Actuarial Geometry

Stephen J. Mildenhall*

March 12, 2006

Abstract

The literature on capital allocation is biased towards an asset modeling framework rather than an actuarial framework. The asset modeling framework leads to the proliferation of inappropriate assumptions about the effect of insurance line of business growth on aggregate loss distributions. This paper explains why an actuarial analog of the asset volume/return model should be based on a Lévy process. It discusses the impact of different loss models on marginal capital allocations. It shows that Lévy process-based models provide a better fit to the NAIC annual statement data. Finally, it shows the NAIC data suggest a surprising result regarding the form of insurance parameter risk.
Keywords: Capital determination, capital allocation, risk measure, game theory, Lévy process, parameter risk, diversification.

1 INTRODUCTION

Geometry is the study of shape and change in shape. Actuarial geometry is the study of the shape and evolution of shape of actuarial variables, in particular the distribution of aggregate losses, as portfolio volume and composition changes. It is also the study of the shape and evolution paths of variables in the space of all risks. Actuarial variables are curved across both

*e-mail: steve_mildenhall@aon.com. Tel: 312-381-5880. Aon Re Services, 200 East Randolph Street, 16th floor, Chicago, IL 60601. Filename: rts.tex (1254) a volumetric dimension as well as a temporal dimension[1]. Asset variables, on the other hand, are flat in the volumetric dimension and only curved in the temporal dimension. Risk, and hence economic quantities like capital, are intimately connected to the shape of the distribution of losses, and so actuarial geometry has an important bearing on capital determination and allocation.

Actuarial geometry is especially important today because risk and probability theory, finance, and actuarial science have begun to converge after prolonged development along separate tracks. In risk and probability theory, the basic building-block stochastic processes we discuss—the compound Poisson process and Brownian motion—have been studied intensively for over 100 years. All of the basic theoretic results in this paper were known by 1940s, as were risk theoretic results, such the Cramer-Lundberg theorem giving the relationship between probability of eventual ruin and safety loading, Bühlmann (1970); Bowers et al. (1986). In addition, risk theoretic approaches led to a plethora of different distribution-based pricing methods such as the standard deviation and utility principles, Borch (1962); Goovaerts et al. (1984)

Actuarial[2] science in the US largely ignored this body of theoretical knowledge in its day-to-day work because of the dominance of bureau-based rates. Property rates historically were made to include a 5% profit provision and a 1% contingency provision; they were priced to a 96% combined ratio. Casualty lines were priced to a 95% combined ratio[3]. Regulators and actuaries started to consider improvements to these long-standing conventions in

[1]Volume refers to expected losses per year, x. Time is the duration, t, fow which a given volume of insurance is written. Total expected losses are xt—just as distance = speed × time. In the paper, volumetric is to volume as temporal is to time.

[2]References to actuaries and actuarial methods always refer to *property-casualty* actuaries.

[3]In 1921 the National Board of Fire Underwriters adopted a 5% profit plus 1% catastrophe (conflagration) load. This approach survived the South East Underwriters Supreme Court case and was reiterated in 1955 by the Inter-Regional Insurance Conference, Magrath (1958). In general liability a 5% provision for underwriting and contingencies is described as "constant for all liability insurance lines in most states" by Lange (1966). Kallop (1975) states that a 2.5% profit and contingency allowance for workers' compensation has been in use for at least 25 years and that it "contemplates additional profits from other sources to realize an adequate rate level". The higher load for property lines was justified by the possibility of catastrophic losses—meaning large conflagration losses rather than today's meaning of hurricane or earthquake related, frequency driven events.

the late 1960's. Bailey (1967) introduced actuaries to the idea of including investment income in profit. Ferrari (1968) was the first actuarial paper to include investment income and to consider return on investor equity as well as margin on premium. During the following dozen years actuaries developed the techniques needed to include investment income in ratemaking. At the same time, finance began to consider how to determine a fair rate of return on insurance capital. The theoretical results they derived, summarized as of 1987 in Cummins and Harrington (1987), focused on the use of discounted cash flow models using CAPM-derived discount rates for each cash flow, including taxes. Since CAPM only prices systematic risk, a side-effect of the financial work was to de-emphasize details of the distribution of ultimate losses in setting the profit provision.

At the same time option and contingent claim theoretic methods, Doherty and Garven (1986); Cummins (1988), were developed as another approach to determining fair premiums. This was spurred in part by the difficulty of computing appropriate β's. These papers applied powerful results from option pricing theory by using a geometric Brownian motion to model losses, possibly with a jump component—an assumption that was based more on mathematical expediency than reality. Cummins and Phillips (2000) and D'Arcy and Doherty (1988) contain a summary of the CAPM and contingent claims approaches from a finance perspective and D'Arcy and Dyer (1997) contains a more actuarial view.

The CAPM-based theories failed to take into consideration the observed fact that insurance companies charged for specific risk. A series of papers, beginning in the early 1990's, developed a theoretical explanation of this based around certainty in capital budgeting, costly external capital for opaque intermediaries, contracting under asymmetric information, and adverse selection, see Froot et al. (1993); Froot and Stein (1998); Merton and Perold (2001); Perold (2001); Zanjani (2002); Froot (2003)

At the same time banking regulation led to the development of robust risk measures. The most important was the idea of a coherent measure of risk, Artzner et al. (1999). Risk measures are obviously sensitive to the particulars of idiosyncratic firm risk, unlike the CAPM-based pricing methods which are only concerned with correlations.

The next theoretical step was to develop a theory of product pricing for a multiline insurance company within the context of costly firm-specific risk and robust risk measures. This proceeded down two paths. Phillips et al. (1998) considered pricing in a multiline insurance company from an option theoretic perspective, modeling losses with a geometric Brownian motion and without allocating capital. They were concerned with the effect of firm-wide insolvency risk on individual policy pricing. The second path, based around explicit allocation of capital, was started by Myers and Read (2001). They used expected default value as a risk measure, determined surplus allocations by line, and presented a gradient vector, Euler theorem based, allocation assuming volumetrically homogeneous losses—but making no other distributional assumptions. This thread was further developed by Tasche (2000); Denault (2001); Fischer (2003); Sherris (2006). Kalkbrener (2005) and Delbaen (2000a) used directional derivatives to clarify the relationship between risk measures and allocations.

With the confluence of these different theoretical threads, and, in particular, in light of the importance of firm-specific risk to insurance pricing, the missing link—and the link considered in this paper—is a careful examination of the underlying actuarial loss distribution assumptions. Unlike traditional *static* distribution-based pricing models, such as standard deviation and utility, modern marginal and differential methods require explicit *volumetric* and *temporal* components. The volumetric and temporal geometry are key to the differential calculations required to perform risk and capital allocations. All of the models used in the papers cited are, implicitly or explicitly, volumetrically homogeneous and geometrically flat in one dimension[4]. Mildenhall

[4] In a geometric Brownian motion model, losses at time t, S_t are of the form $S_t = S_0 \exp(\mu t + \sigma B_t)$ where B_t is a Brownian motion. Changing volume, S_0, simply scales the whole distribution and does not affect the shape of the random component. The jump-diffusion model in Cummins (1988) is of the same form. There are essentially no other explicit loss models in the papers cited. One other interesting approach was introduced by Cummins and Geman (1995) who model the rate of claims payment as a geometric Brownian motion. Claims paid to time t has the form $\int_0^t S_0 \exp(\mu\tau + \sigma B_\tau) d\tau$, but, again, this is volumetrically homogeneous.

(2004); Meyers (2005a) have shown this is not an appropriate assumption. This paper provides further evidence and uses the NAIC annual statement database to explore alternative, more appropriate, models.

With this background we now consider the details of capital determination and allocation. Insurance company capital is determined using a risk measure. The risk measure, ρ, is a real valued function defined on a space of risks $\mathbf{L} = L^0(\Omega, \mathcal{F}, \mathbb{P})$.[5] Given a risk $X \in \mathbf{L}$, $\rho(X)$ defines the amount of capital required to support the risk. The risk measure usually has one free parameter α which is interpreted as a level of safety or security. Examples of risk measures include value at risk, VaR, $\rho(X) = \mathrm{VaR}_\alpha(X) - \mathrm{E}(X)$, tail value at risk, TVaR, $\rho(X) = \mathrm{E}(X \mid X \geq \mathrm{VaR}_\alpha(X)) - \mathrm{E}(X)$, and standard deviation $\rho(X) = \alpha \mathrm{SD}(X)$. By definition $\mathrm{VaR}_\alpha(X)$ is the inverse of the distribution of X, so $\mathrm{VaR}_\alpha(X) = \inf\{x \mid \Pr(X < x) \geq \alpha\}$.

At the firm level, total risk X can be broken down into a sum of parts X_i corresponding to business units, line of business, policies, etc[6]. Since it is costly to hold capital, it is natural to ask for an attribution of total capital $\rho(X)$ to each line X_i. One way to do this is to consider the effect of a marginal change in the volume of line i on total capital. For example, if the marginal profit from line i divided by the marginal change in total capital resulting from a small change in volume in line i exceeds the average profit margin of the firm then it makes sense to expand line i. This is a standard economic optimization that has been discussed in the insurance context by many authors including Tasche (2000), Myers and Read (2001), Denault (2001), Meyers (2003) and Fischer (2003).

[5] Ω is the sample space, $\mathcal{F}$ is a sigma-algebra of subsets of Ω, and $\mathbb{P}$ is a probability measure on $\mathcal{F}$. The space L^0 consists of all real valued random variables, that is, measurable functions $X : \Omega \to \mathbb{R}$, defined up to equivalence (identify random variables which differ on a set of measure zero). As Delbaen (2000b) points out there are only two L^p spaces which are invariant under equivalent measures, L^0 and L^∞, the space of all essentially bounded random variables. Since it is desirable to work with a space invariant under change of equivalent measure, but not to be restricted to bounded variables, we work with L^0. Delbaen explains that the notion of a coherent measure of risk must be extended to allow for infinite values in order to be defined on all of L^0. Kalkbrener (2005) works on L^0.

[6] To be specific and concrete we will always describe the components, X_i, as lines.

The marginal approach leads us to consider

$$\frac{\partial \rho}{\partial X_i} \tag{1}$$

which (vaguely and generically) represents the change in ρ as a result of a change in "the direction of line i". This paper explains how the usual approach to $\partial\rho/\partial X_i$ makes an assumption regarding the actuarial geometry of X_i that is not appropriate, proposes a more "actuarial" assumption, and gives supporting empirical evidence for it.

At root, the difference between asset geometry and actuarial geometry boils down to a fundamental difference between an individual security, or asset, and a line of insurance. A line is analogous to a mutual fund specializing in an asset class and not to an individual asset. When modeling assets it is usual to pick a set of asset return processes $X_1(t), \dots, X_n(t)$ and to focus attention on $\mathbf{L}' \subset \mathbf{L}$ the real vector space they span—see Tasche (2000); Fischer (2003). The asset return processes need not be linearly independent. Each $X_i(t)$ represents the return from asset i over a time period t. To streamline notation, let $X_i = X_i(1)$. Thus $x_i X_i$, $x_i \in \mathbb{R}$ the real numbers, equals the end-of-period 1 value of a holding of x_i units of asset i. We will call this an asset volume/return model. By linear extension, there is a linear map between n-tuples $(x_1, \dots, x_n) \in \mathbb{R}^n$ and portfolio value distributions. The end-of-period value of a portfolio has distribution $\sum_i x_i X_i$. Clearly portfolios are linear because $\sum(x_i + y_i)X_i = \sum_i x_i X_i + \sum_i y_i X_i$. In this context, a risk measure ρ on $\mathbf{L}$ induces a map $\mathbb{R}^n \to \mathbf{L}' \subset \mathbf{L} \to \mathbb{R}$. Computing and interpreting Eq. 1 is straightforward for functions $\mathbb{R}^n \to \mathbb{R}$.

For insurance losses there are two reasons why we have to be more circumspect about interpreting "xX". First, it is easy to reject the notion that xX is an appropriate model for a growing or shrinking book of business—i.e. to show that an asset volume/return model does not apply. Consider two extreme cases: a single auto policy and a large portfolio of auto policies. For the single policy the probability of no claims in one year is around 90%. For a large portfolio the probability of no claims will be very close to zero. Thus there cannot be a random variable X such that the single policy losses have distribution x_1X and the portfolio has distribution x_2X. The loss distribution changes shape with volume. This is a crucial distinction between insurance losses and asset returns. The possibility of modeling liabilities as xX for x in a small range is discussed in Section 8. However, as Mildenhall (2004) shows, for x in the range where capital allocation and optimization will usually be applied, the diversification effects are material, a result supported by the data in Section 7.2.

The second reason for circumspection concerns the exact meaning of xX. Let X denote the end-of-period losses for a given book of business. Then xX for $0 \le x \le 1$ can be interpreted as a quota share of total losses, or as a coinsurance provision. However, xX for $x < 0$ or $x > 1$ is generally meaningless due to policy provisions, laws on over-insurance, and the inability to short insurance. The natural way to interpret a doubling in volume ("$2X$") is as $X_1 + X_2$ where X, X_1, X_2 are identically distributed random variables, rather than as a policy paying \$2 per \$1 of loss. This interpretation is consistent with doubling volume since $\mathrm{E}(X_1 + X_2) = 2\mathrm{E}(X)$. Clearly $X + X$ has a different distribution to $X_1 + X_2$ unless X_1 and X_2 are perfectly correlated.

A safer notation is to regard insurance risks as probability measures μ on $\mathbb{R}$—μ corresponds to a random variable X with distribution $\Pr(X < x) = \mu(-\infty, x)$—because there is no natural way to interpret 2μ. Let $M(\mathbb{R})$ denote the set of probability measures on $\mathbb{R}$. Then $M(\mathbb{R})$ is an abelian semigroup under convolution $\star$ of measures. Now $2X$ in our insurance interpretation, $X_1 + X_2$, corresponds to $\mu \star \mu := \mu^{\star 2}$.

We still have to define "directions" in $\mathbf{L}$ and $M(\mathbb{R})$. These should correspond to straight lines, or strictly to rays. In $\mathbb{R}^n$ there are several possible ways to characterize a straight line or ray $\alpha : [0, 1] \to \mathbb{R}^n$, each of which uses a different aspect of the rich mathematical structure[7] of $\mathbb{R}^n$.

Table 1 shows several possible characterizations of a ray in $\mathbb{R}^n$ each of which could also be used as characterizations in $\mathbf{L}$. The first two use properties of $\mathbb{R}^n$ which would require putting a differential structure on $\mathbf{L}$ which is

[7] $\mathbb{R}^n$ is a real vector space, a differentiable manifold, an abelian group, an inner product space, etc.

Table 1: Possible characterizations of a ray in $\mathbb{R}^n$

Characterization of ray	Required structure on $\mathbb{R}^n$
α is the shortest distance between $\alpha(0)$ and $\alpha(1)$	Notion of distance in $\mathbb{R}^n$, differentiable manifold
$\alpha''(t) = 0$, constant velocity, no acceleration	Very complicated on a general manifold.
$\alpha(t) = t\mathbf{x}$, $\mathbf{x} \in \mathbb{R}^n$.	Vector space structure
$\alpha(s+t) = \alpha(s) + \alpha(t)$	Can add in domain and range, semigroup structure only.

unnecessarily complicated. The third corresponds to the asset volume/return model and uses the $\mathbb{R}$ vector space structure on $\mathbb{R}^n$. This leaves the fourth approach: a ray is characterized by $\alpha(s+t) = \alpha(s) + \alpha(t)$. This definition only requires a semigroup structure (ability to add) for the range space. It is the definition adopted in Stroock (2003). In $\mathbf{L}$, regarded as a convolution semigroup, the condition becomes $\alpha(s+t) = \alpha(s) \star \alpha(t)$. Thus we define directions in $\mathbf{L}$ using families of random variables satisfying $X_s + X_t = X_{s+t}$ (or, equivalently, in $M(\mathbb{R})$ using families of measures μ_s satisfying $\mu_s \star \mu_t = \mu_{s+t}$). This condition implies that X_1 (resp. μ_1) is infinitely divisible, that is, for all integers $n \geq 1$ there exists $X_{1/n}$ (resp. $\mu_{1/n}$) so that $X = X_{1/n,1} + \cdots + X_{1/n,n}$, (resp. $\mu_1 = \mu_{1/n}^{\star n}$). A general result from probability theory then says that there exists a Lévy process X_t extending X_1 to $t \in \mathbb{R}$, $t \geq 0$, and that μ_t is the distribution of X_t. A Lévy process is an additive process with independent and stationary increments. By providing a basis of directions in $\mathbf{L}$, Lévy processes provide an insurance analog of individual asset return variables.

The insurance analog of an asset portfolio basis becomes a set of Lévy processes representing losses in each line. This definition reflects the fact that insurance grows account-by-account and that each account adds new idiosyncratic risk to the total, whereas growing an asset position magnifies risk which is perfectly correlated with the existing position. Growing in insurance is equivalent to adding a new asset to a mutual fund, not increasing an existing stock holding.

This leads us to explore how the total insured loss random variable evolves volumetrically and temporally. Let the random variable $A(x,t)$ denote aggregate losses from a line with expected annual loss x insured for a time period t years. Thus $A(x,1)$ is the annual loss. The central question of this paper is to describe appropriate models for $A(x,t)$. A Lévy process $X(t)$ provides a good basis for modeling $A(x,t)$. We consider four alternative insurance models.

IM1. $A(x,t) = X(xt)$. This model assumes there is no difference between insuring a given insured for a longer period of time and insuring more insureds for a shorter period.

IM2. $A(x,t) = X(xZ(t))$, for a subordinator $Z(t)$ with $\mathrm{E}(Z(t)) = t$. Z is an increasing Lévy process which measures random operational time, rather than calendar time. It allows for systematic time-varying contagion effects, such as weather patterns, inflation and level of economic activity, affecting all insureds.

IM3. $A(x,t) = X(xCt)$, where C is a mean 1 random variable capturing heterogeneity and non-diversifiable parameter risk across an insured population of size x. C could reflect different underwriting positions by firm, which drive systematic and permanent differences in results. The variable C is sometimes called a mixing variable.

IM4. $A(x,t) = X(xCZ(t))$.

All models assume severity has been normalized so that $\mathrm{E}(A(x,t)) = xt$. Two other models suggested by symmetry, $A(x,t) = X(Z(xt))$ and $A(x,t) = X(Z(xCt))$, are already included in this list because $X(Z(t))$ is also a Lévy process.

An important statistic describing the behavior of $A(x,t)$ is the coefficient of variation

$$\upsilon(x,t) := \frac{\sqrt{\mathrm{Var}(A(x,t))}}{xt}. \tag{2}$$

Since insurance is based on the notion of diversification, the behavior of $\upsilon(x,t)$ as $x \to \infty$ and as $t \to \infty$ are both of interest. The variance of a Lévy process either grows with t or is infinite for all t. If $X(\cdot)$ has a variance, then for IM1, $\upsilon(x,t) \propto (xt)^{-1/2} \to 0$ as t or $x \to \infty$. When $\upsilon(x,t) \to 0$ as $t \to \infty$ (resp. $x \to \infty$) we will call $A(x,t)$ temporally (resp. volumetrically) diversifying. A process which is both temporally and volumetrically diversifying will be called diversifying. If $X(x)$ is a standard compound Poisson process whose severity component has a variance then IM1 is diversifying. Meyers (2005a) gives evidence that insurance losses are *not* volumetrically diversifying. In Section 7 we analyze NAIC annual statement data from 1993-2004 and come to the same conclusion. We discuss the behavior of υ across lines of business as a function of x, and estimate the explicit (and surprising) shape of the distribution C.

Given a basis of Lévy processes we can compute partial derivatives as

$$\frac{\partial \rho}{\partial X_i} = \lim_{\epsilon \to 0} \frac{\rho(X_1(x_1) + \cdots + X_i(x_i + \epsilon) + \cdots + X_n(x_n)) - \rho(X_1(x_1) + \cdots + X_n(x_n))}{\epsilon} \tag{3}$$

which will generally give different results to the asset model—see Section 4.

Models IM1-4 are all very different to the asset model

AM1. $A(x,t) = xX(t)$

where $X(t)$ is a return process. X is often modeled using a geometric Brownian motion Hull (1983); Karatzas and Shreve (1988). AM1 is volumetrically homogeneous, meaning $A(kx,t) = kA(x,t)$. Therefore it has no volumetric diversification effect whatsoever, since $\Pr(A(kx,t) < ky) = \Pr(A(x,t) < y)$ and

$$\upsilon(x,t) = \frac{\sqrt{\operatorname{Var}(X(t))}}{t} \tag{4}$$

is independent of x. We discuss some important implications of this in Section 8.

The remainder of the paper is organized as follows. Section 2 shows how the gradient of a risk measure appears naturally in portfolio optimization and capital allocation. This highlights the importance of knowing exactly how partial derivatives with respect to changing volume by line should be computed. Section 3 discusses Gâteaux, or directional, derivatives and discusses Kalkbrener's axiomatic allocation. Section 4 shows that the asset view and actuarial views of portfolio growth give different gradients, using Meyers' example of "economic" vs. "axiomatic" capital. Section 5 provides an overview of Lévy processes for actuaries. Then Section 6 develops the notion that Lévy processes provide an actuarial straight-line to replace the asset volume/return model. It describes the set of directions emanating from the zero variable and computes the directions corresponding to IM1-4 and AM1, further highlighting the difference between the approaches. Section 7 uses NAIC annual statement data to provide empirical evidence supporting the models introduced here. Finally Section 8 revisits the homogeneity assumption of AM1 and Myers and Read (2001) and provides an illustration of how it is a peculiar special case.

2 THE UBIQUITOUS GRADIENT

Tasche (2000) makes the simple marginal analysis given in the introduction more rigorous. He shows that the gradient vector of the risk measure ρ is the only vector suitable for performance measurement, in the sense that it gives the correct signals to grow or shrink a line of business based on its marginal profitability and marginal capital consumption[8]. Tasche's framework is unequivocally financial. As described in the introduction, Tasche considers a set of basis asset return variables X_i, $i = 1, \dots, n$ and then determines a portfolio as a vector of asset position sizes $x = (x_1, \dots, x_n) \in U \subset \mathbb{R}^n$. The portfolio value distribution corresponding to x is simply

$$X = X(x) = X(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i X_i. \tag{5}$$

[8] Tasche's approach is sometimes called a RORAC method, or return on risk-adjusted capital.

A risk measure on $\mathbf{L}$ induces a function $\rho : \mathbb{R}^n \to \mathbb{R}$. Rather than being defined on a space of random variables, the induced ρ is defined on (a subset of) usual Euclidean space using the correspondence between x and a portfolio. In this context $\partial\rho/\partial X_i$ is simply the usual limit

$$\frac{\partial\rho}{\partial x_i} = \lim_{\epsilon\to 0} \frac{\rho(x_1,\dots,x_i+\epsilon,\dots,x_n)-\rho(x_1,\dots,x_n)}{\epsilon}. \tag{6}$$

Eq. 6 is a powerful mathematical notation and it contains two implicit assumptions. First, the fact that we can write $x_i + \epsilon$ requires that we can add in the domain. If ρ were defined on a more general space this may not possible—or it may involve the convolution of measures rather than addition of real numbers. Second, and more importantly, adding ϵ to x in the ith coordinate unambiguously corresponds to an increase "in the direction" of the ith asset. This follows directly from the definition in Eq. 5 and is unquestionably correct in a financial context.

Myers and Read (2001) also assume an asset return/volume model. They model losses in each line a as $\tilde{L}_a = L_a\tilde{R}_a$ where L_a is the present value of losses at time 0 and $\tilde{L}_a$ is the outcome at $t = 1$ (their notation). This assumption is essential to their finding that marginal surplus requirements "add-up". Irrespective of this assumption, Myers and Read work in the same marginal return to marginal capital framework as Tasche and as Meyers (2003). They measure risk and determine capital using a default value, and point out that marginal default values depend on marginal surplus allocations. Following Phillips et al. (1998) they argue that, since a firm defaults on all lines if it defaults on one, marginal surplus requirements should be set so that all lines' marginal contributions to default are the same. Again this results in a surplus requirement proportional to the gradient vector. We re-visit their result in Section 8

Denault (2001) discusses an axiomatic, game theoretic approach to allocation. In game theory, the essential property of an allocation is the no-undercut condition: no collection of lines can be allocated more capital than it would require on a stand-alone basis. Obviously, if this condition were not satisfied then the affected lines would want to leave the firm; they are being "charged" for a supposed diversification benefit. Panjer (2001) also discusses this approach. Denault assumes the risk measure ρ is coherent, that is, it is subadditive, positive homogeneous $\rho(\lambda x) = \lambda\rho(x)$ for $\lambda \geq 0$, monotonic, and translation invariant. Combining the homogeneous assumption with Euler's theorem he shows that the Aumann-Shapley value is a gradient-based per unit allocation[9].

A theorem of Aubin then shows that if ρ is differentiable its gradient is the unique per-unit allocation in the core of the coalitional game. The core is the set of solutions satisfying the no-undercut condition. Thus, under these assumptions, the gradient provides the unique fair per-unit allocation. A coalitional game is one where different units can combine in fractional parts. For insurance, this would correspond to allowing units to quota share together in different proportions. Since the weights are all between zero and one, such fractional coalitions still make sense in our insurance setting.

Delbaen (2000a) considers generalizations where ρ is convex but not differentiable. Then, elements of the subgradient of ρ (the set of all hyperplanes which support the graph of ρ) determine fair allocations. When the subgradient contains a single element, the allocation is also given by a Gâteaux derivative, as described in the next section.

Fischer (2003) also works with a portfolio base and uses it to induce a risk measure on $\mathbb{R}^n$ from a measure on L^p. He shows that risk measures which are everywhere differentiable degenerate to become linear, but shows that non-trivial risk measures result if the differentiability requirement is weakened slightly.

It is worth reiterating that risk can be appropriately measured with a homogeneous risk measure even though the risk process itself is not homogeneous. The risk measure must be homogeneous to allow for innocuous things like changing currency. The problem arises when "risks" are regarded as an $\mathbb{R}$ vector space, because then positive homogeneity becomes close to a linear requirement: risk is independent of position size. Insurance risk is absolutely

[9]The Aumann-Shapley value is the integral of the partial derivative of ρ as volume increases. Since ρ is 1-homogeneous, Euler's theorem shows it is constant. The constant derivative is illustrated in Figure 17.

not independent of position size because of diversification between insureds as volume increases. The scalar in a homogeneous risk measure has a more restricted meaning for insurance where there is no $\mathbb{R}$ vector space structure. In this context, Fischer's results are not so surprising. Recent asset modeling papers, Föllmer and Schied (2002); Cheridito et al. (2003), reconsider the desirability of strictly homogeneous risk measures. They introduce the notion of convex risk measures which include a penalty function for large positions.

3 GATEAUX DERIVATIVES AND KALKBRENER'S ALLOCATION

The differential represents the best linear approximation to a function in a given direction. Thus the differential to a function f, at a point x in its domain, can be regarded as a linear map Df_x which takes a direction, i.e. tangent vector, at x to a direction at $f(x)$. Exactly what this means will be discussed more in Section 6. For now it is enough to know that, under appropriate assumptions, the differential of f at $\mathbf{x}$ in direction $\mathbf{v}$, $D_{\mathbf{x}}f(\mathbf{v})$, is defined by the property

$$\lim_{\mathbf{v}\to\mathbf{0}} \frac{f(\mathbf{x}+\mathbf{v}) - f(\mathbf{x}) - D_{\mathbf{x}}f(\mathbf{v})}{\|\mathbf{v}\|} = 0. \tag{7}$$

The vector $\mathbf{v}$ is allowed to tend to $\mathbf{0}$ from any direction, and Eq. 7 must hold for all of them. This is called Fréchet differentiability. There are several weaker forms of differentiability defined by restricting the convergence of $\mathbf{v}$ to $\mathbf{0}$. These include the Gâteaux differential, where $\mathbf{v} = t\mathbf{w}$ with $t \in \mathbb{R}$, $t \to 0$, the directional differential, where $\mathbf{v} = t\mathbf{w}$ with $t \in \mathbb{R}$, $t \downarrow 0$, and the Dini differential, where $\mathbf{v} = t\mathbf{w}'$ for $t \in \mathbb{R}$, $t \downarrow 0$, and $\mathbf{w}' \to \mathbf{w}$. The function $f(x,y) = 2x^2y/(x^4+y^4)$ if $(x,y) \neq (0,0)$ and $f(0,0) = 0$ is not differentiable at $(0,0)$, in fact it is not even continuous, but all directional derivatives exist at $(0,0)$, and f is Gâteaux differentiable.

Kalkbrener (2005) applied Gâteaux differentiability to capital allocation. The Gâteaux derivative can be computed without choosing a set of basis return variables, that is, without setting up a map from $\mathbb{R}^n \to \mathbf{L}$, provided it is possible to add in the domain. This is the case for $\mathbf{L}$. The Gâteaux derivative of ρ at $Y \in \mathbf{L}$ in the direction $X \in \mathbf{L}$ is defined as

$$\frac{\partial \rho}{\partial X} = D\rho_Y(X) = \lim_{\epsilon \to 0} \frac{\rho(Y + \epsilon X) - \rho(Y)}{\epsilon}. \tag{8}$$

Kalkbrener (2005) shows that if the risk measure ρ satisfies certain axioms then it can be associated with a unique reasonable capital allocation. He shows that the allocation is covariance-based if risk is measured using standard deviation and a co-measure approach when risk is measured by expected shortfall—so his method is very natural.

Kalkbrener requires an allocation satisfy linear aggregation, diversification and continuity axioms, which we describe next. Given a risk measure ρ, he defines a capital allocation with respect to ρ to be a function Λ of two variables satisfying $\Lambda(X, X) = \rho(X)$. $\Lambda(X_i, X)$ is the capital allocated to X_i as a sub-portfolio of X. An allocation Λ is linear if it is linear in its first variable: $\Lambda(aX + bY, Z) = a\Lambda(X, Z) + b\Lambda(Y, Z)$. An allocation is diversifying if $\Lambda(X, Y) \le \Lambda(X, X)$ for all X, so including X in any portfolio does not increase its risk over a stand-alone portfolio. Finally, Λ is continuous at Y if

$$\lim_{\epsilon \to 0} \Lambda(X, Y + \epsilon X) = \Lambda(X, Y) \tag{9}$$

for all X. Kalkbrener proves that if Λ is a linear, diversifying capital allocation with respect to ρ and if Λ is continuous at Y then

$$\Lambda(X, Y) = \lim_{\epsilon \to 0} \frac{\rho(Y + \epsilon X) - \rho(Y)}{\epsilon} \tag{10}$$

is the Gâteaux derivative of ρ at Y in the direction X. Next he uses the Hahn-Banach theorem to prove that any positively homogeneous, sub-additive risk measure ρ can be represented as

$$\rho(X) = \max\{h(X) \mid h \in H_\rho\} \tag{11}$$

where H_ρ is the set of linear functionals on the space of risks which are dominated by ρ, $H_\rho = \{h \mid h(X) \le \rho(X) \text{ for all } X\}$. The surprising part of this result is that the max suffices; the supremum is not needed. Using this result, Kalkbrener defines an allocation associated with a positively homogeneous, sub-additive risk measure ρ as

$$\Lambda_\rho(X, Y) := h_Y(X)$$

where $h_Y \in H_\rho$ satisfies $h_Y(Y) = \rho(Y)$. He shows Λ_ρ is linear and diversifying. Finally he shows that for a positively homogeneous, sub-additive risk measure ρ and risk Y the following are equivalent: (1) that Λ_ρ is continuous at Y, (2) that the directional derivative

$$\lim_{\epsilon \to 0} \frac{\rho(Y + \epsilon X) - \rho(Y)}{\epsilon} \tag{12}$$

exists for all X, and (3) that there exists a unique $h \in H_\rho$ with $h(Y) = \rho(Y)$. When these three conditions hold,

$$\Lambda_\rho(X, Y) = \lim_{\epsilon \to 0} \frac{\rho(Y + \epsilon X) - \rho(Y)}{\epsilon} \tag{13}$$

for all X. Applying this theory Kalkbrener shows that the unique allocation for a standard deviation risk measure produces a CAPM-like covariance allocation, and that an expected shortfall risk measure produces a Phillips-Sherris like allocation, Phillips et al. (1998); Sherris (2006). Eq. 13 combines Theorem 18 and Proposition 5 from Delbaen (2000a).

This and the previous section have shown that notions of differentiability are central to capital allocation. The next section will show that not all the different notions agree, setting up the need for a better understanding of "direction" for actuarial random variables that Lévy processes will provide.

4 ECONOMIC AND AXIOMATIC CAPITAL

Meyers (2005b) gives an example where Kalkbrener's "axiomatic" allocation produces a different result than a marginal business written approach that is based on a more actuarial set of assumptions. Meyers calls his approach "economic" since it is motivated by the marginal business added philosophy discussed in the introduction and Section 2.

The example works with $n = 2$ independent lines of business and allocates capital to line 1 in order to keep the notation as simple as possible. Losses from both lines follow model IM3 with $t = 1$. The risk measure is standard deviation $\rho(X) = \mathrm{SD}(X)$ for $X \in \mathbf{L}$. $X_i(x_i)$ is a mixed compound Poisson variable

$$X_i(x_i) = S_{i,1} + \cdots + S_{i,N_i(x_i)} \tag{14}$$

where $N_i = N_i(x_i)$ is a C_i-mixed Poisson, so the conditional distribution $N \mid C_i$ is Poisson with mean $x_i C_i$ and the mixing distribution C_i has mean 1 and variance c_i. Meyers calls c_i the contagion. The mixing distributions are often taken to be gamma variables, in which case each N_i has a negative binomial distribution. The $S_{i,j}$, $i = 1, 2$ are independent, identically distributed severity random variables. For simplicity, assume that $\mathrm{E}(S_i) = 1$, so that $\mathrm{E}(X_i(x_i)) = \mathrm{E}(N_i(x_i))\mathrm{E}(S_i) = x_i$. Since $t = 1$ the model only considers volumetric diversification and not temporal diversification.

We can compute $\rho(X_i(x_i))$ as follows:

$$\begin{aligned}
\rho(X_i(x_i))^2 &= \mathrm{Var}(X_i(x_i)) \\
&= \mathrm{Var}(N_i)\mathrm{E}(S_i)^2 + \mathrm{E}(N_i)\mathrm{Var}(S_i) \\
&= x_i(1 + c_i x_i)\mathrm{E}(S_i)^2 + x_i(\mathrm{E}(S_i^2) - \mathrm{E}(S_i)^2) \\
&= c_i x_i^2 \mathrm{E}(S_i)^2 + x_i \mathrm{E}(S_i^2) \\
&= c_i x_i^2 + g_i x_i
\end{aligned}$$

where $g_i = \mathrm{E}(S_i^2)$. Note that $\rho(kX) = k\rho(X)$ for any constant k.

Kalkbrener's axiomatic capital is computed using the Gâteaux directional derivative, Eq. 13. Let $\rho_i(x_i) = \rho(X_i(x_i))$ and note that $\rho((1 + \epsilon)X_i(x_i)) = (1 + \epsilon)\rho_i(x_i)$. Then, by definition and the independence of X_1 and X_2, the

Gâteaux derivative of ρ at $X_1(x_1)+X_2(x_2)$ in the direction $X_1(x_1)$ is

$$\begin{aligned}
\frac{\partial\rho}{\partial X_1} &= \Lambda_\rho(X_1(x_1), X_1(x_1)+X_2(x_2)) \\
&= \lim_{\epsilon\to 0}\frac{\rho(X_1(x_1)+X_2(x_2)+\epsilon X_1(x_1)) - \rho(X_1(x_1)+X_2(x_2))}{\epsilon} \\
&= \lim_{\epsilon\to 0}\frac{\sqrt{(1+\epsilon)^2\rho_1(x_1)^2+\rho_2(x_2)^2} - \sqrt{\rho_1(x_1)^2+\rho_2(x_2)^2}}{\epsilon} \\
&= \lim_{\epsilon\to 0}\frac{\sqrt{\rho_1(x_1)^2+\rho_2(x_2)^2+2\epsilon\rho_1(x_1)^2} - \sqrt{\rho_1(x_1)^2+\rho_2(x_2)^2}}{\epsilon} \\
&= \lim_{\epsilon\to 0}\frac{\sqrt{\rho_1(x_1)^2+\rho_2(x_2)^2+2\epsilon\rho_1(x_1)^2} - \sqrt{\rho_1(x_1)^2+\rho_2(x_2)^2}}{2\epsilon\rho_1(x_1)^2} \\
&\qquad\qquad \times \lim_{\epsilon\to 0}\frac{2\epsilon\rho_1(x_1)^2}{\epsilon} \\
&= \frac{\rho_1(x_1)^2}{\rho(X_1(x_1)+X_2(x_2))} \\
&= \frac{c_1x_1^2+g_1x_1}{\rho(X_1(x_1)+X_2(x_2))}. \qquad (15)
\end{aligned}$$

This whole calculation has been performed without picking an asset return basis, but it can be replicated if we do. Specifically, use the $X_i(x_i)$ as a basis and define a linear map of $\mathbb{R}$-vector spaces $k:\mathbb{R}^n\to\mathbf{L}$, by $(y_1,\dots,y_n)\mapsto\sum_i y_iX_i(x_i)$. Let ρ_k be the composition of k and ρ,

$$\begin{aligned}
\rho_k(y_1,\dots,y_n) = \rho(k(y_1,\dots,y_n)) &= \rho\left(\sum_i y_iX_i(x_i)\right) \\
&= \sqrt{\sum_i y_i^2(c_ix_i^2+g_ix_i)}.
\end{aligned}$$

Then

$$\left.\frac{\partial\rho_k}{\partial y_1}\right|_{(1,1)} = \frac{c_1x_1^2+g_1x_1}{\rho(X_1(x_1)+X_2(x_2))} \qquad (16)$$

agreeing with Eq. 15. It is important to remember that $yX_i(x_i)\neq X_i(yx_i)$ for $y\neq 1$.

Given the definition of $X_i(x_i)$, we can also define an embedding $m : \mathbb{R}^n \to \mathbf{L}$, by $(x_1, \dots, x_n) \mapsto \sum_i X_i(x_i)$. The map m is a homomorphism of abelian semigroups but it is *not* a linear map of real vector spaces because $m(kx) \neq km(x)$. In fact, the image of m will generally be an infinite dimensional real vector subspace of $\mathbf{L}$. The lack of homogeneity is precisely what produces a diversification effect. As explained in the introduction and Section 2, an economic view of capital requires an allocation proportional to the gradient vector at the margin. Thus capital is proportional to $x_i \partial\rho_m/\partial x_i$ where $\rho_m : \mathbb{R}^n \to \mathbb{R}$ is the composition of m and ρ,

$$\rho_m(x_1, x_2) = \rho(m(x_1, x_2)) = \sqrt{\sum_i c_i x_i^2 + g_i x_i}. \tag{17}$$

Since ρ_m is a function on the reals, we can compute its partial derivative using standard calculus:

$$\frac{\partial \rho}{\partial x_1} = \frac{2c_1 x_1 + g_1}{2\rho(X_1(x_1) + X_2(x_2))}. \tag{18}$$

There are two important conclusion: (1) the partial derivatives of ρ_m and ρ_k (which is also the Gâteaux derivative of ρ) give very different answers, Equations 15 and 18, and (2) the implied allocations

$$\frac{c_1 x_1^2 + g_1 x_1}{\rho(X_1(x_1) + X_2(x_2))} \quad \text{and} \quad \frac{2c_1 x_1^2 + g_1 x_1}{2\rho(X_1(x_1) + X_2(x_2))} \tag{19}$$

are also different. This is Meyers' example.

We now think about derivatives in a more abstract way. Working with functions on $\mathbb{R}^n$ obscures some of the complication involved in working on more general spaces (like $\mathbf{L}$) because the set of directions at any point in $\mathbb{R}^n$ can naturally be identified with $\mathbb{R}^n$. In general this is not the case; the directions live in a different space. A familiar non-trivial example of this is the sphere in $\mathbb{R}^3$. At each point on the sphere the set of directions, or tangent vectors, is given by a plane. The collection of different planes, together with the original sphere, can be combined to give a new object, called the tangent bundle over the sphere. A point in the tangent bundle consists of a point on the sphere and a direction, or tangent vector, at that point.

There are several different ways to define the tangent bundle. For the sphere, an easy method is to set up a local chart, which is a differentiable bijection from a subset of $\mathbb{R}^2$ to a neighborhood of the point. This moves questions of tangency and direction back into $\mathbb{R}^2$ where they are well understood. Charts must be defined at each point on the sphere in such a way that they overlap consistently, producing an atlas, or differentiable structure, on the sphere. This is called the coordinate approach.

Another way of defining the tangent bundle is to use curves to define tangent vectors: a direction becomes the derivative, or velocity vector, of a curve. The tangent space can be defined as the set of curves through a point, with two curves identified if they are tangent (agree to degree 1). In Section 6 we will apply this approach to $\mathbf{L}$. A good general reference on the construction of the tangent bundle is Abraham et al. (1988).

In this context we see that Kalkbrener and Meyers are computing derivatives with respect to different directions. Kalkbrener is using a direction defined by the velocity vector of the curve

$$x \mapsto xX \tag{20}$$

whereas Meyers is using

$$x \mapsto X(x) \tag{21}$$

for some process X. Note also that the former is a linear map of real vector spaces whereas the latter is simply a homomorphism of abelian semigroups. The extra vector space structure makes sense for assets, where it is possible to change position size and short assets, but not for insurance liabilities. The appropriate mathematical structure must be driven by the financial realities of each situation.

5 LEVY PROCESSES FOR ACTUARIES

Lévy processes correspond naturally to the set of direction from 0 in **L**, as we explained in the introduction. We will now define Lévy processes and then discuss some of their important properties. The next section will explain the correspondence with directions in more detail. Lévy processes are fundamental to actuarial science, but they are rarely discussed explicitly in text books. For example, there is no explicit mention of Lévy processes in Beard et al. (1969), Bowers et al. (1986), Daykin et al. (1994), Klugman et al. (1998), or Panjer and Willmot (1992). However, the fundamental building block of all Lévy processes, the compound Poisson process, is well known to actuaries. It is instructive to learn about Lévy processes in an abstract manner as they provide a very rich source of examples for modeling actuarial processes. There are many good textbooks covering the topics described here, including Feller (1971) volume 2, Breiman (1992), Stroock (1993), Bertoin (1996), Sato (1999), Barndorff-Nielsen et al. (2001), and Applebaum (2004).

Definition 1 *A Lévy process is a stochastic process $X(t)$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying*

LP1. $X(0) = 0$ *almost surely;*

LP2. X *has independent increments, so for* $0 \le t_1 \le \cdots \le t_{n+1}$ *the variables* $X(t_{j+1}) - X(t_j)$ *are independent;*

LP3. X *has stationary increments, so* $X(t_{j+1}) - X(t_j)$ *has the same distribution as* $X(t_{j+1} - t_j)$*; and*

LP4. X *is stochastically continuous, so for all* $a > 0$ *and* $s \ge 0$

$$\lim_{t \to s} \Pr(|X(t) - X(s)| > a) = 0. \tag{22}$$

Lévy processes are in one-to-one correspondence with the set of infinitely divisible distributions. Recall that X is infinitely divisible if, for all integers $n \geq 1$, there exist independent, identically distributed random variables Y_i so that X has the same distribution as $Y_1 + \cdots + Y_n$. If $X(t)$ is a Lévy process then $X(1)$ is infinitely divisible since $X(1) = X(1/n) + (X(2/n) - X(1/n)) + \cdots + (X(1) - X(n-1/n))$, and conversely if X is infinitely divisible there is a Lévy process with $X(1) = X$. In an idealized world, losses should follow an infinitely divisible distribution because annual losses are the sum of monthly, weekly, daily, hourly losses[10] etc. The Poisson, normal, lognormal, gamma, Pareto, and Student t distributions are infinitely divisible. The uniform is not infinitely divisible, nor is any distribution with finite support, nor any whose moment generating function takes the value zero.

Examples

1. $X(t) = kt$ for a constant k is a trivial Lévy process.
2. The Poisson process $N(t)$ with intensity λ has

$$\Pr(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \tag{23}$$

for $n = 0, 1, \ldots$.

3. The compound Poisson process $X(t)$ with severity component Z is defined as

$$X(t) = Z_1 + \cdots + Z_{N(t)} \tag{24}$$

where $N(t)$ is a Poisson process.

4. Brownian motion.
5. Let $B(t)$ be a Brownian motion. Then $T(t) = \inf\{s > 0 \mid B(s) = t/\sqrt(2)\}$ defines a Lévy process called the Lévy subordinator, which only has positive jumps. This process is not a compound Poisson process because it has an infinite number of jumps in any finite period of time. In general, an increasing process (hence supported on $[0, \infty)$) is called a subordinator.
6. The sum of two Lévy processes is a Lévy process.

[10] Predictable fluctuations in frequency from seasonal or daily patterns can be accommodated using operational time—see Example 7.

7. Lundberg introduced the notion of operational time transforms in order to maintain stationary increments for compound Poisson distributions. Operational time is a risk-clock which runs faster or slower in order to keep frequency constant. It allows seasonal and daily effects (rush hours, night-time lulls, etc.) without losing stationary increments. Symbolically, operational time is an increasing function $k : [0, \infty) \to [0, \infty)$ chosen so that $X(k(t))$ becomes a Lévy process. Any operational time adjustments needed below are implicit.
8. Let $X(t)$ be a Lévy process and let $Z(t)$ be a subordinator, that is, a Lévy process with non-decreasing paths. Then $Y(t) = X(Z(t))$ is also a Lévy process. This process is called subordination and Y is subordinate to X. Z is called the directing process. Z is a random operational time.
9. Brownian motion with a drift is an example of a Lévy process that is not a martingale. The process $X_t = \max_{s\le t} B_s$ where B_t is not a Lévy process. It has an infinite number of jumps in some time intervals and none in others.

The characteristic function of a random variable X with distribution μ is defined as $\phi(z) = \mathrm{E}(e^{izX}) = \int e^{izx}\mu(dx)$ for $z \in \mathbb{R}$. The characteristic function of a Poisson variable with mean λ is $\phi(z) = \exp(\lambda(e^{iz} - 1))$. The characteristic function of a compound Poisson process is

$$\begin{aligned}
\phi(z) &= \mathrm{E}(e^{izX(t)}) = \mathrm{E}(\mathrm{E}(e^{izX(t)} \mid N(t))) &(25)\\
&= \mathrm{E}\exp\left(N(t)\log\int e^{izw}d\nu(dw)\right) &(26)\\
&= \exp\left(\lambda t\int(e^{izw}-1)d\nu(dw)\right) &(27)
\end{aligned}$$

where ν is the distribution of severity Z. The characteristic equation of a normal random variable is $\phi(z) = \exp(i\mu z - \sigma^2 z^2/2)$.

We now quote three of the many important results in the theory of Lévy processes. For simplicity we state these in one dimension. See Sato (1999) for proofs, and for precise statements in higher dimensions. The first theorem, the famous Lévy-Khintchine formula, describes the characteristic function of an infinitely divisible distribution function μ. The characteristic function of a general Lévy process follows from this.

Theorem 1 (Lévy-Khintchine) *If the probability distribution μ is infinitely divisible then its characteristic function has the form*

$$\exp\left(-\sigma^2 z^2 + \int_{\mathbb{R}} (e^{izw} - 1 - izw1_{\{|w|\le 1\}}(w))\nu(dw) + i\gamma z\right) \tag{28}$$

where ν is a measure on $\mathbb{R}$ satisfying $\nu(0) = 0$ and $\int_{\mathbb{R}} \min(|w|^2, 1)\nu(dw) < \infty$, and $\gamma \in \mathbb{R}$. The representation by (σ, ν, γ) is unique. Conversely given any such triple (σ, ν, γ) there exists a corresponding infinitely divisible distribution.

In Eq. 28, σ is the standard deviation of a Brownian motion component, and ν is called the Lévy measure. The indicator function $1_{\{|w|\le 1\}}$ is present for technical convergence reasons and is only needed when there are a very large number of very small jumps. If $\int_{\mathbb{R}} \min(|w|, 1)\nu(dw) < \infty$ this term can be omitted. The resulting γ can then be interpreted as a drift. In the general case γ does not have a clear meaning as it is impossible to separate drift from small jumps. The indicator can therefore also be omitted if $\nu(\mathbb{R}) < \infty$, and in that case the inner integral can be written as

$$\nu(\mathbb{R}) \int_{\mathbb{R}} (e^{izw} - 1)\tilde{\nu}(dw) \tag{29}$$

where $\tilde{\nu} = \nu/\nu(\mathbb{R})$ is a distribution. Comparing with Eq. 27 shows this term corresponds to a compound Poisson process.

The triples (σ, ν, γ) in the Lévy-Khintchine formula are called Lévy triples. The Lévy process $X(t)$ corresponding to the Lévy triple (σ, ν, γ) has triple $(t\sigma, t\nu, t\gamma)$. Define $\Psi(z)$ to be the term in the exponential in Eq. 28. The characteristic function of $X(t)$ is then $\exp(t\Psi(z))$.

The Lévy-Khintchine formula helps characterize all subordinators. A subordinator must have a Lévy triple $(0, \nu, \gamma)$ with no diffusion component and the Lévy measure ν must satisfy $\nu((-\infty, 0)) = 0$ (no negative jumps) and $\int_0^\infty \min(x, 1)\nu(dx) < \infty$. In particular, this shows there are no non-trivial continuous increasing Lévy processes.

The next theorem describes a decomposition of the sample paths of a Lévy process.

Theorem 2 (Lévy-Itô) *Let $X(t)$ be a Lévy process generated by the Lévy triple (σ, ν, γ). For any measurable $G \subset (0, \infty) \times \mathbb{R}$ let $J(G) = J(G, \omega)$ be the number of jumps at time s with height $X(s)(\omega) - X(s-)(\omega)$ such that $(s, X(s)(\omega) - X(s-)(\omega)) \in G$. Then $J(G)$ has Poisson distribution with mean $\tilde{\nu}(G)$ where $\tilde{\nu}$ is the measure induced by $\tilde{\nu}((0,t) \times B) = t\nu(B)$. If $G_1, \dots, G_n$ are disjoint then $J(G_1), \dots, J(G_n)$ are independent. Define*

$$X_1(t)(\omega) = \lim_{\epsilon \downarrow 0} \int_{(0,t] \times \{\epsilon < |x| \le 1\}} \{xJ(d(s,x),\omega) - x\tilde{\nu}(d(s,x))\} + \int_{(0,t] \times \{|x| > 1\}} xJ(d(s,x),\omega). \quad (30)$$

The process X_1 is a Lévy process with Lévy triple $(0, \nu, 0)$. Let

$$X_2(t) = X(t) - X_1(t). \quad (31)$$

Then X_2 is a Lévy process with Lévy triple $(\sigma, 0, \gamma)$. The two processes X_1, X_2 are independent. If $\int_{\mathbb{R}} \min(|w|, 1)\nu(dw) < \infty$ then we can define

$$X_3(t)(\omega) = \int_{(0,t] \times \mathbb{R}} xJ(d(s,x),\omega) \quad (32)$$

and

$$X_4(t) = X(t) - X_3(t). \quad (33)$$

X_3 and X_4 are independent Lévy processes, with Lévy triples $(0, \nu, 0)$ and $(\sigma, 0, \gamma)$, and γ is a deterministic drift. X_3 is the jump part and X_4 is the continuous part of X.

In Eq. 30 the first term is the compensated sum of small jumps and the second term is the sum of large jumps. Obviously the cut-off at 1 is arbitrary. The final theorem we quote shows that compound Poisson processes are the fundamental building block of Lévy processes.

Theorem 3 *The class of infinitely divisible distributions coincides with the class of limit distributions of compound Poisson distributions.*

Many properties of a Lévy process are time independent, in the sense that if they are true for one t they are true for all t. For example, the existence of a moment of order n, being continuous, symmetric, or positive are time independent. In particular if $X(1)$ has a variance then $X(t)$ has a variance for all t, and by independent, time homogeneous increments, the variance must grow with t. This is well-known for Brownian motion, where $\mathrm{Var}(B(t)) = t$, and a compound Poisson process, where $\mathrm{Var}(X(t)) = \lambda t \mathrm{E}(Z^2)$ and Z is the severity component.

Next we consider some properties of the models IM1-4. Given a Lévy process $X(t)$ we defined four models for aggregate losses $A(x,t)$ from a volume x of insurance, insured for t years:

IM1. $A(x,t) = X(xt)$;

IM2. $A(x,t) = X(xZ(t))$, for a subordinator $Z(t)$ with $\mathrm{E}(Z(t)) = t$;

IM3. $A(x,t) = X(xCt)$ for a random variables C with $\mathrm{E}(C) = 1$ and $\mathrm{Var}(C) = c$; and

IM4. $A(x,t) = X(xCZ(t))$, C and $Z(t)$ independent.

We also defined the asset return/volume model $A(x,t) = xX(t)$. In all cases severity is normalized so that $E(A(x,t)) = xt$. Define σ and τ so that $\mathrm{Var}(X(t)) = \sigma^2 t$ and $\mathrm{Var}(Z(t)) = \tau^2 t$. Underwriters tend to avoid risks with undefined variance, so the assumption of a variance is not onerous!

Models IM3 and IM4 no longer define Lévy processes because of the common C term. Each process has conditionally independent increments, given C. Thus, these two models no longer assume that each new insured has losses independent of the existing cohort. We will discuss the impact C has on the "direction" defined by A in the next section. Example 8 shows that IM2 is a Lévy process.

Table 5 lays out the variance and coefficient of variation υ of these five models. It also shows whether each model is volumetrically (resp. temporally) diversifying, that is whether $\upsilon(x,t) \to 0$ as $x \to \infty$ (resp. $t \to \infty$).

Table 2: Variance of IM1-4 and AM

Model	Variance	$\upsilon(x,t)$	Diversifying $x \to \infty$	Diversifying $t \to \infty$
$X(xt)$	$\sigma^2 xt$	$\dfrac{\sigma}{\sqrt{xt}}$	Yes	Yes
$X(xZ(t))$	$xt(\sigma^2 + x\tau^2)$	$\sqrt{\dfrac{\sigma^2}{xt} + \dfrac{\tau^2}{t}}$	No	Yes
$X(xCt)$	$xt(\sigma^2 + cxt)$	$\sqrt{\dfrac{\sigma^2}{xt} + c}$	No	No
$X(xCZ(t))$	$x^2t^2\left(\dfrac{(c+1)\tau^2}{t} + c\right) + \sigma^2 xt$	$\sqrt{\dfrac{\sigma^2}{xt} + \dfrac{\tau'^2}{t} + c}$	No	No
$xX(t)$	$x^2\sigma^2 t$	$\sigma/\sqrt{t}$	Const.	Yes

$\tau' = (1+c)\tau$

The calculations follow easily by conditioning. For example

$$\begin{aligned}
\mathrm{Var}(X(xZ(t))) &= \mathrm{E}_{Z(t)}(\mathrm{Var}(X(xZ(t)))) + \mathrm{Var}_{Z(t)}(\mathrm{E}(X(xZ(t)))) \\
&= \mathrm{E}(\sigma^2 xZ(t)) + \mathrm{Var}(xZ(t)) \\
&= \sigma^2 xt + x^2\tau^2 t = xt(\sigma^2 + x\tau^2).
\end{aligned}$$

The characteristics of each model will be tested against NAIC annual statement data in Section 7. We will show that IM1 and AM are not consistent with the data. With one year of data it is impossible to distinguish between models IM2-4; for fixed t they all have the same form. Using twelve years of NAIC data suggests that IM2 and IM4 are not consistent with the data, leaving model IM3.

The models presented here are one-dimensional. A multi-dimensional version would use multi-dimensional Lévy processes. This allows for the possibility of correlation between lines. In addition, correlation between lines can be induced by using correlated mixing variables C. This is the common-shock model, described in Meyers (2005a).

5.1 LEVY PROCESSES AND CATASTROPHE MODEL PMLs

Since Hurricane Andrew in 1992 computer simulation models of hurricane and earthquake insured losses have developed into an essential tool for industry risk assessment and management. The output from these models specifies a compound Poisson Lévy process. Given the interest surrounding the use of these models, this section will translate some common industry terms, such as return period and probable maximum loss (PML), into precise formulae.

PML and maximum foreseeable loss (MFL) are terms from individual risk property underwriting which far pre-date modern computer simulation models. The PML is an estimate of the largest loss that a building or a business in the building is likely to suffer, considering the existing mitigation features, because of a single fire. The PML assumes that critical protection systems are functioning as expected. The MFL is an estimate of the largest fire loss likely to occur if loss-suppression systems fail. For a large office building the PML is sometimes estimated as a total loss of 4 to 6 floors, assuming the fire itself would be contained to one or two floors. The MFL can be estimated as a total loss "within four walls", so a single structure burns down. These terms are now over-loaded with different meanings related to catastrophe modeling. McGuinness (1969) discusses meanings of the term PML.

Simulation models produce a sample of n loss events, each with an associated annual frequency λ_i and expected event loss l_i, $i = 1, \dots, n$. Each event is assumed to have a Poisson occurrence frequency distribution. The associated Lévy measure ν is concentrated on the set $\{l_1, \dots, l_n\}$ with $\nu(\{l_i\}) = \lambda_i$. Since the models only simulate a finite number of events, $\nu(\mathbb{R}) = \sum_i \lambda_i < \infty$. Let $\lambda = \nu(\mathbb{R})$ be the total modeled event frequency. We can normalize to get an event severity distribution $\tilde{\nu} = \nu/\lambda$ because $\lambda < \infty$. Let X_t be the Lévy process associated with the Lévy triple $(0, \nu, 0)$.

Catastrophe risk is usually managed using reinsurance purchased on an occurrence basis; it covers all losses from a single event. Therefore companies are interested in the annual frequency of events above a threshold. Using the Lévy measure ν, the annual frequency of losses greater than or equal to x is simply $\lambda(x) := \nu([x,\infty))$, which, by definition, is

$$\lambda(x) = \nu([x,\infty)) = \sum_{l_i \ge x} \lambda_i. \tag{34}$$

Since $\lambda(x)$ is the annual frequency for a loss of size $\ge x$, and each event frequency has a Poisson distribution, the (exponentially distributed) waiting time for such a loss has mean $1/m(x)$. Surprisingly, this is not what is usually referred to as the "return period".

Using the Poisson count distribution, the annual probability of one or more single events in a year causing loss $\ge x$ is the probability that a Poisson variable N with mean $\lambda(x)$ has value 1 or more, that is

$$\Pr(N \ge 1) = 1 - \Pr(N = 0) = 1 - \exp(-\lambda(x)). \tag{35}$$

Therefore the probability of one or more occurrences causing $\ge x$ loss is $1 - \exp(-\lambda(x))$. The return period of this loss is then generally quoted as $1/(1 - \exp(-\lambda(x)))$. For large enough x, $\lambda(x)$ is very small and $1/(1 - \exp(-\lambda(x)))$ is approximately $1/\lambda(x)$.

Insurance companies are also interested in percentiles of X_1—especially after the hurricane seasons of 2004 and 2005! These percentiles are called "aggregate PML" points. In this context, the return period is again the reciprocal of the percentile point. Aggregate PMLs have two disadvantages: they are hard to compute and they do not correspond to how insurers usually manage risk. They are, however, clearly important for Enterprise Risk Management.

It should now be clear that a statement such as "the 100 year PML loss is \$200M" is inherently ambiguous. Insurance companies have used this to their benefit. For example, one company announced

> Hurricane Katrina is the costliest catastrophe in the United States history. The single event probability of the industry experiencing losses at or above the Katrina level as a result of a hurricane *hitting in the specific area where Katrina made landfall* was 0.2% based on the AIR Worldwide Corporations catastrophe model, which is equivalent to a one in 500 year event. [Emphasis added]

This is a true statement—regarding a subset of the total simulated events. Perhaps losses $x_1, \dots, x_m$ of the n simulated hit the "specific area where Katrina made landfall", for some $m \ll n$. The quoted PML statement refers to using a Lévy measure ν' supported just on these events. Other press releases spoke of Katrina being a 250 year event, or higher, for the industry. Again, these statements could be true if the sample space of events is suitably constrained, but they are manifestly not true of Katrina as an industry hurricane event. On an industry basis Katrina is around a 25-50 year event—one can count Andrew, the Long Island storm of 1938, and Katrina as broadly comparable events in the last 100 years alone.

6 "DIRECTIONS" IN THE SPACE OF ACTUARIAL RANDOM VARIABLES

This section combines three threads we have already discussed to produce a description of "directions" in the space $\mathbf{L}$. The first is the notion that directions, or tangent vectors, live in a separate space called the tangent bundle. The tangent bundle can be identified with the original space in the case of $\mathbb{R}^n$—a simplification which confuses intuition in more involved examples. The second thread comes from regarding tangent vectors as velocity vectors of curves. The third uses the idea presented in the introduction that Lévy processes, characterized by the additive relation $X(s+t) = X(s)+X(t)$, provide the appropriate analog of straight lines or rays. The program is, therefore, to compute the derivative of the curve $t \mapsto X(t)$ at $t = 0$. The ideas presented here are part of an elegant general theory of Markov processes. The presentation follows the beautiful book by Stroock (2003). Before getting into the details we review a schematic of the difference between the Meyers and Kalkbrener maps ρ_m and ρ_k, and then describe a finite sample space version of $\mathbf{L}$ which illustrates the difficulties involved in regarding it as a differentiable manifold. Throughout the section it is more convenient to work with distribution functions μ_t than random variables $X(t)$; the connection between the two is $\mu_t(B) = \Pr(X(t) \in B)$ for a measurable set B.

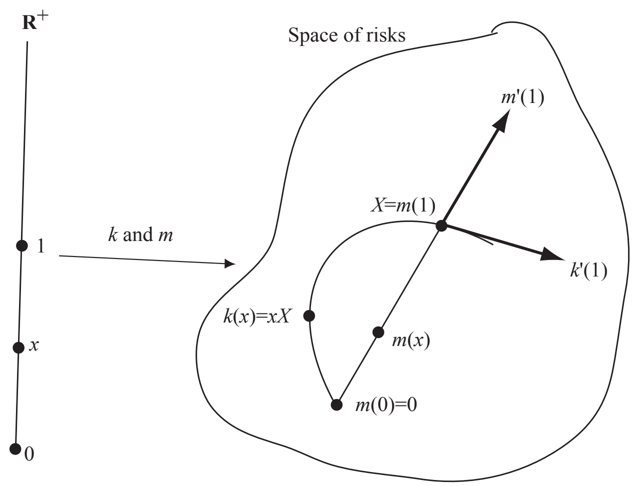

Figure 1: Lévy process and homogeneous embeddings of $\mathbb{R}_+$ into $\mathbf{L}$. The Lévy process embedding corresponds to the straight line.

The difference between the maps $m(x) = X(x)$ and $k(x) = xX$ lies in the meaning of "in the direction X". Kalkbrener takes it to mean adding a small extra amount ϵX_i, perfectly correlated to the existing X_i, to the portfolio. This makes sense for an asset portfolio but not for an portfolio of insurance risks. Each insured risk is unique; laws and contractual provisions forbid insuring the exact same risk twice—or even $(1+\epsilon)$ times. Meyers' model and IM1-4 recognize that insurance volume increases either by insuring more individuals (increase x), or insuring the same individuals for a longer period of time (increasing t). Eq. 14 recognizes the increase by increasing the claim count.

Figure 1 is a schematic of the difference between m and k. It shows the map m from $\mathbb{R}^+ \to \mathbf{L}$, the space of risks; the image of m is shown as a straight line. $X := m(1)$ is the image of 1. The tangent vector, $m'(1)$, to the embedding m is shown extending along the direction of the line; this is the natural direction to interpret as "growth in the direction X". In the schematic, the embedding k is shown as the curved line. The tangent vector $k'(1)$ is not pointing in the same direction as $m'(1)$. For a risk measure ρ, $\partial\rho/\partial X$ is the evaluation of the linear differential map $D\rho$ on a tangent vector in the direction X. Meyer's embedding m corresponds to $\partial(\rho \circ m)/\partial t|_{t=1} = D\rho_X(m'(1))$ whereas Kalkbrener's corresponds to $\partial(\rho \circ k)/\partial t|_{t=1} = D\rho_X(k'(1))$. As demonstrated in Section 4 these are not the same—just as the schematic leads us to expect. The difference between $k'(1)$ and $m'(1)$ is a measure of the diversification benefit given by m compared to k. k maps $x \mapsto xX$ and so offers no diversification to an insurer. Again, this is correct for an asset portfolio (you don't diversify a portfolio by buying more of the same stock) but it is *not* true for an insurance portfolio.

In order to see that the construction of tangent directions in $\mathbf{L}$ may not be trivial, consider the space M of probability measures on $\mathbb{Z}/n$, the integers $\{0, 1, \dots, n-1\}$ with $+$ given by addition modulo n. An element $\mu \in M$ can be identified with an n-tuple of non-negative real numbers $p_0, \dots, p_{n-1}$ satisfying $\sum_i p_i = 1$. Thus elements of M are in one to one correspondent with elements of the n dimensional simplex $\Delta_n = \{(x_0, \dots, x_{n-1}) \mid \sum_i x_i = 1\} \subset \mathbb{R}^n$. Δ_n inherits a differentiable structure from $\mathbb{R}^n$; we already know how to think about directions and tangent vectors in Euclidean space. However, even thinking about $\Delta_3 \subset \mathbb{R}^3$ shows M is not an easy space to work with. Δ_3 is a plane triangle; it has a boundary of three edges and each edge has a boundary of two vertices. The tangent spaces at each of these boundary points is different and different again from the tangent space in the interior of Δ_3. As n increases the complexity of the boundary increases and, to compound the problem, every point in the interior gets closer to the boundary. For measures on $\mathbb{R}$ the boundary is dense.

Now let M be the space of probability measures on $\mathbb{R}$ and let $\delta_x \in M$ be the measure giving probability 1 to $x \in \mathbb{R}$. We will describe the space of tangent vectors to M at δ_0. By definition, all Lévy processes $X(t)$ have distribution δ_0 at $t = 0$. Measures $\mu_t \in M$ are defined by their action on functions f on $\mathbb{R}$. Let $\langle f, \mu \rangle = \int_{\mathbb{R}} f(x)\mu(dx)$. In view of the fundamental theorem of calculus, the derivative $\dot{\mu}_t$ of μ_t should satisfy

$$\langle f, \mu_t \rangle - \langle f, \mu_0 \rangle = \int_0^t \dot{\mu}_\tau f d\tau, \tag{36}$$

indicating $\dot{\mu}_t$ is a functional acting on f. Converting Eq. 36 to its differential form suggests that

$$\dot{\mu}_t f(0) = \lim_{t\downarrow 0} \frac{\langle f, \mu_t\rangle - \langle f, \mu_0\rangle}{t} \tag{37}$$

$$= \lim_{t\downarrow 0} \frac{\mathrm{E}(f(X(t))) - \mathrm{E}(f(X(0)))}{t} \tag{38}$$

where $X(t)$ has distribution μ_t.

We now consider how this formula works when $X(t)$ is related to a Brownian motion or a compound Poisson—the two building block Lévy processes. Suppose first that $X(t)$ is a Brownian motion with drift γ and standard deviation σ, so $X(t) = \gamma t + \sigma B(t)$ where $B(t)$ is a standard Brownian motion. Let f be a function with a Taylor's expansion about 0. We can compute

$$\dot{\mu}_0 f(0) = \lim_{t\downarrow 0} [\mathrm{E}(f(0) + X(t)f'(0) + \frac{X(t)^2 f''(0)}{2} + O(t^2)) - f(0)]/t \tag{39}$$

$$= \lim_{t\downarrow 0} [\gamma t f'(0) + \frac{\sigma^2 t f''(0)}{2} + O(t^2))]/t \tag{40}$$

$$= \gamma f'(0) + \frac{\sigma^2 f''(0)}{2}, \tag{41}$$

because $\mathrm{E}(X(t)) = \mathrm{E}(\gamma t + \sigma B(t)) = \gamma t$, $\mathrm{E}(X(t)^2) = \gamma^2 t^2 + \sigma^2 t$, $\mathrm{E}(B(t)) = 0$ and $\mathrm{E}(B(t)^2) = t$. Thus $\dot{\mu}_0$ acts as a second order differential operator

$$\dot{\mu}_0 = \gamma \frac{d}{dx} + \frac{\sigma^2}{2}\frac{d^2}{dx^2}. \tag{42}$$

Next suppose that $X(t)$ is a compound Poisson distribution with Lévy measure ν, $\nu(\{0\}) = 0$ and $\lambda = \nu(\mathbb{R}) < \infty$. Let J be a variable with distribution ν/λ, so, in actuarial terms, J is the severity. The number of jumps of $X(t)$ follows a Poisson distribution with mean λt. If t is very small then the axioms characterizing the Poisson distribution imply that there is a single jump with probability λt and no jump with probability $1 - \lambda t$. Conditioning on a jump $\mathrm{E}(f(X(t))) = (1-\lambda t)f(0) + \lambda t\mathrm{E}(f(J))$ and so

$$\begin{aligned}
\dot{\mu}_0 f(0) &= \lim_{t\downarrow 0} \frac{\mathrm{E}(f(X(t))) - \mathrm{E}(f(X(0)))}{t} && (43)\\
&= \lim_{t\downarrow 0} \frac{\lambda t(\mathrm{E}(f(J)) - f(0))}{t} && (44)\\
&= \lambda(\mathrm{E}(f(J)) - f(0)) && (45)\\
&= \int f(y) - f(0)\nu(dy) && (46)
\end{aligned}$$

This analysis side-steps some technicalities by assuming that $\nu(\mathbb{R}) < \infty$. Combining these two results makes the following theorem plausible.

Theorem 4 (Stroock (2003) Thm 2.1.11) *There is a one-to-one correspondence between directions, or tangent vectors, in M and Lévy processes. If μ_t is a Lévy process characterized by Lévy triple (σ, ν, γ) then $\dot{\mu}_t$ is the linear operator acting on f by*

$$\dot{\mu}_t f(z) = \gamma \frac{df}{dx} + \frac{1}{2}\sigma^2 \frac{d^2 f}{dx^2} + \int_{\mathbb{R}} f(y+z) - f(z) - \frac{df}{dx}\frac{y}{1+|y|^2}\,\nu(dy) \qquad (47)$$

As in the Lévy-Khintchine theorem, the extra term in the integral is needed for technical convergence reasons when there is a large number of very small jumps.

We can now compute the difference between the directions implied by each of IM1-4 and AM. This quantifies the difference between $m'(1)$ and $k'(1)$ in Figure 1. In order to focus on the models that realistically can be used to model insurance losses we will assume $\gamma = \sigma = 0$ and $\nu(\mathbb{R}) < \infty$. Assume the Lévy triple for the subordinator Z is $(0, \rho, 0)$. Also $\mathrm{E}(C) = 1$, $\mathrm{Var}(C) = c$, and C, X and Z are all independent.

For each model we can consider the time derivative or the volume derivative. There are obvious symmetries between these two for IM1 and IM3. For IM2 the temporal derivative is the same as the volumetric derivative of IM3 with $C = Z(t)$.

Theorem 4 gives the direction for IM1 as corresponding to the operator Eq. 47 multiplied by x or t as appropriate. For example, the time direction is given by the operator

$$\dot{\mu}_0 f(z) = x \int f(z+y) - f(z)\nu(dy). \tag{48}$$

The temporal derivative of IM2, $X(xZ(t))$, is more tricky. Let K have distribution $\rho/\rho(\mathbb{R})$, the severity of Z. For small t, $Z(t) = 0$ with probability $1 - \rho(\mathbb{R})t$ and $Z(t) = K$ with probability $\rho(\mathbb{R})t$. Thus

$$\begin{aligned} \dot{\mu}_0 f(z) &= \rho(\mathbb{R})\mathrm{E}(f(z + X(xK)) - f(z)) & (49)\\ &= \int_{(0,\infty)} \int_{(0,\infty)} f(z+xy) - f(z)\nu^{xk}(dy)\rho(dk) & (50)\end{aligned}$$

where ν^k is the distribution of $X(k)$. This has the same form as IM1, except the underlying Lévy measure ν has been replaced with

$$\nu'(B) = \int_{(0,\infty)} \nu^k(B)\rho(dk). \tag{51}$$

For IM3, $X(xCt)$, the direction is the same as for model IM1. This is not a surprise because the effect of C is to select, once and for all, a random speed along the ray; it does not affect its direction. By comparison, in model IM2 the "speed" is proceeding by jumps, but again, the direction is fixed. If $\mathrm{E}(C) \neq 1$ then the derivative would be multiplied by $\mathrm{E}(C)$.

Finally the volumetric derivative of the asset model is simply

$$\mu_0' f(z) = X(t)\frac{df}{dx}. \tag{52}$$

Thus the derivative is the same as for a deterministic drift Lévy process. This should be expected since once $X(t)$ is known it is fixed regardless of volume x. Comparing with the derivatives for IM1-4 expresses the different directions represented schematically in Figure 1 analytically. The result is also reasonable in light of the different shapes of tZ and $\sqrt{t}Z$ as $t \to 0$, for a random variable Z with mean and standard deviation equal to 1. For very

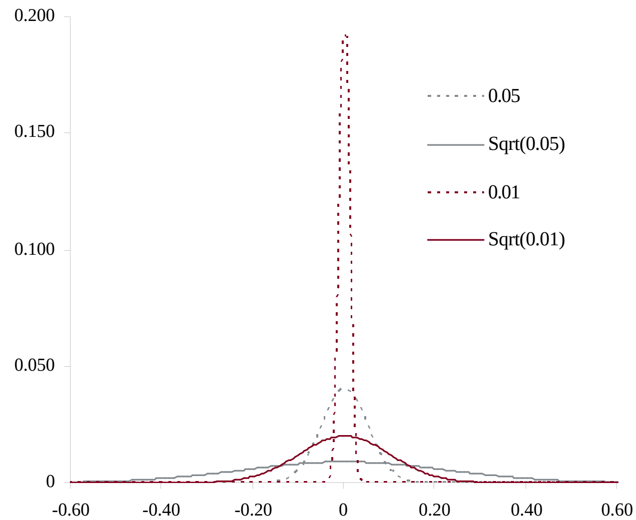

Figure 2: Illustration of the difference between tZ and $\sqrt{t}Z$ for Z a standard normal as $t \to 0$.

small t, tZ is essentially the same as a deterministic $t\mathrm{E}(Z)$, whereas $\sqrt{t}Z$ has a standard deviation $\sqrt{t}$ which is much larger than the mean t. Its coefficient of variation $1/\sqrt{t} \to \infty$ as $t \to 0$. The relative uncertainty in $\sqrt{t}Z$ grows as $t \to 0$ whereas for tZ it disappears. This is illustrated in Figure 2.

Severity uncertainty is another interesting form of uncertainty. Suppose that claim frequency is still λ but that severity is given by a family of measures $\tilde{\nu}_V$ for a random V. Now, in each state, the Lévy process proceeds along a random direction defined by $V(\omega)$, so the resulting direction is a mixture

$$\dot{\mu}_0 = \int \dot{\mu}_{0,v} d\mu(v) \tag{53}$$

where μ is the distribution of V.

It is interesting to interpret these results from the perspective of credibility theory. Credibility is usually associated with repeated observations of a given insured, so t grows but x is fixed. For models IM1-4 severity (direction) is implicitly known. For IM2-4 credibility determines information about the fixed (C) or variable ($Z(t)$) speed of travel in the given direction. If there is severity uncertainty, V, then repeated observation resolves the direction of travel, rather than the speed. Obviously both direction and speed are uncertain in reality.

It may be possible and enlightening for actuaries to model directly with a Lévy measure ν and hence avoid the artificial distinction between frequency and severity. Catastrophe models already work in this way. Several aspects of actuarial practice could benefit from avoiding the frequency/severity dichotomy. Explicitly considering the count density of losses by size range helps clarify the effect of loss trend. In particular, it allows different trend rates by size of loss. Risk adjustments become more transparent. The theory of risk adjusted probabilities for compound Poisson distributions, Delbaen and Haezendonck (1989); Meister (1995), is more straightforward if loss rate densities are adjusted without the constraint of adjusting a severity curve and frequency separately. This approach can be used to generate state price densities directly from catastrophe model output. Finally, the Lévy measure is equivalent to the log of the aggregate distribution, so convolution of aggregates corresponds to a pointwise addition of Lévy measures. This simplification is clearer when frequency and severity are not split. It facilitates combining losses from portfolios with different policy limits.

This section has shown there are important local differences between the maps k and m. They may agree at a point, but the agreement is not first order—the two maps define different directions. Since capital allocation relies on derivatives—the ubiquitous gradient—it is not surprising that different allocations result. This is shown in practice by Meyer's example and by the failure of Myers and Read's formula to add-up for diversifying Lévy processes.

7 EMPIRICAL EVIDENCE: VOLUMETRIC AND TEMPORAL EVOLUTION OF PORTFOLIOS

NAIC annual statement accident year schedule P data from 1993 to 2004 by line of business can be modeled using IM1-4. The model fits can differentiate company effects from accident year pricing cycle effects, and the parameters show considerable variation by line of business. Importantly, the fits can capture information about the mixing distribution C, based on Proposition 1, below.

Three hypotheses will be tested from the previous sections: (1) that the asymptotic coefficient of variation or volatility[11] as volume grows is strictly positive; (2) that time and volume are symmetric in the sense that $v(x,t) = \mathrm{CV}(A(x,t))$ only depends on xt; and (3) that the data is consistent with model IM3. IM3 implies that diversification over time follows a symmetric modified square root rule, $v(x,t) = \sqrt{(\sigma^2/xt) + c}$.

7.1 THEORY

Consider an aggregate loss distribution with a C-mixed Poisson frequency distribution, per Eq. 14. If the expected claim count is large and if the severity has a variance then particulars of the severity distribution diversify away in the aggregate. Any severity from a policy with a limit has a variance; unlimited directors and officers on a large corporation may not have a variance. Moreover the variability from the Poisson claim count component also diversifies away and the shape of the normalized aggregate loss distribution, aggregate losses divided by expected aggregate losses, converges in distribution to the mixing distribution C.

These assertions can be proved using moment generating functions. Let X_n be a sequence of random variables with distribution functions F_n and let X another random variable with distribution F. If $F_n(x) \to F(x)$ as $n \to \infty$ for every point of continuity of F then we say F_n converges weakly to F and that X_n converges in distribution to X.

Convergence in distribution is a relatively weak form of convergence. A stronger form is convergence in probability, which means for all $\epsilon > 0$ $\Pr(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$. If X_n converges to X in probability then X_n also converges to X in distribution. The converse is false. For example, let $X_n = Y$ and X be binomial 0/1 random variables with

[11] Volatility of loss ratio is used to mean coefficient of variation of loss ratio.

$\Pr(Y = 1) = \Pr(X = 1) = 1/2$. Then X_n converges to X in distribution. However, since $\Pr(|X - Y| = 1) = 1/2$, X_n does not converge to X in probability.

X_n converges in distribution to X if the moment generating functions (MGFs) M_n of X_n converge to the MGF of M of X for all z: $M_n(z) \to M(z)$ as $n \to \infty$, see (Feller, 1971, Volume 2, Chapter XV.3 Theorem 2). We can now prove the following proposition.

Proposition 1 *Let N be a C-mixed Poisson distribution with mean n, C with mean 1 and variance c, and let X be an independent severity with mean x and variance $x(1+\gamma^2)$. Let $A = X_1 + \cdots + X_N$ and $a = nx$. Then the normalized loss ratio A/a converges in distribution to C, so*

$$\Pr(A/a < \alpha) \to \Pr(C < \alpha) \tag{54}$$

as $n \to \infty$. Hence

$$\sigma(A/a) = \sqrt{c + \frac{x(1+\gamma^2)}{a}} \to \sqrt{c}. \tag{55}$$

Proof: The moment generating function $M_A(z)$ of A is

$$M_A(z) = M_C(n(M_X(z) - 1)) \tag{56}$$

where M_C and M_X are the moment generating functions of C and X. Using Taylor's expansion we can write

$$\begin{aligned}
\lim_{n\to\infty} M_{A/a}(z) &= \lim_{n\to\infty} M_A(z/a) \\
&= \lim_{n\to\infty} M_C(n(M_X(z/nx) - 1)) \\
&= \lim_{n\to\infty} M_C(n(M'_X(0)z/nx + R(z/nx))) \\
&= \lim_{n\to\infty} M_C(z + nR(z/nx))) \\
&= M_C(z)
\end{aligned}$$

for some remainder function $R(z) = O(z^2)$. The assumptions on the mean and variance of X guarantee $M'_X(0) = x = \mathrm{E}(X)$ and that the remainder term in Taylor's expansion is $O(z^2)$. The second part is trivial. ∎

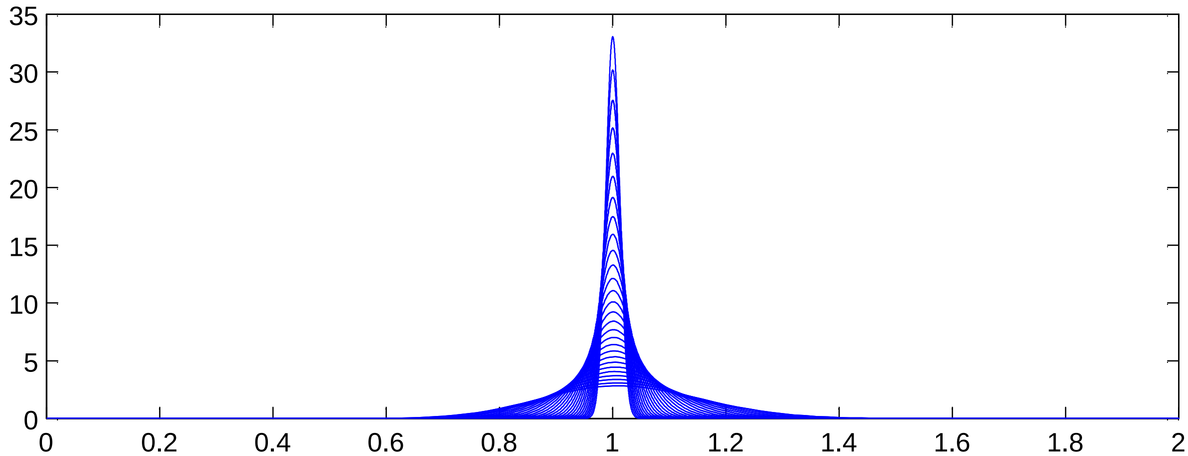

Figure 3: Theoretical distribution of scaled aggregate losses with no parameter or structure uncertainty and Poisson frequency.

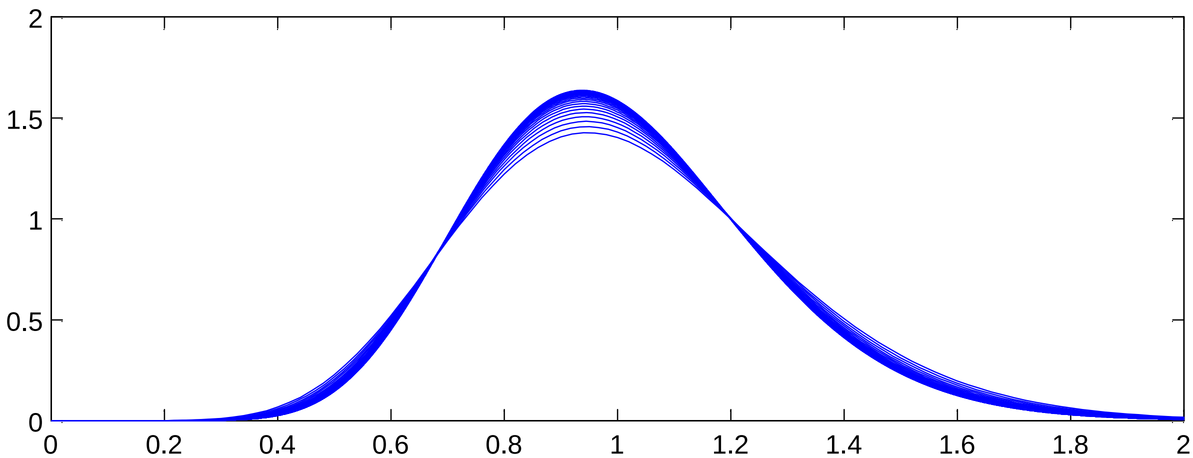

Figure 4: Theoretical distribution envelope of scaled aggregate losses with a gamma mixed Poisson frequency with mixing variance $c = 0.0625$.

The proposition implies that if the frequency distribution is actually a Poisson, so the mixing distribution C is $C = 1$ with probability 1, then the loss ratio distribution of a very large book will tend to the distribution concentrated at the expected. Hence the expression "with no parameter risk the process risk completely diversifies away."

Figures 3 and 4 illustrate the proposition, showing how the aggregate distributions change shape as expected counts increase. In Figure 3, $C = 1$ and the claim count is Poisson. Here the scaled distributions get more and more concentrated about the expected value (scaled to 1.0). In Figure 4, C has a gamma distribution with variance 0.0625 (asymptotic CV of 25%). Now the scaled aggregate distributions converge to C.

The proposition shows that in many realistic insurance situations severity is irrelevant to the shape of the distribution of aggregate losses for a large book of business. This is an irritating but important result. Severity distributions are relatively easy to estimate, particularly when occurrence severity is limited by policy terms and conditions. Frequency distributions, on the other hand, are much more difficult to estimate. This proposition shows that the single most important variable for estimating the shape of A is the mixing distribution C. Problematically, C is never independently observed! The power of the proposition is to suggest a method for determining C: consider the loss ratio distribution of large books of business. We now do this using the NAIC annual statement data.

7.2 EMPIRICAL EVIDENCE—VOLUMETRIC

We use data from property-casualty NAIC annual statements to apply the above theory and determine an appropriate distribution for C (or $Z(1)$). This provides new insight into the exact form of "parameter risk". In the absence of empirical information, mathematical convenience usually reigns and a gamma distribution is used for C; the unconditional claim count is then a negative binomial. The distribution of C is called the structure function in credibility theory, Bühlmann (1970).

Schedule P in the property-casualty annual statement includes a ten accident-year history of gross, ceded and net premiums and ultimate losses

Table 3: Characteristics of Various Lines of Insurance

Line	Homogeneity	Regulation	Limits	Cycle	Cats
Pers Auto	High	High	Low	Low	No
Comm Auto	Moderate	Moderate	Moderate	High	No
Work Comp	Moderate	High	Statutory	High	Possible
Med Mal	Moderate	Moderate	Moderate	High	No
Comm MP	Moderate	Moderate	Moderate	High	Moderate
Other Liab Occ	Low	Low	High	High	Yes
Homeowners	Moderate	High	Low	Low	High
Other Liab CM	Low	Low	High	High	Possible

by major line of business. We focus on gross ultimate losses. The major lines include private passenger auto liability, homeowners, commercial multi-peril, commercial auto liability, workers compensation, other liability occurrence (premises and operations liability), other liability claims made (including directors and officers and professional liability but excluding medical), and medical malpractice claims made. These lines have many distinguishing characteristics that are subjectively summarized in Table 3. In the table, homogeneity refers to the consistency in terms and conditions within the line. The two other liability lines are catch-all classifications including a wide range of insureds and policies. Regulation indicates the extent of rate regulation by the states' insurance departments. Limits refers to the typical policy limit. Personal auto liability limits rarely exceed $300,000 per accident and are characterized as low. Most commercial lines policies have a primary limit of $1M, possibly with excess liability policies above that. Workers compensation policies do not have a limit but the benefit levels are statutorily prescribed by each state. Cycle is an indication of the extent of the pricing cycle in each line; it is simply split personal (low) and commercial (high). Finally, cats covers the extent to which the line is subject to multi-claimant, single occurrence catastrophe losses such as hurricanes, earthquakes, mass tort, securities laddering, terrorism and so on. The data is interpreted in the light of these characteristics.

In order to apply Proposition 1 we proxy a "large" book as one with more than $100M of premium in each accident year. Figure 5 shows how

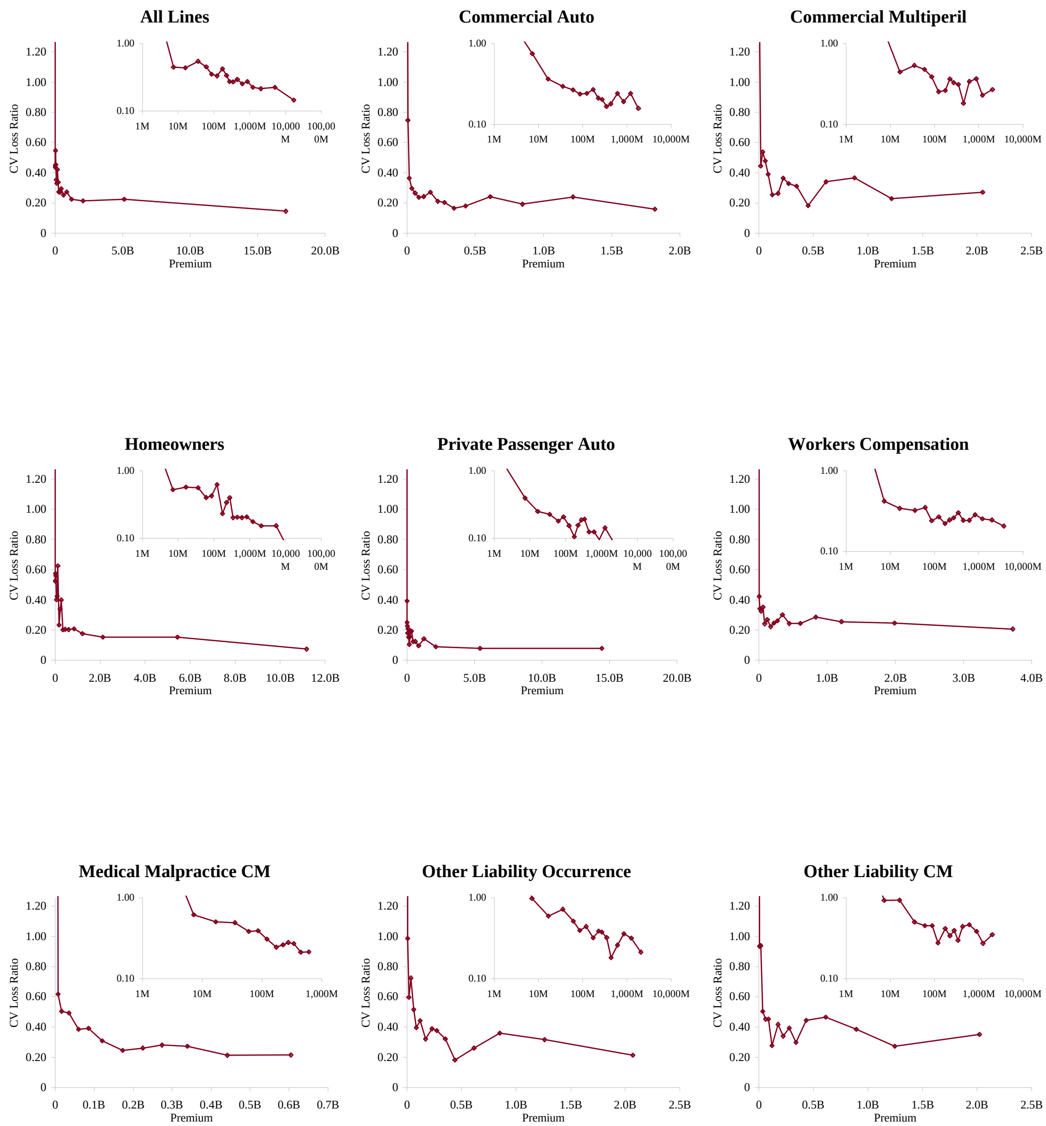

Figure 5: The relationship between raw loss ratio volatility, measured as coefficient of variation of loss ratios, and premium volume, using data from accident years 1993-2004. Each inset graph plots the same data on a log/log scale, showing that the volatility continues to decrease materially for premium volumes in the \$100M's. The total line is distorted by changing mix of business by volume; the largest companies are dominated by private passenger auto liability which is the lowest volatility line.

the volatility of loss ratio by line varies with premium size. It is computed by bucketing schedule P loss ratios by premium size band and computing the volatilities in each bucket. Each inset chart shows the same data on a log/log scale. The figure shows three things: (1) that the loss processes are not volumetrically diversifying, that is the volatility does not decrease to zero with volume; (2) that below a range \$100M-1B (varying by line) there are material changes in volatility with premium size; and (3) that \$100M is a reasonable threshold for large, in the sense that there is less change in volatility beyond \$100M. The second point means that the inhomogeneity in a loss portfolio is material in the range where most companies would try to set profit targets by line or business unit. This is consistent with Mildenhall (2004) and is also discussed in Section 8.

We now determine C by line by applying Proposition 1. The data consists of observed schedule P gross ultimate loss ratios $\lambda_{c,y}$ by company c and accident year $y = 1993, \dots, 2004$. The observation $\lambda_{c,y}$ is included if company c had gross earned premium $\geq$ \$100M in year y. The data is in the form of an unbalanced two-way ANOVA table with at most one observation per cell. Let $\lambda_{.,.}$ denote the average loss ratio over all companies and accident years, and $\lambda_{c,.}$ (resp. $\lambda_{.,y}$) the average loss ratio for company c over all years (resp. accident year y over all companies). Each average can be computed as a straight arithmetic average of loss ratios or as a premium-weighted average. With this data we will determine four different measures of volatility.

Res1. Raw loss ratio volatility across all twelve years of data for all companies. This volatility includes a pricing cycle effect, captured by accident year, and a company effect.

Res2. Control for the accident year effect $\lambda_{.,y}$. This removes the pricing cycle. It also removes some of the catastrophic loss effect for a year. In the sample data this causes problems with the results for homeowners in 2004.

Res3. Control for the company effect $\lambda_{c,.}$. This removes spurious loss ratio variation caused by differing expense ratios, distribution costs, profit targets, classes of business, limits, policy size and so forth.

Res4. Control for both company effect and accident year, i.e. perform an unbalanced two-way ANOVA with zero or one observation per cell. This can be done additively, modeling the loss ratio $\lambda_{c,y}$ for company c in year y as

$$\hat{\lambda}_{c,y} = \lambda_{\cdot,\cdot} + (\lambda_{c,\cdot} - \lambda_{\cdot,\cdot}) + (\lambda_{\cdot,y} - \lambda_{\cdot,\cdot}), \tag{57}$$

or multiplicatively as

$$\hat{\lambda}_{c,y} = \lambda_{\cdot,\cdot}(\lambda_{c,\cdot}/\lambda_{\cdot,\cdot})(\lambda_{\cdot,y}/\lambda_{\cdot,\cdot}). \tag{58}$$

The multiplicative approach is generally preferred as it never produces negative fit loss ratios. The statistical properties of the residual distributions are similar for both forms.

Using Proposition 1 we obtain four estimates for the distribution of C from the empirical distributions of $\lambda_{c,y}/\hat{\lambda}_{\cdot,\cdot}$, $\lambda_{c,y}/\hat{\lambda}_{\cdot,y}$, $\lambda_{c,y}/\hat{\lambda}_{c,\cdot}$ and $\lambda_{c,y}/\hat{\lambda}_{c,y}$ for suitably large books of business. The additive residuals $\lambda_{c,y} - \hat{\lambda}_{c,y}$ also have a similar distribution (not shown).

Figures 6-8 show analyses of variance for the model described by Eq. 57. Because the data is unbalanced, consisting of at most one observation per cell, it is necessary to perform a more subtle ANOVA than in the balanced case. We follow the method described in (Ravishanker and Dey, 2002, Section 9.2.2). The idea is to adjust for one variable first and then to remove the effect of this adjustment before controlling for the other variable. For example, in the extreme case where there is only one observation for a given company, that company's loss ratio is fit exactly with its company effect and the loss ratio observation should not contribute to the accident year volatility measure. Both the accident year effect and the company effect are highly statistically significant in all cases, except the unadjusted company effect for homeowners and the adjusted company effect for other liability claims made. The R^2 statistics are in the 50-70% range for all lines except homeowners. As discussed above, the presence of catastrophe losses in 2004 distorts the homeowners results.

Additive ANOVA for Commercial Auto, $100M Threshold

Source of Variation	Sum of Squares	D of F	Mean Squares	F Ratio	p Value	
Unadjusted Accident Year	6.3446	11	0.5768	37.2271	6.70E-55	***
Adjusted Company Effect	4.9147	56	0.0878	5.6645	3.44E-26	***
Residual	5.7658	407	0.0142			
Std. Deviation			11.9%			
Total (about mean)	17.0251	474	0.0359			
Std. Deviation			19.0%			
R2	0.6613					
Adjusted Accident Year	5.0773	11	0.4616	29.7914	8.48E-46	***
Unadjusted Company Effect	6.1819	56	0.1104	7.1250	3.14E-34	***
Tukey's Test for Interactions						
SSA	0.0035		F statistic	0.2443		
SSB	5.7658		p Value	0.621		
SSR	5.7624					

Additive ANOVA for Commercial Multiperil, $100M Threshold

Source of Variation	Sum of Squares	D of F	Mean Squares	F Ratio	p Value	
Unadjusted Accident Year	7.3649	11	0.6695	21.3469	1.85E-34	***
Adjusted Company Effect	7.9119	67	0.1181	3.7650	5.72E-17	***
Residual	12.0741	420	0.0287			
Std. Deviation			17.0%			
Total (about mean)	27.3509	498	0.0549			
Std. Deviation			23.4%			
R2	0.5585					
Adjusted Accident Year	9.4834	11	0.8621	27.4873	4.09E-43	***
Unadjusted Company Effect	5.7934	67	0.0865	2.7569	3.68E-10	***
Tukey's Test for Interactions						
SSA	0.0277		F statistic	0.9668		
SSB	12.0741		p Value	0.326		
SSR	12.0464					

Additive ANOVA for Homeowners, $100M Threshold

Source of Variation	Sum of Squares	D of F	Mean Squares	F Ratio	p Value	
Unadjusted Accident Year	3.0722	11	0.2793	3.1671	3.67E-04	***
Adjusted Company Effect	12.9407	78	0.1659	1.8813	3.39E-05	***
Residual	42.7530	488	0.0876			
Std. Deviation			29.6%			
Total (about mean)	58.7659	577	0.1018			
Std. Deviation			31.9%			
R2	0.2725					
Adjusted Accident Year	12.9724	11	1.1793	13.3729	1.97E-22	***
Unadjusted Company Effect	3.0405	78	0.0390	0.4420	1.00E+00	
Tukey's Test for Interactions						
SSA	0.0001		F statistic	0.0008		
SSB	42.7530		p Value	0.977		
SSR	42.7529					

Figure 6: Adjusted ANOVA for commercial auto, commercial multiperil and homeowners.

Additive ANOVA for Medical Malpractice CM, $100M Threshold

Source of Variation	Sum of Squares	D of F	Mean Squares	F Ratio	p Value	
Unadjusted Accident Year	5.7299	11	0.5209	8.9928	7.43E-11	***
Adjusted Company Effect	3.2898	29	0.1134	1.9584	7.88E-03	***
Residual	3.9561	97	0.0408			
Std. Deviation			20.2%			
Total (about mean)	12.9758	137	0.0947			
Std. Deviation			30.8%			
R2	0.6951					
Adjusted Accident Year	4.9819	11	0.4529	7.8189	1.49E-09	***
Unadjusted Company Effect	4.0377	29	0.1392	2.4037	7.46E-04	***
Tukey's Test for Interactions						
SSA	0.1502		F statistic	3.8683		
SSB	3.9561		p Value	0.052		
SSR	3.8059					

Additive ANOVA for Other Liability CM, $100M Threshold

Source of Variation	Sum of Squares	D of F	Mean Squares	F Ratio	p Value	
Unadjusted Accident Year	8.6215	11	0.7838	17.0633	6.97E-21	***
Adjusted Company Effect	2.1244	31	0.0685	1.4919	6.22E-02	
Residual	5.0512	138	0.0366			
Std. Deviation			19.1%			
Total (about mean)	15.7971	180	0.0878			
Std. Deviation			29.6%			
R2	0.6802					
Adjusted Accident Year	2.9908	11	0.2719	5.9192	7.73E-08	***
Unadjusted Company Effect	7.7551	31	0.2502	5.4463	1.49E-12	***
Tukey's Test for Interactions						
SSA	0.0566		F statistic	1.5756		
SSB	5.0512		p Value	0.211		
SSR	4.9945					

Additive ANOVA for Other Liability Occurrence, $100M Threshold

Source of Variation	Sum of Squares	D of F	Mean Squares	F Ratio	p Value	
Unadjusted Accident Year	10.5218	11	0.9565	16.9844	4.71E-27	***
Adjusted Company Effect	9.6707	59	0.1639	2.9104	5.08E-10	***
Residual	19.5477	362	0.0540			
Std. Deviation			23.2%			
Total (about mean)	39.7401	432	0.0920			
Std. Deviation			30.3%			
R2	0.5081					
Adjusted Accident Year	10.4953	11	0.9541	16.9416	5.47E-27	***
Unadjusted Company Effect	9.6972	59	0.1644	2.9184	4.56E-10	***
Tukey's Test for Interactions						
SSA	0.0042		F statistic	0.0779		
SSB	19.5477		p Value	0.780		
SSR	19.5435					

Figure 7: Adjusetd ANOVA for medical malpractice claims made and other liability claims made and occurrence.

Additive ANOVA for Private Passenger Auto, $100M Threshold

Source of Variation	Sum of Squares	D of F	Mean Squares	F Ratio	p Value	
Unadjusted Accident Year	1.3640	11	0.1240	20.9565	1.55E-37	***
Adjusted Company Effect	6.1637	101	0.0610	10.3137	1.11E-90	***
Residual	4.5636	786	0.0058			
Std. Deviation			7.6%			
Total (about mean)	12.0913	898	0.0135			
Std. Deviation			11.6%			
R2	0.6226					
Adjusted Accident Year	6.2189	11	0.5654	95.5466	9.70E-137	***
Unadjusted Company Effect	1.3088	101	0.0130	2.1900	2.86E-09	***

Tukey's Test for Interactions			
SSA	0.0022	F statistic	0.3720
SSB	4.5636	p Value	0.542
SSR	4.5615		

Additive ANOVA for Workers Compensation, $100M Threshold

Source of Variation	Sum of Squares	D of F	Mean Squares	F Ratio	p Value	
Unadjusted Accident Year	13.9945	11	1.2722	68.2576	1.94E-96	***
Adjusted Company Effect	5.1661	86	0.0601	3.2229	1.00E-16	***
Residual	9.5719	569	0.0168			
Std. Deviation			13.0%			
Total (about mean)	28.7325	666	0.0431			
Std. Deviation			20.8%			
R2	0.6669					
Adjusted Accident Year	6.1297	11	0.5572	29.8973	1.08E-49	***
Unadjusted Company Effect	13.0309	86	0.1515	8.1295	2.58E-57	***

Tukey's Test for Interactions			
SSA	0.0133	F statistic	0.7954
SSB	9.5719	p Value	0.373
SSR	9.5586		

Figure 8: Adjusted ANOVA for private passenger auto liability and workers compensation.

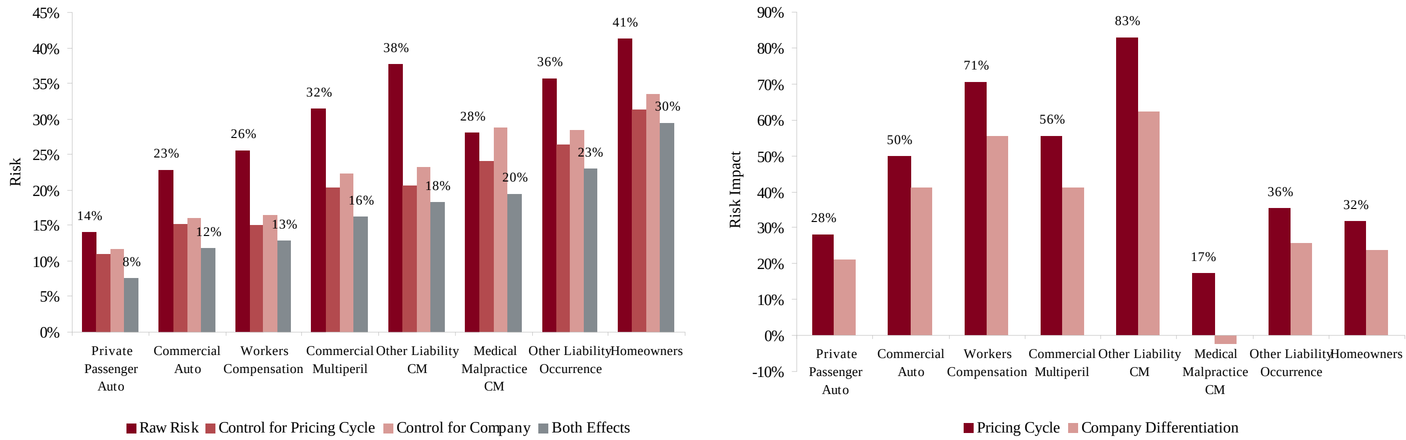

Figure 9: Left plot shows the loss ratio volatility by line for companies writing \$100M or more premium each year based on Schedule P accident year ultimate booked gross loss ratios, from 1993-2004. The graph shows the effect of adjusting the loss ratio for an accident year pricing effect, a company effect, and both effects (i.e. Res1-4). The right hand plot shows the differential impact of the pricing effect and company effect by line. Each bar shows the increase in volatility of the unadjusted loss ratios compared to the adjusted.

Tukey's test for interactions in an ANOVA with one observation per cell, (Miller and Wichern, 1977, Section 4.11) does not support an interaction effect for any line at the 5% level. This is consistent with a hypothesis that all companies participate in the pricing cycle to some extent.

Figure 9 shows the indicated volatilities for commercial auto, commercial multi-peril, homeowners, other liability occurrence, private passenger auto liability and workers compensation for the four models Res1-4 and Eq. 58. The right hand plot shows the impact of the pricing (accident year) effect and the firm effect on total volatility. This chart shows two interesting things. On the left it gives a ranking of line by volatility of loss ratio from private passenger auto liability, 14% unadjusted and 8% adjusted, to homeowners and other liability occurrence, 41% and 36% unadjusted and 30% and 23% adjusted, respectively. The right hand plot shows that personal lines have a lower pricing cycle effect (28% and 32% increase in volatility from pricing) than the commercial lines (mostly over 50%). This is reasonable given the highly regulated nature of pricing and the lack of underwriter schedule credits and debits. These results are consistent with the broad classification in Table 3.

Figures 10–13 show the histograms of normalized loss ratio distributions corresponding to Res1-4 for the same eight lines of business. These give a direct estimate of the distribution of C. There are four plots shown for each line.

The top left plot shows the distribution of normalized Schedule P accident year ultimate booked gross loss ratios for companies writing \$100M or more premium, for 1993-2004. The distributions are shown for each of the four models Res1-4. LR indicates the raw model, Co Avg adjusts for company effect, AY Avg adjusts for accident year or pricing cycle effect, and Mult Both Avg adjusts for both, per Eq. 58. All residuals are computed using the multiplicative model.

The top right hand plot shows various fits to the raw residuals. The fits are: Wald or inverse Gaussian, EV Frechet-Tippet extreme value, gamma distribution, LN lognormal, and SLN the shifted lognormal. Gamma, EV and SLN are fit using method of moments, Wald and lognormal are fit using

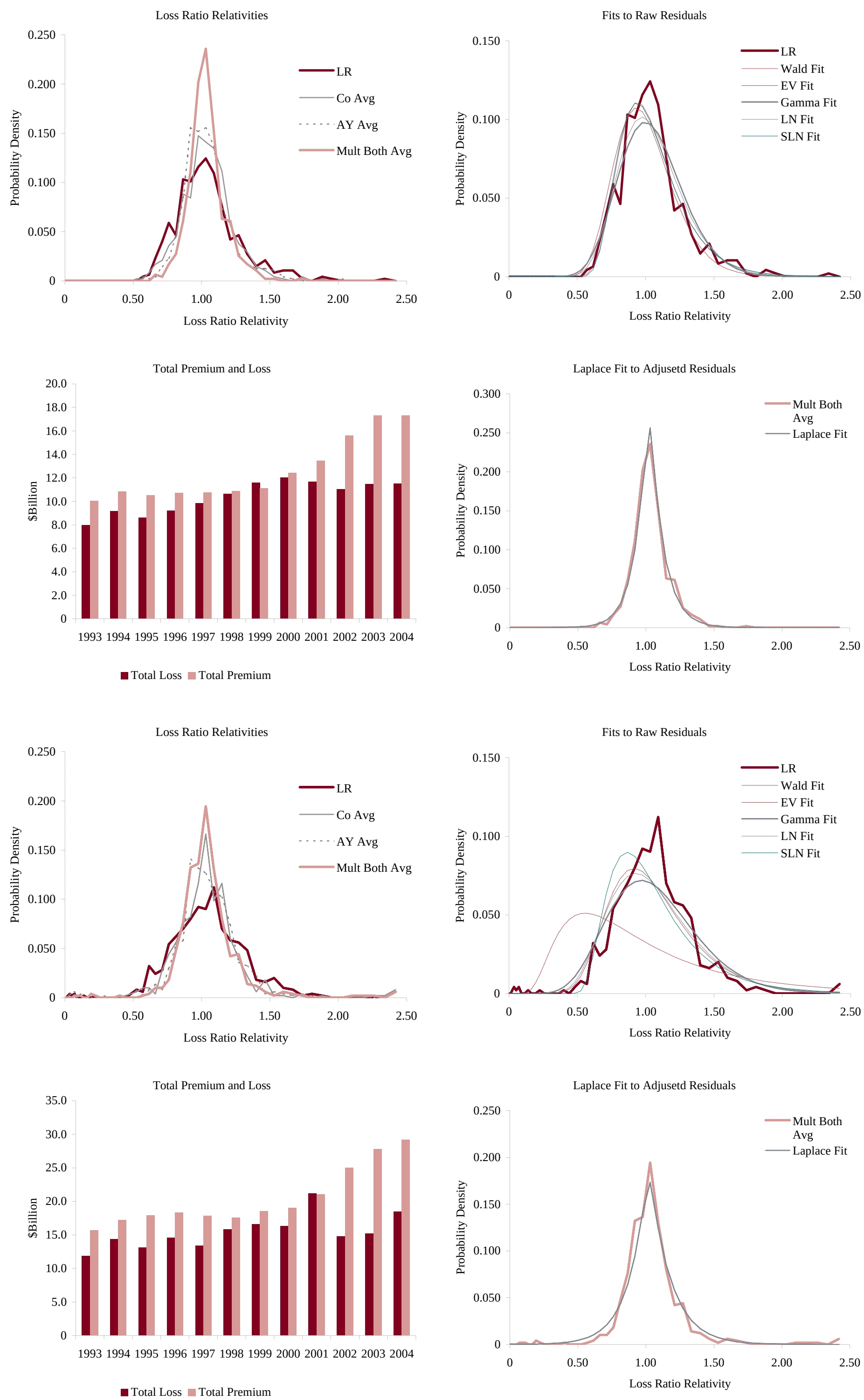

Figure 10: Commercial auto liability (top four plots) and commercial multiperil volatility (bottom four plots). Note 9/11 loss effect in the lower-left plot. See text for a description of the plots.

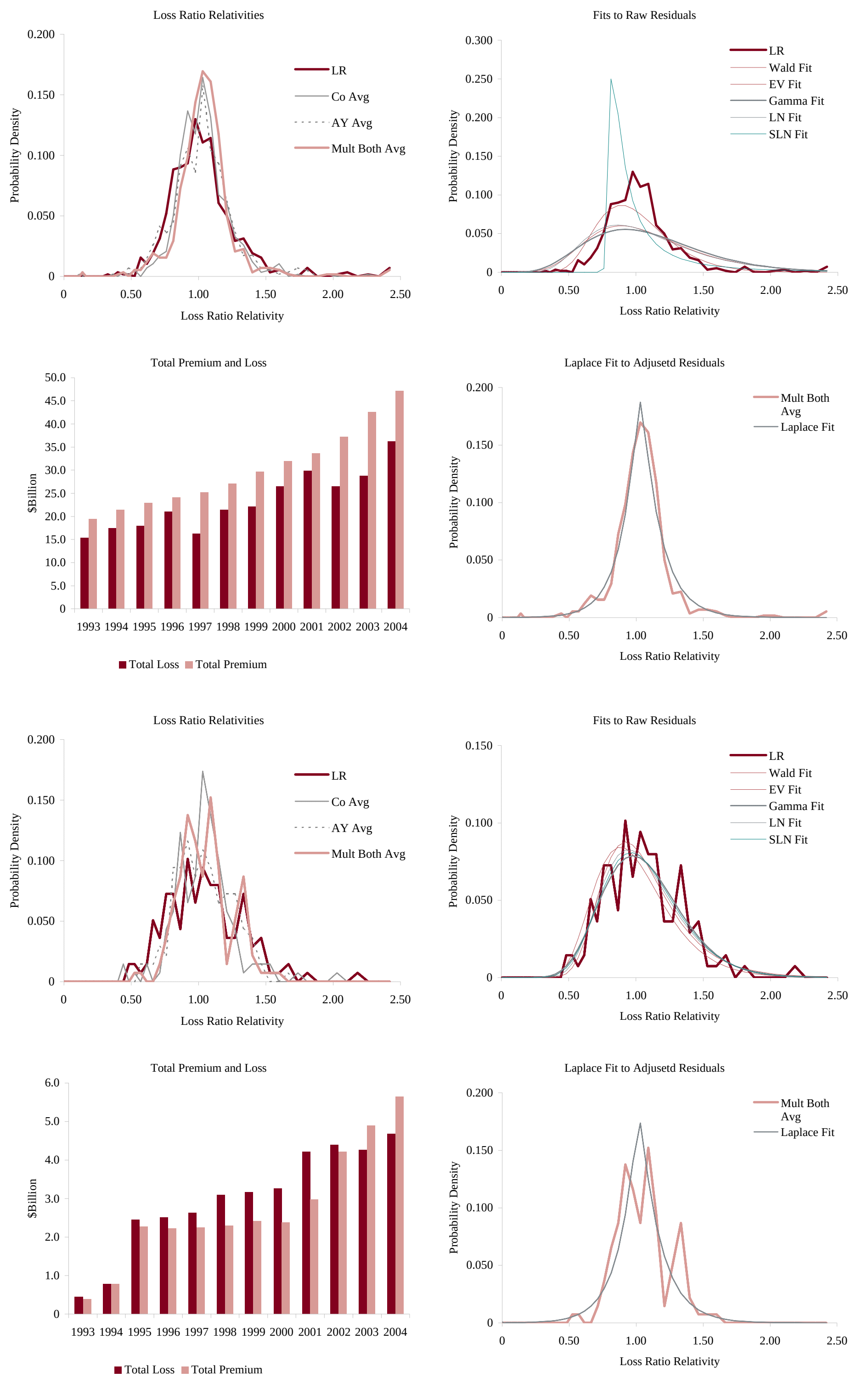

Figure 11: Homeowners (top four plots) and medical malpractice claims made volatility (bottom four plots). Note the 2004 homowners catastrophe losses. See text for a description of the plots.

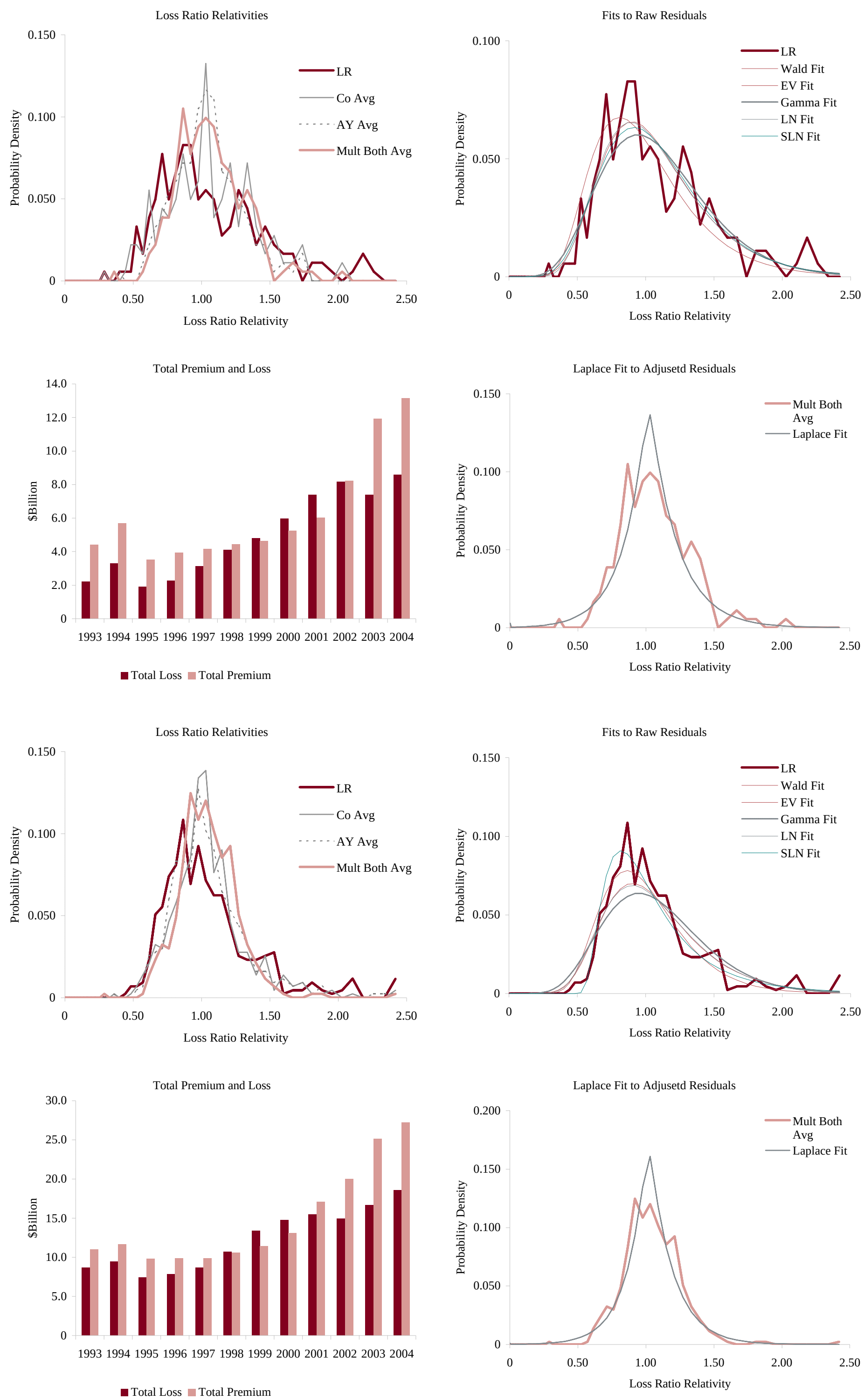

Figure 12: Other liability claims made (top four plots) and occurrence volatility (bottom four plots). See text for a description of the plots.

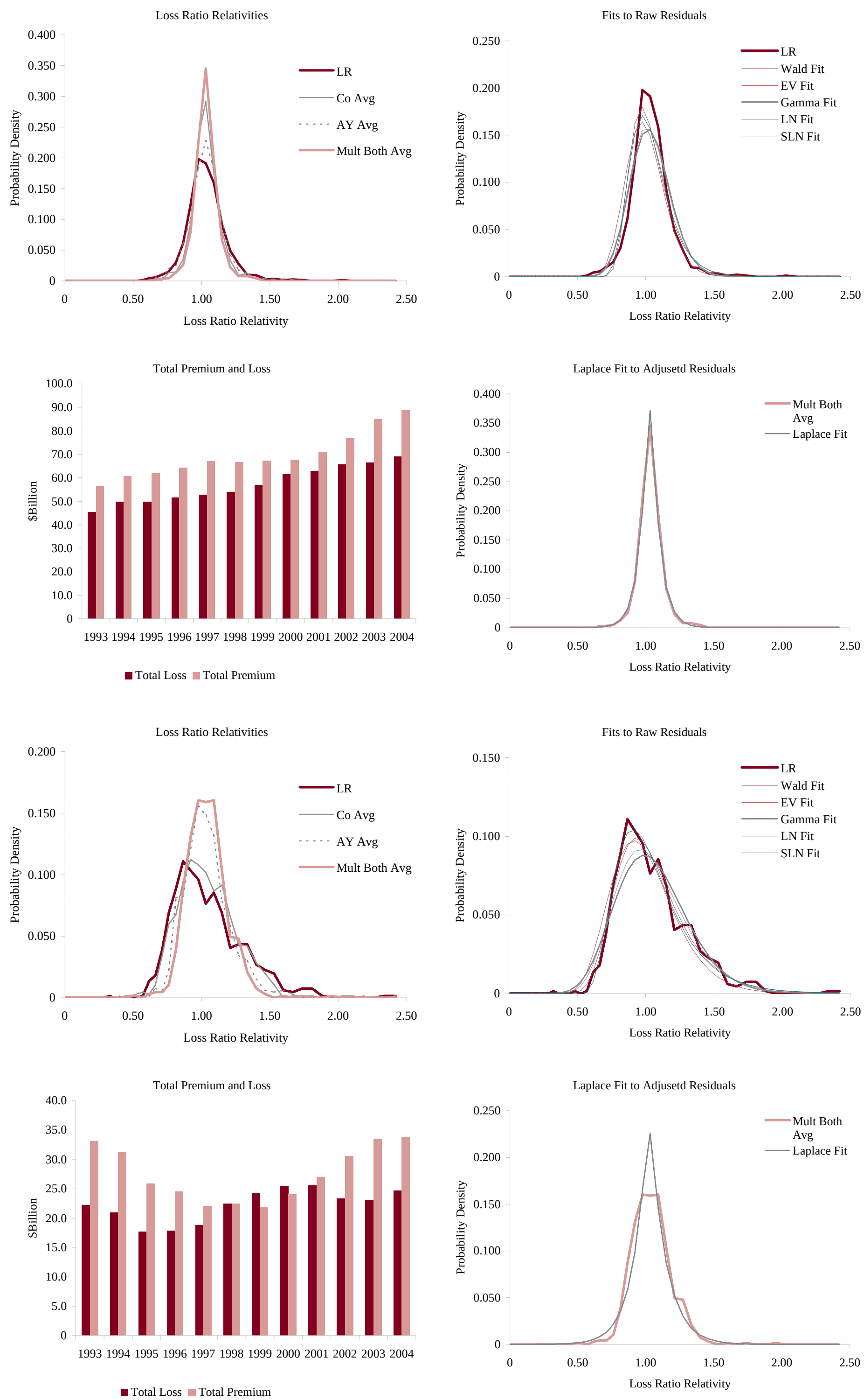

Figure 13: Private passenger auto liability (top four plots) and workers compensation volatility (bottom four plots). Note vertical scale on private passenger auto loss ratios and the visibly higher volatility of premium than loss in the lower left hand plot. See text for a description of the plots.

maximum likelihood. The shifted lognormal distribution has three parameters and so would be expected to fit better. The raw residuals, Res1, are typically more skewed than Res4 and do not have the same peaked shape. The gamma distribution fit is shown in bold grey; the adequacy of its fit varies from line to line.

The lower right hand plot shows the residuals adjusted for both pricing cycle and company effects and it includes a maximum likelihood Laplace fit to the multiplicative model Eq. 58. This plot strongly supports the choice of a Laplace distribution for C in the adjusted case. This is a very unexpected result as the Laplace is symmetric and leptokurtic (peaked). The Laplace distribution is to the absolute value difference what the normal distribution is to squared difference; median replaces mean. One could speculate that a possible explanation for the Laplace is the tendency of insurance company management to discount extreme outcomes and take a more median than mean view of losses. The Laplace can be represented as a subordinated Brownian motion, introducing operational time as in IM2 and IM4. The subordinator has a gamma distribution. The Laplace is also infinitely divisible and its Lévy measure has density $\nu(x) = |x|^{-1}e^{-|x|/s}$. See Kotz et al. (2001).

The lower left hand plot shows the premium and loss volume by accident year. It shows the effect of the pricing cycle and the market hardening since 2001 present in all lines!

The analysis in this section assumes $t = 1$. Therefore it is impossible to differentiate models IM2-4. However, the data shows that losses are not volumetrically diversifying, Figure 5. The data suggests that C (or $Z(1)$) has a right-skewed distribution when it includes a company and pricing cycle effect and strongly suggests a Laplace distribution when adjusted for company and pricing cycle effects.

7.3 EMPIRICAL EVIDENCE—TEMPORAL

This section investigates the behavior of $\upsilon(x, t)$ for different values of t. The analysis is complicated by the absence of long-term, stable observations. Multi-year observations include strong pricing cycle effects, results from different companies, different terms and conditions (for example the change from occurrence to claims made in several lines), and the occurrence of infrequent shock losses. Moreover, management actions, including reserve setting and line of business policy form and pricing decisions, will affect observed volatility.

Reviewing Table 5, and comparing with Figure 5, shows IM2-4 are consistent with the data analyzed so far. The difference between IM2 and IM4 compared to IM3 is the presence of a separate time effect in $\upsilon(x,t)$. Both models IM2 and IM4 should show a lower volatility from a given volume insurance when that insurance comes from multiple years, whereas model IM3 will not. This suggests a method to differentiate IM2/4 from IM3. First, compute $\upsilon(x,1)$, the data underling Figure 5. Then combine two years of premium and losses and recompute volatilities. This essentially computes $\upsilon(x/2,2)$—total volume is still x but it comes from two different years. Similarly, combining 4, 6 or 12 years of data (divisors of the total 12 years of data available) gives an estimate of $\upsilon(x/4,4)$, $\upsilon(x/6,6)$, and $\upsilon(x/12,12)$. Normalizing the data to a constant loss ratio across accident years prior to performing the analysis will remove potentially distorting pricing-cycle effects.

Figure 14 shows the results of performing this analysis for private passenger auto liability. Private passenger auto liability is used because it has very low inherent process risk and low parameter risk, and so provides the best opportunity for the delicate features we are analyzing to emerge. In the figure, the second column shows $\upsilon(x,1)$ and the last four show $\upsilon(x/t,t)$ for $t = 2,4,6,12$. The average volume in each band is shown as average premium in the first column. Below the data we show the averages and standard deviations of υ for broader volume bands. Clearly the differences in means are insignificant relative to the standard deviations, and so a crude analysis of variance accepts the hypothesis that $\upsilon(x/t,t)$ is independent of t. This data implies that models IM2 and IM4 do not provide a good fit to the data—unless τ is very small. However, if τ is small then IM2 and IM4 degenerate to IM1, which has already been rejected since it is volumetrically diversifying.

Average Premium	Coefficient of Variation Loss Ratio Computed From				
	1 Year	2 Years	4 Years	6 Years	12 Years
473	1.085	0.819	0.520	0.471	0.550
1,209	0.580	0.428	0.449	0.419	0.438
1,680	0.448	1.455	0.684	0.342	0.245
2,410	1.927	0.451	1.238	0.423	0.383
3,458	0.294	0.299	0.204	0.187	0.376
4,790	0.369	0.286	0.347	0.312	0.346
6,809	0.475	0.292	0.310	0.267	0.350
9,526	0.272	0.346	0.311	0.236	0.248
13,501	0.290	0.623	0.246	0.521	0.212
19,139	0.191	0.227	0.303	0.204	0.211
26,649	0.244	0.195	0.183	0.292	0.196
37,481	0.188	0.191	0.223	0.171	0.155
54,287	0.173	0.183	0.297	0.239	0.264
73,882	0.191	0.154	0.166	0.167	0.219
108,762	0.158	0.169	0.170	0.122	0.159
153,233	0.137	0.185	0.147	0.204	0.175
213,224	0.127	0.152	0.172	0.146	0.102
307,833	0.186	0.129	0.141	0.116	0.152
439,136	0.117	0.125	0.146	0.174	0.085
606,457	0.110	0.182	0.090	0.136	0.137
845,813	0.092	0.102	0.145	0.126	0.137
1,215,551	0.132	0.103	0.124	0.112	0.101
1,725,327	0.115	0.088	0.111	0.125	0.071
2,362,126	0.068	0.130	0.101	0.089	0.135
3,597,590	0.042	0.111	0.080	0.085	0.082
8,430,433	0.079	0.073	0.094	0.087	0.079
Avg. $3M-20M	0.315	0.345	0.287	0.288	0.291
Std.Dev. $3M-20M	0.097	0.141	0.052	0.123	0.075
Avg. $21M-200M	0.182	0.179	0.198	0.199	0.195
Std.Dev. $21M-200M	0.037	0.015	0.055	0.060	0.042
Avg >$200M	0.107	0.120	0.120	0.120	0.108
Std.Dev. >$200M	0.040	0.032	0.030	0.029	0.030

Figure 14: Coefficient of variation of loss ratio by premium volume for private passenger auto liability, computed using bucketed xt for $t = 1, 2, 4, 6, 12$.

Figure 15: Fit of $\sqrt{\dfrac{\sigma^2}{xt}+c}$ to volatility by volume, xt, for homeowners, private passenger auto and commercial auto. Left hand plot shows data based on a single year $t = 1$; right hand plot shows the same data for $t = 1, 2, 4, 6, 12$.

Figure 16: Fit of $\sqrt{\dfrac{\sigma^2}{xt} + c}$ to volatility by volume, xt, for workers compensation, commercial multiperil and other liability occurrence. Left hand plot shows data based on a single year $t = 1$; right hand plot shows the same data for $t = 1, 2, 4, 6, 12$.

Finally, Figures 15 and 16 provide a graphical representation of the same data for homeowners, private passenger auto, commercial auto, workers' compensation, commercial multi-peril and other liability occurrence (other liability claims made and medical malpractice lack the necessary volume). The left hand plot shows the same data as Figure 5 on a log/linear scale and a fit of $\upsilon(x,t)$ by $\sqrt{(\sigma^2/xt)+c}$. In the fit, c is estimated from the observed asymptotic volatility and σ is estimated using minimum squared distance. The right hand plot overlays $\upsilon(x/t,t)$ for $t=2,4,6,12$ using the method described above. Thus the private passenger auto liability plot shows the data in Figure 14. These plots support the hypothesis that $\upsilon(x/t,t)$ is independent of t as there is no clear trend with t. (The case $t=12$ is subject to higher estimation error owing to the lower number of observations.)

We conclude that only model IM2 of IM1-4 and AM1 has volumetric and temporal properties consistent with the data in the NAIC annual statement database.

8 OPTIMIZATION AND HOMOGENEITY

This section sets up an explicit optimization problem which has a non-trivial solution if and only if losses are homogeneous, $A(x,t)=xX(t)$. It is the optimization derived by Myers and Read (2001), who also assume homogeneous losses. The example is used to show how the homogeneity assumption leads to a peculiar solution. We assume that $t=1$ and drop t from the notation.

Assume there are n lines of business with loss processes $X_i(x)$, $\mathrm{E}(X_i(x)) = x$. The loss processes are *not* assumed to be homogeneous. Assume there are no expenses. The profit margin in line i is π_i, so premium is $(1+\pi_i)x_i$. Capital is denoted y. The cost of capital in excess of the risk free rate for insurance capital is ν. The product market determines π_i and the capital markets determine ν. The company takes these variables as given. Assume that we have a risk measure $\rho:\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}$ where $\rho(x_1,\dots,x_n,\alpha)$ determines the amount of capital corresponding to a safety level $\alpha\ge 0$. Using the implicit function theorem gives a function $\alpha:\mathbb{R}^{n+1}\to\mathbb{R}$ so that

$$\rho(x_1,\dots,x_n,\alpha(x_1,\dots,x_n,y)) = y. \tag{59}$$

Consider the optimization problem of minimizing α subject to the cost of capital constraint:

$$\begin{cases} \min_{x_1,\dots,x_n,y} \ \alpha(x_1,\dots,x_n,y) \quad \text{subject to} \\ \sum_i \pi_i x_i = \nu y. \end{cases} \tag{60}$$

This problem is easy to solve using Lagrangian multipliers. The Lagrangian is

$$L = \alpha(x_1,\dots,x_n,y) - \lambda(\sum_i \pi_i x_i - \nu y). \tag{61}$$

Differentiating with respect to x_i and y gives

$$\frac{\partial L}{\partial x_i} = \frac{\partial \alpha}{\partial x_i} - \lambda \pi_i = 0 \tag{62}$$

$$\frac{\partial L}{\partial y} = \frac{\partial \alpha}{\partial y} + \lambda \nu = 0. \tag{63}$$

Thus, at the optimal solution

$$\lambda = -\frac{\partial \alpha}{\partial x_i}/\pi_i = \frac{\partial \alpha}{\partial y}/\nu. \tag{64}$$

Using the implicit function theorem gives

$$\frac{\pi_i}{\nu} = \frac{\partial y}{\partial x_i} = \frac{\partial \rho}{\partial x_i} \tag{65}$$

since $y = \rho(x)$. Once again, at the optimal volume and capital levels, pricing is proportional to the gradient vector. All optimization problems have a dual which swaps the constraint and objective. Here, the dual problem maximizes the return over cost of capital subject to a constraint on α. The dual can be used to compute implied margins by line assuming the existing book of business $x_1,\dots,x_n$ and level of capital y are optimal. Again the margins are proportional to the gradient vector.

Substituting Eq. 65 into the cost of capital constraint Eq. 60 gives

$$\nu y = \sum_i \pi_i x_i = \sum_i \nu x_i \frac{\partial y}{\partial x_i} \tag{66}$$

and so

$$y = \sum_i x_i \frac{\partial y}{\partial x_i}, \tag{67}$$

recovering the Myers-Read adds-up result.

Mildenhall (2004) shows that Eq. 67 holds if and only if the loss processes are homogeneous: $X_i(x) = xX_i$. The above derivation *appears* to have proved the same result without assuming homogeneity. The wrinkle is that unless losses are homogeneous the solution is trivial: $\alpha = 0$ and write no business. It is clear this is always a solution of the optimization. The reason there are other non-trivial solutions in the homogeneous case is the lack of diversification. If $X_i(x)$ diversifies then another solution to the optimization involves growing infinitely large.

This is illustrated in Figure 17 when $n = 2$. Each plot shows volume x_1, x_2 in lines 1 and 2 in the horizontal plane and required capital $\rho(x_1, x_2, \alpha)$ at a particular α level on the vertical axis. The flat green surface is a budget constraint given by the cost of capital, $\pi_1 x_1 + \pi_2 x_2 = \nu y$. The left hand plot shows diversifying insurance losses. The diversification effect can be seen along the planes $x_1 = 0$ and $x_2 = 0$, which the capital surface intersects in a concave down line. The Lagrangian requirement for a solution is that the flat green plane be tangent to the capital surface. Since the surface is concave down (reflecting diversification with volume in each line and diversification between the two lines) the only possible tangent point with a plane through the origin—such as the budget constraint—is the origin 0.[12] The right hand plot corresponds to a homogeneous model. The surface is "ruled": it is straight along rays from the origin rather than concave down. Now it is possible for the budget plane to be tangent to the capital surface along an entire line, including the origin. Such a line is illustrated on the plot. The

[12]If we use Lagrangian multipliers to maximize the difference $\sum_i \pi_i x_i - \nu y$ at a given α then the plane no longer goes through the origin. In this case there will be a unique solution in the diversifying case, but it will be a minimum and not a maximum. If X_1 and X_2 are correlated it is possible for the surface to be flat or convex along lines of the form $x_1 + x_2 =$ constant, but it will always be concave along rays from the origin for diversifying insurance losses.

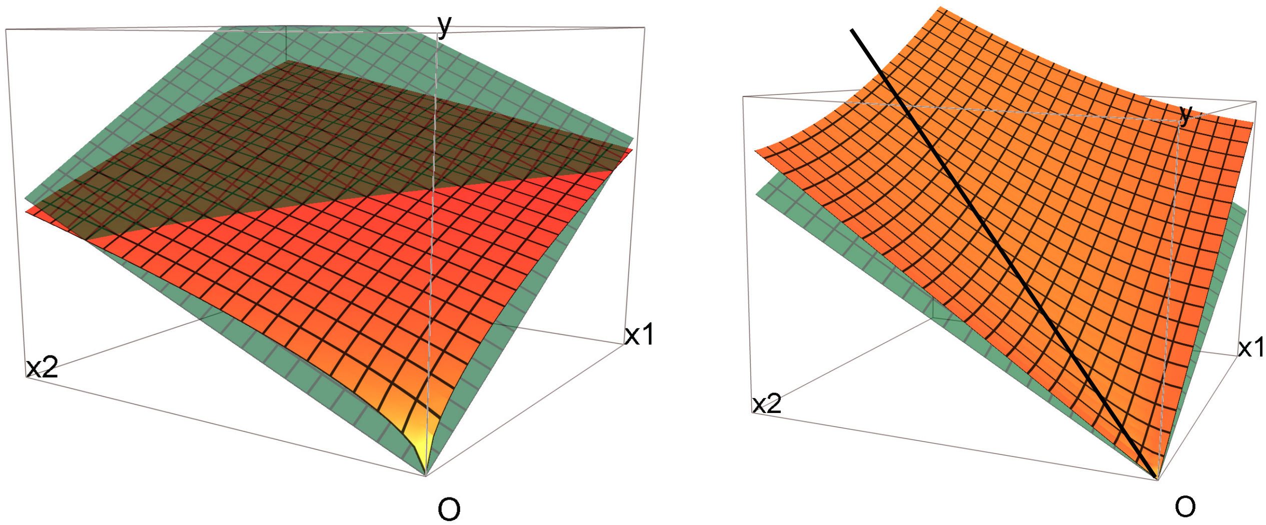

Figure 17: Diversifying, inhomogeneous (left) and non-diversifying, homogeneous (right) capital surfaces. The axes x_1 and x_2 represent the volume in each line and y represents capital at a fixed safety level α. The left plot was produced assuming that $X_i(x) = x + \sqrt{x}Z_i$ for Z_i independent standard normal variables, and the right that $X_i(x) = x + xZ_i$. Thus $\mathrm{Var}(X_i(x)) = x$ on the left and $\mathrm{Var}(X_i(x)) = x^2$ on the right. This model can be extended to model losses as $x + \sqrt{x}Z_1 + xZ_2$ to produce an example with behavior similar to IM2-4. It is tractable enough to have a closed form solution and yet varied enough to model all the different types of behavior of interest.

Myers-Read solution corresponds to a line of solutions along the ruled-capital plane.

Some have tried to salvage Myers-Read's adds-up result by arguing that aggregate losses are "approximately homogeneous". However, inhomogeneity is material in realistically sized portfolios, Mildenhall (2004). The approximation argument fails because Euler's theorem uses derivatives and hence requires a first order approximation rather than just pointwise equality. This failure is supported by Figure 5 and is also illustrated in Figure 1 where k and m cross at an angle. The fact that $k(1) = m(1)$ does not guarantee $k'(1) = m'(1)$. Capital allocation and profit targets are usually set for segments of business generating $10-100M of premium. It is clear from Figure 5 that the volatility of losses decreases materially in this range—resulting in a change in the geometry of the aggregate loss process—and invalidating the adds-up result.

9 CONCLUSIONS

This paper provides an introduction to the actuarial use of Lévy processes to model aggregate losses. The Lévy process model, reflecting the realities of insurance, is an abelian convolution semigroup that is curved in both the volume and time dimensions. Asset returns, in contrast, naturally form a real vector space with scalar asset weights and are volumetrically flat. The paper describes the relationship between Lévy processes and catastrophe model output. It clarifies the notion of a "direction" in the space of risks and uses it to explain two different allocation results derived using the gradient of a risk measure.

NAIC annual statement data is used to demonstrate that insurance liabilities do not diversify volumetrically or temporally. This paper reviews four models of aggregate losses based on Lévy processes—models with a long risk-theoretic pedigree, it should be noted—and shows that only model IM2 is consistent with the NAIC data. It also shows how parameter risk can be quantified at a *distributional* level. Volume-related parameter risk, adjusted for company and pricing cycle effects, is shown to have a Laplace distribution—a surprising result.

In conclusion, this paper is a call-to-arms. Finance now provides a theoretical justification for pricing company-specific risk. Risk theory provides a rigorous approach to evaluating and attributing risk to line using risk measure gradients. Enterprise Risk Management, which depends crucially on an accurate quantification of aggregate loss distributions, demands accurate and realistic modeling. It is time to satisfy that demand with a fully data-grounded model for losses, including appropriate parameter risk.

References

Abraham, R., Marsden, J. E., and Ratiu, T. (1988). *Manifolds, Tensor Analysis, and Applications*. Springer Verlag, second edition.

Applebaum, D. (2004). *Lévy Processes and Stochastic Calculus*. Cambridge University Press.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Math. Finance*, pages 203–28.

Bailey, R. A. (1967). Underwriting profits from investments. *Proceedings of the Casualty Actuarial Society*, LIV:1–8.

Barndorff-Nielsen, O. E., Mikosch, T., and Resnick, S. I. (2001). *Lévy Proceeses - theory and applications*. Birkhäuser.

Beard, R. E., Pentikäinen, T., and Pesonen, E. (1969). *Risk Theory—The Stochastic Basis of Insurance*. Chapman and Hall, third edition.

Bertoin, J. (1996). *Lévy Processes*. Cambridge University Press.

Borch, K. (1962). Equilibrium in a reinsurance market. *Econometrica*, 30(3):424–444.

Bowers, N., Gerber, H., Hickman, J., Jones, D., and Nesbitt, C. (1986). *Actuarial Mathematics*. Society of Actuaries.

Breiman, L. (1992). *Probability*. Classics in Applied Mathematics. SIAM.

Bühlmann, H. (1970). *Mathematical Models in Risk Theory*. Springer Verlag.

Cheridito, P., Delbaen, F., and Kupper, M. (2003). Coherent and convex risk measures for bounded càdlàg processes. *preprint*.

Cummins, J. D. (1988). Risk-based premiums for insurance guarantee funds. *Journal of Finance*, 43:823–839.

Cummins, J. D. and Geman, H. (1995). Pricing catastrophe insurance futures and call spreads: An arbitrage approach. *J. of Fixed Income*, 4(4):46–57.

Cummins, J. D. and Harrington, S. A. (1987). *Fair Rate of Return in Property-Liability Insurance*. Kluwer-Nijhoff.

Cummins, J. D. and Phillips, R. D. (2000). Applications of financial pricing models in property-liability insurance. In Dionne, G., editor, *Handbook of Insurance*. Kluwer Academic.

D'Arcy, S. and Dyer, M. A. (1997). Ratemaking: A financial economics approach. *Proceedings of the Casualty Actuarial Society*, LXXXIV:301–390.

D'Arcy, S. P. and Doherty, N. A. (1988). *The Financial Theory of Pricing Property Liability Insurance Contracts*. Huebner Foundation/Irwin, Homewood IL.

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

Delbaen, F. (2000a). Coherent risk measures. *Monograph, Scoula Normale Superiore*.

Delbaen, F. (2000b). Coherent risk measures on general probability spaces. In *Essays in Honour of Dieter Sondermann*. Springer Verlag.

Delbaen, F. and Haezendonck, J. (1989). A martingale approach to premium calculation principles in an arbitrage free market. *Insurance: Mathematics and Economics*, 8:269–277.

Denault, M. (2001). Coherent allocation of risk capital. *preprint*.

Doherty, N. A. and Garven, J. R. (1986). Price regulation in property-liability insurance: A contingent claims approach. *Journal of Finance*, XLI(5):1031–1050.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications, two volumes*. J. Wiley and Sons, second edition.

Ferrari, J. R. (1968). The relationship of underwriting, investment, leverage, and exposure to total return on owners' equity. *Proceedings of the Casualty Actuarial Society*, LV:295–302.

Fischer, T. (2003). Risk capital allocation by coherent risk measures based on one-sided moments. *Insurance: Mathematics and Economics*, 32:135–146.

Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4).

Froot, K. A. (2003). Risk management, capital budgeting and capital structure policy for insurers and reinsurers. *NBER Working Paper Series*, 10184.

Froot, K. A., Scharfstein, D. S., and Stein, J. C. (1993). Risk management: Coordinating corporate investment and financing policies. *Journal of Finance*, XLVIII (5):1629–1658.

Froot, K. A. and Stein, J. C. (1998). Risk management, capital budgeting, and capital structure policy for inancial institutions: an integrated approach. *Journal of Financial Economics*, 47:52–82.

Goovaerts, M. J., de Vylder, F., and Haezendonck, J. (1984). *Insurance Premiums*. North-Holland.

Hull, J. C. (1983). *Options Futures and Other Derivative Securities*. Prentice-Hall, second edition.

Kalkbrener, M. (2005). An axiomatic approach to capital allocation. *Mathematical Finance*, to appear.

Kallop, R. H. (1975). A current look at workers' compensation ratemaking. *Proceedings of the Casualty Actuarial Society*, LXII:62–81.

Karatzas, I. and Shreve, S. (1988). *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (1998). *Loss Models, From Data to Decisions*. John Wiley and Sons.

Kotz, S., Kozubowski, T., and Podgorski, K. (2001). *The Laplace Distribution and Generalizations*. Birkhauser.

Lange, J. T. (1966). General liability insurance ratemaking. *Proceedings of the Casualty Actuarial Society*, LIII:26–53.

Magrath, J. J. (1958). Ratemaking for fire insurance. *Proceedings of the Casualty Actuarial Society*, XLV:176–195.

McGuinness, J. S. (1969). Is "probable maximal loss" (PML) a useful concept? *Proceedings of the Casualty Actuarial Society*, LVI:31–40.

Meister, S. (1995). Contributions to the mathematics of catastrophe insurance futures. *ETH*.

Merton, R. C. and Perold, A. (2001). Theory of risk capital in financial firms. In Donald H. Chew, J., editor, *The New Corporate Finance, Where Theory Meets Practice*, pages 438–454. McGraw-Hill.

Meyers, G. (2003). The economics of capital allocation. *preprint*. http://www.aria.org/rts/rts2005/proceedings05.htm.

Meyers, G. (2005a). The common shock model for correlated insurance losses. *Proceedings of the Risk Theory Society*.

Meyers, G. (2005b). Distributing capital: Another tactic. *The Actuarial Review*, November:25–26. with on-line technical appendix.

Mildenhall, S. (2004). A note on the Myers and Read capital allocation formula. *North American Actuarial Journal*, 8(2):32–44.

Miller, R. B. and Wichern, D. W. (1977). *Intermediate Business Statistics: Analysis of Variance, Regression and Time Series*. HRW.

Myers, S. C. and Read, J. A. (2001). Capital allocation for insurance companies. *Journal of Risk and Insurance*, 68 no. 4:545–580.

Panjer, H. H. (2001). Measurement of risk, solvency requirements and allocation of capital within financial congolomerates. *preprint*.

Panjer, H. H. and Willmot, G. E. (1992). *Insurance Risk Models*. Society of Actuaries.

Perold, A. F. (2001). Capital allocation in financial firms. *HBS Competition and Strategy Working Paper Series*, 98-072.

Phillips, R. D., Cummins, J. D., and Allen, F. (1998). Financial pricing of insurance in the multiple-line insurance company. *The Journal of Risk and Insurance*, 65(4):597–636.

Ravishanker, N. and Dey, D. K. (2002). *A First Course in Linear Model Theory*. Chapman & Hall/CRC.

Sato, K.-I. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press.

Sherris, M. (2006). Solvency, capital allocation and fair rate of return in insurance. *Journal of Risk and Insurance*, 73(1):71–96.

Stroock, D. W. (1993). *Probability Theory, An Analytic View.* Cambridge University Press.

Stroock, D. W. (2003). *Markov Processes from K. Itô's Perspective.* Annals of Mathematics Studies. Princeton University Press.

Tasche, D. (2000). Risk contributions and performance measurement. *preprint.*

Zanjani, G. (2002). Pricing and capital allocation in catastrophe insurance. *Journal of Financial Economics*, 65:283–305.

Biography of the Author

Stephen Mildenhall is an Executive Vice President at Aon Re Services in Chicago Illinois. Aon Re Services is the technical arm of Aon Re, the world's largest reinsurance intermediary. At Aon Re, Stephen heads a team of actuaries providing a variety of actuarial services for clients, including reinsurance pricing and structure evaluation, dynamic financial analysis, capital allocation and predictive modeling. Prior to joining Aon Re, Stephen worked at Kemper Insurance and CNA Insurance. He has an undergraduate degree in Mathematics from the University of Warwick, England, and a Masters and Doctorate in Mathematics from the University of Chicago. Stephen is a Fellow of the Casualty Actuarial Society, an Associate of the Society of Actuaries and a Member of the American Academy of Actuaries. His research interests include risk theory, capital determination, allocation and optimization, and applications of statistics to reserving and ratemaking problems.