

Actuarial Geometry

Stephen J. Mildenhall

St. John's University, Peter J. Tobin College of Business, 101 Astor Place, New York, NY 10003, USA; mildenh@stjohns.edu; Tel.: +1-312-961-8781

Academic Editor: Albert Cohen

Received: 12 April 2017; Accepted: 2 June 2017; Published: 16 June 2017

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 US statutory accounting data, and identifies how parameter risk scales with volume and increases with time. Finally, it shows the data suggest a surprising result regarding the form of insurance parameter risk.

Keywords: capital; capital allocation; capital determination; diversification; homogeneous; insurance; insurance pricing; Lévy process; parameter risk; risk measure; risk theory

1. Introduction

Geometry is the study of shape and change in shape. Actuarial Geometry¹ studies the shape and evolution of shape of actuarial variables, in particular the distribution of aggregate losses, as portfolio volume and composition changes. It also studies the shape and evolution paths of variables in the space of all risks. Actuarial variables are curved across both a volumetric dimension as well as a temporal dimension. Volume here refers to expected losses per year, x , and temporal to the duration, t , for which a given volume of insurance is written. Total expected losses are xt —just as distance = speed $\times$ time. Asset variables are determined by a curved temporal return distribution but are flat in the volumetric (position size) dimension. Risk, and hence economic quantities like capital, are intimately connected to the shape of the distribution of losses, and so actuarial geometry is inextricably linked to capital determination and allocation.

Actuarial geometry is especially important today because risk and probability theory, finance, and actuarial science are converging after prolonged development along separate tracks. There is now general agreement that idiosyncratic insurance risk matters for pricing, and as a result we need to appropriately understand, model, and reflect the volumetric and temporal diversification of insurance risk. These are the central topics of the paper.

The paper makes two research contributions both linked to the use of Lévy processes in actuarial science. The first contribution is theoretical. It is to explain precisely how insurance losses diversify as

¹ Actuarial Geometry was originally presented to the 2006 Risk Theory Seminar in Richmond, Virginia, Mildenhall (2006). This version is largely based on the original, with some corrections and clarifications, as well as more examples to illustrate the theory. Since 2006 the methodology it described has been successfully applied to a very wide variety of global insurance data in Aon Benfield's annual Insurance Risk Study, ABI (2007, 2010, 2012, 2013, 2014, 2015), now in its eleventh edition. The findings have remained overwhelmingly consistent. Academically, the importance of the derivative and the gradient allocation method has been re-confirmed in numerous papers since 2006. Applications of Lévy processes to actuarial science and finance have also greatly proliferated. However, the new literature has not touched on the clarification between "direction" in the space of asset return variables and in the space of actuarial variables presented here.

volume increases and to compute the impact of this diversification compared to an asset portfolio model where risk is independent of position size. In particular we show that even when insurance losses and an asset portfolio have the same distribution of outcomes for a particular volume, the agreement is that of two lines crossing at an angle. It is not a first order tangency and so any risk allocation involving derivatives—which almost all do—will produce different results. The picture we make precise is shown in Figure 1. In the figure k is the distribution of values of an asset portfolio with initial value x , modeled as $k(x) = xX$ for a fixed return variable X . The map m represents aggregate losses from an insurance portfolio with expected losses x . Even though $X = m(1) = k(1)$ the tangent vector, $\dot{m}(1)$, to the embedding m at $x = 1$ is not the same as the tangent vector $\dot{k}(1)$. We have drawn m as a straight line because it will naturally capture the idea of “growth in the direction X ”. The full rationale behind Figure 1 is described in Section 6.

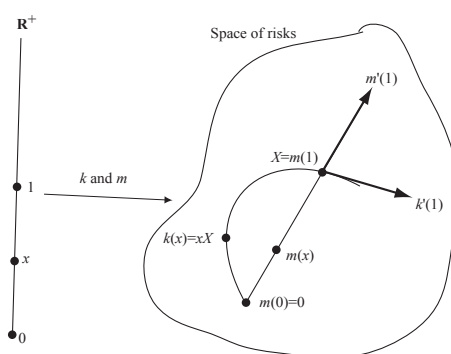


Figure 1. Lévy process and homogeneous embeddings of $\mathbb{R}_+$ into the space of risks, $\mathbf{L}$. The Lévy process embedding corresponds to the straight line m and the asset embedding to the curved line k .

The second contribution is empirical. It uses US statutory accounting data to determine a Lévy process based model for insurance losses by line of business that reflects their observed volumetric and temporal properties. The analysis compares four potential models and determines that only one is consistent with the data. The analysis produces specific line of business measures of non-diversifiable parameter risk that vary substantially but that have been consistent over time. It also provides an explicit form for the distribution of parameter risk, even though parameter risk cannot be directly observed. Most papers on risk measures take the actual distribution of losses as given. And much work done by companies to quantify risk is regarded as proprietary and is not published. The explicit quantification we provide should therefore be useful as a benchmark for both academics and practicing actuaries.

The remainder of the paper is organized as follows.

Section 2 describes the how actuaries and academics came to agree, over the last century, that idiosyncratic insurance risk matters for pricing. This agreement provides an important motivation for our theoretical and empirical work.

Section 3 defines a risk measure and explains how the allocation problem naturally leads to the derivative and gradient of a risk measure.

Section 4 presents two motivating examples that it is instructive to keep in mind through the rest of the paper, and which also illustrate Figure 1.

Section 5 defines Lévy processes and gives some basic examples. It then defines four Lévy process-based loss models that will be used as candidate models for aggregate losses, as well as an alternative asset-based model, and it establishes some of their basic properties.

Section 6 is the technical heart of the paper. It investigates the definition of derivative for a real function and considers how it could be defined on more general spaces, such as the space of random

variables. It explains how Lévy processes can be used to define “direction” and how the infinitesimal generator of a Lévy process relates to derivatives. This allows us to pin-point the difference between the derivatives of an insurance process and of an asset process.

Section 7 contains all the empirical results in the paper. It shows how we can effectively quantify parameter risk, even though it cannot be observed directly. It then determines the amount and shape of parameter risk across many lines of business. Finally, it addresses the differences between temporal and volumetric growth.

The paper covers topics from a variety of viewpoints befitting an article in this special edition celebrating the connections between actuarial science and mathematical finance. As a result it is quite long. Readers more interested in the theoretical findings can focus on Sections 4, 5.1, 5.2 and 6. Readers more interested in the empirical analysis can focus on Sections 5.1, 5.2 and 7.

2. Why Idiosyncratic Insurance Risk Matters

In its early years property-casualty actuarial science in the US largely ignored risk theory in rate making because of the dominance of bureau-based rates. Property rates were made to include a 5% profit provision and a 1% contingency provision; they were priced to a 94% combined ratio Magrath (1958). Lange (1966) describes a 5% provision for underwriting and contingencies as “constant for all liability insurance lines in most states”. 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, severity driven events.

Regulators and actuaries started to consider improvements to these long-standing conventions in the late 1960s. 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. Interest in option theoretic models was motivated in part by the difficulty of computing appropriate β s. These papers applied powerful results from option pricing theory using a geometric Brownian motion to model losses, possibly with a jump component. 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 explain the observed fact that insurance companies charged for specific risk. A series of papers, beginning in the early 1990s, developed a theoretical explanation of this based around agency, taxation and regulatory costs of capital, certainty in capital budgeting, costly external capital for opaque intermediaries, contracting under asymmetric information, and adverse selection, see Cummins (2000); Froot and O’Connell (2008); Froot and Stein (1998); Froot et al. (1993); Froot (2007); Merton and Perold (2001); Perold (2001); Zanjani (2002).

At the same time banking regulation led to the development of robust risk measures and an axiomatic theory of risk measures, including the idea of a coherent measure of risk Artzner et al. (1999). Risk measures are sensitive to the particulars of idiosyncratic firm risk, unlike the CAPM-based pricing methods which are only concerned with systemic risks.

The next 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 a complete-market 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 also worked in a complete market setting and 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 continued by Tasche (1999), Denault (2001) and Fischer (2003). Sherris (2006) takes the view that, in a complete market setting, only the default put has a canonical allocation and that there is no natural allocation of the remaining capital—a view echoed by Gründl and Schmeiser (2007). Kalkbrener (2005) and Delbaen (2000a) used directional derivatives to clarify the relationship between risk measures and allocations.

Concepts from banking regulation, including an own risk solvency assessment, have been adopted by insurance regulators and have led to increased academic interest in technical aspects of risk measurement, capital allocation and risk based pricing. A focus on catastrophe reinsurance pricing following the US hurricanes of 2004 and 2005 and the development of a robust capital market alternative to traditional reinsurance has also motivated research. As a result there is now a very rich literature around this nexus, including the following.

- Technical and axiomatic characterization of risk measures: (Dhaene et al. 2003; Furman and Zitikis 2008; Laeven and Stadje 2013).
- Capital allocation and its relationship with risk measurement: (Dhaene et al. 2003; Venter et al. 2006; Bodoff 2009; Buch and Dorfleitner 2008; Dhaene et al. 2012; Erel et al. 2015; Furman and Zitikis 2008; Powers 2007; Tsanakas 2009).
- The connection between purpose and method in capital allocation: (Dhaene et al. 2008; Zanjani 2010; Bauer and Zanjani 2013b; Goovaerts et al. 2010).
- Questioning the need for capital allocation in pricing: Gründl and Schmeiser 2007.

Recent summaries include Venter (2009) and Bauer and Zanjani (2013a).

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. For example, in a geometric Brownian motion model losses at time 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. Mildenhall (2004) and Meyers (2005b) show volumetric homogeneity is not an appropriate assumption. This paper provides further evidence and uses insurance regulatory data to explore more appropriate models.

3. Risk Measures, Risk Allocation and the Ubiquitous Gradient

3.1. Definition and Examples of Risk Measures

A risk measure, ρ , is a real valued function defined on a space of risks $\mathbf{L} = L^0(\Omega, \mathcal{F}, \mathbb{P})$. Here 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 $X : \Omega \rightarrow \mathbb{R}$, 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 . [Kalkbrener \(2005\)](#) works on L^0 . Risk measures are a large and important topic, but their details are not central this paper. For more details see [Föllmer and Schied \(2011\)](#) and [Dhaene et al. \(2006\)](#).

Given a risk $X \in \mathbf{L}$, $\rho(X)$ is the amount of capital required to support the risk. Examples of risk measures include value at risk at a percentile α (the inverse of the distribution of X , defined as $\inf\{x \mid \Pr(X \leq x) \geq \alpha\}$), tail value at risk (the average of the worst $1 - \alpha$ outcomes), and standard deviation $\rho(X) = \alpha \text{SD}(X)$.

3.2. Allocation and the Gradient

At the firm level, total risk X can be broken down into a sum of parts X_i corresponding to different lines of business. Since it is costly for insurers to hold capital [Froot and O'Connell \(2008\)](#) 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 \(1999\)](#), [Myers and Read \(2001\)](#), [Denault \(2001\)](#), [Meyers \(2005b\)](#) and [Fischer \(2003\)](#).

The need to understand marginal capital leads us to consider

$$\frac{\partial \rho}{\partial X_i} \quad (1)$$

which, in a sense to be made precise, represents the change in ρ as a result of a change in the volume of line i , or more generally the gradient vector of ρ representing the change across all lines. Much of this paper is an examination of exactly what this equation means.

[Tasche \(1999\)](#) 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. Tasche's framework is unequivocally financial. He 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

$$X(x) = X(x_1, \dots, x_n) = \sum_{i=1}^n x_i X_i. \quad (2)$$

A risk measure on $\mathbf{L}$ induces a function $\rho : \mathbb{R}^n \rightarrow \mathbb{R}$, $\rho(x_1, \dots, x_n) \mapsto \rho(\sum_i x_i X_i)$. Rather than being defined on a space of random variables, the induced ρ is defined on (a subset of) Euclidean space $\mathbb{R}^n$ using the correspondence between x and a portfolio. In this context $\partial \rho / \partial X_i$ is simply the usual limit

$$\frac{\partial \rho}{\partial x_i} = \lim_{\epsilon \rightarrow 0} \frac{\rho(x_1, \dots, x_i + \epsilon, \dots, x_n) - \rho(x_1, \dots, x_n)}{\epsilon}. \quad (3)$$

Equation (3) 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 be 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 i th coordinate unambiguously corresponds to an increase “in the direction” of the i th asset. This follows directly from the definition in Equation (2) and is unquestionably correct in a financial context.

Numerous papers work in an asset/volume model framework, either because they are working with assets or as a simplification of the real insurance world, for example (Myers and Read 2001; Panjer 2001; Erel et al. 2015; Fischer 2003). The resulting risk process homogeneity is essential to all Euler-based “adds-up” results: in fact the two are equivalent for homogeneous risk measures (Mildenhall 2004; Tasche 2004). However, it is important to realize that risk can be measured appropriately with a homogeneous risk measure, that is one satisfying $\rho(\lambda X) = \lambda \rho(X)$, even if the risk process itself is not homogeneous, that is $X(\lambda x) \neq \lambda X(x)$. The compound Poisson process and Brownian motion are examples of non-homogeneous processes.

In order to consider alternatives to the asset/return framework, we now discuss the meaning of the differential and examine other possible definitions. The differential represents the best linear approximation to a function at a particular point 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., a tangent vector at x , to a direction at $f(x)$. Under appropriate assumptions, the differential of f at x in direction $\mathbf{v}$, $D_x f(\mathbf{v})$, is defined by the property

$$\lim_{\mathbf{v} \rightarrow 0} \frac{\|f(\mathbf{x} + \mathbf{v}) - f(\mathbf{x}) - D_x f(\mathbf{v})\|}{\|\mathbf{v}\|} = 0, \quad (4)$$

see Abraham et al. (1988) or Borwein and Vanderwerff (2010). The vector $\mathbf{v}$ is allowed to tend to 0 from any direction, and Equation (4) 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 0. These include the Gâteaux differential, where $\mathbf{v} = t\mathbf{w}$ with $t \in \mathbb{R}$, $t \rightarrow 0$, 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 $\mathbf{w}' \rightarrow \mathbf{w}$. 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. The Gâteaux differential need not be linear in its direction argument.

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

$$\frac{\partial \rho}{\partial X} = D\rho_Y(X) = \lim_{\epsilon \rightarrow 0} \frac{\rho(Y + \epsilon X) - \rho(Y)}{\epsilon}. \quad (5)$$

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

We have shown that notions of differentiability are central to capital allocation. The next section will present two archetypal examples and that show the asset/return and insurance notions of growth do not agree, setting up the need for a better understanding of “direction” for actuarial random variables. We will see that Lévy processes provide that understanding.

4. Two Motivating Examples

This section presents two examples illustrating the difference between an asset/return model and a realistic insurance growth model.

Let $X(u)$ be a Poisson random variable with mean u . Consider two functions, $k(u) = uX(1)$ and $m(u) = X(u)$. The function k defines a random variable with mean u and standard deviation u . The function m also defines a random variable with has mean u , but it has standard deviation $u^{1/2}$. The variable k defines a homogeneous family, that is $k(\lambda u) = \lambda k(u)$, and correctly models the returns from a portfolio of size u in an asset with an (unlikely) Poisson(1) asset return distribution. The variable m is more realistic for a growing portfolio of insurance risks with expected annual claim count u .

If we measure risk using the standard deviation risk measure $\rho(X) = \text{SD}(X)$, this example shows that although $k(1) = m(1) = \text{Poisson}(1)$ have the same distribution the marginal risk for k is $\partial\rho(k(u))/\partial u = 1$ whereas the marginal risk for m is $\partial\rho(m(u))/\partial u = 1/(2u^{1/2})$. For m risk decreases as volume increases owing to portfolio effects whereas for k there is no diversification.

Next we present a more realistic example, due to Meyers (2005a), 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 increase in business philosophy discussed in Section 3.2. This example has also been re-visited recently by Boonen et al. (2017).

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

$$X_i(x_i) = S_{i,1} + \dots + S_{i,N_i(x_i)} \quad (6)$$

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 $E(S_i) = 1$, so that $E(X_i(x_i)) = E(N_i(x_i))E(S_i) = x_i$. Since $t = 1$ the model only considers volumetric diversification and not temporal diversification.

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

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

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

Kalkbrener's axiomatic capital is computed using the Gâteaux directional derivative. 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

$$\begin{aligned} \frac{\partial \rho}{\partial X_1} &= \lim_{\epsilon \rightarrow 0} \frac{\rho(X_1(x_1) + X_2(x_2) + \epsilon X_1(x_1)) - \rho(X_1(x_1) + X_2(x_2))}{\epsilon} \\ &= \lim_{\epsilon \rightarrow 0} \frac{\sqrt{(1 + \epsilon)^2 \rho_1(x_1)^2 + \rho_2(x_2)^2} - \sqrt{\rho_1(x_1)^2 + \rho_2(x_2)^2}}{\epsilon} \\ &= \frac{\rho_1(x_1)^2}{\rho(X_1(x_1) + X_2(x_2))} \\ &= \frac{c_1 x_1^2 + g_1 x_1}{\rho(X_1(x_1) + X_2(x_2))}. \end{aligned} \quad (7)$$

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 $k : \mathbb{R}^n \rightarrow \mathbf{L}$, by $(y_1, \dots, y_n) \mapsto \sum_i y_i X_i(x_i)$. Let ρ_k be the composition of k and ρ ,

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

Then

$$\left. \frac{\partial \rho_k}{\partial y_1} \right|_{(1,1)} = \frac{c_1 x_1^2 + g_1 x_1}{\rho(X_1(x_1) + X_2(x_2))} \quad (8)$$

agreeing with Equation (7). 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 $m : \mathbb{R}^n \rightarrow \mathbf{L}$, by $(x_1, \dots, x_n) \mapsto \sum_i X_i(x_i)$. The map m satisfies $m(x + y) = m(x) + m(y)$ 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 Section 3.2, 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 $\rho_m : \mathbb{R}^n \rightarrow \mathbb{R}$ is the composition of m and ρ ,

$$\rho_m(x_1, x_2) = \rho(m(x_1, x_2)) = \sqrt{\sum_i c_i x_i^2 + g_i x_i}. \quad (9)$$

Since $\rho_m : \mathbb{R}^2 \rightarrow \mathbb{R}$ a real function, we can compute its partial derivative using standard calculus:

$$\frac{\partial \rho_m}{\partial x_1} = \frac{2c_1 x_1 + g_1}{2\rho(X_1(x_1) + X_2(x_2))}. \quad (10)$$

There are two important conclusion: (1) the partial derivatives of ρ_m and ρ_k (which is also the Gâteaux derivative of ρ) give different answers, Equations (7) and (10), and (2) the implied allocations

$$\frac{c_1 x_1^2 + g_1 x_1}{\rho(X_1(x_1) + X_2(x_2))} \quad \text{and} \quad \frac{c_1 x_1^2 + g_1 x_1 / 2}{\rho(X_1(x_1) + X_2(x_2))} \quad (11)$$

are also different. This is Meyers' example.

5. Lévy process Models of Insurance Losses

We define Lévy processes and discuss some of their important properties. We then introduce four models of insurance risk which we will analyze in the rest of the paper.

5.1. Definition and Basic Properties of Lévy processes

Lévy processes are fundamental to actuarial science, but they are rarely discussed explicitly in basic actuarial text books. For example, there is no explicit mention of Lévy processes in Bowers et al. (1986); Beard et al. (1969); Daykin et al. (1994); Klugman et al. (1998); 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), and 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 $0 \leq t_1 \leq \dots \leq t_{n+1}$ 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 $s \geq 0$

$$\lim_{t \rightarrow s} \Pr(|X(t) - X(s)| > a) = 0. \quad (12)$$

Based on the definition it is clear that the sum of two Lévy processes is a Lévy process. Lévy processes are in one-to-one correspondence with the set of infinitely divisible distributions, where 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 $Y_1 + \dots + Y_n$. If $X(t)$ is a Lévy process then $X(1)$ is infinitely divisible since $X(1) = X(1/n) + (X(2/n) - X(1/n)) + \dots + (X(1) - X(n-1/n))$, and conversely if X is infinitely divisible there is a Lévy process with $X(1) = X$. In an idealized world, insurance losses should follow an infinitely divisible distribution because annual losses are the sum of monthly, weekly, daily, or hourly losses. Bühlmann (1970) discusses infinitely divisible distributions and their relationship with compound Poisson processes. 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, see Sato (1999).

Example 1 (Trivial process). $X(t) = kt$ for a constant k is a trivial Lévy process.

Example 2 (Poisson process). The Poisson process $N(t)$ with intensity λ has

$$\Pr(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad (13)$$

for $n = 0, 1, \dots$ is a Lévy process.

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

$$X(t) = Z_1 + \dots + Z_{N(t)} \quad (14)$$

where $N(t)$ is a Poisson process with intensity λ . The compound Poisson processes is the fundamental building block of Lévy processes in the sense that any infinitely divisible distribution is the limit distribution of a sequence of compound Poisson distributions, see Sato (1999) Corollary 8.8

Example 4 (Brownian motion). Brownian motion is an example of a continuous Lévy process.

Example 5 (Operational time). 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 claim frequency constant. It allows seasonal and daily effects (rush hours, night-time lulls, etc.) without losing stationary increments. Operational time is an increasing function $\tau : [0, \infty) \rightarrow [0, \infty)$ chosen so that $X(\tau(t))$ becomes a Lévy process.

Example 6 (Subordination). 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.

The characteristic function of a random variable X with distribution μ is defined as $\phi(z) = E(e^{izX}) = \int e^{izx} \mu(dx)$ 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 $X(t) = Z_1 + \dots + Z_{N(t)}$ is

$$\phi(z) = E(e^{izX(t)}) = E(E(e^{izX(t)} | N(t))) \quad (15)$$

$$= E \exp \left(N(t) \log \int e^{izw} \nu(dw) \right) \quad (16)$$

$$= \exp \left(\lambda t \int (e^{izw} - 1) \nu(dw) \right) \quad (17)$$

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

We now quote an important result in the theory of Lévy processes that allows us to identify an infinitely divisible distribution, and hence a Lévy process, with a measure ν on $\mathbb{R}$, and two constants $\sigma > 0$ and γ .

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

$$\exp \left(-\sigma^2 z^2/2 + \int_{\mathbb{R}} (e^{izw} - 1 - izw 1_{\{|w| \leq 1\}}(w)) \nu(dw) + i\gamma z \right) \quad (18)$$

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

See Breiman (1992) or Sato (1999) a proof. In Equation (18), σ is the standard deviation of a Brownian motion component, and ν is called the Lévy measure. The indicator function $1_{\{|w| \leq 1\}}$ is present for technical convergence reasons and is only needed when there are a very large number of very small jumps. If $\int_{-1}^1 \min(|w|, 1) \nu(dw) < \infty$ it can be omitted and the resulting γ can 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

$$\nu(\mathbb{R}) \int_{\mathbb{R}} (e^{izw} - 1) \tilde{\nu}(dw) \quad (19)$$

where $\tilde{\nu} = \nu/\nu(\mathbb{R})$ is a distribution. Comparing with Equation (17) 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 $(t\sigma, t\nu, t\gamma)$.

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

The insurance analog of an asset return portfolio basis becomes a set of Lévy processes representing losses in each line of business and “line” becomes synonymous with the Lévy measure that describes the frequency and severity of the jumps, i.e., of the losses. Unless the Lévy process has an infinite number of small jumps the Lévy measure can be separated into a frequency component and a severity component. Patrik et al. (1999) describes modeling with Lévy measures, which the authors call a loss frequency curve.

5.2. Four Temporal and Volumetric Insurance Loss Models

We now define four models describing 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 that is insured for a time period t years. Thus $A(x, 1)$ is the distribution of annual losses. The central question of the paper is to describe appropriate models for $A(x, t)$ as x and t vary. A Lévy process $X(t)$ provides the appropriate 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 given insureds 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 $E(Z(t)) = t$. 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. Z could be a deterministic drift or it could combine a deterministic drift with a stochastic component.
- 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 $E(A(x, t)) = xt$. 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

$$v(x, t) := \frac{\sqrt{\text{Var}(A(x, t))}}{xt}. \quad (20)$$

Since insurance is based on the notion of diversification, the behavior of $v(x, t)$ as $x \rightarrow \infty$ and as $t \rightarrow \infty$ 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, $v(x, t) \propto (xt)^{-1/2} \rightarrow 0$ as t or $x \rightarrow \infty$ or as $t \rightarrow \infty$.

Definition 2. For $v(x, t)$ in Equation (20):

1. If $v(x, t) \rightarrow 0$ as $t \rightarrow \infty$ we will call $A(x, t)$ temporally diversifying.
2. If $v(x, t) \rightarrow 0$ as $x \rightarrow \infty$ we will call $A(x, t)$ volumetrically diversifying.
3. 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.

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

$$\text{AM1. } A(x, t) = xX(t)$$

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

$$v(x, t) = \frac{\sqrt{\text{Var}(X(t))}}{t} \quad (21)$$

is independent of x .

Next we consider some properties of the models IM1-4 and AM1. In all cases severity is normalized so that $E(A(x, t)) = xt$. Define σ and τ so that $\text{Var}(X(t)) = \sigma^2 t$ and $\text{Var}(Z(t)) = \tau^2 t$.

Practical underwritten loss distributions will have a variance or will have limits applied so the distribution of insured losses has a variance, so this is not a significant restriction.

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. Example 6 shows that IM2 is a Lévy process.

Table 1 lays out the variance and coefficient of variation v of these five models. It also shows whether each model is volumetrically (resp. temporally) diversifying, that is whether $v(x, t) \rightarrow 0$ as $x \rightarrow \infty$ (resp. $t \rightarrow \infty$). The calculations follow easily by conditioning. For example

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

Table 1. Variance of IM1-4 and AM.

Model	Variance	$v(x, t)$	Diversifying	
			$x \rightarrow \infty$	$t \rightarrow \infty$
IM1: $X(xt)$	$\sigma^2 xt$	$\frac{\sigma}{\sqrt{xt}}$	Yes	Yes
IM2: $X(xZ(t))$	$xt(\sigma^2 + x\tau^2)$	$\sqrt{\frac{\sigma^2}{xt} + \frac{\tau^2}{t}}$	No	Yes
IM3: $X(xCt)$	$xt(\sigma^2 + cxt)$	$\sqrt{\frac{\sigma^2}{xt} + c}$	No	No
IM4: $X(xCZ(t))$	$x^2 t^2 \left(\frac{(c+1)\tau^2}{t} + c \right) + \sigma^2 xt$	$\sqrt{\frac{\sigma^2}{xt} + \frac{\tau'^2}{t} + c}$	No	No
AM1: $xX(t)$	$x^2 \sigma^2 t$	$\sigma / \sqrt{t}$	Const.	Yes

In IM4, $\tau' = (1 + c)\tau$.

The characteristics of each model will be tested against regulatory insurance data in Section 7.

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 (2005b).

6. Defining the Derivative of a Risk Measure and Directions in the Space of Risks

This section is the technical heart of the paper. It investigates the definition of derivative for a real function and considers how it could be defined on more general spaces, such as the space of random variables. It explains how Lévy processes can be used to define "direction" and how the infinitesimal generator of a Lévy process relates to derivatives. This allows us to pin-point the difference between the derivatives of an insurance process and of an asset process.

6.1. Defining the Derivative

When $\rho : \mathbb{R}^n \rightarrow \mathbb{R}$ the meaning of $\partial \rho / \partial X_i$ is clear. However we want to consider $\rho : \mathbf{L} \rightarrow \mathbb{R}$ where $\mathbf{L}$ is the more complicated space of random variables. We need to define the derivative mapping $D\rho_X$

as a real-valued linear map on tangent vectors or “directions” at $X \in \mathbf{L}$. Meyers’ example shows the asset/return model and an insurance growth model correspond to different directions.

A direction in $\mathbf{L}$ can be identified with the derivative of a coordinate path $x : U \rightarrow \mathbf{L}$ where $U \subset \mathbb{R}$. Composing ρ and x results in a real valued function of a real variable $\rho_x := \rho \circ x : U \rightarrow \mathbb{R}$, $u \mapsto \rho(x(u))$, so standard calculus defines $d\rho_x/du$. The derivative of ρ at $x(u)$ in the direction defined by the derivative $\dot{x}(u)$ of $x(u)$ is given by

$$D\rho_{x(u)}(\dot{x}(u)) := \frac{d\rho_x}{du} \quad (22)$$

The surprise of Equation (22) is that the two complex objects on the left combine to the single, well-understood object on the right. The exact definitions of the terms on the left will be discussed below.

Section 4 introduced two important coordinates. The first is $k : \mathbb{R} \rightarrow \mathbf{L}$, $k(u) = uX$ for some fixed random variable $X \in \mathbf{L}$. It is suitable for modeling assets: u represents position size and X represents the asset return. The second coordinate is $m : [0, \infty) \rightarrow \mathbf{L}$, $m(u) = X_u$, where X_u is a compound Poisson distribution with frequency mean $u\lambda$ and severity component Z . It is suitable for modeling aggregate losses from an insurance portfolio. (There is a third potential coordinate path $w(u) = B_u$ where B_u is a Brownian motion, but because it always takes positive and negative values it is of less interest for modeling losses.)

An issue with the asset coordinate in an insurance context is the correct interpretation of uX . For $0 < u \leq 1$, uX can be interpreted as a quota share of total losses, or as a coinsurance provision. However, uX for $u < 0$ or $u > 1$ 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 $E(X_1 + X_2) = 2E(X)$. Clearly $X + X$ has a different distribution to $X_1 + X_2$ unless X_1 and X_2 are perfectly correlated. The insurance coordinate has exactly this property: $m(2) = X_{2u}$ is the sum of two independent copies of X_u because of the additive property of the Poisson distribution.

To avoid misinterpreting uX it is safer to regard insurance risks as probability measures (distributions) μ on $\mathbb{R}$. The measure μ corresponds to a random variable X with distribution $\Pr(X \leq x) = \mu(-\infty, x]$. Now there is no natural way to interpret 2μ . Identify $\mathbf{L}$ with $M(\mathbb{R})$, the set of probability measures on $\mathbb{R}$. We can combine two elements of $M(\mathbb{R})$ using convolution: the distribution of the sum of the corresponding random variables. Since the distribution if $X + Y$ is the same as the distribution of $Y + X$ order of convolution does not matter. Now $2X$ in our insurance interpretation, $X_1 + X_2$, corresponds to $\mu \star \mu := \mu^{*2}$, where $\star$ represents convolution, and we are not led astray.

We still have to define “directions” in $\mathbf{L}$ and $M(\mathbb{R})$. Directions should correspond to the derivatives of curves. The simplest curves are straight lines. A straight line through the origin is called a ray. Table 2 shows several possible characterizations of a ray $\mathbb{R}^n$ each of which uses a different aspect of the rich mathematical structure of $\mathbb{R}^n$, and which could be used as characterizations in $\mathbf{L}$.

Table 2. 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.

The first two use properties of $\mathbb{R}^n$ that require putting a differential structure on $\mathbf{L}$, which is very complicated. The third corresponds to the asset volume/return model and uses the identification of the set of possible portfolios with the $\mathbb{R}$ vector space $\mathbb{R}^n$. This leaves the fourth approach: a ray is

characterized by the simple relationship $\alpha(s+t) = \alpha(s) + \alpha(t)$. This definition only requires the ability to add for the range space, which we have on $\mathbf{L}$. It is the definition adopted in Stroock (2003).

Therefore rays in $\mathbf{L}$ should correspond to families of random variables satisfying $X_s + X_t = X_{s+t}$ (or, equivalently, in $M(\mathbb{R})$ to families of measures μ_s satisfying $\mu_s \star \mu_t = \mu_{s+t}$, i.e., to Lévy processes. Since $X_0 = X_{0+0} = X_0 + X_0$ a ray must start at 0, the random variable taking the value 0 with probability 1. Straight lines correspond to translations of rays: a straight line passing through the point $Y \in \mathbf{L}$ is a family $Y + X_t$ where X_t is a ray (resp. passing through $\nu \in M(\mathbb{R})$ is $\nu \star \mu_t$ where μ_t is a ray.) Directions in $\mathbf{L}$ are determined by rays. By providing a basis of directions in $\mathbf{L}$, Lévy processes provide the insurance analog of individual asset return variables.

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 a point in $\mathbb{R}^n$. 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 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 family of local charts, where a chart is a differentiable bijection from a subset of $\mathbb{R}^2$ to a neighborhood of each point. 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. Charts move questions of tangency and direction back to functions on $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ where they are well understood. This is called the coordinate approach.

Another way of defining the tangent bundle is to use curves, or coordinate paths, to define tangent vectors: a direction becomes the derivative 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 the next section we will apply this approach to $\mathbf{L}$. A good general reference on the construction of the tangent bundle is Abraham et al. (1988).

Figure 2 is an illustrative schematic. The sphere S is used as a proxy for $\mathbf{L}$, an object with more complex geometry than flat Euclidean space. The two paths m and k are shown as the red and blue lines, passing through the same point (distribution) x on the sphere at $t = 1$. The red line is part of a great circle geodesic—the analog of a straight line on a sphere—whereas the blue line is not. Above x is the tangent plane (isomorphic to $\mathbb{R}^2$) to the sphere at x , TS_x ; π is the projection from the tangent bundle TS to S . The derivative of ρ at x is a linear map $D\rho : TS_x \rightarrow T\mathbb{R}$. For Euclidean spaces we can identify the tangent bundle with the space so $T\mathbb{R} = \mathbb{R}$. Although $k(1) = m(1) = x$ they have different derivatives (define different vectors in TS_x), $\dot{m} \neq \dot{k}$ at $t = 1$.

The derivative of a risk measure ρ , $\partial\rho/\partial X$, is the evaluation of the linear differential $D\rho$ on a tangent vector in the direction X . Meyer's embedding m corresponds to $\partial(\rho \circ m)/\partial t|_{t=1} = D\rho_X(\dot{m}(1))$ whereas Kalkbrenner's corresponds to $\partial(\rho \circ k)/\partial t|_{t=1} = D\rho_X(\dot{k}(1))$. As demonstrated in Section 4 these derivatives are not the same—just as the schematic leads us to expect—because the direction $\dot{m}(1)$ is not the same as the direction $\dot{k}(1)$.

The difference between $\dot{k}(1)$ and $\dot{m}(1)$ is a measure of the diversification benefit given by m compared to k . The embedding 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. We will describe the analog of TS_x next.

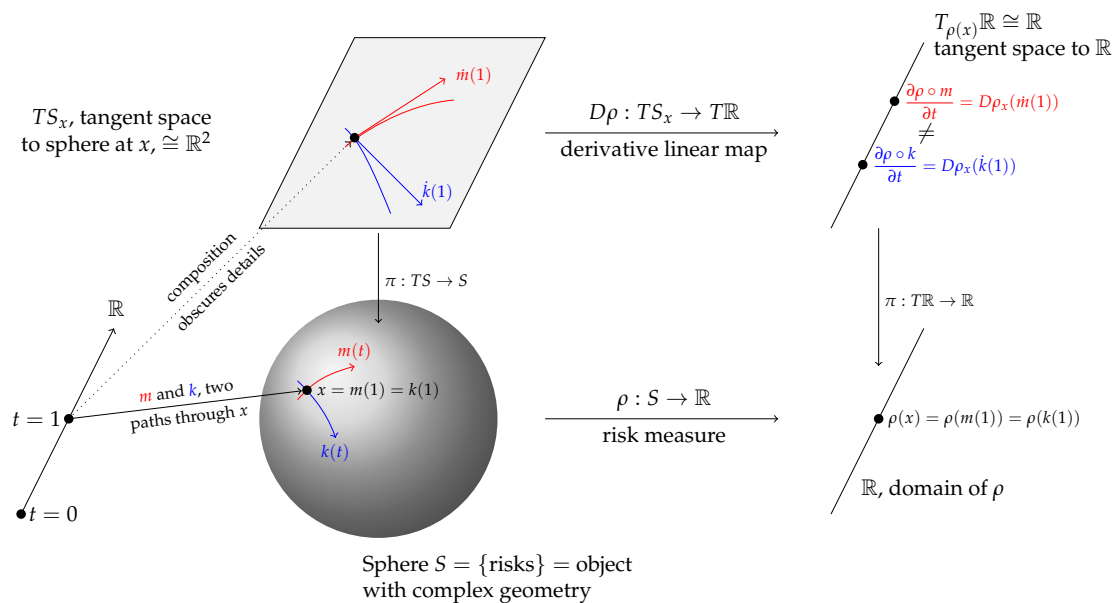


Figure 2. Schematic of the Kalkbrener-Meyers example, using the sphere to illustrate the more complex space $\mathbf{L}$.

6.2. Directions in the Space of Actuarial Random Variables

We now show how Lévy processes provide a description of “directions” in the space $\mathbf{L}$. The analysis combines three threads:

1. The notion that directions, or tangent vectors, live in a separate space called the tangent bundle.
2. The identification of tangent vectors as derivatives of curves.
3. The idea that Lévy processes, characterized by the additive relation $X(s+t) = X(s) + X(t)$, provide the appropriate analog of rays to use as a basis for insurance risks.

The program is to compute the derivative of the curve $t \mapsto X(t) \in \mathbf{L}$ defined by a Lévy process family of random variables (or $t \mapsto \mu_t \in M(\mathbb{R})$ defined by an additive family of probability distributions on $\mathbb{R}$). The ideas presented here are part of a general theory of Markov processes. The presentation follows the beautiful book by [Stroock \(2003\)](#). We begin by describing a finite sample space version of $\mathbf{L}$ which illustrates the difficulties involved in regarding it as a differentiable manifold.

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 $n-1$ dimensional simplex $\Delta_{n-1} = \{(x_0, \dots, x_{n-1}) \mid \sum_i x_i = 1\} \subset \mathbb{R}^n$. Δ_n inherits a differentiable structure from $\mathbb{R}^{n+1}$ and we already know how to think about directions and tangent vectors in Euclidean space. However, even thinking about $\Delta_2 \subset \mathbb{R}^3$ shows M is not an easy space to work with. Δ_2 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_2 . 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.

Let $\delta_x \in M(\mathbb{R})$ be the measure giving probability 1 to $x \in \mathbb{R}$. We will describe the space of tangent vectors to $M(\mathbb{R})$ at δ_0 . By definition, all Lévy processes $X(t)$ have distribution δ_0 at $t = 0$. Measures $\mu_t \in M(\mathbb{R})$ are defined by their action on functions f on $\mathbb{R}$.

Let $\langle f, \mu \rangle = \int_{\mathbb{R}} f(x) \mu(dx) = E(f(X))$, where X has distribution μ . In view of the fundamental theorem of calculus, the derivative $\dot{\mu}_t$ of μ_t should satisfy

$$\langle f, \mu_t \rangle - \langle f, \mu_0 \rangle = \int_0^t \dot{\mu}_\tau f d\tau, \quad (23)$$

with $\dot{\mu}_t$ a linear functional acting on f , i.e., $\dot{\mu}_t(f) \in \mathbb{R}$ and $f \mapsto \dot{\mu}_t(f)$ is linear in f . Converting Equation (23) to its differential form suggests that

$$\dot{\mu}_0 f = \lim_{t \downarrow 0} \frac{\langle f, \mu_t \rangle - \langle f, \mu_0 \rangle}{t} \quad (24)$$

$$= \lim_{t \downarrow 0} \frac{E(f(X(t))) - E(f(X(0)))}{t} \quad (25)$$

where $X(t)$ has distribution μ_t .

We now consider how Equation (25) 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. Then

$$\dot{\mu}_0 f = \lim_{t \downarrow 0} [E(f(0) + X(t)f'(0) + \frac{X(t)^2 f''(0)}{2} + o(t)) - f(0)]/t \quad (26)$$

$$= \lim_{t \downarrow 0} [\gamma t f'(0) + \frac{\sigma^2 t f''(0)}{2} + o(t)]/t \quad (27)$$

$$= \gamma f'(0) + \frac{\sigma^2 f''(0)}{2}, \quad (28)$$

because $E(B(t)) = 0$ and $E(B(t)^2) = t$ and so $E(X(t)) = E(\gamma t + \sigma B(t)) = \gamma t$ and $E(X(t)^2) = \gamma^2 t^2 + \sigma^2 t$. Thus $\dot{\mu}_0$ acts as a second order differential operator evaluated at $x = 0$ (because we assume $\mu_0 = \delta_0$):

$$\dot{\mu}_0 f = \gamma \frac{df}{dx}(0) + \frac{\sigma^2}{2} \frac{d^2 f}{dx^2}(0). \quad (29)$$

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 in the time interval $[0, t]$ there is a single jump with probability λt and no jump with probability $1 - \lambda t$. Conditioning on the occurrence of a jump, $E(f(X(t))) = (1 - \lambda t)f(0) + \lambda t E(f(J))$ and so

$$\dot{\mu}_0 f = \lim_{t \downarrow 0} \frac{E(f(X(t))) - E(f(X(0)))}{t} \quad (30)$$

$$= \lim_{t \downarrow 0} \frac{\lambda t (E(f(J)) - f(0))}{t} \quad (31)$$

$$= \lambda (E(f(J)) - f(0)) \quad (32)$$

$$= \int (f(y) - f(0)) \nu(dy) \quad (33)$$

This analysis side-steps some technicalities by assuming that $\nu(\mathbb{R}) < \infty$. For both the Brownian motion and the compound Poisson if we are interested in tangent vectors at $\mu_0 = \delta_x$ for $x \neq 0$ then we replace 0 with x because $\langle f, \mu_0 \rangle = E(f(x + X_0)) = f(x)$. Thus Equation (33) becomes

$$\dot{\mu}_0 f = \int (f(x + y) - f(x)) \nu(dy) \quad (34)$$

for example. Combining these two results makes the following theorem plausible.

Theorem 2 (Stroock (2003) Thm 2.1.11). *There is a one-to-one correspondence between Lévy triples and rays (continuous, additive maps) $t \in [0, \infty) \rightarrow \mu_t \in M(\mathbb{R})$. The Lévy triple (σ, ν, γ) corresponds to the infinitely divisible map $t \mapsto \mu_t^{(\sigma, \nu, \gamma)}$ given by the Lévy process with the same Lévy triple. The map is $t \mapsto \mu_t^{(\sigma, \nu, \gamma)}$ is differentiable and*

$$\dot{\mu}_t^{(\sigma, \nu, \gamma)} f = \dot{\mu}_t f = \langle L^{(\sigma, \nu, \gamma)} f, \mu_t \rangle. \quad (35)$$

where $L^{(\sigma, \nu, \gamma)}$ is a pseudo differential operator, see Applebaum (2004); Jacob (2001 2002 2005), given by

$$L^{(\sigma, \nu, \gamma)} f(x) = \gamma \frac{df}{dx} + \frac{1}{2} \sigma^2 \frac{d^2 f}{dx^2} + \int_{\mathbb{R}} f(y+x) - f(x) - \frac{df}{dx} \frac{y}{1+|y|^2} \nu(dy). \quad (36)$$

If $t \mapsto \mu_t \in \mathbf{L}$ is a differentiable curve and $\mu_0 = \delta_x$ for some $x \in \mathbb{R}$ then there exists a unique Lévy triple (σ, ν, γ) such that $\dot{\mu}_0$ is the linear operator acting on f by

$$\dot{\mu}_0 f = L^{(\sigma, \nu, \gamma)} f(x) = \langle L^{(\sigma, \nu, \gamma)} f, \mu_0 \rangle. \quad (37)$$

Thus $T_{\delta_x}(M(\mathbb{R}))$, the tangent space to $M(\mathbb{R})$ at δ_x , can be identified with the cone of linear functionals of the form $f \mapsto L^{(\sigma, \nu, \gamma)} f(x)$ where (σ, ν, γ) is a Lévy triple.

Just as in the Lévy-Khintchine theorem, the extra term in the integral is needed for technical convergence reasons when there is an infinite number of very small jumps. Note that $\dot{\mu}_0 f \in \mathbb{R}$ is a number, $L^{(\sigma, \nu, \gamma)} f$ is a function and its value at x , $L^{(\sigma, \nu, \gamma)} f(x) \in \mathbb{R}$ is a number. The connection between μ_t and x is $\mu_0 = \delta_x$ is the measure concentrated at $x \in \mathbb{R}$.

At this point we have described tangent vectors to $\mathbf{L} = M(\mathbb{R})$ at degenerate distributions δ_x . To properly illustrate Figures 1 and 2 we need a tangent vector at a more general μ . Again, following (Stroock 2003, sct. 2.11.4), define a tangent vector to $M(\mathbb{R})$ at a general μ to be a linear functional of the form $\Lambda_\mu f = \langle Lf, \mu \rangle$ where $L = L^{(\sigma(x), \nu(x), \gamma(x))}$ is a continuous family of operators L determined by x . We will restrict attention to simpler tangent vectors where $L = L^{(\sigma, \nu, \gamma)}$ does not vary with x . If μ_t is the Lévy process corresponding to the triple (σ, ν, γ) and f is bounded and has continuously bounded derivatives, then, by independent and stationary increments

$$\dot{\mu}_u f = \langle L^{(\sigma, \nu, \gamma)} f, \mu_t \rangle \quad (38)$$

$$= \lim_{s \downarrow 0} \mathbb{E} \left(\frac{f(X_{s+u}) - f(X_u)}{s} \right) \quad (39)$$

$$= \lim_{s \downarrow 0} \mathbb{E}_{X_u} \left(\frac{\mathbb{E}_{X_s}(f(X_s + X_u) - f(X_u))}{s} \right) \quad (40)$$

$$= \mathbb{E}_{X_u} \left(\lim_{s \downarrow 0} \frac{\mathbb{E}_{X_s}(f(X_s + X_u) - f(X_u))}{s} \right) \quad (41)$$

$$= \mathbb{E}_{X_u} \left(\langle L^{(\sigma, \nu, \gamma)} f, \mu_u \rangle \right) \quad (42)$$

by dominated convergence. The tangent vector is an average of the direction at all the locations that X_u can take.

6.3. Examples

We present a number of examples to illustrate the theory. Test functions f are usually required to be bounded and twice continuously and boundedly differentiable to ensure that all relevant integrals exist. However, we can apply the same formulas to unbounded differentiable functions for particular μ_t if we know relevant integrals converge. Below we will use $f(x) = x^2$ as an example, with distributions having a second moment.

Example 7 (Brownian motion). Let X_t be a standard Brownian motion, corresponding to Lévy triple $(\sigma, \nu, \gamma) = (1, 0, 0)$ and $f(x) = x^2$. The density of X_t is $g(x, t) = (2\pi t)^{-1/2} \exp(-x^2/2t)$ and let l_t be the associated measure. We can compute $\dot{l}_t f$ in three ways. First, using Stroock's theorem Equation (37)

$$\dot{l}_t f = \langle L^{(1,0,0)} f, l_t \rangle = \left\langle \frac{1}{2} \frac{\partial^2 f}{\partial x^2}, l_t \right\rangle = \langle 1, l_t \rangle = 1. \quad (43)$$

Second, using $\dot{l}_t f = d/dt \langle f, l_t \rangle$ and the fact that $\langle f, l_t \rangle = E(X_t^2) = t$ again gives $\dot{l}_t f = 1$. Thirdly, differentiating $d/dt \langle f, l_t \rangle$ through the integral gives

$$\frac{d}{dt} \langle f, l_t \rangle = \int_{\mathbb{R}} f(x) \frac{\partial g}{\partial t}(x, t) dx \quad (44)$$

$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} x^2 \left(\frac{x^2}{2t^2} - \frac{1}{2t} \right) dx \quad (45)$$

$$= \frac{E(X_t^4)}{2t^2} - \frac{E(X_t^2)}{2t} \quad (46)$$

$$= 1 \quad (47)$$

since $E(X_t^4) = 3t^2$ and $E(X_t^2) = t$.

Example 8 (Gamma process). Let X_t be a gamma process, (Sato 1999, p. 45), Barndorff-Nielsen (2000), meaning X_t has law $l_t = \Gamma(t\lambda, \alpha)$ with density $\alpha^{t\lambda} / \Gamma(t\lambda) x^{t\lambda-1} e^{-\alpha x}$ and Lévy measure $\nu(dx) = (\lambda/x) e^{-\alpha x} dx$, $\gamma = \sigma = 0$ on $x \geq 0$. We have

$$E(X_t) = \frac{\lambda t}{\alpha}, E(X_t^2) = \frac{\lambda t(\lambda t + 1)}{\alpha^2}, \text{ and } \text{Var}(X_t) = \frac{\lambda t}{\alpha^2}. \quad (48)$$

Notice that $\nu([0, \infty)) = \infty$ so this example is not a compound Poisson process because it has infinitely many small jumps. But it is a limit of compound Poisson processes. For $f(x) = x^2$, $\langle f, l_t \rangle = E(X_t^2) = \lambda t(\lambda t + 1)/\alpha^2$ so

$$\frac{d}{dt} \langle f, l_t \rangle = \frac{2\lambda^2 t + \lambda}{\alpha^2}. \quad (49)$$

On the other hand, using Equation (37) with $L^{(0,\nu,0)}$ gives

$$L^{(0,\nu,0)} f(x) = \int (f(x+y) - f(x)) \frac{\lambda}{y} e^{-\alpha y} dy \quad (50)$$

$$= \frac{2\lambda x}{\alpha} + \frac{\lambda}{\alpha^2}, \quad (51)$$

so $\dot{l}_t f = \langle L^{(0,\nu,0)} f, l_t \rangle = (2\lambda^2 t + \lambda)/\alpha^2$, agreeing with Equation (49).

Example 9 (Laplace process). Let X_t be a Laplace process with law l_t , (Sato 1999, p. 98), and (Kotz et al. 2001, p. 47). X_1 has density $\alpha \exp(-\alpha|x|)/2$ and Lévy measure $\nu(dx) = \exp(-\alpha|x|)/|x| dx$. X_t can be represented as the difference of two $\Gamma(t, \alpha)$ variables. $E(X_t) = 0$, $\text{Var}(X_t) = E(X_t^2) = \langle f, l_t \rangle = 2t/\alpha^2$ and hence

$$\frac{d}{dt} \langle f, l_t \rangle = \frac{2}{\alpha^2}. \quad (52)$$

On the other hand

$$L^{(0,\nu,0)}f(x) = \int (f(x+y) - f(x)) \frac{1}{|y|} e^{-\alpha|y|} dy \quad (53)$$

$$= \int (2xy + y^2) \frac{1}{|y|} e^{-\alpha|y|} dy \quad (54)$$

$$= 2 \int_0^\infty ye^{-\alpha y} dy = \frac{2}{\alpha^2} \quad (55)$$

as the first term in the middle equation is odd and hence zero.

6.4. Application to Insurance Risk Models IM1-4 and Asset Risk Model AM1

We now compute the difference between the directions implied by each of IM1-4 and AM1 to quantify the difference between $m(1)$ and $k(1)$ in Figures 1 and 2. In order to focus on realistic insurance loss models we will assume $\gamma = \sigma = 0$ and $\nu(\mathbb{R}) < \infty$. Assume the Lévy triple for the subordinator Z is $(0, \rho, 0)$. Also assume $E(C) = 1$, $\text{Var}(C) = c$, and that 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 2 gives the direction for IM1 as corresponding to the operator Equation (37) multiplied by x or t as appropriate. If we are interested in the temporal derivative then losses evolve according to the process $\tilde{X}_t = X_{xt}$, which has Levy triple $(0, x\nu, 0)$. Therefore, if $\mu_0 = \delta_z$, $z \in \mathbb{R}$, then the time direction is given by the operator

$$\dot{\mu}_0 f = \int f(z+y) - f(z) (x\nu)(dy) = x \int f(z+y) - f(z) \nu(dy). \quad (56)$$

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

$$\dot{\mu}_0 f = \rho(\mathbb{R}) E(f(z + X(xK)) - f(z)) \quad (57)$$

$$= \int_{(0,\infty)} \int_{(0,\infty)} f(z + xy) - f(z) \nu^{xk}(dy) \rho(dk) \quad (58)$$

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 the mixture

$$\nu'(B) = \int_{(0,\infty)} \nu^k(B) \rho(dk). \quad (59)$$

See (Sato 1999, chp. 6, Thm 30.1) for more details and for the case where X or Z includes a deterministic drift.

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 $E(C) \neq 1$ then the derivative would be multiplied by $E(C)$.

Finally the volumetric derivative of the asset model is simply

$$\dot{\mu}_0 f = X(t) \frac{df}{dx}(z). \quad (60)$$

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 2 analytically. The result is also reasonable in light of the different shapes of tZ and $\sqrt{t}Z$ as $t \rightarrow 0$, for a random variable Z with mean and standard deviation equal to 1. For very small t , tZ is essentially the same as a deterministic $tE(Z)$, whereas $\sqrt{t}Z$ has a standard deviation $\sqrt{t}$ which is much larger than the mean t . Its coefficient of variation $1/\sqrt{t} \rightarrow \infty$ as $t \rightarrow 0$. The relative uncertainty in $\sqrt{t}Z$ grows as $t \rightarrow 0$ whereas for tZ it disappears see Figure 3.

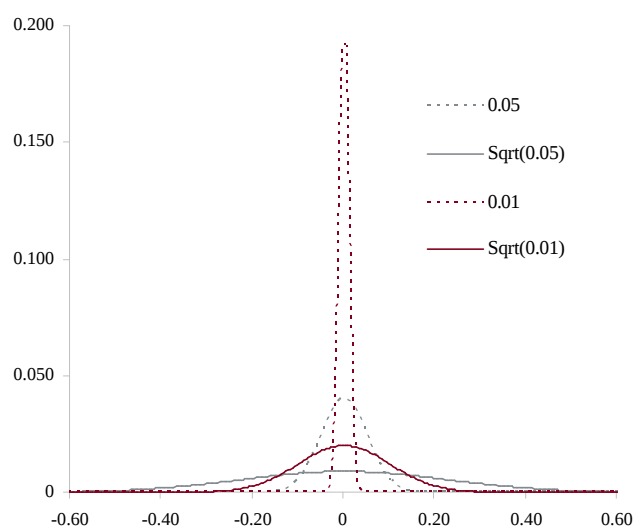


Figure 3. Illustration of the difference between tZ and $\sqrt{t}Z$ for Z a standard normal as $t \rightarrow 0$.

Severity uncertainty is also interesting. 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

$$\dot{\mu}_0 = \int \dot{\mu}_{0,v} dv. \quad (61)$$

We can 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.

Actuaries could model directly with a Lévy measure ν and hence avoid the artificial distinction between frequency and severity as Patrik et al. (1999) suggested. Catastrophe models already work in this way. Several aspects of actuarial practice could benefit from avoiding the artificial frequency/severity dichotomy. The dichotomy is artificial in the sense it depends on an arbitrary choice of one year to determine frequency. Explicitly considering the claim 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, facilitating combining losses from portfolios with different policy limits. This simplification is clearer when frequency and severity are not split.

6.5. Higher Order Identification of the Differences Between Insurance and Asset Models

We now consider whether Figure 1 is an exaggeration by computing the difference between the two operators $\dot{k}$ and $\dot{m}$ acting on test functions. We first extend k slightly by introducing the idea of a homogeneous approximation.

Let X be an infinitely divisible distribution with associated Lévy process $u \mapsto X_u$. As usual, we consider two coordinate maps $\mathbb{R}^+ \rightarrow \mathbf{L}$: the asset return model $k : \mathbb{R}^+ \rightarrow \mathbf{L}, u \mapsto k(u) = uX_1$, and the insurance model $m : \mathbb{R}^+ \rightarrow \mathbf{L}, u \mapsto m(u) = X_u$. These satisfy $k(1) = m(1) = X_1$, but in general $\dot{k}_u \neq \dot{m}_u$ at $u = 1$. We use u rather than t as the argument to avoid giving the impression that the index represents time: remember it represents the combination of time and volume.

Obviously there is no need to restrict k to be an approximation of X_1 . For general $u = u_0$ we can construct a homogeneous approximation to $m(u_0)$ at u_0 by $k(u) = (u/u_0)X_{u_0}$. The name homogeneous approximation is apt because $k(u_0) = m(u_0)$, and k is homogeneous: $k(su) = (su/u_0)X_{u_0} = sk(u)$, for real $s > 0$. For a general Lévy process X , X_{xt} does not have the same distribution as xX_t , so m is not homogeneous. For example if X is stable with index α , $0 < \alpha < 2$ then $X_{xt} \stackrel{d}{=} x^{1/\alpha} X_t$ (Brownian motion is stable with $\alpha = 2$). This section will compare $\dot{k}$ and $\dot{m}$ by computing the linear maps corresponding to $\dot{k}$ and $\dot{m}$ and showing they have a different form. We will compute the value of these operators on various functions to quantify how they differ. In the process we will recover Meyers' Meyers (2005a) example of "axiomatic capital" vs. "economic capital" from Section 4.

Suppose the Lévy triple defining X_u is (σ, ν, γ) , where σ is the standard deviation of the continuous term and ν is the Lévy measure. Let $L^{(\sigma, \nu, \gamma)}$ be the pseudo-differential operator defined by Theorem 2 and let μ_u be the law of X_u . Using the independent and additive increment properties of an Lévy process and (Sato 1999, Theorem 31.5) we can write

$$\dot{\mu}_u f = \langle L^{(\sigma, \nu, \gamma)} f, \mu_u \rangle = E_{X_u} \left[\lim_{s \downarrow 0} s^{-1} E_{X_s} (f(X_u + X_s) | X_u) - f(X_u) \right] \quad (62)$$

where $f : \mathbb{R} \rightarrow \mathbb{R}$ is a doubly differentiable, bounded function with bounded derivatives.

Regarding k as a deterministic drift at a random (but determined once and for all) speed X_{u_0}/u_0 , we can apply Equation (62) with $\gamma = X_{u_0}/u_0$ and average over X_{u_0} to get

$$\dot{k}_u f = E_{X_{u_0}} (\langle Lf, k_u \rangle) = E[X_{u_0}/u_0 f'(uX_{u_0}/u_0)]. \quad (63)$$

We can see this equation is consistent with Equation (23):

$$\int_0^u \dot{x}_v dv = \int_0^u E[X_{u_0}/u_0 f'(vX_{u_0}/u_0)] dv \quad (64)$$

$$= \int_0^u \int x/u_0 f'(vx/u_0) \mu_{u_0}(dx) dv \quad (65)$$

$$= \int \int_0^{ux/u_0} f'(y) dy \mu_{u_0}(dx) \quad (66)$$

$$= \int f(ux/u_0) - f(0) \mu_{u_0}(dx) \quad (67)$$

$$= E[f(uX_{u_0}/u_0)] - f(0), \quad (68)$$

where l_t is the law of X_t .

Suppose that the Lévy process X_t is a compound Poisson process with jump intensity λ and jump component distribution J . Suppose the jump distribution has a variance. Then, using Equation (62), and conditioning on the presence of a jump in time s , which has probability λs , gives

$$\dot{m}_u f = \lambda E_{X_u} [E(f(X_u + J)) - f(X_u)]. \quad (69)$$

Now let $f(x) = x$. Usually test functions are required to be bounded. We can get around this by considering $\min(f, n)$ for fixed n and letting $n \rightarrow \infty$ and only working with relatively thin tailed distributions—which we do through our assumption the severity J has a variance. Since $E(X_u) = \lambda u E(J)$, Equations (63) and (69) give $\dot{k}_u f = E(X_{u_0}/u_0) = \lambda E(J)$ and $\dot{m}_u f = \lambda E(J)$ respectively so the homogeneous approximation has the same derivative in this case.

If $f(x) = x^2$ then since $E(X_u^2) = \lambda^2 u^2 E(J)^2 + \lambda u E(J^2)$ we get

$$\dot{k}_u f = 2\lambda^2 u E(J)^2 + 2\lambda u E(J^2)/u_0. \quad (70)$$

On the other hand

$$\dot{m}_u f = 2\lambda^2 u E(J)^2 + \lambda E(J^2). \quad (71)$$

Thus

$$\dot{k}_{u_0} f = 2\lambda^2 u_0 E(J)^2 + 2\lambda E(J^2) = 2\lambda E(X_{u_0}) E(J) + 2\lambda E(J^2) \quad (72)$$

$$\dot{m}_{u_0} f = 2\lambda^2 u_0 E(J)^2 + \lambda E(J^2) = 2\lambda E(X_{u_0}) E(J) + \lambda E(J^2). \quad (73)$$

The difference $\dot{k}_{u_0} f - \dot{m}_{u_0} f = \lambda E(J^2)$ is independent of u_0 and so the relative difference decreases as u_0 increases, corresponding to the fact that X_{u_0} changes shape more slowly as u_0 increases. If J has a second moment, which we assume, then the relative magnitude of the difference depends on the relative size of $E(J^2)$ compared to $\lambda E(J)^2$, i.e., the variance of J offset by the expected claim rate λ .

In general, if $f(x) = x^n$, $n \geq 3$ then

$$\dot{k}_{u_0} f = E(n X_{u_0}^n / u_0). \quad (74)$$

On the other hand

$$\dot{m}_u f = \lambda E_{X_u} \left[\sum_{i=1}^n \binom{n}{i} X_u^{n-i} E(J^i) \right] \quad (75)$$

$$= \lambda \sum_{i=1}^n \binom{n}{i} E(X_u^{n-i}) E(J^i). \quad (76)$$

Let $\kappa_n(u)$ be the n th cumulant of X_u and $\mu'_n(u) = E(X_u^n)$ be the n th moment. Recall $\kappa_n(u) = u \int x^n \nu(dx) = \lambda u E(J^n)$ and the relationship between cumulants and moments

$$\kappa_n = \mu'_n - \sum_{k=1}^{n-1} \binom{n-1}{k-1} \kappa_k \mu'_{n-k}. \quad (77)$$

Combining these facts gives

$$E(X_u^n) = \lambda u E(J^n) + \sum_{i=1}^{n-1} \binom{n-1}{i-1} \lambda u E(J^i) E(X_u^{n-i}) \quad (78)$$

and hence

$$\begin{aligned}
 \dot{k}_{u_0}f - \dot{m}_{u_0}f &= E(nX_{u_0}^n/u_0) - \lambda \sum_{i=1}^n \binom{n}{i} E(J^i)E(X_{u_0}^{n-i}) \\
 &= \lambda(nE(J^n) + n \sum_{i=1}^{n-1} \binom{n-1}{i-1} E(J^i)E(X_{u_0}^{n-i}) - \sum_{i=1}^n \binom{n}{i} E(J^i)E(X_{u_0}^{n-i})) \\
 &= \lambda((n-1)E(J^n) + \sum_{i=1}^{n-1} \left(n \binom{n-1}{i-1} - \binom{n}{i} \right) E(J^i)E(X_{u_0}^{n-i})) \\
 &= \lambda \sum_{i=2}^n (i-1) \binom{n}{i} E(J^i)E(X_{u_0}^{n-i}).
 \end{aligned} \tag{79}$$

As for $n = 2$, the $E(J^n)$ term is independent of u_0 whereas all the remaining terms grow with u_0 . For $n = 3$ the difference is $3\lambda E(J^2)E(X) + 2E(J^3)$.

In the case of the standard deviation risk measure we recover same results as Section 4. Let $\rho(\mu_u) = (\langle x^2, \mu_u \rangle - \langle x, \mu_u \rangle^2)^{1/2}$ be the standard deviation risk measure. Using the chain rule, the derivative of ρ in direction $\dot{k}$ at $u = u_0$, where $m(u_0) = k(u_0)$, is

$$D\rho_{m(u_0)}(\dot{k}) = \frac{\dot{k}_{u_0}(x^2) - 2\langle x, x_{u_0} \rangle \dot{k}_{u_0}(x)}{2\rho(k_{t_0})} \tag{80}$$

and similarly for direction $\dot{m}$. Thus

$$D\rho(\dot{k}) - D\rho(\dot{m}) = \frac{\lambda E(J^2)}{2\rho(m_{u_0})}, \tag{81}$$

which is the same as the difference between Equations (7) and (10) because here $c_i = 0$, $g = E(J^2)$ and, since we are considering equality at u_0 where frequency is $x = \lambda u_0$, and we are differentiating with respect to u we pick up the additional λ in Equation (10).

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. Meyer's example and the failure of gradient based formulas to add-up for diversifying Lévy processes are practical manifestations of these differences.

The commonalities we have found between the homogeneous approximation k and the insurance embedding m are consistent with the findings of Boonen et al. (2017) that although insurance portfolios are not linearly scalable in exposure the Euler allocation rule can still be used in an insurance context. Our analysis pinpoints the difference between the two and highlights particular ways it could fail and could be more material in applications. Specifically, it is more material for smaller portfolios and for portfolios where the severity component has a high variance: these are exactly the situations where aggregate losses will be more skewed and will change shape most rapidly.

7. Empirical Analysis

7.1. Overview

Next we test different loss models against US statutory insurance data. Aon Benfield's original Insurance Risk Study ABI (2007) was based on the methodology described in this paper and the exhibits below formed part of its backup. The Risk Study has been continued each year since, see ABI (2012, 2013, 2014, 2015) for the most recent editions. The 2015 Tenth Edition provides a high-level analysis using regulatory insurance data from 49 countries that together represent over 90% of global P & C premium. The conclusions reported here hold across a very broad range of geographies and lines of business.

The original analysis [ABI \(2007\)](#) focused on the US and used National Association of Insurance Commissioners (NAIC) data. The NAIC is an umbrella organization for individual US state regulators. The NAIC statutory annual statement includes an accident year report, called Schedule P, showing ten years of premium and loss data by major line of business. This data is available by insurance company and insurance group. The analysis presented here will use data from 1993 to 2004 by line of business. We will model the data using IM1-4 from Section 5.2. The model fits can differentiate company effects from accident year pricing cycle effects, and the parameters show considerable variation by line of business. The fits also capture information about the mixing distribution C .

We will show the data is consistent with two hypotheses:

- H1.** The asymptotic coefficient of variation or volatility as volume grows is strictly positive.
- H2.** Time and volume are symmetric in the sense that the coefficient of variation of aggregate losses for volume x insured for time t only depends on xt .

H1 implies that insurance losses are *not* volumetrically diversifying. Based on Table 1, H1 is only consistent with IM3 or IM4. H2 is only consistent with IM1 and IM3. Therefore the data is only consistent with model IM3 and not consistent with the other models. IM3 implies that diversification over time and volume follows a symmetric modified square root rule, $v(x, t) = \sqrt{(\sigma^2/xt) + c}$.

7.2. Isolating the Mixing Distribution

We now show that the mixing distribution C in IM3 and IM4 can be inferred from a large book of business even though it cannot be directly observed.

Consider an aggregate loss distribution with a C -mixed Poisson frequency distribution, per Equation (6) or IM3, 4. 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 obviously has a variance. Moreover the variability from the Poisson claim count component also diversifies away, because the coefficient of variation of a Poisson distribution tends to zero as the mean increases. Therefore the shape of the normalized aggregate loss distribution, aggregate losses divided by expected aggregate losses, converges in distribution to the mixing distribution C .

This assertion can be proved using moment generating functions. Let X_n be a sequence of random variables with distribution functions F_n and let X be another random variable with distribution F . If $F_n(x) \rightarrow F(x)$ as $n \rightarrow \infty$ 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 $\epsilon > 0$ $\Pr(|X_n - X| > \epsilon) \rightarrow 0$ as $n \rightarrow \infty$. 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 $\Pr(Y = 1) = \Pr(X = 1) = 1/2$. 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) $M_n(z) = E(e^{zX_n})$ of X_n converge to the MGF of M of X for all z : $M_n(z) \rightarrow M(z)$ as $n \rightarrow \infty$, see ([Feller 1971](#), vol. 2, chp. 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 $A_n = X_1 + \dots + X_N$ and $a = nx$. Then the normalized loss ratio A_n/a converges in distribution to C , so

$$\Pr(A_n/a < \alpha) \rightarrow \Pr(C < \alpha) \quad (82)$$

as $n \rightarrow \infty$. Hence the standard deviation of A_n/a satisfies

$$\sigma(A_n/a) = \sqrt{c + \frac{x(1 + \gamma^2)}{a}} \rightarrow \sqrt{c}. \quad (83)$$

Proof. The moment generating function $M_{A_n}(z)$ of A_n is

$$M_{A_n}(z) = M_C(n(M_X(z) - 1)) \quad (84)$$

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

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

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

Proposition 1 is equivalent to a classical risk theory result of Lundberg describing the stabilization in time of portfolios in the collective, see (Bühlmann 1970, sct. 3.3). It also implies that if the frequency distribution is actually Poisson, so the mixing distribution 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.

Figures 4 and 5 illustrate the proposition, showing how the aggregate distributions change shape as expected counts increase. In Figure 4, $C = 1$ 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 5, C has a gamma distribution with variance 0.0625 (asymptotic coefficient of variation of 0.25). Now the scaled aggregate distributions converge to C .

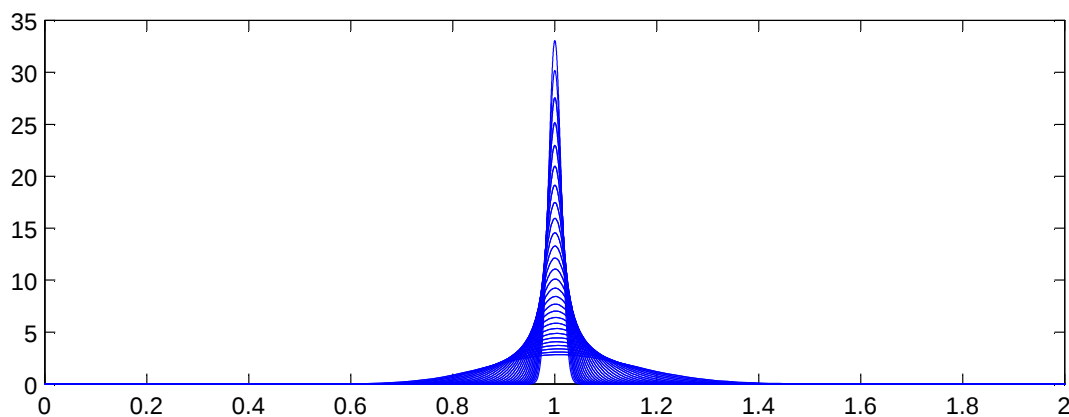


Figure 4. Theoretical distribution of scaled aggregate losses with no parameter or structure uncertainty and Poisson frequency.

Proposition 1 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. Proposition 1 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.

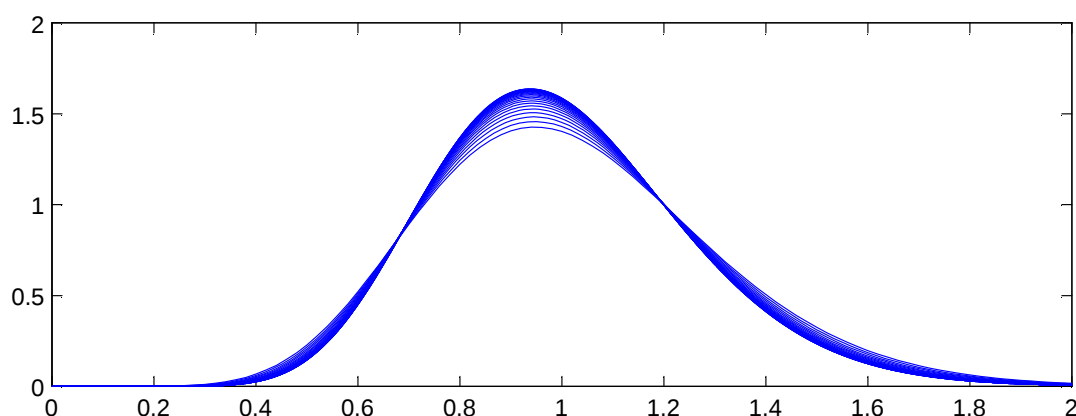


Figure 5. Theoretical distribution envelope of scaled aggregate losses with a gamma mixed Poisson frequency with mixing variance $c = 0.0625$.

The mixing distribution C can be thought of as capturing parameter risk or systematic insurance risks since its effect does not diversify away in a large book of business. In our context C is capturing a number of non-diversifiable risk elements, including variation the type of insured or coverage within a given classification, variation in the weather or other macro-risk factor over a long time frame (for example, the recent rise in distracted driving or changes in workplace injuries driven by the business cycle) as well as changes in the interpretation of policy coverage. We will estimate expected losses using premium and so the resulting C also captures inter-company pricing effects, such as different expense ratios, profit targets and underwriting appetites, as well as insurance pricing cycle effects (both of which are controlled for in our analysis). Henceforth we will refer to C as capturing parameter risk rather than calling it the mixing distribution.

7.3. Volumetric Empirics

We use NAIC annual statement data to determine an appropriate distribution for C (or $Z(1)$), providing 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 NAIC annual statement includes a ten accident-year history of gross, ceded and net premiums and ultimate losses 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](#) as follows.

- **Heterogeneity** refers to the level of consistency in terms and conditions and types of insureds within the line, with high heterogeneity indicating a broad range. 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 state insurance departments.
- **Limits** refers to the typical policy limit. Personal auto liability limits rarely exceed \$300,000 per accident in the US 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).

- **Cats** (i.e., catastrophes) 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.

Table 3. Characteristics of Various Lines of Insurance

Insurance Line	Heterogeneity	Regulation	Limits	Cycle	Cats
Personal Auto	Low	High	Low	Low	No
Commercial Auto	Moderate	Moderate	Moderate	High	No
Workers Compensation	Moderate	High	Statutory	High	Possible
Medical Malpractice	Moderate	Moderate	Moderate	High	No
Commercial Multi-Peril	Moderate	Moderate	Moderate	High	Moderate
Other Liability Occurrence	High	Low	High	High	Yes
Homeowners Multi-Peril	Moderate	High	Low	Low	High
Other Liability Claims Made	High	Low	High	High	Possible

In order to apply Proposition 1 we proxy a “large” book as one with more than \$100M of premium in each accident year. Figure 6 shows how 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. The loss processes are not volumetrically diversifying, that is the volatility does not decrease to zero with volume.
2. Below a range \$100M–1B (varying by line) there are material changes in volatility with premium size.
3. \$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 very material in the \$10–100M premium range where most companies would try to set profit targets by line or business unit. This is consistent with Mildenhall (2004).

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 but it also removes some of the catastrophic loss effect for a year—an issue 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

$$\hat{\lambda}_{c,y} = \lambda_{..} + (\lambda_{c.} - \lambda_{..}) + (\lambda_{.,y} - \lambda_{..}), \quad (85)$$

or multiplicatively as

$$\hat{\lambda}_{c,y} = \lambda_{.,y}(\lambda_{c,.}/\lambda_{.,.})(\lambda_{.,y}/\lambda_{.,.}). \quad (86)$$

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.

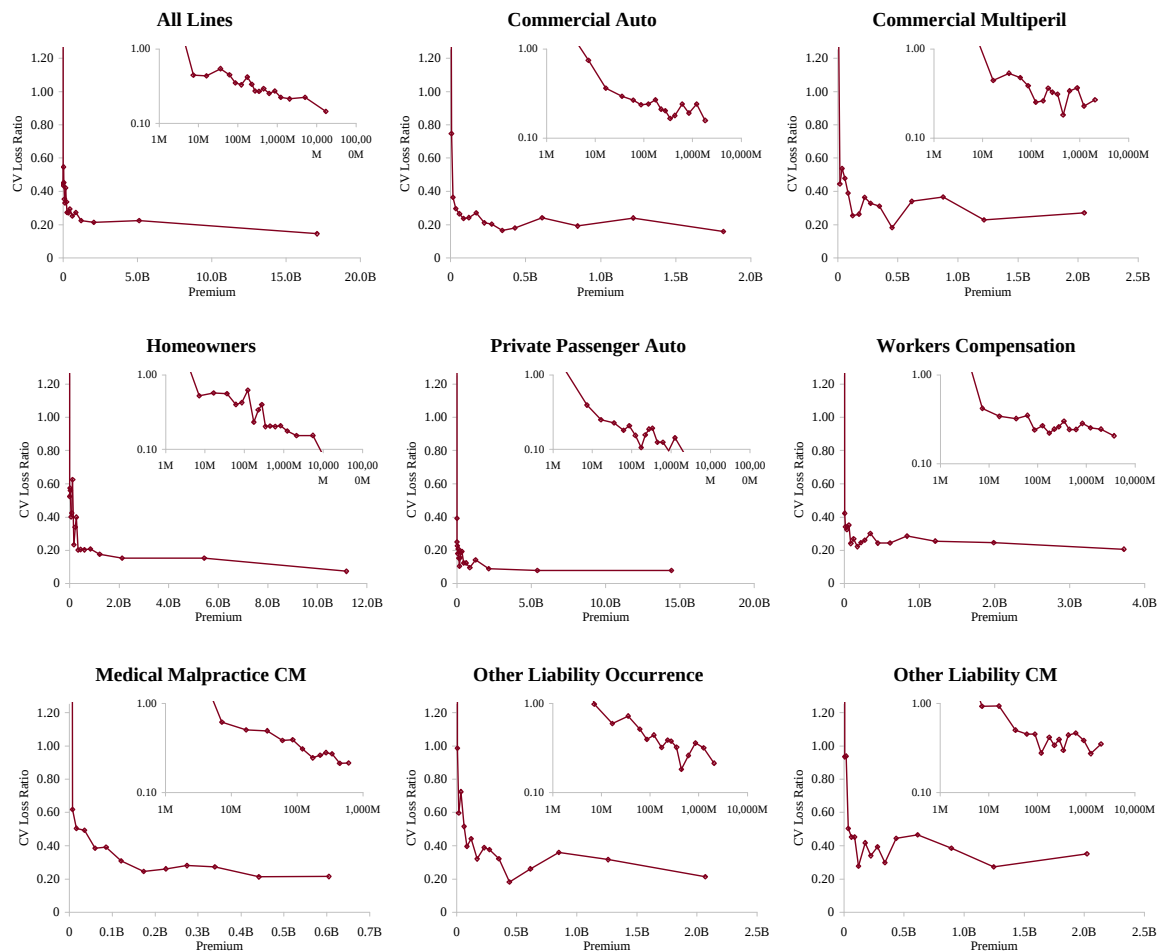


Figure 6. The relationship between raw loss ratio volatility, measured as coefficient of variation of loss ratios, and premium volume, using data from accident years 1993–2004. Each inset graph plots the same data on a log/log scale, showing that the volatility continues to decrease materially for premium volumes in the \$100Ms. 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.

Using Proposition 1 we obtain four estimates for the distribution of C from the empirical distributions of $\lambda_{c,y}/\hat{\lambda}_{.,y}$, $\lambda_{c,y}/\hat{\lambda}_{.,y}$, $\lambda_{c,y}/\hat{\lambda}_{c,.}$ 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 7–9 show analyses of variance for the model described by Equation (85). 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, sect. 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 7. Adjusted analysis of variance (ANOVA) for commercial auto, commercial multiperil and homeowners.

Tukey's test for interactions in an ANOVA with one observation per cell (Miller and Wichern 1977, sct. 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.

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 8. Adjusted ANOVA for medical malpractice claims made and other liability claims made and occurrence.

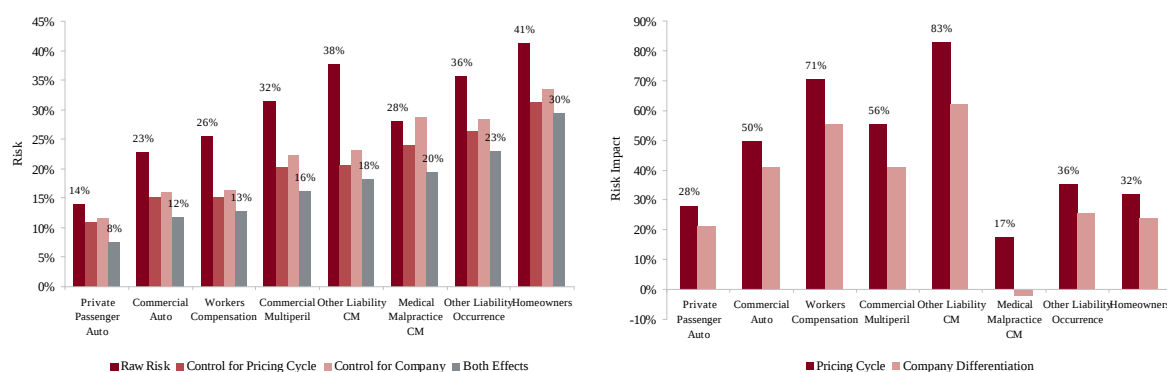
Figure 10 shows the indicated volatilities for commercial auto, commercial multi-peril, homeowners, other liability occurrence, private passenger auto liability and workers compensation for the four models Res1-4 and Equation (86). The right hand plot shows the impact of the pricing (accident year) effect and the firm effect on total volatility. This Figure 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.

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 9. Adjusted ANOVA for private passenger auto liability and workers compensation.**Figure 10.** 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.

Figures 11–14 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 Res1, AY Avg adjusts for

accident year or pricing cycle effect Res 2, Co Avg adjusts for company effect Res 3, and Mult Both Avg adjusts for both Res 4, per Equation (86). All residuals are computed using the multiplicative model.

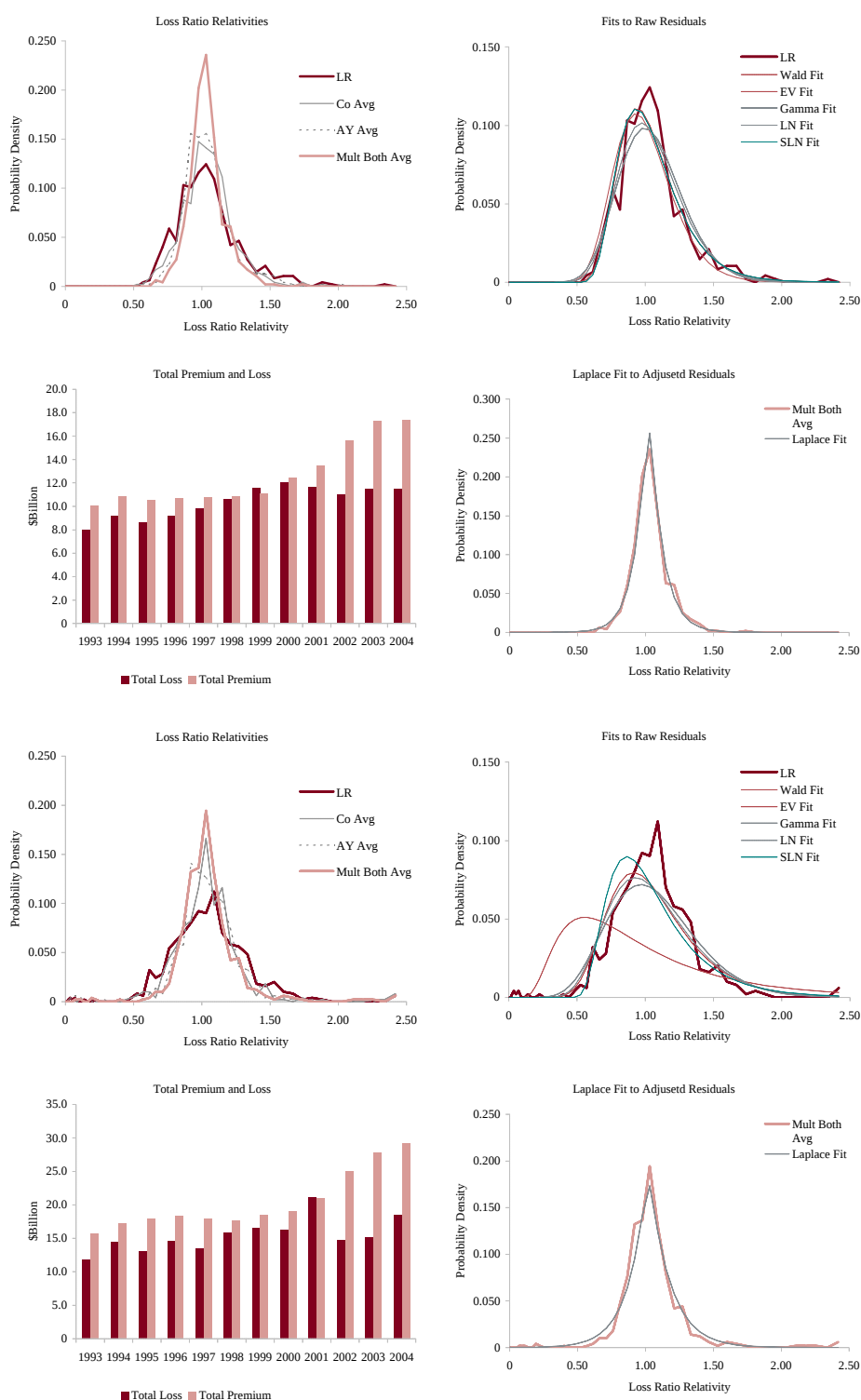


Figure 11. Commercial auto liability (top four plots) and commercial multiperil volatility (bottom four plots). Note 9/11 loss effect in the lower-left plot. See text for a description of the plots.

The top right hand plot shows five parametric distribution fits to the raw residuals, Res1. The distributions are described in Table 4. 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 commonly-assumed gamma distribution fit is shown in bold grey; the adequacy of its fit varies from line to line.

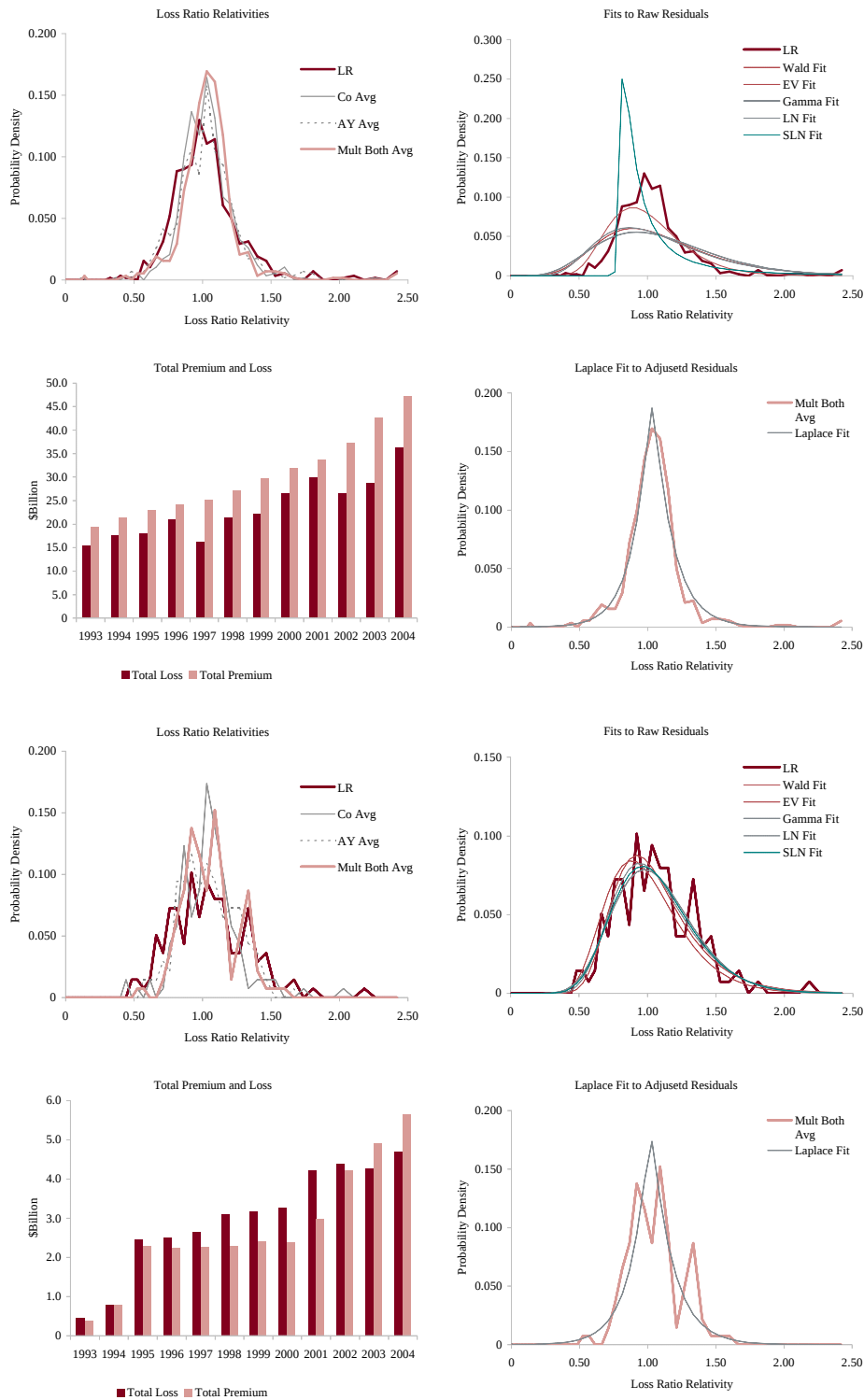


Figure 12. Homeowners (top four plots) and medical malpractice claims made volatility (bottom four plots). Note the 2004 homeowners catastrophe losses. See text for a description of the plots.

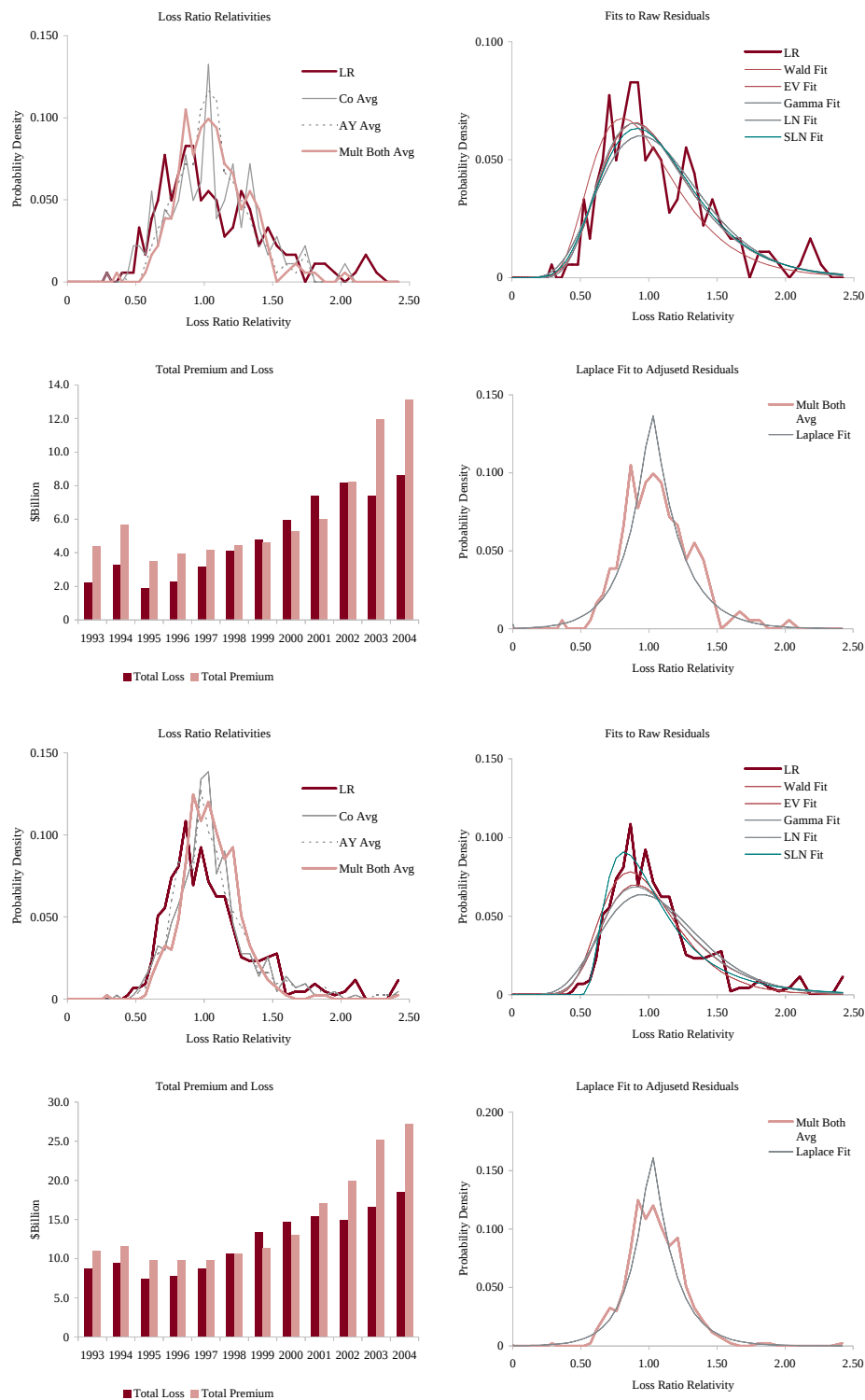


Figure 13. Other liability claims made (top four plots) and occurrence volatility (bottom four plots). See text for a description of the plots.

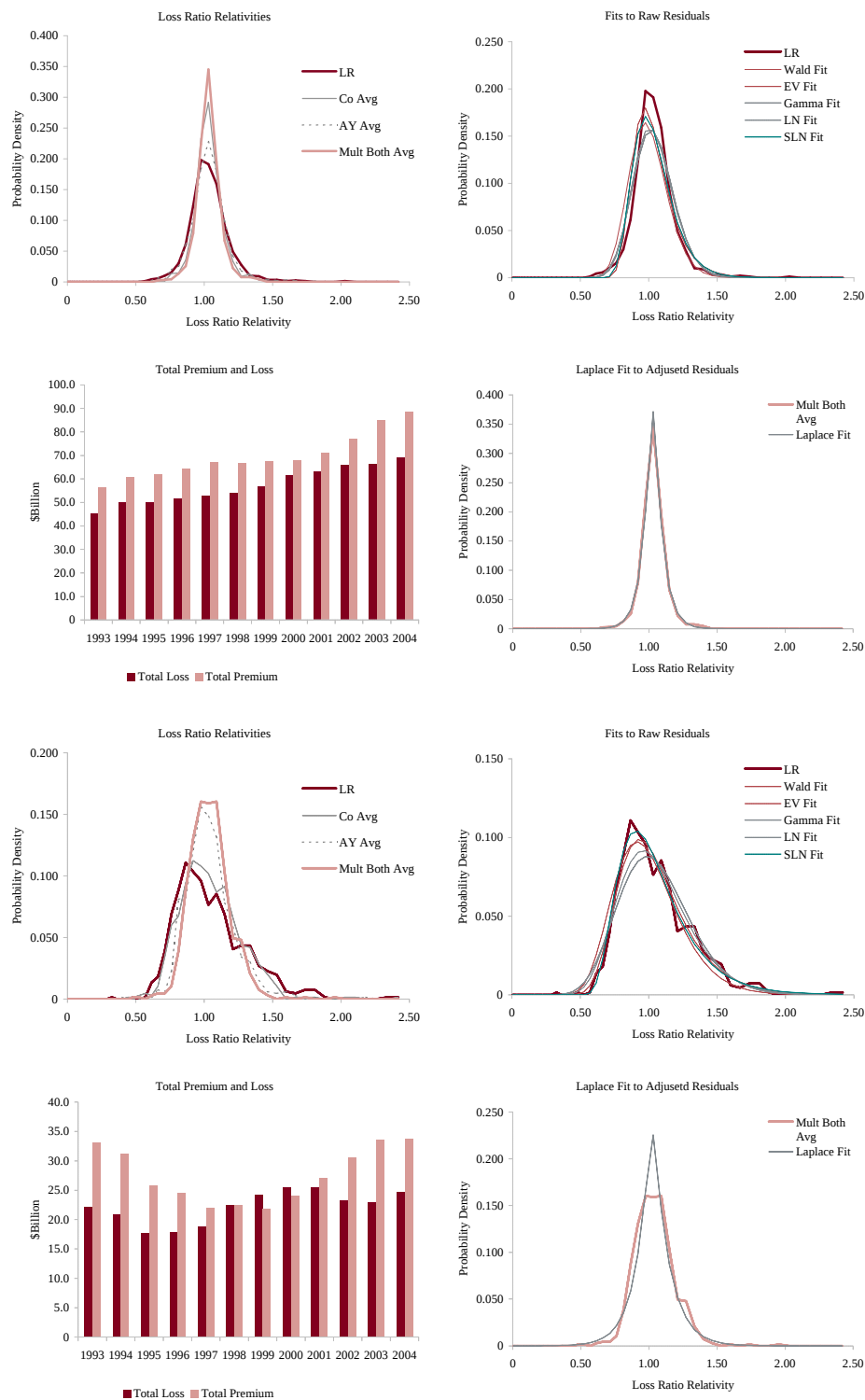


Figure 14. Private passenger auto liability (top four plots) and workers compensation volatility (bottom four plots). Note vertical scale on private passenger auto loss ratios and the visibly higher volatility of premium than loss in the lower left hand plot. See text for a description of the plots.

Table 4. Summary of distributions fit to C in Figures 11–14.

Abbreviation	Parameters	Distribution	Fitting Method
Wald	2	Wald (inverse Gaussian)	Maximum likelihood
EV	2	Frechet-Tippet extreme value	Method of moments
Gamma	2	Gamma	Method of moments
LN	2	Lognormal	Maximum likelihood
SLN	3	Shifted lognormal	Method of moments

The lower right hand plot shows the residuals adjusted for both pricing cycle and company effects, Res4, and it includes a maximum likelihood Laplace fit to the multiplicative model Equation (86). 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 has the same relationship to the absolute value difference that the normal distribution has 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 $\nu(x) = |x|^{-1}e^{-|x|/s}$ explored in Example 13. See Kotz et al. (2001) for a comprehensive survey of the Laplace distribution.

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 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 6. 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.

Subsequent analyses, conducted after 2006 when the bulk of this paper was written, confirm the parameter estimates shown in Figure 10 are reasonably stable over time. Volatility for liability lines has increased since 2004 driven by loss development from the soft market years that has dispersed loss ratios further as they emerged to ultimate, but the relative ordering is unchanged. Interestingly the Global Financial Crisis had very little impact on insurance volatility other than for Financial Guarantee.

Table 5 and (ABI 2010, p. 6) show a comparison of Solvency II premium risk factors with the risk factors computed here. Finally, Table 6 and (ABI 2012, p. 6) show a comparison of the individual line of business parameters based on data 1992–2011 vs. the original study 1992–2004. See (ABI 2015, p. 52) for a further update of the underwriting cycle effect on volatility by line.

Table 5. Comparison of risk factors with Solvency II premium risk factors.

Dimension	Actuarial Geometry	Solvency II
Time horizon	to ultimate	one year
Catastrophe risk	included	excluded
Size of company	large	average

Table 6. Coefficient of variation of gross loss ratio, Source: Aon Benfield Insurance Risk Study, 7th Edition, used with permission.

Line	1st Edition	7th Edition	Change
Private Passenger Auto	14%	14%	0%
Commercial Auto	24%	24%	0%
Workers' Compensation	26%	27%	1%
Commercial Multi Peril	32%	34%	2%
Medical Malpractice: Claims-Made	33%	42%	9%

Table 6. Cont.

Line	1st Edition	7th Edition	Change
Medical Malpractice: Occurrence	35%	35%	0%
Other Liability: Occurrence	36%	38%	2%
Special Liability	39%	39%	0%
Other Liability: Claims-Made	39%	41%	2%
Reinsurance Liability	42%	67%	25%
Products Liability: Occurrence	43%	47%	4%
International	45%	72%	27%
Homeowners	47%	48%	1%
Reinsurance: Property	65%	85%	20%
Reinsurance: Financial	81%	93%	12%
Products Liability: Claims-Made	102%	100%	−2%

7.4. Temporal Empirics

We now investigate the behavior of the coefficient of variation of a book with volume x insured for t years, $v(x, t)$ 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 or catastrophe losses. Moreover, management actions, including reserve setting and line of business policy form and pricing decisions, will affect observed volatility.

Reviewing Table 1, and comparing with Figure 6, 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 $v(x, t)$. 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 $v(x, 1)$, from the data underlying Figure 6. Then combine two years of premium and losses, from the same company and line, and recompute volatilities. This computes $v(x/2, 2)$ —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 estimates of $v(x/4, 4)$, $v(x/6, 6)$, and $v(x/12, 12)$. Normalizing the data to a constant loss ratio across accident years prior to performing the analysis will remove potentially distorting pricing-cycle effects.

Figure 15 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 $v(x, 1)$ and the last four show $v(x/t, t)$ 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 v for broader volume bands. Clearly the differences in means are insignificant relative to the standard deviations, and so a crude analysis of variance would not reject the hypothesis that $v(x/t, t)$ 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 15. Coefficient of variation of loss ratio by premium volume for private passenger auto liability, computed using bucketed xt for $t = 1, 2, 4, 6, 12$.

Finally, Figures 16 and 17 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 6 on a log/linear scale and a fit of $v(x, t)$ 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 $v(x/t, t)$ for $t = 2, 4, 6, 12$ using the method described above. Thus the private passenger auto liability plot shows the data in Figure 15. These plots are consistent with the hypothesis that $v(x/t, t)$ 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 of the models IM1-4 and AM1 only model IM2 of has volumetric and temporal properties consistent with the data in the NAIC annual statement database.



Figure 16. Fit of $\sqrt{\frac{\sigma^2}{xt}} + c$ 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 17. Fit of $\sqrt{\frac{\sigma^2}{xt}} + c$ 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 $t = 1, 2, 4, 6, 12$.

8. Conclusions

The difference between asset geometry and actuarial geometry reflects 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. The choice of coordinate used to differentiate risk measures must reflect these differences.

We have provided an introduction to the actuarial use of Lévy processes to model aggregate losses. The Lévy process model reflects the realities of insurance: it is curved in both the volume and time dimensions. Asset returns, in contrast, are volumetrically flat. We have clarified the notion of a

“direction” in the space of risks and used it to explain two different allocation results derived using the gradient of a risk measure.

US NAIC annual statement data is used to demonstrate that insurance liabilities do not diversify volumetrically or temporally. We reviewed four models of aggregate losses based on Lévy processes—models with a long risk-theoretic pedigree, it should be noted—and showed that only model IM2 is consistent with the NAIC data. We also show how parameter risk can be explicitly quantified at a *distributional* level even though it is unobservable. 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. Regulation and Enterprise Risk Management, both of which depend crucially on an accurate quantification of aggregate loss distributions, demand accurate and realistic modeling. It is time to satisfy that demand with a fully data-grounded model for losses, including appropriate parameter risk.

Conflicts of Interest: The author declares no conflict of interest.

References

- Insurance Risk Study*. First Edition. Chicago: Aon Re Global, 2007.
- Insurance Risk Study*. Fifth Edition. Chicago: Aon Benfield, 2010.
- Insurance Risk Study*. Seventh Edition. Chicago: Aon Benfield, 2012.
- Insurance Risk Study*. Eighth Edition. Chicago: Aon Benfield, 2013.
- Insurance Risk Study*. Ninth Edition. Chicago: Aon Benfield, 2014.
- Insurance Risk Study*. Tenth Edition. Chicago: Aon Benfield, 2015.
- Abraham, Ralph, Jerrold E. Marsden, and Tudor Ratiu. 1988. *Manifolds, Tensor Analysis, and Applications*, 2nd ed. New York: Springer Verlag.
- Applebaum, David. 2004. *Lévy Processes and Stochastic Calculus*. Cambridge: Cambridge University Press.
- Artzner, Philippe, F. Delbaen, J.M. Eber, and D. Heath. 1999. Coherent Measures of Risk. *Mathematical Finance* 9: 203–28.
- Bühlmann, Hans. 1970. *Mathematical Models in Risk Theory*. Berlin: Springer Verlag.
- Bailey, Robert A. 1967. Underwriting Profits From Investments. *Proceedings of the Casualty Actuarial Society* LIV: 1–8.
- Barndorff-Nielsen, Ole E., Thomas Mikosch, and Sidney I. Resnick, eds. 2001. *Lévy Processes—Theory and Applications*. Boston: Birkhäuser.
- Barndorff-Nielsen, Ole Eiler. 2000. *Probability Densities and Lévy Densities*. Aarhus: University of Aarhus. Centre for Mathematical Physics and Stochastics (MaPhySto), vol. MPS RR 2000-8.
- Bauer, Daniel, and George Zanjani. 2013a. The Marginal Cost of Risk in a Multi-Period Risk Model. *preprint*, November.
- Bauer, Daniel, and George H. Zanjani. 2013b. Capital allocation and its discontents. In *Handbook of Insurance*. New York: Springer, pp. 863–80.
- Beard, R. E., T. Pentikäinen, and E. Pesonen. 1969. *Risk Theory—The Stochastic Basis of Insurance*, 3rd ed. London: Chapman and Hall.
- Bertoin, Jean. 1996. *Lévy Processes*. Cambridge: Cambridge University Press.
- Bodoff, Neil M. 2009. Capital Allocation by Percentile Layer. 2007 ERM Symposium. Available online: www.ermssymposium.org/2007/pdf/papers/Bodoff.pdf (accessed on 13 June 2017).
- Boonen, Tim J., Andreas Tsanakas, and Mario V. Wüthrich. 2017. Capital allocation for portfolios with non-linear risk aggregation. *Insurance: Mathematics and Economics* 72: 95–106.
- Borwein, Jonathan M., and Jon D. Vanderwerff. 2010. *Convex Functions: Constructions, Characterizations and Counterexamples*. Cambridge: Cambridge University Press, vol. 109.
- Bowers, Newton, Hans Gerber, James Hickman, Donald Jones, and Cecil Nesbitt. 1986. *Actuarial Mathematics*. Schaumburg: Society of Actuaries.

- Breiman, Leo. 1992. *Probability, Volume 7 of Classics in Applied Mathematics*. Philadelphia: Society for Industrial and Applied Mathematics (SIAM).
- Buch, Arne, and Gregor Dorfleitner. 2008. Coherent risk measures, coherent capital allocations and the gradient allocation principle. *Insurance: Mathematics and Economics* 42: 235–42.
- Cummins, J. David, and Scott E. Harrington, eds. 1987. *Fair Rate of Return in Property-Liability Insurance*. Boston: Kluwer-Nijhoff.
- Cummins, J. David, and Richard D. Phillips. 2000. Applications of Financial Pricing Models in Property-Liability Insurance. In *Handbook of Insurance*. Edited by G. Dionne. Boston: Kluwer Academic.
- Cummins, J. David. 1988. Risk-Based Premiums for Insurance Guarantee Funds. *Journal of Finance* 43: 823–39.
- Cummins, J. David. 2000. Allocation of Capital in the Insurance Industry. *Risk Management and Insurance Review* 3: 7–27.
- D'Arcy, Stephen P., and Neil A. Doherty. 1988. *The Financial Theory of Pricing Property Liability Insurance Contracts*. SS Huebner Foundation for Insurance Education, Wharton School, University of Pennsylvania; Homewood: Irwin.
- D'Arcy, Stephen P., and Michael A. Dyer. 1997. Ratemaking: A Financial Economics Approach. *Proceedings of the Casualty Actuarial Society* LXXXIV: 301–90.
- Daykin, Chris D., Teivo Pentikäinen, and Martti Pesonen. 1994. *Practical Risk Theory for Actuaries*. London: Chapman and Hall.
- Delbaen, F., and J. Haezendonck. 1989. A martingale approach to premium calculation principles in an arbitrage free market. *Insurance: Mathematics and Economics* 8: 269–77.
- Delbaen, Freddy. 2000a. Coherent Risk Measures. *Monograph, Pisa: Scuola Normale Superiore*.
- Delbaen, Freddy. 2000b. *Coherent Risk Measures on General Probability Spaces*. Advances in finance and stochastics. Berlin: Springer, 2002. 1–37.
- Denault, Michel. 2001. Coherent allocation of risk capital. *Journal of Risk* 4: 1–34.
- Dhaene, Jan, Mark J. Goovaerts, and Rob Kaas. 2003. Economic capital allocation derived from risk measures. *North American Actuarial Journal* 7: 44–56.
- Dhaene, Jan, Steven Vanduffel, Marc J. Goovaerts, Rob Kaas, Qihe Tang, and David Vyncke. 2006. Risk Measures and Comonotonicity: A Review. *Stochastic Models* 22: 573–606.
- Dhaene, Jan, Roger JA Laeven, Steven Vanduffel, Gregory Darkiewicz, and Marc J. Goovaerts. 2008. Can a coherent risk measure be too subadditive? *Journal of Risk and Insurance* 75: 365–86.
- Dhaene, Jan, Andreas Tsanakas, Emiliano A. Valdez, and Steven Vanduffel. 2012. Optimal Capital Allocation Principles. *Journal of Risk and Insurance* 79: 1–28.
- Doherty, Neil A., and James R. Garven. 1986. Price Regulation in Property-Liability Insurance: A Contingent Claims Approach. *Journal of Finance* XLI: 1031–50.
- Erel, Isil, Stewart C. Myers, and James A. Read. 2015. A theory of risk capital. *Journal of Financial Economics* 118: 620–635.
- Föllmer, Hans, and Alexander Schied. 2011. *Stochastic Finance: An Introduction in Discrete Time*. Berlin: Walter de Gruyter.
- Feller, William. 1971. *An Introduction to Probability Theory and Its Applications, Two Volumes*, 2nd ed. New York: John Wiley and Sons.
- Ferrari, J. Robert. 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, Tom. 2003. Risk Capital Allocation by Coherent Risk Measures Based on One-Sided Moments. *Insurance: Mathematics and Economics* 32: 135–146.
- Froot, Kenneth A., and Paul GJ O'Connell. 2008. On the pricing of intermediated risks: Theory and application to catastrophe reinsurance. *Journal of Banking and Finance* 32: 69–85.
- Froot, Kenneth A., and Jeremy C. Stein. 1998. Risk management, capital budgeting, and capital structure policy for financial institutions: an integrated approach. *Journal of Financial Economics* 47: 52–82.
- Froot, Kenneth A., David S. Scharfstein, and Jeremy C. Stein. 1993. Risk Management: Coordinating Corporate Investment and Financing Policies. *Journal of Finance* XLVIII: 1629–58.
- Froot, Kenneth A. 2007. Risk management, capital budgeting, and capital structure policy for insurers and reinsurers. *Journal of Risk and Insurance* 74: 273–299.

- Furman, Edward, and Riardas Zitikis. 2008. Weighted risk capital allocations. *Insurance: Mathematics and Economics* 43: 263–69.
- Goovaerts, Marc J., Rob Kaas, and Roger JA Laeven. 2010. Decision principles derived from risk measures. *Insurance: Mathematics and Economics* 47: 294–302.
- Gründl, Helmut, and Hato Schmeiser. 2007. Capital allocation for insurance companies—what good is it? *Journal of Risk and Insurance* 74: 301–17.
- Hull, John. 1983. *Options Futures and Other Derivative Securities*, 2nd ed. Englewood Cliffs: Prentice-Hall.
- Jacob, Niels. 2001. *Pseudo Differential Operators & Markov Processes: Fourier Analysis and Semigroups*. London: Imperial College Press, vol. I.
- Jacob, Niels. 2002. *Pseudo Differential Operators & Markov Processes: Generators and Their Potential Theory*. London: Imperial College Press, vol. II.
- Jacob, Niels. 2005. *Pseudo Differential Operators & Markov Processes: Markov Processes and Applications*. London: Imperial College Press, vol. III.
- Kalkbrener, Michael. 2005. An axiomatic approach to capital allocation. *Mathematical Finance* 15: 425–37.
- Kallop, R. H. 1975. A current look at workers' compensation ratemaking. *Proceedings of the Casualty Actuarial Society* LXII: 62–81.
- Karatzas, Ioannis, and Steven Shreve. 1988. *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag.
- Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot. 1998. *Loss Models, from Data to Decisions*. New York: John Wiley and Sons.
- Kotz, Samuel, Tomasz Kozubowski, and Krzysztof Podgorski. 2001. *The Laplace Distribution and Generalizations*. Boston: Birkhauser.
- Laeven, Roger JA, and Mitja Stadje. 2013. Entropy coherent and entropy convex measures of risk. *Mathematics of Operations* 38: 265–93.
- Lange, Jeffrey T. 1966. General liability insurance ratemaking. *Proceedings of the Casualty Actuarial Society* LIII: 26–53.
- Magrath, Joseph J. 1958. Ratemaking for fire insurance. *Proceedings of the Casualty Actuarial Society* XLV: 176–95.
- Meister, Steffen. 1995. *Contributions to the Mathematics of Catastrophe Insurance Futures*. Unpublished Diplomarbeit, ETH Zurich.
- Merton, Robert C., and Andre Perold. 2001. Theory of Risk Capital in Financial Firms. In *The New Corporate Finance, Where Theory Meets Practice*. Edited by Donald H. Chew. Boston: McGraw-Hill, pp. 438–54.
- Meyers, G. 2005a. Distributing capital: another tactic. *Actuarial Rev* 32: 25–26. with on-line technical appendix.
- Meyers, G. 2005b. The Common Shock Model for Correlated Insurance Losses. *Proc. Risk Theory Society*. Available online: <http://www.aria.org/rts/proceedings/2005/Meyers%20-%20Common%20Shock.pdf> (accessed on 13 June 2017).
- Mildenhall, Stephen J. 2004. A Note on the Myers and Read Capital Allocation Formula. *North American Actuarial Journal* 8: 32–44.
- Mildenhall, Stephen J. 2006. Actuarial Geometry. *Proc. Risk Theory Society*. Available online: <http://www.aria.org/rts/proceedings/2006/Mildenhall.pdf> (accessed on 13 June 2017).
- Miller, Robert Burnham, and Dean W. Wichern. 1977. *Intermediate Business Statistics: Analysis of Variance, Regression and Time Series*. New York: Holt, Rinehart and Winston.
- Myers, Stewart C., and James A. Read Jr. 2001. Capital Allocation for Insurance Companies. *Journal of Risk and Insurance* 68: 545–80.
- Panjer, Harry H., and Gordon E. Willmot. 1992. *Insurance Risk Models*. Schaumburg: Society of Actuaries.
- Panjer, Harry H. 2001. *Measurement of Risk, Solvency Requirements and Allocation of Capital Within Financial Conglomerates*. Waterloo: University of Waterloo, Institute of Insurance and Pension Research.
- Patrik, Gary, Stefan Bernegger, and Marcel Beat Rüegg. 1999. The Use of Risk Adjusted Capital to Support Business Decision-Making. In *Casualty Actuarial Society Forum, Spring*. Baltimore: Casualty Actuarial Society, pp. 243–334.
- Perold, Andre F. 2001. *Capital Allocation in Financial Firms*. HBS Competition and Strategy Working Paper Series, 98-072. Boston: Harvard Business School.
- Phillips, Richard D., J. David Cummins, and Franklin Allen. 1998. Financial Pricing of Insurance in the Multiple-Line Insurance Company. *The Journal of Risk and Insurance* 65: 597–636.

- Powers, Michael R. 2007. Using Aumann-Shapley values to allocate insurance risk: the case of inhomogeneous losses. *North American Actuarial Journal* 11: 113–27.
- Ravishanker, Nalini, and Dipak K. Dey. 2002. *A First Course in Linear Model Theory*. Boca Raton: Chapman & Hall/CRC.
- Sato, K. I. 1999. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge: Cambridge University Press.
- Sherris, M. 2006. Solvency, Capital Allocation and Fair Rate of Return in Insurance. *Journal of Risk and Insurance* 73: 71–96.
- Stroock, Daniel W. 1993. *Probability Theory, an Analytic View*. Cambridge: Cambridge University Press.
- Stroock, Daniel W. 2003. *Markov Processes from K. Itô's Perspective*. Annals of Mathematics Studies. Princeton: Princeton University Press.
- Tasche, Dirk. 1999. Risk Contributions and Performance Measurement. *Report of the Lehrstuhl für Mathematische Statistik*, . München: TU München.
- Tasche, Dirk. 2004. Allocating portfolio economic capital to sub-portfolios. In *Economic Capital: A Practitioner's Guide*. London: Risk Books, pp. 275–302.
- Tsanakas, Andreas. 2009. To split or not to split: Capital allocation with convex risk measures. *Insurance: Mathematics and Economics* 1: 1–28.
- Venter, Gary G., John A. Major, and Rodney E. Kreps. 2006. Marginal Decomposition of Risk Measures. *ASTIN Bulletin* 36: 375–413.
- Venter, Gary G. 2009. Next steps for ERM: valuation and risk pricing. In *Proceedings of the 2010 Enterprise Risk Management Symposium*, Chicago, IL, USA, 12–15 April.
- Zanjani, George. 2002. Pricing and capital allocation in catastrophe insurance. *Journal of Financial Economics* 65: 283–305.
- Zanjani, George. 2010. An Economic Approach to Capital Allocation. *Journal of Risk and Insurance* 77: 523–49.



© 2017 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).