters of a loss function (which may be nonlinear) over an arbitrary smooth distribution of risk factors. While the derivative of VaR in the case of linear risk models is widely known (see, for example, Gourieroux et. al. [2000]) a theorem of Uryasev [1995a,b; 1999] is applied to derive the general solution applicable to nonlinear parameters, such as retentions and limits, as well as to linear portfolio weights.

---

[1] email: john.a.major@guycarp.com

Specializations to linear and quadratic loss functions and analytic solutions for normally distributed risk factors will be discussed. A normal mixture approximation will be shown to lead to a closed form solution for the incremental impact on VaR or TVaR of adding or removing a contract from a portfolio of piecewise linear excess-of-loss contracts.

Section 2 of the paper provides the theoretical results. First the risk measures of interest are defined, and the role of the gradient in first-order Taylor series expansions and optimization is outlined. Second, the main theorems representing gradients as functions of expectations are given. Section 3 presents the bilinear special case and applications to insurance and reinsurance problems.

Section 4 discusses the efficient implementation of the theoretical formulas within existing Monte Carlo catastrophe simulation models. Section 5 addresses analytical solutions and approximations, including the delta-normal and gamma-normal models as well as conditional normal models. Ultimately the piecewise-linear normal-mixture is revealed as a generic analytic approximation. Section 6 showcases the results of a simulation study comparing "brute force" and theorem-based Monte Carlo methods to the re/insurance applications.

Section 7 provides an overview of some related work. Section 8 concludes. Section 9 is an appendix with a dictionary of mathematical symbols and selected derivations.

2. Theory

2.1 Mathematical framework

Let $\mathbf{X}$ in R^N represent a random variable that underlies a loss process (think of hurricane wind speeds or earthquake peak ground acceleration at various locations). Let $f(\mathbf{x})$ be the (multivariate) probability density function. The pdf f is assumed smooth for the theoretical development, but this will be relaxed in the discussion of Monte Carlo simulation later. $\mathbf{X}$ describes the "hazard" independently of any particular portfolio of exposures.[2] The components of $\mathbf{X}$ are often referred to as "risk factors."

Let $L = g(\mathbf{X},\theta)$ define the net losses to a portfolio of insured exposures, where θ in R^M is a parameter describing the exposure (think of all the financial specifications of a reinsurance deal: exposures at various locations, as well as retentions, limits, sublimits, etc.).

The cumulative probability distribution of net losses is given by the multidimensional integral

$$F(l;\theta) = \Pr\{L \le l\} = \int_{\{x:g(x,\theta)\le l\}} f(\mathbf{x})dV$$

where $dV = dx_1 dx_2 \ldots dx_N$ represents a volume integral element. That is, integrating (or summing) all the probability for events which cause losses less than or equal to l gives the cumulative probability. Note that the set $\{\mathbf{x}|g(\mathbf{x},\theta) = l\}$ is an (N-1)-dimensional submanifold of R^N which divides the space of $\mathbf{X}$ points into two regions: those that result in losses less than l and those that result in losses greater than l. Call this boundary surface an *isoclast*.[3]

---

[2] The reader may regard this discussion as being concerned with severity or per-occurrence distributions only. For an extension to compound processes, see Daykin et. al. [1994].

[3] From the Greek "*isos*" = "equal" and "*klastos*" = "broken."

2.2 Defining risk measures

The cumulative probability distribution of L can be cumbersome to deal with. It is a function and requires an infinite set (theoretically) or at least a very large set (practically) of numbers to describe it. One often seeks scalar measures of risk to characterize the distribution more economically.

2.2.1 Expectations and moments

One class of risk measures consists of mathematical expectations of functions of the random loss E[h(L)]. The simplest nontrivial form of this is the first moment, or mean:

$$\mu[L] = E[L] = \int g(\mathbf{x};\boldsymbol{\theta}) \cdot f(\mathbf{x}) dV$$

The mean is a measure of central tendency or location. In statistical practice, it is usually associated with a companion measure of spread, the second moment (variance) or its square root, the standard deviation:

$$\sigma^2[L] = E\left[(L-\mu)^2\right] = \int (g(\mathbf{x};\boldsymbol{\theta}) - \mu)^2 \cdot f(\mathbf{x}) dV$$

$$\sigma[L] = \sqrt{\sigma^2[L]}$$

2.2.2 Exceedance probabilities

While it is inconvenient to assess the entire distribution, it may be valuable to capture certain points on the curve. For example, the loss of particular dollar amounts corresponding to various levels of financial distress (e.g., insolvency or regulatory thresholds) may be of interest. Define the exceedance probability at loss amount l as:

$$Q(l;\boldsymbol{\theta}) = \Pr\{L > l\} = 1 - F(l) = \int_{g(\mathbf{x};\boldsymbol{\theta})>l} f(\mathbf{x}) dV$$

2.2.3 VaR L_q, and TVaR T_q

Value at Risk (VaR) answers the question: "If this portfolio experiences an adverse outcome, how bad could it be?" An exercise in measuring VaR consists of specifying the time and probability level and quantifying the monetary outcome (Jorion [2000]). Because time and probability must be specified, there are many VaR measurements that can be taken on a particular portfolio. In this paper, it is assumed that one time horizon (not symbolized) and one probability level (symbolized by q) have been specified. In application, many measurements can (and should) be taken for a more complete understanding of the risk.

Define the Value-at-Risk (VaR) $\mathsf{L}_q(\boldsymbol{\theta})$ as the solution to the equation

$$Q\left(\mathsf{L}_q(\boldsymbol{\theta});\boldsymbol{\theta}\right) = q\,.$$

That is, VaR is the level of losses that will only be exceeded with probability q. Note that it is a function of the portfolio description $\boldsymbol{\theta}$.

The typical VaR application in finance has $\mathbf{X}$ representing changes to current values of key interest rates, foreign exchange rates, etc., and L the resulting change in value of an investment portfolio. The parameter $\boldsymbol{\theta}$ might be implicit in the computation of g, especially if the task is to compute VaR on the current

portfolio, without regard to changes in the portfolio composition. When explicit, θ may represent quantities of various securities or classes of securities held.

A pure insurance species of VaR is Probable Maximum Loss (PML). Here, **X** represents the random contingencies affecting a portfolio of insurance liabilities and L the associated loss. The canonical example from property insurance is the use of a catastrophe model to answer the question "What catastrophe loss will not be exceeded, on average, once in a hundred years?" The "100-year return period loss" is the one-year VaR corresponding to $q = 0.01$.[4] Similar questions about net operating results, including underwriting gain as well as investment income, are now routine in insurance applications of Dynamic Financial Analysis (DFA).

Tail Value at Risk (TVaR) is defined as the conditional expectation of the loss (the average loss) given that the loss is greater than the VaR. That is, it is the average loss among only those losses that exceed the VaR:

$$\mathsf{T}_q(\boldsymbol{\theta}) = \frac{1}{q} \cdot \int_{\{\mathbf{x}: g(\mathbf{x},\boldsymbol{\theta}) > \mathsf{L}_q(\boldsymbol{\theta})\}} g(\mathbf{x},\boldsymbol{\theta}) \cdot f(\mathbf{x}) dV$$

TVaR is gaining favor relative to VaR because it possesses a property known as *coherence* (Artzner et. al. [1999]).

2.2.4 The gradient operator ∇

Managing a portfolio means observing and controlling (or at least influencing) its changes over time to achieve desired outcomes. The gradient of a risk measure – formally, the vector of partial derivatives – with respect to changes in parameters defining the portfolio provides a "sensitivity map" that shows explicitly how any small change to the portfolio will change the measure. Coupled with other sensitivity measures, the gradient can be used to design a change strategy aimed at improving the risk/reward characteristics of the portfolio.

If a portfolio described by $\boldsymbol{\theta}$ has a risk measure $\mathsf{R}(\boldsymbol{\theta})$, and the portfolio parameter $\boldsymbol{\theta}$ is perturbed by a small $\Delta\boldsymbol{\theta}$, then the new value of R can be approximated by a first-order Taylor expansion:

$$\mathsf{R}(\boldsymbol{\theta} + \Delta\boldsymbol{\theta}) = \mathsf{R}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}}\mathsf{R} \cdot \Delta\boldsymbol{\theta}$$

where, in the case of an (M=1)-dimensional parameter θ, $\nabla_{\theta}\mathsf{R}$ is the first derivative of $\mathsf{R}(\theta)$ with respect to θ, or, in the general vector case (M>1), the vector of partial derivatives, i.e., the *gradient vector* of $\mathsf{R}(\boldsymbol{\theta})$ with respect to the portfolio vector parameter. Specifically,

$$\nabla_{\boldsymbol{\theta}}\mathsf{R} = \left\langle \frac{\partial \mathsf{R}}{\partial \theta_1}, \frac{\partial \mathsf{R}}{\partial \theta_2}, \frac{\partial \mathsf{R}}{\partial \theta_3}, \ldots, \frac{\partial \mathsf{R}}{\partial \theta_M} \right\rangle$$

Writing out the vector multiplication term-by term, it would look like:

---

[4] Actually, the 100-year return period loss corresponds to VaR at q=0.00995 under Poisson frequency assumptions, but this distinction is often ignored in practice.

$$\nabla_\theta R \cdot \Delta\theta = \sum_j \left(\frac{\partial R(\theta)}{\partial \theta_j} \cdot \Delta\theta_j \right)$$

where the j indexes over all the components of θ.

The significance of the gradient is this: with the gradient in hand, the incremental impact of proposed (small) changes to the portfolio can be evaluated *without recomputing R*. The gradient can be used as part of an optimization procedure to seek out "change directions" for portfolio improvement, thereby reducing the number of computations of R in the search. Without using the gradient, if the dimension M of the portfolio parameter is large, it may be impractical to schedule all the required recomputations of R, even if a single cycle is fairly efficient. Additionally, if the model is analytically tractable, optimal change vectors might be obtained by a closed-form solution. Both of these points will be elaborated below.

2.3 Main theorems – gradients as integrals and expectations

2.3.1 Gradient of moments

The gradients of moments follow from smoothness giving us the ability to interchange differentiation and integration:

$$\nabla_\theta \mu[L] = \nabla_\theta \int g(\mathbf{x};\theta) \cdot f(\mathbf{x}) dV = \int \nabla_\theta g(\mathbf{x};\theta) \cdot f(\mathbf{x}) dV = E[\nabla_\theta g(\mathbf{X},\theta)]$$

where the subscript θ on the gradient symbol ∇ is there to emphasize that the partial derivatives are with respect to the θ arguments, not the $\mathbf{x}$ arguments.[5] This equation, as written, involves the integral of vectors, but it can be interpreted one term at a time by substituting $\partial/\partial\theta_j$ for the ∇ symbol.

The reader can verify that

$$\nabla_\theta \sigma^2[L] = 2 \cdot \int g \cdot \nabla_\theta g \cdot f(\mathbf{x}) dV - \mu \cdot \nabla_\theta \mu = 2 \cdot E[(g-\mu) \cdot \nabla_\theta g(\mathbf{X},\theta)]$$

$$\nabla_\theta \sigma[L] = \frac{1}{2 \cdot \sigma} \cdot \nabla_\theta \sigma^2[L] = E\left[\frac{g-\mu}{\sigma} \cdot \nabla_\theta g(\mathbf{X},\theta) \right]$$

2.3.2 "Integral over the surface formula"

The most direct route to deriving the gradients of exceedance probabilities is to make use of the following result (Uryasev [1995a, b; 1999]). Let the domain of integration be defined by $\rho(\psi) = \{ \mathbf{x} \in R^N \mid \gamma(\mathbf{x},\psi) \le 0 \}$ and by $\partial\rho$ the boundary of this set. Consider the volume integral $H(\psi) = \int_{\rho(\psi)} \phi(\mathbf{x},\psi) dV$ where dV denotes $\mathbf{x}$-integration as before. Note that both the integrand and the region of integration are functions of a free parameter ψ. If the constraint function γ is differentiable and the following integrals exist, then the gradient of H with respect to ψ is given by:

---

[5] Superficially, it appears that Stokes' Theorem – $\int_\rho \nabla_x g dV = \int_{\partial\rho} g dS$ – might be relevant here, but we are dealing with ∇_θ, not ∇_x.

$$\nabla_\psi H(\psi) = \int_{\rho(\psi)} \nabla_\psi \phi(\mathbf{x},\psi) dV - \int_{\partial\rho(\psi)} \frac{\nabla_\psi \gamma(\mathbf{x},\psi)}{\|\nabla_x \gamma(\mathbf{x},\psi)\|} \cdot \phi(\mathbf{x},\psi) dS$$

where dS denotes (hyper)surface element and ||v|| denotes vector norm.[6]

2.3.3 Gradients of exceedance probabilities

The derivative of $Q(L,\theta)$ with respect to L can be obtained by taking $\psi = L$, $H(\psi) = Q(L,\theta)$, $\phi(\mathbf{x},\psi) = f(\mathbf{x})$, and $\gamma(\mathbf{x},\psi) = L - g(\mathbf{x},\theta)$, where θ is treated as a constant. Since ϕ is not a function of ψ, the first integral vanishes and one is left with:

$$\frac{\partial Q(L,\theta)}{\partial L} = -\int_{g(\mathbf{x},\theta)=L} \frac{1}{\|\nabla_x g(\mathbf{x},\theta)\|} \cdot f(\mathbf{x}) dS = -C_N \cdot E\left[\|\nabla_x g(\mathbf{X},\theta)\|^{-1} \middle| g(\mathbf{X},\theta) = L\right] \text{ where the}$$

prefactor $C_N = \int_{g(\mathbf{x};\theta)=L} f(\mathbf{x}) dS$.

The gradient of $Q(L,\theta)$ with respect to θ is obtained similarly by taking $\psi = \theta$, $H(\psi) = Q(L,\theta)$, $\phi(\mathbf{x},\psi) = f(\mathbf{x})$, and $\gamma(\mathbf{x},\psi) = L - g(\mathbf{x},\theta)$, where now L is treated as a constant:

$$\nabla_\theta Q(L,\theta) = -\int_{g(\mathbf{x},\theta)=L} \frac{-\nabla_\theta g(\mathbf{x},\theta)}{\|\nabla_x g(\mathbf{x},\theta)\|} \cdot f(\mathbf{x}) dS = C_N \cdot E\left[\frac{\nabla_\theta g(\mathbf{X},\theta)}{\|\nabla_x g(\mathbf{X},\theta)\|} \middle| g(\mathbf{X},\theta) = L\right]$$

The relevance of $\|\nabla_x g\|$, and how to avoid the necessity for computing it, is discussed in the section on Monte Carlo implementation.

2.3.4 Gradient of Value at Risk

Applying the implicit function theorem to the equation $Q(L_q(\theta),\theta) - q = 0$:

$$\frac{\partial L_q}{\partial \theta} = -\frac{\partial Q}{\partial \theta} \Big/ \frac{\partial Q}{\partial L}.$$

Substituting the previously-obtained derivatives of Q:

$$\frac{\partial L_q}{\partial \theta} = -\frac{\partial Q/\partial \theta}{\partial Q/\partial L} = \frac{\int_{g(\mathbf{x};\theta)=L_q} \frac{\partial g}{\partial \theta} \|\nabla_x g\|^{-1} \cdot f(\mathbf{x}) dS}{\int_{g(\mathbf{x};\theta)=L_q} \|\nabla_x g\|^{-1} \cdot f(\mathbf{x}) dS} = \frac{E\left[\|\nabla_X g(\mathbf{X},\theta)\|^{-1} \cdot \nabla_\theta g(\mathbf{X},\theta) \mid g(\mathbf{X},\theta) = L_q\right]}{E\left[\|\nabla_X g(\mathbf{X},\theta)\|^{-1} \mid g(\mathbf{X},\theta) = L_q\right]}$$

---

[6] See previous footnote.

Corollary: Define the probability density function $h(\mathbf{x})$ on the $g(\mathbf{X};\theta)=\mathsf{L}_q$ isoclastic hypersurface by $h(\mathbf{x})=\dfrac{\|\nabla_x g\|^{-1}\cdot f(\mathbf{x})}{\int_{g(\mathbf{x};\theta)=\mathsf{L}_q}\|\nabla_x g\|^{-1}\cdot f(\mathbf{x})dS}$. Then

$$\nabla_\theta \mathsf{L}_q = \mathsf{E}_h[\nabla_\theta g(\mathbf{X},\theta)]$$

where the expectation is taken with respect to the density $h(X)$, on its domain (the isoclast).

2.3.5 Gradient of Tail Value at Risk

With the result for the gradient of VaR in hand, the gradient of TVaR is obtained by a more-or-less straightforward application of the integral over the surface formula:

$$\nabla \mathsf{T}_q(\theta)=\frac{1}{q}\cdot\int_{\{\mathbf{x}:g(\mathbf{x},\theta)>\mathsf{L}_q(\theta)\}}\nabla_\theta g(\mathbf{x},\theta)\cdot f(\mathbf{x})dV = E[\nabla_\theta g(\mathbf{X},\theta)\mid g(\mathbf{X},\theta)>\mathsf{L}_q(\theta)]=E_k[\nabla_\theta g(\mathbf{X},\theta)]$$

where $k(\mathbf{x})$ is the conditional distribution of $\mathbf{X}$ in the tail where $L > \mathsf{L}_q$. This is proved in the appendix (section 9.2).

Note the conceptual similarity between the gradient of VaR and the gradient of TVaR. Both are conditional expectations of $\nabla_\theta g$; they differ in the underlying conditional distribution on $\mathbf{X}$. For VaR, it is h, the distribution restricted to the isoclast (but reweighted for "thickness"). For TVaR, it is k, the distribution restricted to the tail, the half space where $\mathbf{X}$ puts losses over the VaR.

3. Applications

3.1 Bilinear, delta model

Perhaps the simplest nontrivial form is the bilinear loss function $L=g(\mathbf{X};\theta)=\theta^T\cdot\mathbf{X}$, where the superscript "T" indicates matrix transpose. The bilinear case occurs in property catastrophe loss modeling where elements of θ represent exposed property values at subscripted locations and the elements of $\mathbf{X}$ are corresponding damage rates (in a hurricane, say). L represents "ground-up" loss in CAT modeling parlance. This will be illustrated in the case study later. The bilinear case is by far the most frequently used form in financial modeling. There, $\mathbf{X}$ represents prices or returns of various asset classes and θ represents the quantity of each asset in a portfolio.

3.1.1 Moments

Calculation of moments is nearly trivial.

$\mu[L]=\theta^T\cdot\mu[\mathbf{X}]$ and $\sigma^2[L]=\theta^T\cdot vcv[\mathbf{X}]\cdot\theta$ where $vcv[\mathbf{X}]=E\left[(\mathbf{X}-\mu)\cdot(\mathbf{X}-\mu)^T\right]$, making

$\nabla_\theta\mu[L]=\mu[\mathbf{X}]$ and $\nabla_\theta\sigma^2[L]=2\cdot vcv[\mathbf{X}]\cdot\theta$ and $\nabla_\theta\sigma[L]=\sigma\cdot\theta$

3.1.2 Exceedance probabilities

Recalling the constant C_N from section 2.3.3, the simple form of g results in simple gradients:

$$\partial Q/\partial L = -C_N/\|\theta\| \text{ and } \nabla_\theta Q = (C_N/\|\theta\|)\cdot E[\mathbf{X} \mid \theta^T \cdot \mathbf{X} = L].$$

3.1.3 Value at Risk and Tail Value at Risk

VaR and TVaR follow as: $\nabla_\theta \mathsf{L}_q = E[\mathbf{X} \mid \theta^T \cdot \mathbf{X} = \mathsf{L}_q]$ and $\nabla_\theta \mathsf{T}_q = E[\mathbf{X} \mid \theta^T \cdot \mathbf{X} > \mathsf{L}_q]$. Note in the case of VaR that the "transformed distribution" h is simply the conditional distribution of $\mathbf{X}$ on the isoclast.

The gradient of VaR in the special bilinear case has been proved in many papers. See, e.g., Gourieroux et. al. [2000] and Tasche [1999].

3.1.4 Delta approximation

The bilinear case corresponds to the use of the "delta" approximation and most often emerges as part of the "delta-normal" model of VaR. The delta approximation writes:

$$L = g(\mathbf{x}_0 + \mathbf{X}) \approx g_0 + (\nabla_{\mathbf{X}} g)\cdot \mathbf{X}$$

and considers the vector $\boldsymbol{\delta} = (\nabla_{\mathbf{X}} g)^T$ to comprise the portfolio parameter θ. The delta-normal model will be considered in section 5.1 in the context of analytic approximations.

3.2 Reinsurance layers

The model $L = \Omega^T\mathbf{X}$ can describe *gross* or *ground-up* losses in catastrophe risk assessment, but a bit more is needed to analyze reinsurance. Say a treaty covers the losses up to a limit λ after a retention (attachment point) of α and a co-reinsurance of κ. Specifically, define a reinsurance function as $r(L,\alpha,\lambda,\kappa) = (1-\kappa)\cdot\max(0,\min(L-\alpha,\lambda))$. This makes the net loss, after reinsurance, equal to $N = L - r(L,\alpha,\lambda,\kappa) = \boldsymbol{\Omega}^T\cdot\mathbf{X} - r(\boldsymbol{\Omega}^T\cdot\mathbf{X},\alpha,\lambda,\kappa) \equiv g(\mathbf{X};\theta)$. The symbol Ω is used to represent exposures here, because the parameter vector θ is the composite vector $\langle\Omega_1,\Omega_2,\ldots,\Omega_N,\alpha,\lambda,\kappa\rangle^T$.

Now the question of risk measures and gradients of risk measures of the net loss random variable N can be addressed.

3.2.1 Moments

The moment theorems still apply. While g is no longer smooth (its first derivatives are undefined at the bottom and top of the layer), it is piecewise smooth.

$$\nabla_\theta \mu[L] = E[\nabla_\theta g(\mathbf{X},\theta)] \text{ and } \nabla_\theta \sigma^2[L] = 2\cdot E[(g(\mathbf{X},\theta)-\mu)\cdot\nabla_\theta g(\mathbf{X},\theta)] \text{ and}$$
$$\nabla_\theta \sigma[L] = E\left[\frac{g(\mathbf{X},\theta)-\mu}{\sigma}\cdot\nabla_\theta g(\mathbf{X},\theta)\right].$$

Note, however, that the gradients of g are a bit different than in the bilinear case.

The gradient of net loss with respect to a change to exposure (for a given damage scenario **X**) is given by

$$\nabla_{\Omega} g = \begin{cases} \kappa \cdot \mathbf{X}, & \alpha < \boldsymbol{\Omega}^T \cdot \mathbf{X} < \alpha + \lambda \\ \mathbf{X}, & otherwise \end{cases}.$$

That is, it is the damage rate at that location if the reinsurance is not "in the layer," but only κ times that if the contract is "in the layer."[7]

The partial derivative of net loss with respect to the treaty's attachment point is given by:

$$\frac{\partial g}{\partial \alpha} = \begin{cases} (1-\kappa), & \alpha < \boldsymbol{\Omega}^T \cdot \mathbf{X} < \alpha + \lambda \\ 0, & otherwise \end{cases}.$$

For every dollar that the layer is raised, the net loss increases by (1-κ) dollars, but only if the contract is "in the layer."

The partial derivative of net loss with respect to the limit is given by:

$$\frac{\partial g}{\partial \lambda} = \begin{cases} (\kappa - 1), & \alpha + \lambda < \boldsymbol{\Omega}^T \cdot \mathbf{X} \\ 0, & otherwise \end{cases}.$$

For every dollar the limit is increased, the net loss decreases by (1-κ) dollars, but only if the contract has exhausted (paid out fully).

The partial derivative of net loss with respect to the co-reinsurance is given by:

$$\frac{\partial g}{\partial \kappa} = \begin{cases} 0, & \boldsymbol{\Omega}^T \cdot \mathbf{X} < \alpha \\ \boldsymbol{\Omega}^T \cdot \mathbf{X} - \alpha, & \alpha < \boldsymbol{\Omega}^T \cdot \mathbf{X} < \alpha + \lambda \\ \lambda, & \alpha + \lambda < \boldsymbol{\Omega}^T \cdot \mathbf{X} \end{cases}.$$

Increasing co-reinsurance has no effect until the layer attaches, after which it applies to the amount eligible for reimbursement.

3.2.2 Exceedance probabilities

The application of the "integral over the surface formula" is not stymied by the discontinuous derivatives; one can divide the problem into three regions (below, in, and above the layer) where smoothness prevails. The gradients of Q with respect to N and the components of θ are computed according to the expectation formulas given in section 2.3.3. The gradient of g with respect to the damage variable **x** is given by

$$\nabla_x g(x,\theta) = \begin{cases} \kappa \cdot \boldsymbol{\Omega}, & \alpha < \boldsymbol{\Omega}^T \cdot \mathbf{X} < \alpha + \lambda \\ \boldsymbol{\Omega}, & otherwise \end{cases}.$$

---

[7] Actually, the derivative is undefined in the cases where the gross loss equals α or $\alpha+\lambda$, but that only occurs on a set with probability zero.

Increasing damage rates are borne fully by the reinsured outside the layer, but only in proportion κ inside the layer.

Note, however, that if $\kappa = 0$, then the gradient $\nabla_x g(\mathbf{x},\theta)$ vanishes in the layer, making integrals involving $\|\nabla_x g(\mathbf{x},\theta)\|^{-1}$ nonexistent. In such a case, the probability distribution of N has a "mass point" at N=α, so the left-hand derivative of Q with respect to N, and the right-hand derivative of Q with respect to α, are both infinite there. The gradients of Q become undefined.

3.2.3 Value at Risk and Tail Value at Risk

With the details given in section 3.2.1 for gradients of g above, in, and below the layer, the computation of gradients of VaR and TVaR are straightforward applications of the formulas in 2.3.4-2.3.5. In the case where $\kappa = 0$, the exceedance probability Q is discontinuous and therefore the gradients of VaR and TVaR may be undefined.

3.3 Portfolio of layers

Here the canonical reinsurance pricing problem is addressed.[8] Consider a reinsurer who holds a portfolio of layers $L_c = r\left(\left(\Omega^{<c>}\right)^T \cdot \mathbf{X}, \alpha_c, \lambda_c, \kappa_c\right)$. Here the symbol c indexes contracts, and the superscript <c> is used to denote vector #c (whereas a subscript would denote component #c of the vector).[9]

Ideally, pricing a proposed new layer L_n or considering the renewal of an existing layer L_e would take into account the effect of the change on the capital needs of the firm. The capital needs, in turn, would be driven (at least in part) by the risk characteristics of the portfolio $\{L_c\}$. A risk measure such as VaR or TVaR is typically used to make this assessment.

Reinsurers routinely compute the probability distribution of their payouts (and, hence, VaR and TVaR) by running a catastrophe model. This is usually done quarterly because the process is time-consuming.[10] How is one to assess the impact on VaR of adding a proposed new contract or nonrenewing an existing contract?

To cover either case, refer to the contract under consideration as contract #u. Examining contract u's contribution $\Omega^{<u>}$ to the total exposure vector $\Sigma_c \Omega^{<c>}$ is not helpful because of the nonlinear payoff of each contract – the total exposure $\Sigma\, \Omega$ is virtually meaningless. Nonetheless, the gradient theorem for a risk measure R can be profitably applied to the following:

$$g(\mathbf{X};\theta) \;=\; \sum_c r\left(\left(\Omega^{<c>}\right)^T \cdot \mathbf{X}, \alpha_c, \lambda_c, \kappa_c\right) \;+\; \theta \cdot r\left(\left(\Omega^{<u>}\right)^T \cdot \mathbf{X}, \alpha_u, \lambda_u, \kappa_u\right)$$

where θ is a scalar equal to 1 for the case of new business u = n and -1 for the case of renewals u = e. The partial derivative with respect to θ will reveal (through the conditional expectation theorems) the impact on R of adding or dropping a "small amount" θ of the contract under consideration to or from the portfolio. If

---

[8] It also corresponds to the evaluation of a portfolio of call options at the (future) time of their exercise.

[9] Of course, real reinsurance often involves other complicating factors as well, such as reinstatements, per-occurrence and aggregate retentions and limits, complex "towers" of layers, backup capacity, etc., etc. The abstraction presented here is complex enough to make the point.

[10] Major and Kreps [2002] cite an example of seven hours.

the contract is already "small" compared to the portfolio, then the approximation $R|_{\theta=1} = R|_{\theta=0} + \partial R/\partial\theta$ will be "good."

The partial derivative of g with respect to θ

$$\frac{\partial}{\partial\theta} g(\mathbf{X};\theta) = r\left(\left(\Omega^{<u>}\right)^T \cdot \mathbf{X}, \alpha_u, \lambda_u, \kappa_u\right)$$

is simply the payout of the contract under consideration. The gradient with respect to $\mathbf{X}$ is more complex, however.[11] The Monte Carlo section presented next will discuss an approach to evaluating $E_h[\nabla_\theta g]$ that does not require computing $\nabla_x g$.

4. Monte Carlo implementation

4.1 MC simulation and computation of risk measures

The most general applications of risk measures involve complicated nonlinear loss functions g and relatively intractable probability distributions f, and are usually solved by Monte Carlo methods. The basic method is to generate a sample of risk factor vectors $\mathbf{X}^{<i>}$ (i = 1, 2, ..., 10000, say) and apply the loss function to them sequentially to create a corresponding sample of loss values $L_i = g(\mathbf{X}^{<i>};\boldsymbol{\theta})$. Statistical methods are used to estimate percentage points of the distribution of L, hence risk measures.

Variance reduction techniques (VRTs), usually involving special steps in the construction of the sample $\mathbf{X}^{<i>}$, can be applied to improve the accuracy/speed tradeoff. The most common VRTs create samples where the points are not equally probable and adjust the estimation procedure accordingly. Stratified sampling and importance sampling are two such approaches.[12] If each scenario $\mathbf{x}^{<i>}$ has an associated probability p_i, then one can approximate an expectation integral by a summation as follows:

$$\int \varphi(\mathbf{x}) \cdot f(\mathbf{x}) dV \approx \sum_i \varphi(\mathbf{x}^{<i>}) \cdot p_i$$

4.1.1 Moments

For example, the first and second moments of a loss $L = g(\mathbf{X};\boldsymbol{\theta})$ would be estimated by

$$\mu[L] = \int g(\mathbf{x};\boldsymbol{\theta}) \cdot f(\mathbf{x}) dV \approx \sum_i L_i \cdot p_i \text{ and}$$

$$\sigma^2[L] = \int (g(\mathbf{x};\boldsymbol{\theta}) - \mu)^2 \cdot f(\mathbf{x}) dV \approx \sum_i (L_i - \mu)^2 \cdot p_i$$

---

[11] Technically, there is a problem in that $\nabla_x g$ vanishes where none of the layers attach or where all of the layers have exhausted. As long as the isoclast is in neither region, however, this does not present a practical problem.

[12] The generation of complementary pairs of sample points ("antithetic variates") while useful in estimating means, is generally not so helpful in tail-sensitive risk measure computations. See Rubinstein [1981] for more on the subject.

4.1.2 Exceedance probabilities

The exceedance probabilities are estimated by $Q(l;\theta) \approx \sum_{\{i:L_i>l\}} p_i$.

4.1.3 Value at Risk and Tail Value at Risk

The condition that defines the VaR level L_q, $q = Q(\mathrm{L}_q;\theta)$, is evaluated by a search procedure. In practice, this means computing L_i for each scenario i, sorting the results, and then finding that position in the sort order where the sum of probabilities p_i above that point adds up to the desired exceedance probability q. The numerical value of that cut point is then the VaR L_q.

Having divided scenarios i into those above and those below the VaR, one can then estimate the TVaR as:

$$\mathrm{T}_q \approx \frac{1}{q} \cdot \sum_{\{i:L_i>\mathrm{L}_q\}} L_i \cdot p_i$$

4.2 MC simulation and computation of gradients

The gradient of a risk measure could be obtained by rerunning the model and recomputing the measure after making small changes to each portfolio parameter, in turn.[13] For problems of realistic size, involving hundreds or thousands of parameters (asset classes, insured locations, etc.) and taking hours to run a single risk measure computation, this approach would be impractical. A more efficient solution is to appeal to the appropriate expectation theorem.

4.2.1 Gradients of moments

The first and second moment gradients, because they are expressed as expectations, are easily computed via Monte Carlo:

$$\nabla_\theta \mu[L] = \int \nabla_\theta g(\mathbf{x};\theta) \cdot f(\mathbf{x}) dV \approx \sum_i \nabla_\theta g(\mathbf{x}^{<i>};\theta) \cdot p_i \text{ and}$$

$$\nabla_\theta \sigma^2[L] = 2 \cdot \int (g-\mu) \cdot \nabla_\theta g \cdot f(\mathbf{x}) dV \approx \sum_i \left(g(\mathbf{x}^{<i>};\theta) - \mu\right) \cdot \nabla_\theta g(\mathbf{x}^{<i>};\theta) \cdot p_i$$

The continuous (theoretical) versions follow from the fact that smoothness allows us to interchange differentiation and integration. For the discrete (Monte Carlo) version, no such justification is needed. Differentiation is linear, hence can be interchanged with summation and multiplication by constants.

4.2.2 Gradient of exceedance probabilities

The situation is not so bright with the probability gradient theorems. Strictly speaking, they do not apply to simulation models of this type. The probability distribution for $\mathbf{X}$ is discrete, so the isoclast of $\mathbf{X}$ values solving $g(\mathbf{X};\theta) = \mathrm{L}_q$ is not a smooth manifold.[14] However, if the discrete distribution were modified by

---

[13] It would be helpful to use the same sample of $\mathbf{X}$ for each run, thereby reducing the between-run variance.

[14] Depending on how the quantile is defined, there may be only one or no points $x^{<i>}$ at all in the solution set.

kernel smoothing (Silverman [1986]), so that it became smooth, then the theorems could apply. In particular, each point in the discrete distribution of **X** should be replaced by a patch of density sufficiently large and smooth so that there are no "holes" in the resulting kernel mixture.

Using a normal kernel would accomplish this. Unfortunately, a normal kernel has infinite support, and would give every one of the original points $\mathbf{x}^{<i>}$ some nonzero contribution to the conditional distribution on the isoclast. Therefore, all points would need to be included in the computation of the gradient, even though most would be "far away" from the isoclast, contributing a negligible amount to the expectation. The computations would be needlessly burdensome.

For this reason, a kernel with finite support is recommended. By varying the kernel bandwidth, one can assess the numerical stability of the gradient computation. In particular, one can use the Epanechnikov kernel, defined as:

$$K(t,w) = \begin{cases} 0, & |t| > w \\ 1-(t/w)^2, & |t| \le w \end{cases}$$

where t is the distance from the kernel origin and w is the kernel bandwidth.

The best way to understand why kernel smoothing makes sense in this context is to write the derivatives in the form of their definitions as limits of ratios. For example,

$$\frac{\partial Q}{\partial L} = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(\int_{g(\mathbf{x}) > L+\varepsilon} f(\mathbf{x}) dV - \int_{g(\mathbf{x}) > L} f(\mathbf{x}) dV \right) = -\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{L < g(\mathbf{x}) \le L+\varepsilon} f(\mathbf{x}) dV .$$

The last expression, the N-dimensional volume integral on the hypershell $\{ \mathbf{x} \mid L < g(\mathbf{x};\boldsymbol{\theta}) < L+\varepsilon \}$ has "thickness" $\|\nabla_{\mathbf{x}} g\|^{-1} \cdot \varepsilon$ around the $g(\mathbf{x};\boldsymbol{\theta}) = L$ hypersurface. This is why the limiting expression is $\partial Q / \partial L = -\int_{g(\mathbf{x},\boldsymbol{\theta})=L} \|\nabla_{\mathbf{x}} g(\mathbf{x},\boldsymbol{\theta})\|^{-1} \cdot f(\mathbf{x}) dS$. If a finite ε were used in a Monte Carlo approximation of the hypershell integral, one would get

$$\frac{\partial Q}{\partial L} \approx -\frac{1}{\varepsilon} \sum_{i: L < g \le L+\varepsilon} p_i = -\frac{1}{\varepsilon} \sum_i I(\mathbf{x}^{<i>}, \varepsilon) \cdot p_i$$

where the indicator function I takes on values 1 or 0 according to whether the point $\boldsymbol{x}^{<i>}$ puts g inside the ε-hypershell or not. Taking the next step and using the Epanechnikov kernel, write instead

$$\frac{\partial Q}{\partial L} \approx -\frac{3}{4 \cdot w} \cdot \sum_i K(L_i - L, w) \cdot p_i$$

where now w takes on the role that ε had. (The prefactor is the inverse of the expected value of K, 2/3, times the interval width 2w.)

Similarly, the partial derivative with respect to a component θ is approximated by

$$\frac{\partial Q}{\partial \theta} \approx \frac{3}{4 \cdot w} \cdot \sum_i K(L_i - L, w) \cdot \frac{\partial g}{\partial \theta} \cdot p_i .$$

The numerical implementation here is, in effect, approximating the well-behaved hypershell limit integral by using an ε of the same order of magnitude as the kernel window. Note that the factor $\|\nabla_x g\|^{-1}$ does not explicitly occur in the simulation approach; rather, its effects are handled in the windowing.

In these summations, only points $x^{<i>}$ whose g-values L_i are within the kernel bandwidth of the VaR value need be included. This can yield a substantial efficiency improvement over methods that must evaluate all points.

4.2.3 Gradients of Value at Risk and Tail Value at Risk

The gradient of Value at Risk is computed directly as the ratio of the probability gradients derived previously.

$$\nabla_\theta \mathsf{L}_q = -\frac{\nabla_\theta Q}{\partial Q/\partial L} = \frac{\sum_{i:|L_i-\mathsf{L}_q|<w} K(L_i-\mathsf{L}_q,w)\cdot \nabla_\theta g \cdot p_i}{\sum_{i:|L_i-\mathsf{L}_q|<w} K(L_i-\mathsf{L}_q,w)\cdot p_i}.$$

The steps are as follows:

(1) Run the model once, saving an index to L_i values, and compute VaR as the desired empirical percentile.

(2) Determine which points i are within the kernel bandwidth.

(3) Run the model a second time, computing the vectors of partial derivatives $\partial g/\partial\theta$ at the $\mathbf{x}^{<i>}$ points identified in step 2.

(4) Tabulate the weighted average of the partial derivative vectors to obtain the gradient of VaR $\nabla_\theta \mathsf{L}_q$.

The gradient of TVaR is estimated as

$$\nabla_\theta \mathsf{T}_q(\theta) = E\left[\nabla_\theta g(\mathbf{X},\theta) \mid g(\mathbf{X},\theta) > \mathsf{L}_q(\theta)\right] = \frac{1}{q}\cdot \sum_{i:L_i>\mathsf{L}_q} \nabla_\theta g \cdot p_i$$

Note again the similarity between VaR and TVaR. In the Monte Carlo implementation, both are weighted averages of $\nabla_\theta g$; they differ in the x-sets over which the summation is carried out and in the weights themselves. VaR involves points x near the isoclast and with weights proportional to $K(w)p_i$. TVaR involves points x on one side of the isoclast and with weights proportional to p_i.

4.3 Computational savings over brute force approach

One can get a rough idea,[15] at least in relative terms, of how long it would take a computer to calculate VaR or TVaR. Assume that most of the work is in the computation of losses for one scenario $g(x_i,\theta)$ and this takes, on average, time T_g. Assume there are I scenarios (i=1,2,...,I), so the total time taken in

---

[15] The practical implementation issues of disk caching, overhead of function calls, etc., make this only a rough approximation.

computation of losses is $T_g \times I$. Sorting algorithms typically take time proportional to the log of the size of the list, say $T_s \times \log(I)$. A (relatively) fast linear accumulation search finds the VaR. Assume this time is negligible. The additional step in calculating TVaR consists of multiplying some number $q' \times I$ pairs of losses and probabilities.[16] This is about twice the work as the accumulation step for VaR, so this additional time will also be assumed negligible. To a first approximation, the time it takes to compute VaR or TVaR is $T_g \times I + T_s \times \log(I)$. If I is very large and the computation of g is particularly complex, this is approximately $T_g \times I$.

What about the gradients? Gradient components are given by

$$\frac{\partial}{\partial \theta_j} \mathrm{R}_q \approx \sum_{\{i\}} k_i \frac{\partial}{\partial \theta_j} g(\mathbf{x}^{<i>}, \boldsymbol{\theta}) \cdot p_i$$

where the weights k_i and the summation set $\{i\}$ differ according to whether R is VaR or TVaR. Say there are $q' \times I$ terms in the summation for TVaR and $q'' \times I$ terms for VaR. (Presumably, $q'' < q' << 1$.) Depending on the nature of the particular parameter θ_j, computing the partial derivative at a particular scenario may be difficult or easy. Worst case, it can be computed as

$$\frac{\partial}{\partial \theta_j} g(\mathbf{x}^{<i>}, \boldsymbol{\theta}) \approx \frac{g(\mathbf{x}^{<i>}, \boldsymbol{\theta} + \Delta\theta_j) - g(\mathbf{x}^{<i>}, \boldsymbol{\theta})}{\Delta\theta_j}$$

making it cost approximately $2 \times T_g$ in computer time. If N components j are to be evaluated, the total is $(N+1) \times T_g$ because the base case $g(\mathbf{x}^{<i>}, \boldsymbol{\theta})$ needs only be evaluated once and stored.

Say, in general, one evaluation cost $d \times T_g$. The better cases have very small d. For example, as was seen above, an exposure change is typically linear in the loss. So is a change to a pro-rata share. Limits and retentions on XOL contracts result in dollar-for-dollar changes or no change at all, depending on whether the scenario is below, in, or above the layer. In all of these examples, no further evaluation of g is needed – although there may be some overhead in tracking the status of the original g computation – and d is close to zero.

The "brute force" method of assessing changes in VaR (resp., TVaR) is to start with a base case and then re-evaluate it for each of N perturbations of the $\boldsymbol{\theta}$ parameter, for a total computation time of approximately $T_g \times I \times (1+N)$. In contrast, by using the gradient, this time can be cut to approximately $T_g \times I \times (1+q^0 \times d \times N)$ where q^0 is q' (resp., q''). If N is large, then the ratio of time taken by using the theorems to the time taken by the brute force method is approximately $q^0 \times d$.

5. Analytical solutions and approximations

5.1 Delta-normal model

Perhaps the most widely-used analytical model for VaR, very familiar to financial risk management, is the Delta-normal model alluded to in section 3.1.4. The delta approximation writes:

$$L = g(\mathbf{x}_0 + \mathbf{X}) = g(\mathbf{x}_0) + (\nabla_x g) \cdot \mathbf{X}$$

[16] If scenarios are equally probable, then q'=q. If scenarios are oversampled at the high end, then q' is likely to be larger than q, but still less than one.

and considers the vector $\boldsymbol{\delta} = (\nabla_x g)^T$ to comprise the portfolio parameter $\boldsymbol{\theta}$. When $\mathbf{X}$ is assumed to possess a multivariate normal distribution with mean zero and variance-covariance matrix $\boldsymbol{\Sigma}$, then L has a (univariate) normal distribution with moments

$$E(L) = g(\mathbf{x}_0), \qquad \sigma^2(L) = \boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}$$

Exceedance probabilities can be computed from the cumulative normal distribution Φ (Johnson et. al. [1994]) as:

$$Q(L) = 1 - \Phi\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}}\right)$$

VaR is readily calculated from the appropriate percentile z_q of the normal distribution:

$$\mathsf{L}_q(\boldsymbol{\delta}) = E(L) + z_q \cdot \sqrt{\sigma^2(L)} = g(\mathbf{x}_0) + z_q \cdot \sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}$$

To develop TVaR, a lemma about the conditional expectation of a normal random variable is needed.

Lemma: Given a scalar random variable Y distributed normally with mean μ and variance σ^2, then

$$E[\max(0, Y - t)] = \int_t^\infty (y - t) \cdot f(y) dy =$$

$$\xi(t, \mu, \sigma) \equiv \sigma \cdot \phi\left(\frac{t - \mu}{\sigma}\right) - (t - \mu) \cdot \left(1 - \Phi\left(\frac{t - \mu}{\sigma}\right)\right)$$

where ϕ is the standard normal probability density function and Φ is the standard normal cumulative density function. $\xi(t,\mu,\sigma)$ is related to the *limited expected value* function (Klugman et. al. [1998]).

Proof of Lemma. First, note that $\int_t^\infty z \cdot \phi(z) dz = \phi(t)$, which follows by differentiating $\phi(t)$.

Change variables from Y to $z = \dfrac{Y - \mu}{\sigma}$ and write the expectation as

$$\int_{\frac{t-\mu}{\sigma}}^\infty (\sigma \cdot z + \mu - t) \cdot \phi(z) dz = \sigma \cdot \int_{\frac{t-\mu}{\sigma}}^\infty z \cdot \phi(z) dz + (\mu - t) \cdot \int_{\frac{t-\mu}{\sigma}}^\infty \phi(z) dz$$

$$= \sigma \cdot \phi\left(\frac{t - \mu}{\sigma}\right) - (t - \mu) \cdot \left(1 - \Phi\left(\frac{t - \mu}{\sigma}\right)\right)$$

QED.

TVaR is then obtained as $\mathsf{T}_q = \frac{1}{q}\cdot\int_{\mathsf{L}_q}^{\infty} L\cdot f(L)dL = \mathsf{L}_q + \frac{1}{q}\cdot\xi\left(\mathsf{L}_q, g(\mathbf{x}_0), \sqrt{\delta^T\cdot\Sigma\cdot\delta}\right)$, which simplifies to $\mathsf{T}_q = g(\mathbf{x}_0) + \frac{\phi(z_q)}{q}\cdot\sqrt{\delta^T\cdot\Sigma\cdot\delta}$.

5.1.1 Gradients in the delta-normal model

Above, the symbol δ was used instead of θ to emphasize that its nature is more in describing the portfolio than in controlling it. One cannot change the "delta" of a portfolio directly; it comes about as the result of some other sort of action (e.g., changing the amount of a risk held in the portfolio). Nonetheless, it is meaningful to inquire into the relationship between various risk measures and the delta of the portfolio.

The δ-gradients of the first two moments and the partial derivative of exceedance probability with respect to loss amount follow directly from the bilinear case examined in section 3.1:

$$\nabla_\delta \mu = 0$$
$$\nabla_\delta \sigma^2 = 2\cdot\Sigma\cdot\delta$$
$$\frac{\partial Q}{\partial L} = -\frac{C_N}{\|\delta\|}$$

Normality assumptions add nothing except to specify[17]

$$C_N = \int_{g(x)=L} f(\mathbf{x})dS = \frac{\|\delta\|}{\sqrt{\delta^T\cdot\Sigma\cdot\delta}}\cdot\phi\left(\frac{L-g(\mathbf{x}_0)}{\sqrt{\delta^T\cdot\Sigma\cdot\delta}}\right)$$

The gradient of exceedance probability with respect to δ can be derived in two ways. Differentiation of the delta-normal formula for Q gives the gradient with respect to the δ vector. An alternative is to apply the conditional expectation theorem. Both approaches are outlined in the appendix (section 9.3.1) and achieve the following result: $\nabla_\delta Q = \left(C_N/\|\delta\|\right)\cdot E\left[\mathbf{X}\mid g(\mathbf{x}_0)+\delta^T\cdot\mathbf{X}=L\right] = \frac{C_N}{\|\delta\|}\cdot\frac{L-g(\mathbf{x}_0)}{\delta^T\cdot\Sigma\cdot\delta}\cdot\Sigma\cdot\delta$.

Notice that the result is a vector because Σ is a matrix and δ is a vector. The other terms in the product are scalars.

The gradient of VaR in the delta-normal model is then readily obtained as

$$\nabla_\delta \mathsf{L}_q = \frac{z_q}{\sqrt{\delta^T\cdot\Sigma\cdot\delta}}\cdot\Sigma\cdot\delta.$$

Differentiating the expression for TVaR, or applying the gradient theorem, produces:

---

[17] To see this, rotate to a new orthonormal coordinate system $\mathbf{W}$ where $\mathbf{X} = W_1\mathbf{n}^{<1>}+W_2\mathbf{n}^{<2>}+\ldots+W_N\mathbf{n}^{<N>}$ and $\mathbf{n}^{<1>} = \delta/\|\delta\|$. Since $L = \|\delta\|W_1$, it follows that the marginal pdf of W_1 is $\|\delta\|$ times the marginal pdf of L.

$$\nabla_{\delta} \mathsf{T}_q = \frac{\phi(z_q)}{q} \frac{\Sigma \cdot \delta}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}}$$

These are also detailed in the appendix (sections 9.3.2-9.3.3).

5.2 Gamma-normal model

It has been widely recognized that the delta-normal model does not do a good job when nonlinear instruments (e.g., call options) are present in the portfolio in significant quantities. That is, the first-order Taylor approximation to the loss function g is inadequate over the range of $\mathbf{X}$ values relevant to the computation of VaR or TVaR. To remedy this, various authors have suggested the use of the second-order Taylor expansion in the gamma-normal model. Following Britten-Jones and Schaeffer [1999], consider the model:

$$L = g(\mathbf{X}; \delta, \Gamma) = \nu + \delta^T \cdot \mathbf{X} + \frac{1}{2} \cdot \mathbf{X}^T \cdot \Gamma \cdot \mathbf{X} \qquad \mathbf{X} \sim N(\mu, \Sigma)$$

where δ is the vector of first-degree sensitivities and Γ is the (symmetric) matrix of second-degree sensitivities. $\mathbf{X}$ is assumed multivariate normal. This can be rewritten by "completing the square" to become:

$$L = \left(\nu - \frac{1}{2} \cdot \delta^T \cdot \Gamma^{-1} \cdot \delta \right) + \frac{1}{2} \cdot \sum_{i=1}^{N} \mathbf{D}_i \cdot (Z_i)^2 \qquad \mathbf{Z} \sim N\left(\mathbf{H}^T \cdot \left(\mu + \Gamma^{-1} \cdot \delta\right), \mathbf{I}\right)$$

where $\Sigma^{1/2} \cdot \Gamma \cdot \Sigma^{1/2} = \mathbf{H} \cdot \mathbf{D} \cdot \mathbf{H}^T$, $\mathbf{D}$ is a diagonal matrix of eigenvalues, $\mathbf{H}$ is orthonormal (its columns being eigenvectors), and $\mathbf{I}$ is the identity matrix. Thus L can be seen as a linear combination of non-central chi-square variables.

In particular, the moments can be expressed in closed form. The first two are:

$$E[L] = \nu + \delta^T \cdot \mu + \frac{1}{2} \cdot \mu^T \cdot \Gamma \cdot \mu + \frac{1}{2} \cdot trace(\Gamma \cdot \Sigma)$$

$$\sigma^2[L] = \delta^T \cdot \Sigma \cdot \delta + \frac{1}{2} \cdot (trace(\Gamma \cdot \Sigma))^2$$

Britten-Jones and Schaeffer [1999] discuss how to obtain a good, tractable, approximation to the cdf of L under some circumstances. Rouvinez [1997] numerically inverts the characteristic function of the distribution.[18]

5.3 Conditional (not necessarily normal) moment models

In computing the gradients of VaR or TVaR, it may be possible to obtain the conditional moments of X (μ_q, Σ_q) by simulation or analytical methods.

In particular, for VaR:

---

[18] Britten-Jones and Schaeffer also remark "It is always possible to proceed by simulation."

$$C = \sum_{i:|L_i - \mathsf{L}_q| < w} p_i \cdot K(L_i - \mathsf{L}_q, w)$$
$$\boldsymbol{\mu}_q = C^{-1} \cdot \sum_i \mathbf{x}^{<i>} \cdot p_i \cdot K(L_i - \mathsf{L}_q, w)$$
$$\boldsymbol{\Sigma}_q = C^{-1} \cdot \sum_i (\mathbf{x}^{<i>} - \boldsymbol{\mu}_q) \cdot (\mathbf{x}^{<i>} - \boldsymbol{\mu}_q)^T \cdot p_i \cdot K(L_i - \mathsf{L}_q, w)$$

and for TVaR:

$$C = \sum_{i:L_i > \mathsf{L}_q} p_i$$
$$\boldsymbol{\mu}_q = C^{-1} \cdot \sum_i \mathbf{x}^{<i>} \cdot p_i$$
$$\boldsymbol{\Sigma}_q = C^{-1} \cdot \sum_i (\mathbf{x}^{<i>} - \boldsymbol{\mu}_q) \cdot (\mathbf{x}^{<i>} - \boldsymbol{\mu}_q)^T \cdot p_i$$

For a bilinear model (e.g., the delta model of sections 3.1 and 5.1) μ_q can be substituted for the conditional mean of $\mathbf{X}$ in the VaR and TVaR gradient formulas (section 3.1.3); covariance is of no consequence.

A quadratic g function (the gamma model, specialized in section 5.2) is nearly as simple, because $g(\mathbf{X}) = \nu + \boldsymbol{\delta}^T \cdot \mathbf{X} + \frac{1}{2} \cdot \mathbf{X}^T \cdot \boldsymbol{\Gamma} \cdot \mathbf{X}$ implies $E[\nabla_\delta g] = \boldsymbol{\mu}_q$ and $E[\nabla_\Gamma g] = (1/2)\boldsymbol{\Sigma}$. This follows from term-by-term differentiation. If $\boldsymbol{\delta}$ and $\boldsymbol{\Gamma}$ are themselves functions of a scalar parameter θ, then $E\left[\frac{\partial g}{\partial \theta}\right] = \frac{\partial \boldsymbol{\delta}}{\partial \theta} \cdot \boldsymbol{\mu}_q + \frac{1}{2} \cdot \sum_{i,j} \left(\frac{\partial \Gamma_{i,j}}{\partial \theta} \cdot \Sigma_{i,j} \right)$. This follows from the chain rule. If $\boldsymbol{\theta}$ is a vector parameter, the obvious componentwise extension applies. Here, only the first two moments matter and normality is not needed.

5.4 Conditional normality yields analytic solution for XOL portfolio management

The problem of evaluating the marginal impact of adding or dropping a cat layer yields to a tractable solution under the assumption of conditional normality. Recall the marginal impact model from section 3.3:

$$g(\mathbf{X};\theta) = g_0 + \theta \cdot r\left((\boldsymbol{\Omega}^{<u>})^T \cdot \mathbf{X}, \alpha_u, \lambda_u, \kappa_u \right)$$

with the solution $\nabla_\theta \mathsf{R}_q = E_*\left[r\left((\boldsymbol{\Omega}^{<u>})^T \cdot \mathbf{X}, \alpha_u, \lambda_u, \kappa_u \right) \right]$ where the conditional expectation is taken with respect to the distribution appropriate for R being VaR or TVaR. If $\mathbf{X}$ is assumed conditionally distributed as $N(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q)$, define the following:[19]

[19] Computation of μ_q and Σ_q are discussed in section 5.3.

$$\mu = \left(\Omega^{<u>}\right)^T \cdot \mu_q$$
$$\sigma^2 = \left(\Omega^{<u>}\right)^T \cdot \Sigma_q \cdot \Omega^{<u>}$$
$$\varsigma(t) = \xi(t, \mu, \sigma)$$

where $\xi(t,\mu,\sigma)$ was defined in section 5.1. Then:

$$\nabla_\theta \mathsf{R}_q = E_*[r_u] = (1 - \kappa_u) \cdot (\varsigma(\alpha_u) - \varsigma(\alpha_u + \lambda_u)).$$

This follows from the lemma proved in section 5.1 and the construction of the *r* function.

The normal approximation reduces the calculation of the marginal impact of a contract to a matrix multiplication, two vector multiplications, and a relatively simple scalar formula.[20]

This technique can be refined by extending to a mixture of normals.[21] Divide sample points into groups $\{\mathbf{x}^{<i1>}\}$, $\{\mathbf{x}^{<i2>}\}$,..., $\{\mathbf{x}^{<ik>}\}$,..., $\{\mathbf{x}^{<iG>}\}$, e.g., by a clustering algorithm (Anderberg [1973]). Compute separate moment vectors μ_k and variance-covariance matrices Σ_k for each group *k* as described previously. Instead of a single multivariate normal with density $\phi(\mathbf{X}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, the mixture has density $\sum_{k=1}^{G} \omega_k \cdot \phi(\mathbf{X}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, where the coefficients ω_k are proportional to $\sum_{i_k} p_{i_k} \cdot K(L_{i_k} - \mathsf{L}_q, w)$ for the VaR and to $\sum_{i_k} p_{i_k}$ for the TVaR (and sum to one). The layer expectation is similarly a weighted sum of terms of the previous form. The added cost is more computation;[22] the benefit is improved accuracy of the approximation.

6. Case study

6.1 Background: CAT model

In this section a simplified CAT model illustrates the application of the gradient theorems and helps assess their numerical accuracy in the context of Monte Carlo simulation.

Consider a hypothetical stretch of coastline with locations measured from 0 to 10. An event will consist of a simulated hurricane making landfall at location *t* and producing damage at location *s* of X(*s*,*t*,*r*) where the parameter *r* is the maximum loss rate for that event. Specifically,

---

[20] Good approximations for the normal cdf are available in many mathematics and statistics handbooks (e.g., Abramowitz and Stegun [1964], Johnson et. al. [1994]) and are built-in functions in many PC software packages, e.g. Excel.

[21] One could go further and model arbitrary *g* by a piecewise linear approximation built up from several *r* functions.

[22] Not necessarily proportional to G, however. It may be the case that separate groups reflect a number of shared zero **X**-components, thus reducing the effective dimensions of the group variance-covariance matrices compared to the matrix for the entire isoclast which must be at least as large.

$$X(s,t,r)=\frac{r}{1+(t-s)^2}.$$

Figure 1 shows such a damage profile for t = 0.4, r = 1.

The simulation will create random hurricanes (**X**-vectors indexed by *s*) according to a specific joint distribution for *t* and *r*. The joint distribution is defined as follows. The variable *t* is distributed uniformly between 0 and 10. Conditional on *t*, the variable *r* is distributed exponentially with mean

$$\bar{r}|t=(10+4\cdot\cos(0.2\cdot\pi\cdot(t-1.5))+2\cdot\sin(0.7\cdot\pi\cdot(t-4.5)))^{-1}$$

Figure 2 graphs the conditional mean of *r* as a function of *t*.

From this distribution, 50,000 equal-probability events were generated. This set was resampled by importance sampling (Rubinstein [1981]) on *r* resulting in 5002 events with associated sampling weights (relative frequencies). Figure 3 shows a 20% subsample of these events.

Events generated as (*r,t*) pairs were converted to vectors $\langle X_s: s = 0,1,\ldots,10\rangle$ and stored on disk.

6.2 Ground up losses

Figure 4 shows, and Table 1 enumerates, a hypothetical schedule of insured property exposures along a stretch of coastline, with locations numbered from 0 to 10. This vector corresponds to the parameter $\boldsymbol{\Omega}$ in the equation $L=\boldsymbol{\Omega}^T\cdot\mathbf{X}$. Note the subscripting needs to handle potential as well as actual members of the portfolio. In a real simulation, rather than list thousands of individual insured properties one would create group aggregations, say, by ZIP code and occupancy type. If location #s has no current presence, but a potential future presence, in the portfolio being studied, then $\Omega_s = 0$, but there is still a need to model the risk factor component $\mathbf{X}_s$.

After carrying out the multiplication $L_i=\boldsymbol{\Omega}^T\cdot\mathbf{X}^{<i>}$ for each event *i*, one has a discrete distribution of total (ground up) losses L_i each with associated probability p_i. The cumulative distribution function is shown in figures 5a and 5b. Moments for the loss are $\mu = 0.682$ and $\sigma = 0.805$. The mean is indicated in figure 5a.

Take $q = 0.02$ as the reference point for VaR and TVaR. VaR for losses is 3.139, or just a bit more than 3 standard deviations above the mean. VaR is indicated in figures 5a and 5b. TVaR is 4.251.

The gradients of VaR and TVaR were computed both via the "brute force" approach and by the conditional expectation theorems (section 4.2.3). The brute force approach was to make a discrete change $\Delta\Omega_s$ to the exposure at location *s*, recomputing the statistic R, and then calculating the rate of change by $(R(\Omega+\Delta\Omega)-R(\Omega-\Delta\Omega))/2\Delta\Omega$. This requires a choice for the increment $\Delta\Omega$. Values of 2 and 8 were chosen for illustrative purposes.

The computation of ∇TVaR by the expectation theorem[23] does not require a "bandwidth." The results are shown in figure 6. The vertical axis (for the curves) represents the gradient of TVaR with respect to changes in the exposure at each location. Exposures are shown by small bars at the bottom of the graph. The results from the theorem are shown by the dashed line. The thick and thin lines show the brute force approach with different $\Delta\Omega$. The results are quite consistent between the three methods. Locations 6 and

[23] For ground-up loss, $\nabla_\Omega g=X$.

7, with the most intense event risk, are also the most sensitive to exposure change. Locations 3, 4, and 10 are the least. Location 1 is more sensitive than locations 9 and 10, despite the fact that the event risk at location 1 is less intense than it is at the right-hand end. This is because of the relative concentration of exposures on the left end. Of the 5002 events, 781 (15.6%) entered into the theorem-based calculation of ∇TVaR.

The computation of ∇VaR by the expectation theorem does require a "bandwidth" w. Choices of 0.05 and 0.2 were used.[24] The results are shown in figure 7. They are quite similar to that of the TVaR graph, except that the various methods and parameter picks show less consistency. For w = 0.05, only 66 events were within the kernel, and for w = 0.2, 213 events. Thus by using the expectation theorems only 1.3% or 4.3% of the events needed to enter into the weighted average. Since this is the bilinear model, the vectors being averaged took on a particularly simple form (i.e., $\nabla_\Omega g = \mathbf{X}$).

Coupled with a model for the gradient of profitability, the gradient of VaR or TVaR can map proposed (small) portfolio changes into changes in the risk-reward plane of a portfolio risk management model. For example, if the relative profitability of business varied little over the locations, strategies involving writing business in locations 2-4 and 10, possibly coupled with attrition at locations 1 and 6, would yield the best changes in risk-reward position. While some of this might seem obvious from examination of the exposures and damage distributions alone, in a real application there would be many more locations (and classes of business) with correspondingly more difficult tradeoffs to assess.

6.3 Reinsurance buying

To make the problem a bit more realistic, consider that each location has a per-risk treaty associated with it, with an attachment $\alpha_s = 0.5$, limit $\lambda_s = 1$, and co-reinsurance $\kappa_s = 0.1$.

The mean and standard deviation of net loss are 0.569 and 0.563, respectively. Thus the expected net loss is 84% of the expected gross loss. The 2% VaR is 2.00, only 64% of the gross loss's, and the TVaR is 2.98, about 70% of what it is for the gross loss.

But is this the optimal set of attachments and limits? After all, the exposures vary by a factor of 4:1 from location to location. Perhaps it would make more sense to alter the α and λ parameters. Specifically, how do changes in α and λ vectors translate to changes in VaR and TVaR? Here, the gradient theorems must be applied in their more general form, because the relationship between net loss and the parameters in question is not linear.[25] There are 12 nontrivial parameters to be assessed here: one attachment point and one limit for each of the six locations with a treaty. Attachments and limits for locations at which there is no exposure have no effect on the loss and hence a zero gradient.

Figure 8 shows the ∇TVaR results for the attachment parameters. Values for $\Delta\alpha$ of 0.05 and 0.2 were chosen for the brute-force method. Curiously, location 1 is much less sensitive to a change in attachment than the adjacent locations 0 and 2. Out of the tail events defining TVaR, location 1 is only "in the layer" about 4.4% of the time whereas locations 0 and 2 are in the layer 18% and 21% of the time, respectively. While location 1 is more likely to exhaust the layer, that situation is insensitive to changes in attachment point. In particular, this means that the attachment point can be raised with (relatively) little impact on the net loss. A canny buyer would investigate the possibility of a premium reduction to accompany such a change. For net losses, 812 events (16.2%) entered into the computation of ∇TVaR by the expectation method.

---

[24] These choices are consistent with the $\Delta\Omega$ picks of 2 and 8 because the mean X across all locations is 0.027.

[25] $\nabla_\alpha g$ and $\nabla_\lambda g$ are given in section 3.2.1.

Location 6 has a low sensitivity for the same reason as location 1: in the tail, location 6 usually either does not attach (42.9% of the time) or completely exhausts the layer (55.2%); it is only in the layer 1.7% of the time.

Figure 9 shows the comparable results for ∇VaR with respect to attachment parameters. Again, bandwidth w = 0.05 and 0.2 were chosen for the expectation method.[26] These results are comparable to those of the TVaR graph, and again show less consistency. For the two bandwidth picks, 78 (1.6%) and 338 (6.8%) events, respectively, entered into the expectation calculations.

Figures 10 and 11 show the results for ∇TVaR and ∇VaR with respect to changes in the limits at each location. Since increasing a limit can only reduce net loss, the negative of the gradients is graphed. Changing limit is most effective at location 6 (which exhausts its cover in over half the tail events) and nearly ineffective at location 2 (which only exhausts in about 1% of the tail events).

6.4 Reinsurer portfolio of XOL layers

Consider now the problem of assessing the impact that adding a new contract or nonrenewing an existing contract has on the VaR and TVaR of a portfolio of layers.

Table 2 shows a proposed contract and 25 existing contracts. Each is defined by exposures at the same 11 locations as modeled previously, with attachments and limits applying to the total damages from all locations of a contract combined. With this small an example, it is feasible to rerun the model on the new portfolio, to find the new values of VaR and TVaR, and hence the new contract's marginal contributions to risk. On the other hand, application of the theorems requires evaluating the losses on *only* the new contract, and *only* for the scenarios within the kernel window or tail. In this case, that means one instead of 25 contracts and typically 2% to 20% of the scenarios (depending on bandwidth and risk measure), resulting in hundreds of times fewer loss evaluations.

The mean payout for this portfolio is 0.68, and its standard deviation is 2.908. The mean payout of the proposed new contract is 0.033 and its standard deviation is 0.175. Figure 12 shows the means and standard deviations for the existing contracts, with the statistics for the new contract overlaid as horizontal dotted lines.

The 2% VaR for the existing portfolio is 12.48, a bit more than 4 standard deviations above the mean. The TVaR is 18.21. Of the 5002 simulated events, 825 (16.5%) are in the 2% upper tail.

Direct computation of the change in VaR or TVaR upon adding or dropping a contract does not require a choice for the change amount as in previous examples; the natural unit is one contract. The computation of ∇TVaR by the expectation theorem does not require a bandwidth.

The approximation formula for ∇TVaR (based on the model for g in section 3.3) is:

$$\nabla \mathrm{T}_q(\theta) \approx \frac{1}{q} \cdot \sum_{\{i:L_i > \mathrm{L}_q(\theta)\}} r\left(\left(\Omega^{<u>}\right)^T \cdot \mathbf{X}^{<i>}, \alpha_u, \lambda_u, \kappa_u\right) \cdot p_i$$

where the summation, please note, is over the scenarios where the loss to the *existing* portfolio exceeds the VaR. In the case of dropping existing contracts, this formula is exactly the discrete implementation of the *co-TVaR* measure (Kreps [2003]). In the case of adding a proposed contract, the slight difference here is that the VaR does not include the effect of the proposed contract.

---

[26] These are consistent with $\Delta\alpha$ because in the layers, changes in attachment are approximately one-for-one effective in changing net losses.

The change in TVaR upon adding the new contract is 0.682, whereas it is estimated by the expectation theorem to be 0.683. Figure 13 shows the change in TVaR upon cancelling each of the existing contracts. The estimates are quite close, with a root mean square error of 0.012.

For the computation of ∇VaR by the expectation theorem, bandwidths of 0.5 and 1.5 were assumed. These encompassed 74 and 245 events, respectively. The change in VaR by adding the new contract is 0.198; it is estimated as 0.246 or 0.234 for the two bandwidths, respectively, 24% and 18% higher than the actual. Figure 14 shows the change in VaR upon cancelling each of the existing contracts. The estimates track fairly well, with a root mean square error of 0.083.

6.5 Normal approximation

Using the mean and variance-covariance matrix of the **X** vectors in the 2% tail, one can assume a normal distribution and evaluate ∇TVaR by a closed-form expression (section 5.4). The estimate for the new contract is 0.728, about 7% higher than the actual figure of 0.682. Figure 15 shows the results for the existing contracts. The root mean square error is 0.092.

Again, less accurate results are obtained for closed-form ∇VaR computation. The estimates for the new contract based on the two bandwidths are 0.304 and 0.329, respectively, 53% and 66% higher than actual. Figure 16 shows the results for the existing contracts. These had root mean square errors of 0.121 and 0.114, respectively.

7. Related work

The Value-at-Risk (VaR) approach to risk assessment is becoming more prevalent in the insurance industry as computing power increases and the convergence between finance and insurance disciplines continues. However, the VaR concept is not new to insurance. Actuaries have been familiar with the VaR approach (although not the name) since at least 1969, when McGuinness [1969] proposed a mathematical definition for PML.[27] What is different today is the recognition of the similarities between insurance and finance risk management needs and the consequent cross-fertilization of methodologies for evaluating the risk of a portfolio (whether that be an investment portfolio, a portfolio of insurance policies, or an entire enterprise).

Kreps [1990] links reinsurance pricing to the incremental risk a new contract adds to the reinsurer's portfolio. Kreps [1999] places reinsurance pricing on the same theoretical footing as other investment decisions. Major and Kreps [2002] present a field study on how reinsurers think and behave with respect to pricing catastrophic risk.

Kreps [2003] provides a general method to decompose risk measures into additive component "co-measures." Tasche [1999] develops a theory of capital allocation and portfolio component returns where the first derivative of a (fairly arbitrary) risk capital measure of choice plays a central role. His principle of "suitability" can be interpreted as saying that the rate of return on a component of a portfolio (with respect to allocated capital) should exceed the portfolio rate of return in precisely those situations where an additional increment of that component would increase the overall portfolio rate of return. He shows that in order for suitability to hold, the capital allocation must be equal to the first derivative of the risk measure. Acerbi [2002] presents a "spectral decomposition" of coherent risk measures which includes TVaR as a special case. Major [2003] provides an analysis of suitability for a variety of risk measures and explores the relationship between suitability, spectral measures, and co-measures.

---

[27] Since VaR is essentially a quantile of the cumulative distribution function, it is difficult to assign original credit. Gauss's [1809] work on the normal distribution extended the work of De Moivre [1738] and Laplace [1783], with other distributions having been studied a century earlier.

The literature on VaR is enormous; many papers are accessible through www.gloriamundi.org. Jorion [2000] is a good introduction. Fuglsbjerg [2000] applies variance reduction techniques to estimate VaR. Glasserman et. al. [2000] discuss efficient VaR estimation in the presence of heavy-tail risk factors. Artzner et. al. [1999] present axioms for risk measures which lead many to conclude that TVaR is superior to VaR.

The gradient of Value at Risk in the context of the linear loss model is well known. Garman [1996] analyzes the gradient in the delta-normal model and provides a blueprint for using the gradient in risk management. Hallerbach [1999] focuses on the decomposition of VaR between the risk factors, with a nuanced discussion distinguishing marginal, component, and incremental VaR. Gourieroux et. al. [2000] also deal with the second derivative and discuss kernel smoothing in the context of estimation. Tasche [1999] presents the gradient in a more general case for the distribution of the risk factors. Wang [2002] develops the Monte Carlo sampling distribution of incremental VaR.

Many researchers have gone beyond the derivative to explore VaR and TVaR optimization in the linear loss model. Andersson et. al. [2001] apply their results to the problem of credit risk TVaR. Rockafellar and Uryasev [2000], Uryasev [2000a, b], Krokhmal et. al. [2001], and Larsen et. al. [2002] develop a technique that simultaneously calculates VaR and optimizes TVaR. Krokhmal et. al. [2001] provide an example based on a portfolio of S&P 100 stocks. Larsen et. al. [2002] provide an example based on a portfolio of emerging market bonds. Gaivoronski and Pflug [2000] optimize the VaR and show that mean-VaR-optimal frontiers are quite different from mean-TVaR frontiers. Lemus Rodriguez [1999] proposes a nonparametric statistical technique for the estimation of the gradient and uses it to numerically optimize linear combinations of VaRs. Yamai and Yoshiba [2002] focus on sampling issues in comparing VaR and TVaR optimization.

The literature on the gradient of VaR in the context of nonlinear loss models is less well developed. Britten-Jones and Schaefer [1999] "complete the square" in the quadratic model to show that the distribution of losses is a sum of non-central chi-squares. They also provide formulas for the mean and variance and show how to approximate the cumulative distribution function.[28] Uryasev [1995a, b; 1999] provides the general framework for evaluating derivatives as integrals for arbitrary loss models. Uryasev [2000a] explicitly treats the derivative of VaR in the general case. (The optimization discussed there is with regard to the linear model, however.) Major et. al. [2001] and Belubekian [2003] discuss the use of the gradient in an insurance portfolio management context.

There is also a vast body of literature on other approaches to risk measurement, analysis, and management. Only a few items will be mentioned here. Berger et. al. [1998] optimize risk-adjusted return through the "tabu search" technique. Manganelli et. al. [2002] develop the gradient of GARCH variance in the linear model. Studer [1995] and Studer and Luthi [1996] present "Maximum Loss" as an alternative concept to VaR, and show its relationship with and advantages to VaR.

8. Conclusion

The gradient of a risk measure is a powerful tool for analyzing the high-dimensional problem of directing change in the composition of a portfolio. Whereas thousands of simulation runs might be needed to test all the relevant "what-if" scenarios (assuming they could all be identified), the gradient can quickly point out the maximally advantageous directions for change, whether that means growth in opportunity areas, attrition in risk-concentrated areas, or adjusting other contractual terms. Since Parkinson's Law (Parkinson

[28] See section 5.2 for a discussion

[1958]) seems to apply to simulation modeling,[29] the ability to lever modeling resources in this manner is crucial to a thorough analysis.

This paper reviewed basic risk measures including VaR and TVaR and discussed various closed-form and simulation implementations of the gradient. Insurance examples illustrated their use in CAT modeling, reinsurance program design, and reinsurance portfolio management. The general simulation methods are equally effective with linear parameters such as exposure amounts and non-linear parameters such as retentions and limits. A normal approximation was shown to solve the XOL contract marginal impact problem with a simple formula.

9. Appendix

9.1 Dictionary of mathematical symbolism

9.1.1 Styles and fonts

These standards are typical, but not rigidly enforced:

- **bold** is a vector or matrix
- *italic* is a scalar variable or function
- CAPITAL is a random variable
- γρεεκ is a parameter
- `Special` is a probabilistic operator (risk measure)

9.1.2 Specific symbols

$\mathbf{X}$	random variable in R^N, represents hazard process (section 2.1)
$f(\mathbf{x})$	probability density function of $\mathbf{X}$ (section 2.1)
$L = g(\mathbf{X},\theta)$	defines the losses to a portfolio of insured exposures (section 2.1)
θ	parameter in R^M, describes the transformation of $\mathbf{X}$ to a loss dollar amount (section 2.1)
$\int_t f(\mathbf{x})dV$	volume integral where $\mathbf{x}$ ranges over the N-dimensional region t in R^N
$\int_t f(\mathbf{x})dS$	surface integral where $\mathbf{x}$ ranges over the (N-1)-dimensional submanifold t in R^N
$E[g(\mathbf{X})]$	mathematical expectation of $g(\mathbf{X})$ with respect to pdf $f(\mathbf{x})$
$h(\mathbf{x})$	transformed density on isoclastic hypersurface (section 2.3.4)
$k(\mathbf{x})$	conditional density in tail (section 2.3.5)
$E_h[g(\mathbf{X})]$	mathematical expectation of $g(\mathbf{X})$ with respect to pdf $h(\mathbf{x})$
$R(\theta)$	generic risk measure (section 2.2.4)
L_q	value at risk or VaR (section 2.2.3)
T_q	tail value at risk or TVaR (section 2.2.3)

[29] "Work expands to fill the available time" becomes "requirements expand to absorb available computer resources."

$\partial f/\partial x$	partial derivative of f with respect to x (section 2.2.4)
∇	gradient operator; vector of partial derivatives (section 2.2.4)
$\|\mathbf{v}\|$	norm (magnitude) of vector $\mathbf{v}$
C_N	normalizing factor, probability integrated over the isoclastic surface (section 2.3.3)
α	attachment (retention) of reinsurance layer (section 3.2)
λ	limit of reinsurance layer (section 3.2)
κ	co-reinsurance of reinsurance layer (section 3.2)
Ω	parameter vector defines exposures (section 3.2)
$r(L,\alpha,\lambda,\kappa)$	reinsurance payoff function (section 3.2)
$N = L - r$	net loss after reinsurance (section 3.2)
$\Omega^{<c>}$	vector number c; column c of matrix (section 3.3)
$K(t,w)$	Epanechnikov kernel function (section 4.2.2)
μ	mean operator or mean of $\mathbf{X}$
$\Phi(z)$	cumulative normal distribution function (section 5.1)
$\phi(z)$	normal probability distribution function (section 5.1)
μ_q	conditional mean of $\mathbf{X}$ at the VaR isoclast or TVaR tail (section 5.3)
Σ	summation operator or variance-covariance matrix of $\mathbf{X}$
Σ_q	conditional variance-covariance of $\mathbf{X}$ at the VaR isoclast or TVaR tail (section 5.3)
$\xi(t,\mu,\sigma)$	expected value E[max(0,Y-t)] for normal distribution (section 5.1)

9.2 Gradient of Tail Value at Risk

The gradient of TVaR is derived. Recall

$$\mathrm{T}_q(\boldsymbol{\theta}) = \frac{1}{q} \cdot \int_{g(\mathbf{x},\boldsymbol{\theta}) > \mathrm{L}_q(\boldsymbol{\theta})} g(\mathbf{x},\boldsymbol{\theta}) \cdot f(\mathbf{x}) dV \,.$$

Applying the integral over the surface formula with $\psi = \boldsymbol{\theta}$, $\mathrm{H} = \mathrm{T}_q$, $\phi = g(\mathbf{x},\boldsymbol{\theta})f(\mathbf{x})$, and $\gamma = \mathrm{L}_q - g(\mathbf{x},\boldsymbol{\theta})$,

$$\nabla_\theta \mathsf{T}_q(\theta) = E\left[\nabla_\theta g(\mathbf{x},\theta) \middle| g(\mathbf{X},\theta) > \mathsf{L}_q\right] + \frac{1}{q} \cdot \int_{g(\mathbf{x},\theta)=\mathsf{L}_q} \frac{\nabla_\theta\left(g(\mathbf{x},\theta) - \mathsf{L}_q\right)}{\left\|\nabla_{\mathbf{x}} g(\mathbf{x},\theta)\right\|} \cdot g(\mathbf{x},\theta) \cdot f(\mathbf{x}) dS =$$

$$= E\left[\nabla_\theta g(\mathbf{x},\theta) \middle| g(\mathbf{X},\theta) > \mathsf{L}_q\right] + \mathsf{L}_q \cdot \frac{C_N}{q} \cdot E\left[\frac{\nabla_\theta\left(g(\mathbf{X},\theta) - \mathsf{L}_q\right)}{\left\|\nabla_{\mathbf{x}} g(\mathbf{X},\theta)\right\|} \middle| g(\mathbf{X},\theta) = \mathsf{L}_q\right]$$

where $C_N = \int_{g(\mathbf{x},\theta)=\mathsf{L}_q} f(\mathbf{x}) dS.$

$$= E\left[\nabla_\theta g(\mathbf{x},\theta) \middle| g(\mathbf{X},\theta) > \mathsf{L}_q\right] + \mathsf{L}_q \cdot \frac{C_N}{q} \cdot \left(E\left[\frac{\nabla_\theta g(\mathbf{X},\theta)}{\left\|\nabla_{\mathbf{x}} g(\mathbf{X},\theta)\right\|} \middle| g = \mathsf{L}_q\right] - \nabla_\theta \mathsf{L}_q \cdot E\left[\frac{1}{\left\|\nabla_{\mathbf{x}} g(\mathbf{X},\theta)\right\|} \middle| g = \mathsf{L}_q\right] \right)$$

Recall (section 2.3.4)

$$\nabla_\theta \mathsf{L}_q = \frac{E\left[\frac{\nabla_\theta g(\mathbf{X},\theta)}{\left\|\nabla_{\mathbf{X}} g(\mathbf{X},\theta)\right\|} \middle| g(\mathbf{X},\theta) = \mathsf{L}_q\right]}{E\left[\frac{1}{\left\|\nabla_{\mathbf{X}} g(\mathbf{X},\theta)\right\|} \middle| g(\mathbf{X},\theta) = \mathsf{L}_q\right]}$$

Therefore the expression in parentheses is zero and we are left with

$$\nabla_\theta \mathsf{T}_q(\theta) = E\left[\nabla_\theta g(\mathbf{x},\theta) \middle| g(\mathbf{X},\theta) > \mathsf{L}_q\right].$$

9.3 Derivations of gradients under the delta-normal model

9.3.1 Gradients of exceedance probability

Here, the gradients $\frac{\partial Q}{\partial L} = -\frac{C_N}{\|\boldsymbol{\delta}\|}$ and $\nabla_\delta Q = \frac{C_N}{\|\boldsymbol{\delta}\|} \cdot \frac{L - g(\mathbf{x}_0)}{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}$, presented in section 5.1.1, are derived in two ways.

Method 1: Differentiation of the expression $Q(L;\boldsymbol{\delta}) = 1 - \Phi\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}}\right)$

$$\frac{\partial Q}{\partial L} = -\phi\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}}\right) \cdot \frac{\partial}{\partial L}\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}}\right) = -\frac{1}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}} \cdot \phi\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}}\right)$$

$$\nabla_\delta Q = -\phi\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}}\right) \nabla_\delta\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}}\right) = \frac{L - g(\mathbf{x}_0)}{\left(\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}\right)^{3/2}} \cdot \phi\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}}\right) \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}$$

Recalling (section 5.1.1) that $C_N = \frac{\|\delta\|}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}} \cdot \phi\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}}\right)$, the expressions using C_N follow.

Method 2: Using the conditional expectation theorem for Q (section 2.3.3), provides insight into the nature of the isoclast. As a consequence of the linear case,

$$\frac{\partial Q}{\partial L} = -C_N \cdot E\left[\|\delta\|^{-1} \middle| \delta^T \cdot \mathbf{X} = L - g(\mathbf{x}_0)\right] = -C_N / \|\delta\|.$$

$$\nabla_\delta Q = C_N \cdot E\left[\frac{\nabla_\delta g}{\|\nabla_x g\|} \middle| g = L\right] = C_N \cdot E\left[\frac{\mathbf{X}}{\|\delta\|} \middle| \delta^T \cdot \mathbf{X} = L - g(\mathbf{x}_0)\right].$$

Apply a singular value decomposition $\Sigma = \Lambda \cdot \mathbf{P} \cdot \Lambda^T$ where $\mathbf{P}$ is diagonal (eigenvalues) and Λ (eigenvectors) satisfies $\Lambda^T\Lambda=\mathbf{I}$. Then transform variables to $\mathbf{X} = \Lambda \cdot \mathbf{P}^{1/2} \cdot \mathbf{Z}$ where $\mathbf{Z}$ is a spherical unit normal (i.e. mean zero and identity variance-covariance matrix) random variable. Therefore, one can rewrite

$$\nabla_\delta Q = \frac{C_N}{\|\delta\|} \cdot \Lambda \cdot \mathbf{P}^{1/2} \cdot E\left[\mathbf{Z} \mid \delta^T \cdot \Lambda \cdot \mathbf{P}^{1/2} \cdot \mathbf{Z} = L - g(\mathbf{x}_0)\right].$$

The conditional distribution of $\mathbf{Z}$ is an N-1 dimensional spherical normal with mean=mode being the point $\mathbf{z}_0$ on the conditional hyperplane closest to the origin. The unit vector $\mathbf{u}$ in that direction is given by:

$$\mathbf{u} = \frac{\mathbf{P}^{1/2} \cdot \Lambda^T \cdot \delta}{\sqrt{\delta^T \cdot \Lambda \cdot \mathbf{P} \cdot \Lambda^T \cdot \delta}}$$

Putting $\mathbf{z}_0 = k\mathbf{u}$ and solving for k gives us $k = \frac{L - g(\mathbf{x}_0)}{\sqrt{\delta^T \cdot \Lambda \cdot \mathbf{P} \cdot \Lambda^T \cdot \delta}}$ and therefore

$$\nabla_\delta Q = \frac{C_N}{\|\delta\|} \cdot \Lambda \cdot \mathbf{P}^{1/2} \cdot \frac{L - g(\mathbf{x}_0)}{\sqrt{\delta^T \cdot \Lambda \cdot \mathbf{P} \cdot \Lambda^T \cdot \delta}} \cdot \frac{\mathbf{P}^{1/2} \cdot \Lambda^T \cdot \delta}{\sqrt{\delta^T \cdot \Lambda \cdot \mathbf{P} \cdot \Lambda^T \cdot \delta}} = \frac{C_N}{\|\delta\|} \cdot \frac{L - g(\mathbf{x}_0)}{\delta^T \cdot \Sigma \cdot \delta} \cdot \Sigma \cdot \delta.$$

9.3.2 Gradient of Value at Risk

Here, the gradient

$$\nabla_\delta \mathsf{L}_q = \frac{z_q}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}} \cdot \Sigma \cdot \delta$$

is derived in two ways.

Method 1: Term-by-term differentiation with respect to δ of the expression for $\mathsf{L}_q(\delta)$ gives us

$$\nabla_\delta \mathsf{L}_q(\delta) = \nabla_\delta\left(g(\mathbf{x}_0) + z_q \cdot \sqrt{\delta^T \cdot \Sigma \cdot \delta}\right) = \nabla_\delta g(\mathbf{x}_0) + \nabla_\delta\left(z_q \cdot \sqrt{\delta^T \cdot \Sigma \cdot \delta}\right)$$
$$= 0 + z_q \cdot \nabla_\delta\left(\sqrt{\delta^T \cdot \Sigma \cdot \delta}\right) = z_q \cdot \frac{1}{2 \cdot \sqrt{\delta^T \cdot \Sigma \cdot \delta}} \nabla_\delta\left(\delta^T \cdot \Sigma \cdot \delta\right)$$
$$= \frac{z_q}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}} \Sigma \cdot \delta$$

Method 2: Apply the formula from 2.3.4 expressing $\nabla_\delta \mathsf{L}_q(\delta)$ as the ratio of gradients of Q:

$$\frac{\partial \mathsf{L}_q}{\partial \theta} = -\frac{\partial Q/\partial \theta}{\partial Q/\partial L} = -\frac{\frac{C_N}{\|\delta\|} \cdot \frac{\mathsf{L}_q - g(\mathbf{x}_0)}{\delta^T \cdot \Sigma \cdot \delta} \cdot \Sigma \cdot \delta}{-\frac{C_N}{\|\delta\|}} = \frac{\mathsf{L}_q - g(\mathbf{x}_0)}{\delta^T \cdot \Sigma \cdot \delta} \cdot \Sigma \cdot \delta = \frac{z_q}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}} \cdot \Sigma \cdot \delta .$$

9.3.3 Gradient of Tail Value at Risk

Here, the gradient

$$\nabla_\delta \mathsf{T}_q = \frac{\phi(z_q)}{q} \frac{\Sigma \cdot \delta}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}}$$

is derived in two ways.

Method 1: Term-by-term differentiation with respect to δ of the expression for $\mathsf{T}_q(\delta)$ gives us

$$\nabla_\delta \mathsf{T}_q(\delta) = \nabla_\delta\left(g(\mathbf{x}_0) + \frac{\phi(z_q)}{q} \cdot \sqrt{\delta^T \cdot \Sigma \cdot \delta}\right) = \nabla_\delta g(\mathbf{x}_0) + \nabla_\delta\left(\frac{\phi(z_q)}{q} \cdot \sqrt{\delta^T \cdot \Sigma \cdot \delta}\right)$$
$$= 0 + \frac{\phi(z_q)}{q} \cdot \nabla_\delta\left(\sqrt{\delta^T \cdot \Sigma \cdot \delta}\right) = \frac{\phi(z_q)}{q} \cdot \frac{1}{2 \cdot \sqrt{\delta^T \cdot \Sigma \cdot \delta}} \nabla_\delta\left(\delta^T \cdot \Sigma \cdot \delta\right)$$
$$= \frac{\phi(z_q)}{q} \cdot \frac{\Sigma \cdot \delta}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}}$$

Method 2: We may use the conditional expectation theorem (sections 2.3.5, 9.2) in a manner parallel to the derivation of the gradient of exceedance probability above. As a consequence of linearity,

$$\nabla_\delta \mathsf{T}_q = E\left[\nabla_\delta g(\mathbf{X}, \theta) \mid g > \mathsf{L}_q\right] = E\left[\mathbf{X} \mid \frac{\delta^T \cdot \mathbf{X}}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}} > z_q\right].$$

As before, transform variables to $\mathbf{X} = \Lambda \cdot \mathbf{P}^{1/2} \cdot \mathbf{Z}$ with spherical unit normal $\mathbf{Z}$, and then by a rotation to $\mathbf{Y}$ via $\mathbf{Z} = Y_1 \cdot \mathbf{u}^{<1>} + \sum_{j=2}^{N} Y_j \cdot \mathbf{u}^{<j>}$ where $\mathbf{u}^{<1>}$ is the "**u**" defined in section 9.3.1.

Note that the $\mathbf{u}^{<j>}$ are orthogonal to each other. The random variable $\mathbf{Y}$ is still spherical normal, being related to $\mathbf{Z}$ by a rotation. Therefore, one can rewrite

$$\nabla_\delta \mathsf{T}_q = \boldsymbol{\Lambda}\cdot\mathbf{P}^{1/2}\cdot E\left[Y_1\cdot\mathbf{u}^{<1>}+\sum_{j=2}^{N}Y_j\cdot\mathbf{u}^{<j>}\mid \mathbf{u}^{<1>T}\cdot\mathbf{Z}>z_q\right]$$
$$=\boldsymbol{\Lambda}\cdot\mathbf{P}^{1/2}\cdot\left(E\left[Y_1\cdot\mathbf{u}^{<1>}\mid Y_1>z_q\right]+E\left[\sum_{j=2}^{N}Y_j\cdot\mathbf{u}^{<j>}\mid Y_1>z_q\right]\right)$$

Since components Y_j are independent of Y_1, the second expectation is zero, and therefore (recalling the ξ function from section 5.1):

$$\nabla_\delta \mathsf{T}_q=\boldsymbol{\Lambda}\cdot\mathbf{P}^{1/2}\cdot E\left[Y_1\mid Y_1>z_q\right]\cdot\mathbf{u}^{<1>}=\boldsymbol{\Lambda}\cdot\mathbf{P}^{1/2}\cdot\left(z_q+E\left[Y_1-z_q\mid Y_1>z_q\right]\right)\cdot\mathbf{u}^{<1>}$$
$$=\boldsymbol{\Lambda}\cdot\mathbf{P}^{1/2}\cdot\left(z_q+\frac{1}{q}\cdot\xi\left(z_q,0,1\right)\right)\cdot\mathbf{u}^{<1>}=\boldsymbol{\Lambda}\cdot\mathbf{P}^{1/2}\cdot\left(z_q+\frac{1}{q}\cdot\left(\phi\left(z_q\right)-z_q\cdot\left(1-\Phi\left(z_q\right)\right)\right)\right)\cdot\mathbf{u}^{<1>}$$
$$=\boldsymbol{\Lambda}\cdot\mathbf{P}^{1/2}\cdot\left(z_q+\frac{1}{q}\cdot\left(\phi\left(z_q\right)-z_q\cdot q\right)\right)\cdot\mathbf{u}^{<1>}=\boldsymbol{\Lambda}\cdot\mathbf{P}^{1/2}\cdot\frac{\phi\left(z_q\right)}{q}\frac{\mathbf{P}^{1/2}\cdot\boldsymbol{\Lambda}^T\cdot\boldsymbol{\delta}}{\sqrt{\boldsymbol{\delta}^T\cdot\boldsymbol{\Lambda}\cdot\mathbf{P}\cdot\boldsymbol{\Lambda}^T\cdot\boldsymbol{\delta}}}$$
$$=\frac{\phi\left(z_q\right)}{q}\frac{\boldsymbol{\Lambda}\cdot\mathbf{P}\cdot\boldsymbol{\Lambda}^T\cdot\boldsymbol{\delta}}{\sqrt{\boldsymbol{\delta}^T\cdot\boldsymbol{\Lambda}\cdot\mathbf{P}\cdot\boldsymbol{\Lambda}^T\cdot\boldsymbol{\delta}}}$$
$$=\frac{\phi\left(z_q\right)}{q}\frac{\boldsymbol{\Sigma}\cdot\boldsymbol{\delta}}{\sqrt{\boldsymbol{\delta}^T\cdot\boldsymbol{\Sigma}\cdot\boldsymbol{\delta}}}$$

10. Figures

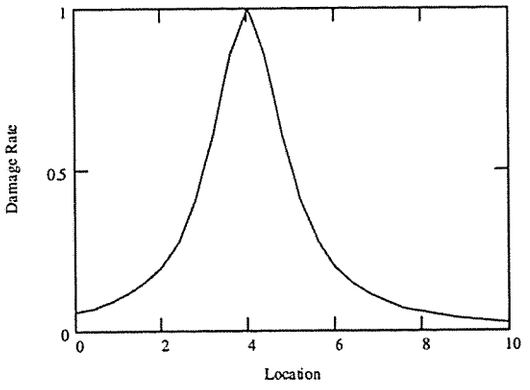

Figure 1: Profile of simulated hurricane damage, landfall at t=4, maximum loss rate r=1.

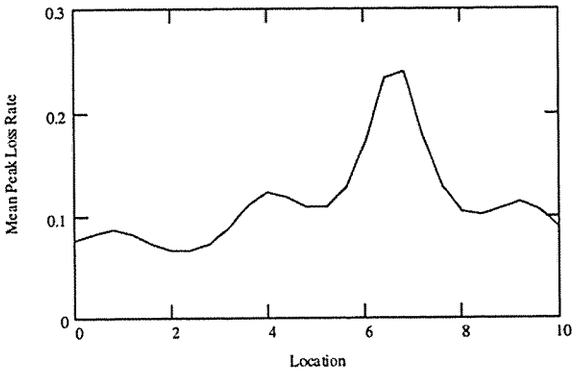

Figure 2: Conditional mean r as a function of location t.

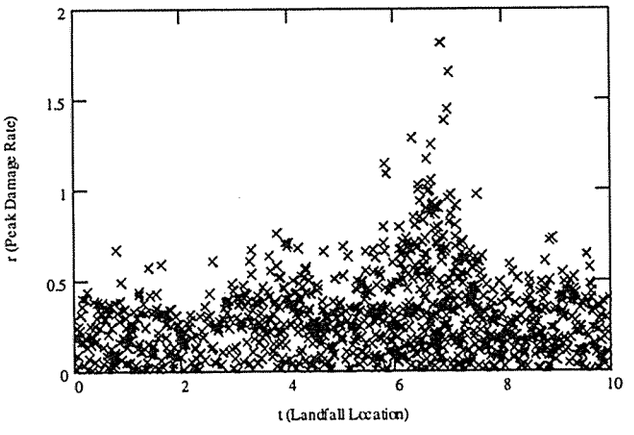

Figure 3: 1000 of the final 5002 events. Note: events with lower r values have higher associated weights.

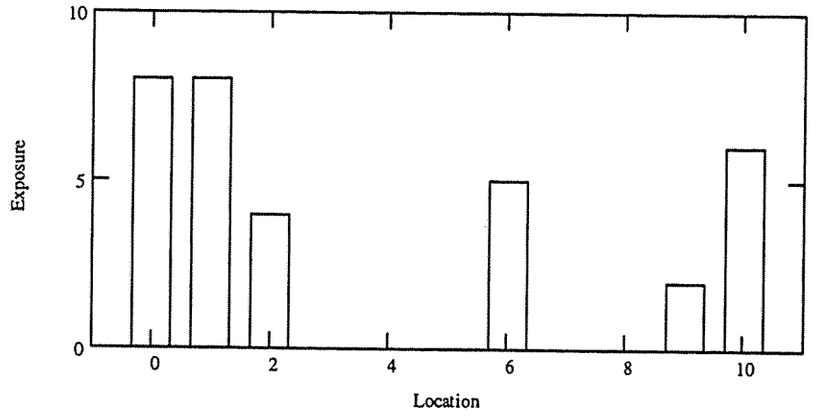

Figure 4: Insured exposures by location for simulation experiment

Location	Exposure
0	8
1	8
2	4
6	5
9	2
10	6

Table 1: Insured exposures by location for simulation experiment

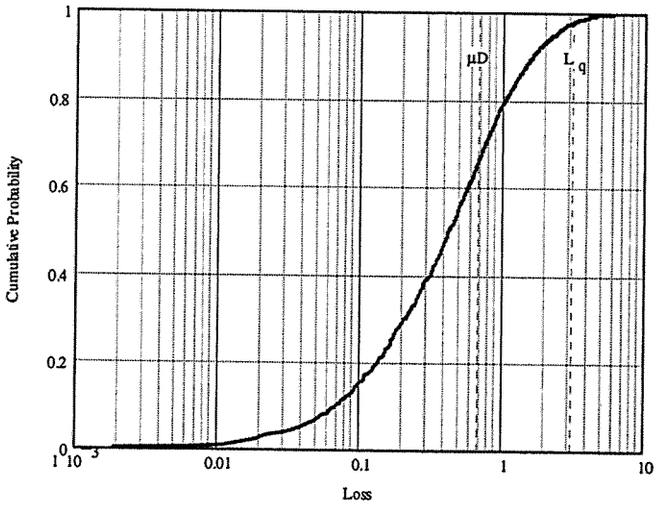

Figure 5a: Cumulative loss distribution for portfolio in figure 4

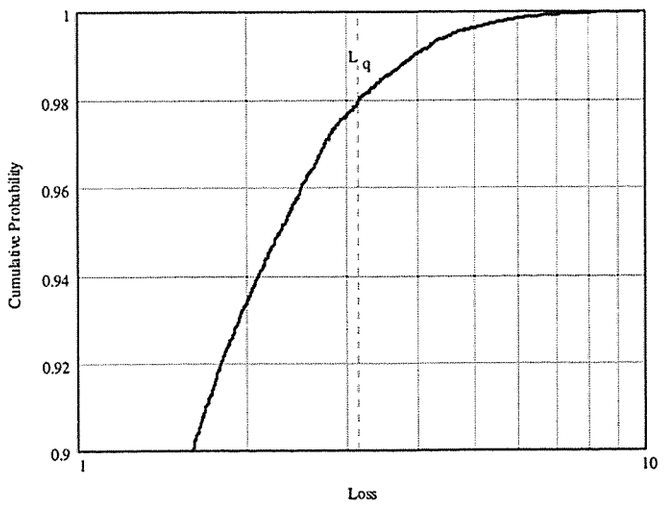

Figure 5b: Cumulative loss distribution for portfolio in figure 4 (closeup)

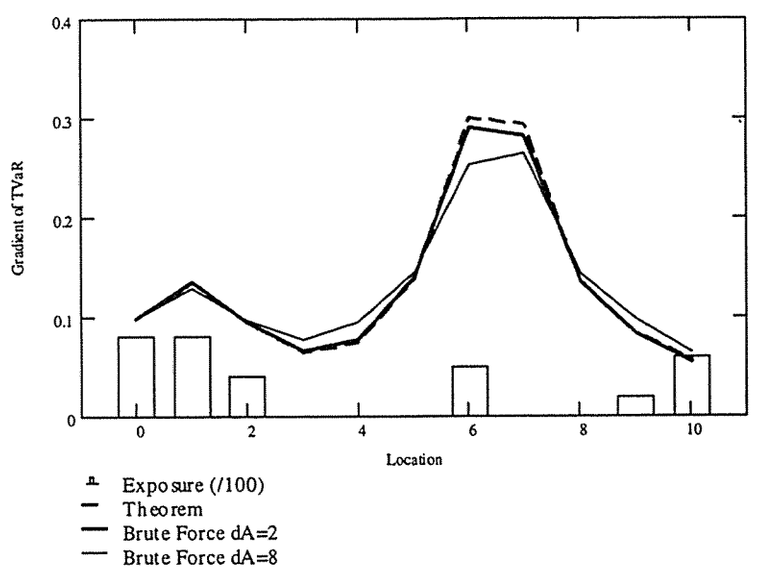

Figure 6: Comparison of methods for computing gradient of Tail Value at Risk with respect to exposures for ground up losses

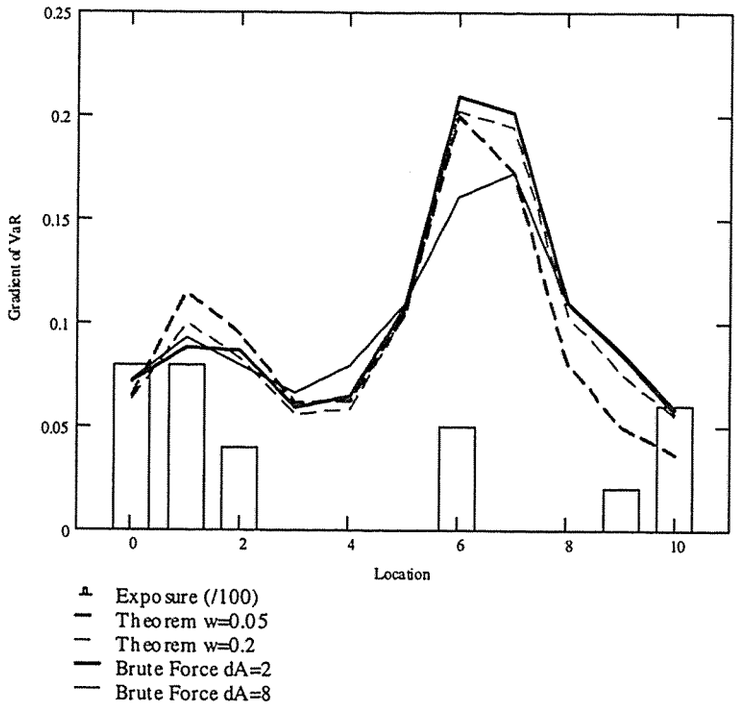

Figure 7: Comparison of methods for computing gradient of Value at Risk with respect to exposures for ground up losses

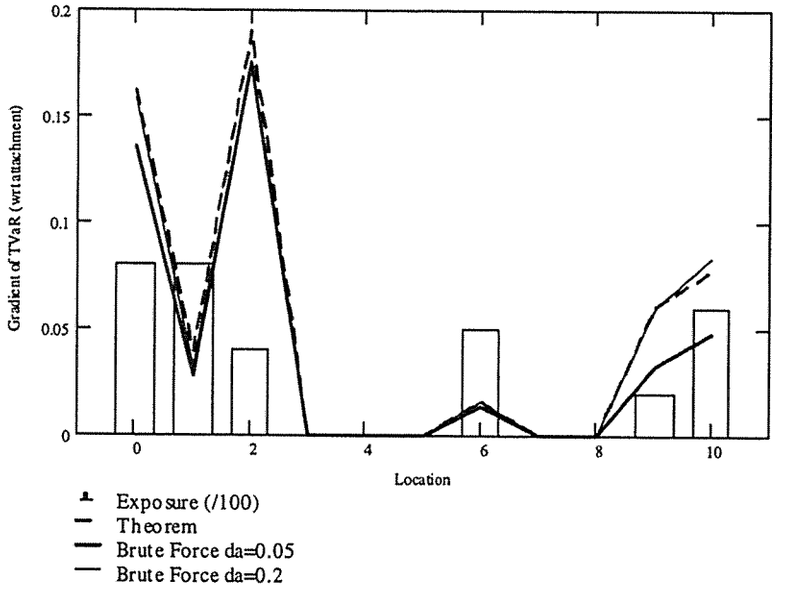

Figure 8: Comparison of methods for computing gradient of Tail Value at Risk with respect to attachment for losses net of treaty reinsurance

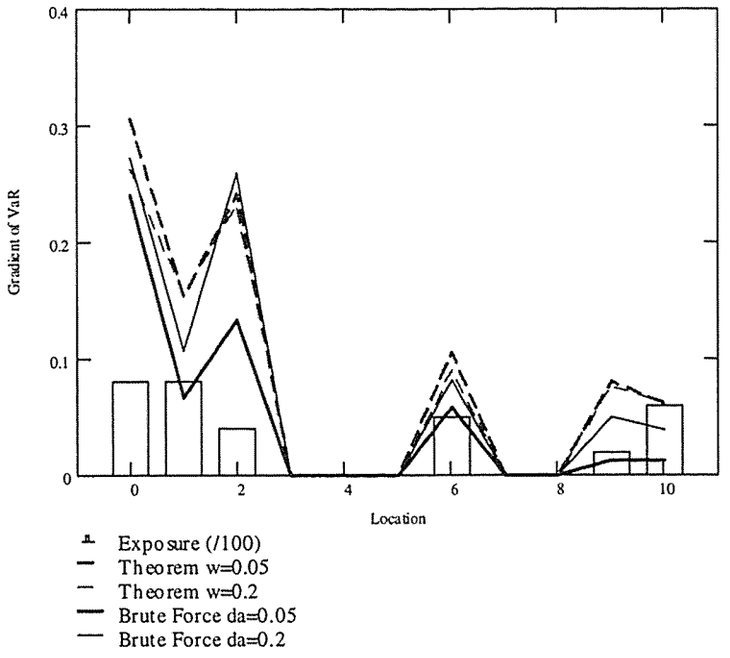

Figure 9: Comparison of methods for computing gradient of Value at Risk with respect to attachment for losses net of treaty reinsurance

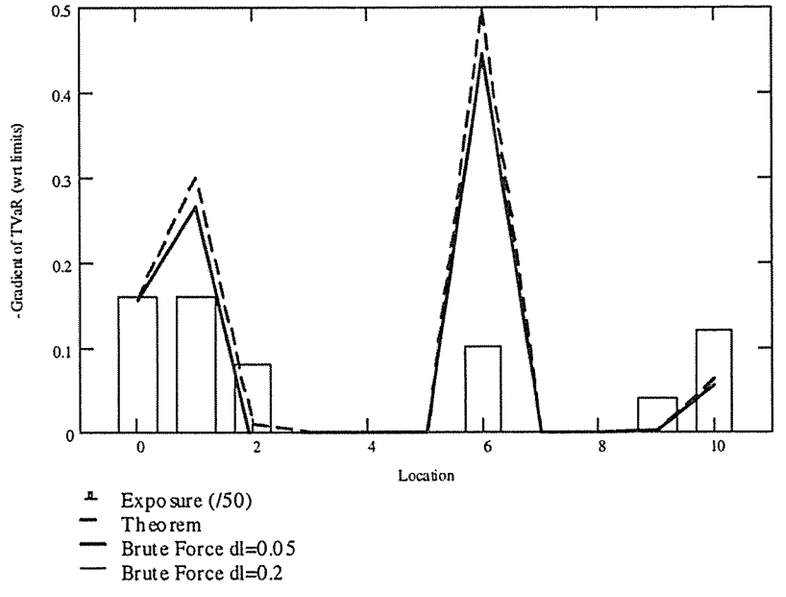

Figure 10: Comparison of methods for computing gradient of Tail Value at Risk with respect to limit for losses net of treaty reinsurance.

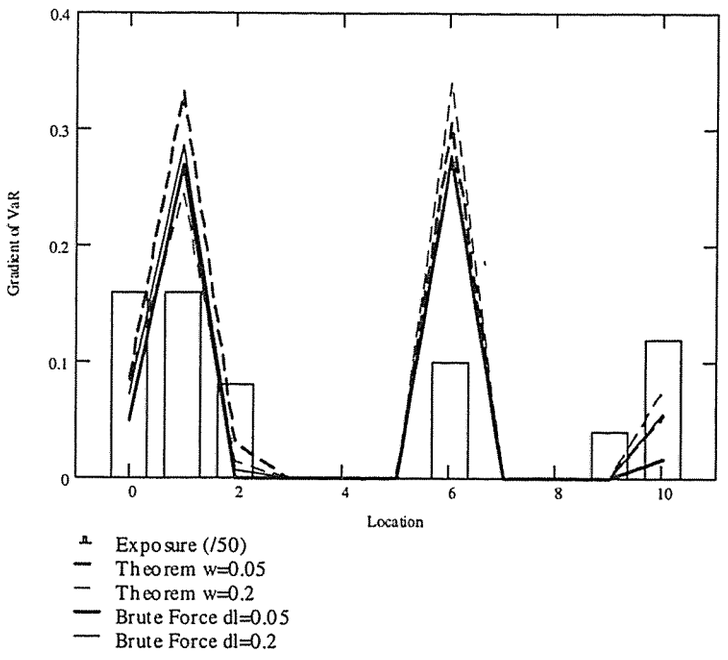

Figure 11: Comparison of methods for computing gradient of Value at Risk with respect to limit for losses net of treaty reinsurance.

	Exposure at Location											Att	Lim
	0	1	2	3	4	5	6	7	8	9	10		
New	8	8	4	0	0	0	5	0	0	2	6	2.3	1.2
Existing													
1	0	0	5	0	3	0	4	0	0	0	6	1.5	1
2	0	0	7	6	0	0	0	5	7	4	4	3	1
3	2	0	6	8	8	6	0	6	0	6	0	4	2
4	0	8	8	0	6	0	6	7	5	4	0	5	1
5	0	3	6	0	0	5	0	0	6	0	2	2	1
6	0	8	0	6	0	3	0	7	7	0	6	4	1
7	6	0	3	7	0	5	0	2	7	0	3	3	1
8	6	0	5	6	4	3	7	0	0	0	0	3	2
9	5	3	0	0	8	0	3	0	2	4	0	3	1
10	6	0	3	4	6	6	0	0	0	0	0	3	1
11	5	6	0	0	5	0	4	9	4	5	0	5	1
12	5	0	4	5	0	6	6	5	8	8	4	6	1
13	7	0	0	0	3	0	0	4	5	0	0	3	1
14	0	3	6	7	0	0	8	6	0	3	0	4	1
15	0	6	5	0	0	7	3	4	0	0	5	3	1
16	7	0	3	3	0	0	5	0	0	0	3	3	1
17	2	0	4	0	0	0	0	8	3	3	0	2	1
18	0	0	0	3	6	0	5	0	4	6	0	5	1
19	0	0	3	0	4	0	6	0	4	0	0	3	1
20	0	4	6	4	0	4	0	5	4	0	0	3	1
21	5	3	0	0	0	3	6	5	5	0	8	4	1
22	0	4	0	6	0	5	0	0	0	0	6	2	1
23	4	0	0	4	0	0	0	0	0	0	0	1	1
24	2	0	6	5	6	0	4	6	4	0	0	2	1
25	5	0	4	0	7	5	0	0	3	2	0	3	1

Table 2: Portfolio of reinsurance layers with proposed new contract.

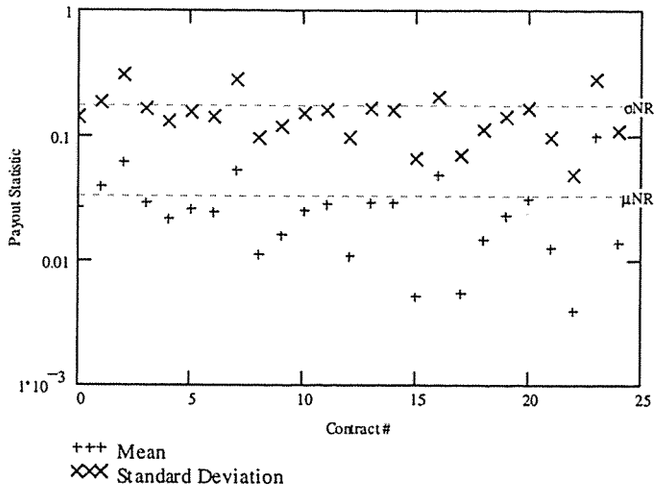

Figure 12: Payout statistics for new and existing contracts.

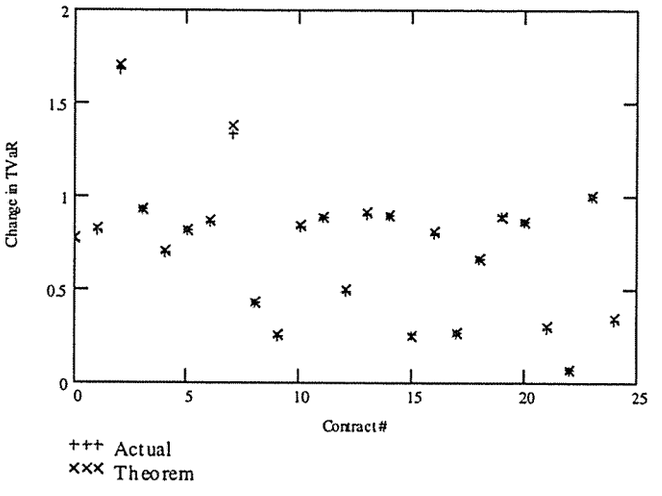

Figure 13: Comparison of actual and estimated change in TVaR upon cancelling each existing policy.

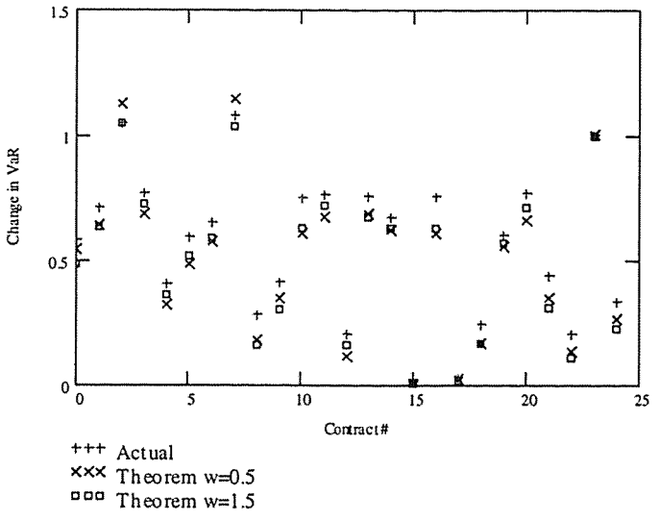

Figure 14: Comparison of actual and estimated change in VaR upon cancelling each existing policy.

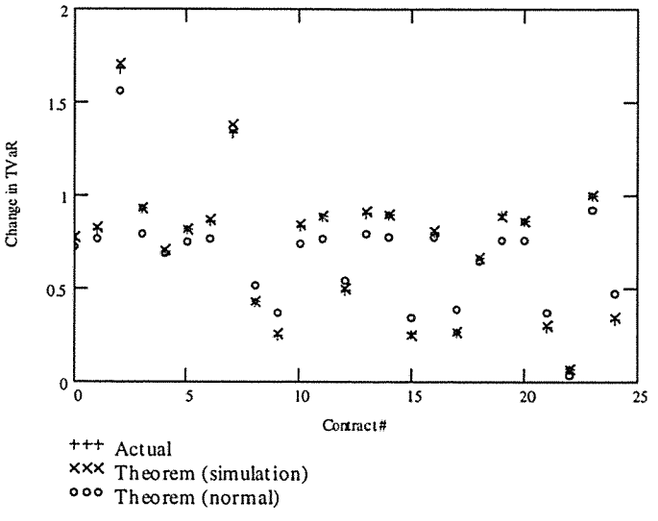

Figure 15: Comparison of actual and estimated change in TVaR upon cancelling each existing policy, including multivariate normal closed-form approximation.

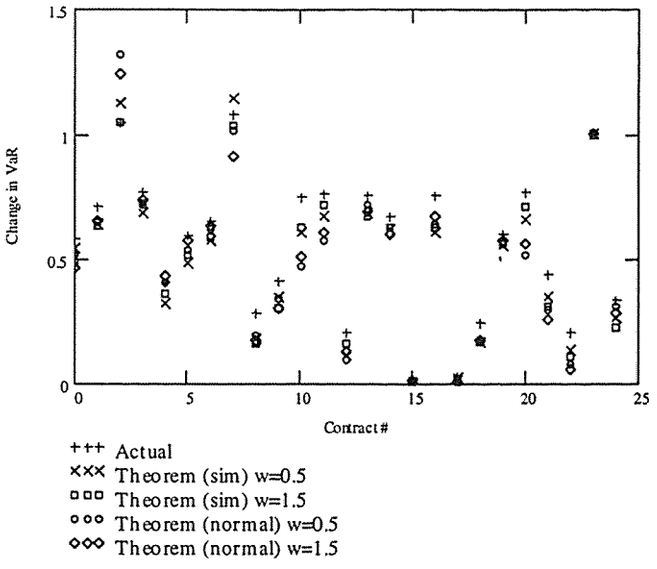

Figure 16: Comparison of actual and estimated change in VaR upon cancelling each existing policy, including multivariate normal closed-form approximation.

11. References

Abramowitz, M. and I. A. Stegun (1964) *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables,* Washington DC: National Bureau of Standards, U. S. Government Printing Office.

Acerbi, Carlo (2002) "Spectral measures of risk: A coherent representation of subjective risk aversion," *Journal of Banking & Finance* 26, pp. 1505–1518.

Anderberg, Michael R. (1973) *Cluster Analysis for Applications,* New York: Academic Press.

Andersson, Fredrik, Helmut Mausser, Dan Rosen and Stanislav Uryasev (2001) "Credit risk optimization with Conditional Value-at-Risk Criterion," *Mathematical Programming* Series B 89, 273-291.

Artzner, P., F. Delbaen, J.-M. Eber and D. Heath (1999) "Coherent measures of risk," *Mathematical Finance* 9(3), 203-228.

Basel Committee on Banking Supervision (1996) "Amendment to the Capital Accord to Incorporate Market Risks," Basel, Switzerland: Bank for International Settlements.

Belubekian, Maya E. (2003) "Getting Results – and Exploiting Them," *Risk Transfer* Vol. 1 No. 5.

Berger, Adam J., John M. Mulvey, Kevin Nish and Robert Rush (1998) "A Portfolio Management System for Catastrophe Property Liabilities," Alexandria VA: Casualty Actuarial Society.

Britten-Jones, Mark and Stephen M. Schaefer (1999) "Non-Linear Value-at-Risk," *European Finance Review* 2: 161–187.

Daykin, C. D., T. Pentikäinen and M. Pesonen (1994) *Practical Risk Theory for Actuaries,* London: Chapman & Hall.

De Moivre, Abraham (1738) *Doctrine of Chances: A Method of Calculating the Probabilities of Events in Play.* Republished 1967 Irvington, Chelsea.

Fuglsbjerg, Brian (2000) "Variance Reduction Techniques for Monte Carlo Estimates of Value at Risk," Copenhagen, Denmark: SimCorp A/S.

Gaivoronski, Alexei A. and Georg Pflug (1999) "Finding Optimal Portfolios with Constraints on Value at Risk," *Proceedings of the III Stockholm Seminar on Risk Behavior and Risk Management,* January.

Gaivoronski, Alexei A. and Georg Pflug (2000) "Value at Risk in Portfolio Optimization: Properties and Computational Approach," Working paper, Department of Industrial Economics and Technology Management, NTNU (Norwegian University of Science and Technology).

Garman, Mark (1996) "Improving on VaR," *Risk* Vol. 9, No.5, pp. 61-63.

Gauss, C. F. (1809) *Theoria Motus Corporum Coelestium,* Hamburg: Perthes & Besser. (English translation by C. H. Davis, 1857, Boston: Little, Brown.)

Glasserman, Paul, Philip Heidelberger and Perwez Shahabuddin (2000) "Variance Reduction Techniques for Value-at-Risk with Heavy-Tailed Risk Factors," *Proceedings of the 2000 Winter Simulation Conference,* J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

Gourieroux, C., J. P. Laurent and O. Scaillet (2000) "Sensitivity Analysis of Values at Risk," *Econometric Society World Congress 2000 Contributed Papers.*

Hallerbach, Winfried G. (1999) "Decomposing Portfolio Value-at-Risk: A General Analysis," Technical Report, Erasmus University Rotterdam and Tinbergen Institute Graduate School of Economics.

Johnson, N. L., S. Kotz and N. Balakrishnan (1994) *Continuous Univariate Distributions Volume 1*, New York: Wiley-Interscience.

Jorion, Philippe (2000) *Value at Risk: The New Benchmark for Managing Financial Risk,* McGraw-Hill.

Klugman, Stuart A., Harry H. Panjer and Gordon E. Willmot (1998) *Loss Models: From Data to Decisions,* New York: Wiley-Interscience.

Kreps, R. E. (1990) "Reinsurer Risk Loads from Marginal Surplus Requirements," *Proceedings of the Casualty Actuarial Society* LXXX.

____ (1999) "Investment-Equivalent Reinsurance Pricing," in *Actuarial Considerations Regarding Risk and Return in Property-Casualty Insurance Pricing*, O. Van Slyke, ed., Alexandria VA: Casualty Actuarial Society.

____ (2003) "Riskiness Leverage Models," Instrat Working Paper, New York: Guy Carpenter & Co., Inc.

Krokhmal, Pavlo, Jonas Palmquist, and Stanislav Uryasev (2001) "Portfolio Optimization with Conditional Value-At-Risk Objective and Constraints" *Journal of Risk* Vol. 4 No. 2, Winter.

Laplace, Pierre S. (1783) "Sur les Naissances, les Mariages, et les Morts," *Mém. Acad. R. Sci. Paris,* pp. 693-702.

Larsen, Nicklas, Helmut Mausser and Stanislav Uryasev (2002) "Algorithms for Optimization of Value-at-Risk" in *Financial Engineering, e-Commerce and Supply Chain,* P. Pardalos and V.K. Tsitsiringos, eds., Kluwer Academic Publishers, 129-157.

Lemus Rodriguez, Gerardo José (1999) "Portfolio Optimization with Quantile-based Risk Measures," Massachusetts Institute of Technology, March.

Major, J. A. (2003) "Risk Contributions, Performance Measurement, and Co-measures," Instrat Working Paper, New York: Guy Carpenter & Co., Inc.

Major, J. A., M. Belubekian and S. Mathai (2001) "Portfolio Risk Management: Using the Gradient of Value-at-Risk," *Financing Risk & Reinsurance,* International Risk Management Institute, May.

Major, J. A. and R. E. Kreps (2002) "Catastrophe Risk Pricing in the Traditional Market," in *Alternative Risk Strategies*, M. Lane, ed., London: Risk Waters Group.

Manganelli, Simone, Vladimiro Ceci and Walter Vecchiato (2002) "Sensitivity Analysis of Volatility: A New Tool for Risk Management," European Central Bank Working Paper No. 194, November.

McGuinness, John S. (1969) "Is 'Probable Maximum Loss' (PML) a Useful Concept?" *Proceedings of the Casualty Actuarial Society* LVI.

Parkinson, C. Northcote (1958) *Parkinson's Law: The Pursuit of Progress,* London: John Murray.

Rockafellar, R. Tyrrell and Stanislav Uryasev (2000) "Optimization of Conditional Value-at-Risk" *Journal of Risk* Volume 2/Number 3, Spring.

Rouvinez, C. (1997) "Going Greek with VaR" *Risk* 10(2): 57-65.

Rubinstein, R. Y. (1981) *Simulation and the Monte Carlo Method*, New York: Wiley-Interscience.

Silverman , B. W. (1986) *Density Estimation for Statistics and Data Analysis*, CRC Press.

Studer, G. (1995) "Value At Risk and Maximum Loss Optimization," RiskLab Technical Report, ETH Zurich, December.

Studer, G. and H. J. Luthi (1996) "Quadratic Maximum Loss for Risk Measurement of Portfolios," RiskLab Technical Report, ETH Zurich, September.

Tasche, Dirk (1999) "Risk Contributions and Performance Measurement," Report of the Lehrstuhl für Mathematische Statistik, T.U. München.

Uryasev, S. (1995a) "Derivatives of Probability Functions and Integrals over Sets Given by Inequalities," *J. Computational and Applied Mathematics*, 56, 197-223.

____ (1995b) "Derivatives of Probability Functions and Some Applications," *Annals of Operations Research* 56, pp. 287-311.

____ (1999) "Derivatives of Probability and Integral Functions: General Theory and Examples," in *Encyclopedia of Optimization,* C.A. Floudas and P. M. Pardalos, eds., Kluwer Academic Publishers.

____ (2000a) "Introduction to the Theory of Probabilistic Functions and Percentiles (Value-at-Risk)," in *Probabilistic Constrained Optimization: Methodology and Applications*, S. P. Uryasev, ed., Kluwer Academic Publishers, pp. 1-25.

____ (2000b) "Conditional Value-at-Risk: Optimization Algorithms and Applications, *Financial Engineering News,* February.

Wang, Zheng (2002) "The Properties Of Incremental VaR In Monte Carlo Simulations, *The Journal of Risk Finance,* Spring.

Yamai, Yasuhiro and Toshinao Yoshiba (2002) "Comparative Analyses of Expected Shortfall and Value-at-Risk: Their Estimation Error, Decomposition, and Optimization," *Monetary And Economic Studies,* January.